

Imperial College of Science, Technology and Medicine
Department of Physics, Astrophysics Group

**Constraining the Anisotropic Expansion of the
Universe with Type Ia Supernovae and Improving
the Treatment of Selection Effects within
Bayesian Hierarchical Models**

Wahidur Rahman

Submitted in part fulfilment of the requirements for the degree of
Doctor of Philosophy to Imperial College London, February 28, 2023

Abstract

In thesis, I aim to apply advanced methods in Bayesian statistical modelling on Type Ia Supernovae (SNIa) data to determine tighter constraints on the fiducial Lambda-Cold-Dark-Matter (Λ CDM) cosmology and improve the modelling of systematic uncertainties in the data. The body of work covered herein can be broadly classified into three main topics:

New Constraints on Anisotropic Expansion from Supernovae Type Ia within Λ CDM

I re-examine the contentious question of constraints on anisotropic expansion from Type Ia supernovae (SNIa) in the light of a novel determination of peculiar velocities, which are crucial to test isotropy with SNe out to distances $\lesssim 200/h$ Mpc. I re-analyze the Joint Light-Curve Analysis (JLA) Supernovae (SNe) data, improving on previous treatments of peculiar velocity corrections and their uncertainties (both statistical and systematic) by adopting state-of-the-art flow models constrained independently via the 2M++ galaxy redshift compilation. I also introduce a novel procedure to account for colour-based selection effects, and adjust the redshift of low- z SNe self-consistently in the light of our improved peculiar velocity model.

The Bayesian hierarchical model **BAHAMAS** is adopted to constrain a dipole in the distance modulus in the context of the Λ CDM model and the deceleration parameter in a phenomenological Cosmographic expansion. I do not find any evidence for anisotropic expansion, and place a tight upper bound on the amplitude of a dipole, $|D_\mu| < 5.93 \times 10^{-4}$ (95% credible interval) in a Λ CDM setting, and $|D_{q_0}| < 6.29 \times 10^{-2}$ in the Cosmographic expansion approach. Using Bayesian model comparison, I obtain posterior odds in excess of 900:1 (640:1) against a constant-in-redshift dipole for Λ CDM (the Cosmographic expansion). In the isotropic case, an accelerating universe is favoured with odds of $\sim 1100 : 1$ with respect to a decelerating one.

Testing Phenomenological Models of Anisotropy with SNIa

The Cosmological Principle, based on the assumptions of spatial homogeneity and isotropy, is one of the founding pillars of contemporary cosmology. In this work, I propose a simple model where isotropy is broken at the time of dark matter–dark energy equality, and I investigate a way to test this proposal by using Type Ia Supernovae data sets. I show that current data from the JLA dataset is not enough to decisively distinguish (according to the Jeffrey’s scale) this model from Λ CDM, with posterior odds of 56:1 for moderate to strong evidence in favour of

the isotropic model. I also place a 95% credible interval for the anisotropy breaking parameter $\epsilon_0 > 0.970$.

Accounting for Selection Effects in SNIa Cosmology using Bayesian Hierarchical Modelling

One of the modern problems of SNIa cosmology is accounting for selection effects caused by Malmquist bias in a principled way. Here, I present a complete formalism for handling selection effects in Type Ia supernova (SNIa) cosmology in the context of Bayesian Hierarchical Modeling. I demonstrate the method on simulated data sets where selection cuts are made on the apparent magnitude and show that previous results by [Rubin et al. \(2015\)](#); [Hinton et al. \(2018\)](#) are incorrect and can lead to biased cosmological parameters reconstruction. I also show how this formalism is easily extended to include the Phillips corrections that are used to standardize SNIa. In addition, the formalism presented exhibits better statistical properties in terms of bias and mean squared error relative to a traditional ad hoc style correction and the model of [\(Rubin et al., 2015\)](#).

Acknowledgements

First and foremost, I graciously thank the enthusiastic efforts of my supervisor, Roberto Trotta, for being a guiding light on this research journey and his tireless support that made this PhD possible. I am still in awe at the depths of your knowledge and how no matter how stuck or lost I seemed, you always seemed to know the way forward.

To my fellow PhD students whose company has kept me sane during this endeavour, it has been a pleasure working together. I thank in particular Harry Mootoovaloo, George Kyriacou, Adélie Gorce, Luke Johnson, Lena Lenz, Kimeel Sooknunan, Ian Hothi, Daniel Jones, Ciarán Coneely, Josh Greenslade and Claude Schmit for the support and encouragement you have given, the endless board game nights and for *totally never* judging me for my regularly scheduled post-lunch naps on my desk.

To everyone else at Imperial, I am forever grateful for everyone's effort to create a welcoming environment for any budding academic that involved both light hearted chats and stimulating discussions at coffee time. I thank in particular Alan Heavens and Andrew Jaffe for inviting me to join the regular Cosmology group meetings to broaden my horizons and to see the insightful work of others in the department.

To my collaborators who chose to work with a young and inexperienced, but still growing, researcher by advising and proofreading my manuscripts, I thank you wholeheartedly. In particular, I thank David van Dyk, Supranta Boruah, Mike Hudson.

Finally, I thank my family and friends outside of Imperial, too numerous to list by name, for their boundless support and encouragement, including my mother, who for a long time thought I was to become a *Physician* because I chose to study Physics. I think she has since figured out the difference.

Declaration

Copyright

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution NoDerivatives 4.0 International Licence (CC BY-ND).

Under this licence, you may copy and redistribute the material in any medium or format for both commercial and non-commercial purposes. This on the condition that; you credit the author and do not distribute modified versions of the work.

When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text.

Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

Statement of Originality

The work presented in this thesis is my own. Where it is not wholly my own work I note these below

Contributions

Below I list contributions from others in this thesis, and my own, where relevant.

Chapter 5

- This work was based on the work presented in [Rahman et al. \(2021\)](#)
- The majority of the work, including modifications to BAHAMAS, and numerical analysis is my own.

- S.S. Boruah and M. Hudson were responsible for generating the new peculiar velocity systematic uncertainty covariance matrix of the analysis and provided input on how to handle the statistical uncertainty.
- General concepts and discussions were done with all listed co-authors.

Chapter 6

- Based on a paper in prep.
- The cosmological model including the design of the transition function was proposed by collaborators at the International School for Advanced Studies (SISSA), Dimitar Ivanov, Stefano Liberati and Matteo Viel.
- Clarifying the derivation of luminosity distance formula that arises from the proposed model as well as making the functional form appropriate to use with our SNIa data was done by me.
- Implementation of the model within BAHAMAS, applying it to SNIa data and the resultant numerical analysis and discussion are my own.

Chapter 7

- Based on a paper prep
- Appropriate parameterizations of the selection function was first explored by H. Shariff and H. Z. Chen.
- The initial methodology, implementation and the inclusion of errors in the model was done by H. Z. Chen.
- My contribution included further refinements to the model as well as software engineering aspects of implementing the model within BAHAMAS.
- Testing the model ‘at scale’ on simple standard candle-like simulations, the resultant plots, numerical analysis and discussion are my own.

Publications

Publications that contribute to this Thesis

Rahman W., Trotta R., Boruah S. S., Hudson M. J., van Dyk D. A., 2021, *New Constraints on Anisotropic Expansion from Supernovae Type Ia*, Arxiv eprint: 2108.12497, *Submitted to MNRAS - Reviewed and awaiting minor revisions.*

Publications that do not contribute to this Thesis

Haworth T. J. , Clarke C. J., **Rahman W.**, Winter A. J., Facchini S., *The FRIED grid of mass loss rates for externally irradiated protoplanetary discs*, Arxiv eprint: 1808.07484, *Accepted at MNRAS*

Rahman W., Xu Y., Pu F., Xuan J., Jia X., Basios M., Kanthan L., Li L., Wu F., Xu B., *Clone Detection on Large Scala Codebases*, Arxiv eprint: 2204.04247, *Presented at proceedings of IWSC SANER 2020*

...for my mother and in loving memory of my father.

“Cosmologists are often in error, but never in doubt”

- *Lev Landau*

“Doubt is not a pleasant condition, but certainty is absurd”

- *Voltaire*

Contents

Abstract	i
Acknowledgements	iii
Declaration	iii
Publications	vi
1 Introduction	1
1.1 Motivations and Objectives	1
1.2 Outline of Thesis	3
2 Cosmology Background	4
2.1 Hubble-Lemaître Law and the Cosmological Principle	4
2.2 The Friedmann Equations	7
2.2.1 Density Evolution	11
2.3 Distance Determinations	13
2.3.1 Luminosity Distance	14
2.3.2 Angular Diameter Distance	15
2.4 Observational Probes	16

2.4.1	The Cosmic Microwave Background	17
2.4.2	Baryon Acoustic Oscillations	21
3	Type Ia Supernovae	24
3.1	A Historical Overview of Supernovae	24
3.2	Supernovae Origins and Classification	26
3.2.1	Superluminous Supernovae	28
3.3	Core-Collapse Mechanism	29
3.4	Thermonuclear Runaway Reaction and Type Ia Supernova	30
3.4.1	Progenitor Scenarios	31
3.4.2	Peculiar Type Ia Supernovae	32
3.4.3	Type Ia Supernova Rates	33
3.5	Type Ia Supernovae Cosmology	34
3.6	Light Curve Fitting	34
3.7	Concordance Cosmology	39
3.7.1	Precision Cosmology and Systematic Uncertainties	42
3.7.2	Future Surveys	49
4	Bayesian Statistics	53
4.1	Bayesian Analysis	54
4.1.1	Interpretations of Probability	55
4.1.2	Bayes' Theorem	56
4.1.3	Priors	59
4.1.4	Parameter Inference	62

4.1.5	Bayesian Evidence and Model Comparison	63
4.2	MCMC Algorithms and Nested Sampling	65
4.2.1	Nested Sampling	67
4.3	Bayesian Hierarchical Modelling of SNIa	70
4.3.1	The χ^2 Statistic	71
4.3.2	BAHAMAS	72
4.3.3	Comparison with the χ^2 method	75
5	New Constraints on Anisotropic Expansion from SNIa	78
5.1	Introduction	78
5.2	Methodology and Data	82
5.2.1	Cosmological model and dipole modulation	82
5.2.2	Bayesian Hierarchical Model	86
5.2.3	Data	87
5.2.4	Accounting for Colour-dependent Selection Effects	89
5.2.5	A New Derivation of Peculiar Velocity Corrections	96
5.2.6	Peculiar Velocity Correction Uncertainties	104
5.2.7	Choice of Priors	107
5.3	Simulations and Tests of Methodology	110
5.3.1	Simulated Data	110
5.3.2	Parameter Reconstruction from Simulations	115
5.3.3	Parameter Inference	118
5.3.4	Bayesian Model Comparison	124

5.4	Summary and Conclusions	132
6	A Cosmological Model for Anisotropic Expansion	134
6.1	A Simple Model for Anisotropic Expansion	135
6.2	Constraints from JLA SNIa data	143
6.2.1	Overview of Statistical Model	143
6.2.2	Tests on Simulated Data	146
6.2.3	Parameters Constraints and Model Selection from the JLA Data	148
6.3	Summary and Conclusions	151
7	Selection Effects	152
7.1	Selection Effects Review	152
7.1.1	Previous Approaches to Handling Selection Effects	154
7.2	A Bayesian Formalism for Selection Effects	156
7.3	Applications to the SNIa-like Standard Candles	163
7.3.1	Known Selection Function and No Measurement Error	163
7.3.2	Including Measurement Noise	172
7.4	Summary and Conclusion	182
8	Conclusion	184
8.1	Summary of Thesis Achievements	184
8.2	Future Work	186
A	Derivation and Test of Method of Moments	188

B Selection Effects for SNIa	194
B.1 Simulations of Standard Candles	194
B.2 Derivation of the Observed Data Likelihood	194
B.3 Connection Between our Result and R15	199
B.4 Including Phillips' Corrections	201
 Bibliography	 202

List of Tables

2.1	Evolution of the density terms in the Friedmann equations	11
3.1	Summary of the classification criteria for SNe subtypes	28
4.1	Jeffrey’s scale for evaluating the Bayes Factor	64
5.1	Supernovae parameters and prior choices for dipole inference	111
5.2	Assessment of bias in parameter recovery from simulations for the dipole models	116
5.3	Parameter constraints from the 1D marginal posterior distribution for the various dipole models considered on the JLA dataset	121
5.4	Comparison of our dipole model results with that of previous searches	127
6.1	Population parameters assumed in BAHAMAS for the anisotropic model and their priors	147
6.2	Assessment of bias on parameter recovery on simulated datasets for the isotropic model	148
7.1	Assessment of mean bias and mean squared error over 100 realisations of simu- lated data with measurement error for an unknown selection function	183

List of Figures

2.1	Hubble’s 1829 distance/velocity diagram for extragalactic nebulae	6
2.2	The CMB Temperature-Temperature angular power spectrum from Planck . . .	19
2.3	The CMB temperature map from Planck	20
2.4	Power spectrum from BOSS for BAO’s	23
3.1	The remnants of Tycho’s ‘De Stella Nova’ or SN1572 as seen in X-ray light . . .	25
3.2	Spectra of the four major subtypes of SNe	27
3.3	Best fit Type Ia Supernovae rates	33
3.4	Example light curves showing before after applying Phillips corrections	35
3.5	Example SNIa light curve that was fit by SALT2	36
3.6	Cosmological constraints from the Pantheon sample	40
3.7	Constraints on an evolving w CDM model by Pantheon	41
3.8	Comparison of Hubble measurements at early and late times	43
3.9	Hubble diagram from the Pantheon sample of SNIa	44
3.10	Summary of sources of SNIa taken from the decadal survey whitepaper	52
4.1	The DAG for BAHAMAS	74
5.1	Mollweide plot of the JLA data	88

5.2	Forward simulated JLA-like data with colour selection effects.	97
5.3	Comparison of CMB restframe redshifts derived in this work compared to original JLA redshifts	102
5.4	Comparison of the rederived peculiar velocities used in this work	103
5.5	Evolution of peculiar velocity statistical and systematic uncertainties	108
5.6	Example of forward simulated JLA-like data with a local dipole	113
5.7	Example of forward simulated JLA-like data with a dipole constant in redshift	113
5.8	Posterior plots of the dipole models on simulated data	117
5.9	Impact of peculiar velocity treatment on cosmological parameters	120
5.10	Posteriors of the dipole models on the improved treatment of the JLA data with Λ CDM	122
5.11	Posteriors of the dipole models on the improved treatment of the JLA data with a Cosmographic expansion	125
5.12	Posteriors of the dipole models on the improved treatment of the JLA data with a Cosmographic expansion and dipole on the deceleration parameter	126
5.13	Distribution of Bayes Factors for our simulated data sets	131
6.1	Diagram showing the difference between how b is measured in galactic coordinates compared to θ in spherical coordinates.	139
6.2	Chart detailing the evolution of $\epsilon(t)$ with time	140
6.3	Chart detailing the evolution of $\epsilon(z)$ with redshift	144
6.4	The evolution of the luminosity distance under this anisotropic model	145
6.5	Posterior inference on simulated data for the anisotropic model	149
6.6	Results for the JLA when using the anisotropic model for parameter inference	150

7.1	Simulated dataset of 5000 SNIa-like standard candles with selection effects applied and no measurement noise	167
7.2	Distribution of observed and missed data with no measurement noise and the ad hoc correction	168
7.3	Posterior of the mean inference over 100 realisations of simulated data with no measurement error and selection effects and known selection function	170
7.4	Posterior of the mean inference over 100 realisations of simulated data with no measurement error and selection effects but using a biased selection function . .	171
7.5	Posterior of the mean inference over 100 realisations of simulated data with no measurement error and the selection function is unknown	173
7.6	Simulated dataset of 5000 SNIa-like standard candles with selection effects applied and with measurement noise	175
7.7	Distribution of observed and missed data with measurement noise and the ad hoc correction	176
7.8	Posterior of the mean inference over 100 realisations of simulated data with measurement error and the selection function is unknown	181
A.1	Reconstructions of the selection function for various values of c_{sj}^{obs} and σ_{sj}^{obs} . . .	192
A.2	Posterior marginal distributions on simulated data with colour based selection effects	193
B.1	Simulated dataset of 5000 SNIa-like standard candles with selection effects applied and no measurement noise	195
B.2	Simulated dataset of 5000 SNIa-like standard candles with selection effects applied and with measurement noise	196

Chapter 1

Introduction

1.1 Motivations and Objectives

The Nobel Prize winning discovery of the accelerated expansion of the Universe in the late 1990's by [Riess et al. \(1998\)](#); [Perlmutter et al. \(1999\)](#) is considered one of the monumental leaps of our understanding of Cosmology. The critical instrument of this discovery was the Type Ia Supernova (SNIa), one of the Universe's brightest firework, formed during the violent dying throws of a White Dwarf.

The discovery of this accelerated expansion has caused us to rework our understanding of the universe, but many open questions remain and heated debates take place. Is the universe composed primarily of a 'dark energy' that drives this expansion? Or is our understanding of General Relativity inadequate and alternate modified theories of Gravity need to be proposed? What is the cause of the so-called 'Hubble Tension' between measurements of the Hubble-Lemaître constant by SNIa and the Cosmic Microwave Background? In addition, some observations call into question the foundational assumptions of our Cosmology, the Cosmological Principle (the notion that the universe is homogeneous and isotropic at sufficiently large scales) such as the large-scale alignment of quasars out to Gpc scales ([Friday et al., 2022](#)) and local bulk flows of galaxies that do not appear to decay at large distances ([Boruah et al., 2020b](#)) as they are expected to (a matter we return to in Chapter 5).

SNIa have matured as cosmological probes from when [Riess et al. \(1998\)](#); [Perlmutter et al. \(1999\)](#) discovered the accelerated expansion of the Universe. At that time, they made use of only a paltry (by today's standards) ~ 50 SNIa. As the number of detected SNIa rose to the ~ 1000 we have measured today, the sophistication of the machinery used to analyse and measure them have increased to take into account the growing importance of systematic uncertainties in SNIa cosmology, causing the current period to be colloquially called the era of 'precision cosmology'.

However, in the next decade, new surveys are being planned, with facilities such as the Large Synoptic Survey Telescope (LSST) ([Ivezić et al., 2019b](#)), the recently launched James Webb Space Telescope (JWST) ([Gardner et al., 2006](#)), Wide-Field Infrared Survey Telescope (WFIRST) ([Spergel et al., 2015](#)) and more. One of the common goals of these facilities is to gather data on Supernovae that is several orders of magnitudes larger than what is currently available to the Astrophysics community. We currently stand on the precipice of discovery once again, as with this boost in statistics provided by these larger datasets, we will gain the ability to answer some of the fundamental questions about our Universe and reconcile observational discrepancies. This new increase in data will pose fresh challenges however. To take advantage of this massive influx of data, appropriate tools need to be developed to handle the data including methods to model the specific systematics of them in a coherent and principled way.

During this transition to the 'big data' era of Cosmology, Bayesian methods and tooling have become increasingly popular, both in SNIa analysis and Cosmology as a field ([Trotta, 2008](#); [Hobson et al., 2009](#)). Useful features associated with them, include the ability to handle the systematics of data in a principled manner ([Kelly, 2007](#)) and [March et al. \(2011\)](#) shows that traditional non-Bayesian methods can introduce biases in SNIa cosmology analysis that are more easily resolved within a Bayesian framework. Methods in Bayesian statistics therefore provide a useful tool in handling the problems associated with the next generation of surveys and are employed heavily in this thesis.

The objectives of the thesis can be split into two components. The first objective of this thesis is to contribute to these open questions on our Cosmology by exploring whether potential

deviations from our fiducial Lambda Cold Dark Matter (Λ CDM) cosmology can be detected using SNIa and whether alternative cosmological models that allow for anisotropies can be used to describe our Universe. The second objective is to help build the Bayesian statistical tooling that will allow us to analyse the data from the next generation of surveys to unravel the mysteries of the Universe.

1.2 Outline of Thesis

To elucidate the path taken in answering the aforementioned objectives, this thesis is structured in the following manner. In Chapter 2, I provide the historical backdrop to our cosmology as well outline the necessary background formalism required to understand the cosmological models discussed throughout the thesis. In Chapter 3, I discuss SNIa and the instrumental role they play in constraining cosmological parameters and determining the accelerated expansion of the Universe. With the required cosmological background defined, in Chapter 4, I go on to discuss the background statistical formalism that is required to understand our tool of choice, Bayesian Hierarchical Models, and how we can use them in the context of Cosmology to constrain cosmological parameters with SNIa.

I then move to on to discuss the research contributions of this thesis. In Chapter 5 I explore whether any signatures of anisotropy exist in the distance modulus of SNIa within the context of our Λ CDM cosmology and the deceleration parameter in a Cosmographic expansion. I then go on to consider a phenomenological model that does not assume isotropy and analyse whether the anisotropy parameters in this model can be constrained with SNIa and whether this model can be discounted with respect to our fiducial cosmology using Bayesian model comparison. In Chapter 7, I outline a fully Bayesian formalism for handling the selection effects of SNIa data caused by brightness limited observational effects. I test this model on simulated data and show how it is easily extended to a more realistic complex model for real data to make it compatible with the future of SNIa surveys.

Chapter 2

Cosmology Background

Before we begin discussing aspects of Type Ia Supernovae, and the statistical methods we use in conjunction with them, it is necessary to first provide a brief history and introductory prescription of Cosmology and related equations we will need when analysing them which we provide in the following sections. For a more complete historical account, see e.g. [Rowan-Robinson \(1985\)](#); [Mitton \(2020\)](#). The majority of the material presented here also uses [Liddle \(1999\)](#); [Peacock \(1999\)](#); [Dodelson \(2021\)](#); [Ryden \(2016\)](#) as a source as well. We discuss at the end, the major historical probes of cosmological parameters (not including SNIa).

2.1 Hubble-Lemaître Law and the Cosmological Principle

Einstein’s seminal work in 1915 on the Theory of General Relativity ([Einstein, 1915](#)) forms the foundation of modern Cosmology. The ideas contained therein, have lead to the currently widely accepted ‘Cosmological Principle’ which states that on sufficiently large scales, the Universe is isotropic (looks the same in all directions) and homogeneous (looks the same at each point), meaning that we occupy no special place in the Universe at these scales. This represent a large paradigm shift from the anthropocentric principle first used by ancient Greeks when Alexander Ptolemy first posited that the Earth lay at the center of the Universe. Harlow Shapley was the first to discover we lay approximately two-thirds of the way from the galactic center by

measuring the distribution of globular clusters (Shapley, 1918), but he still believed the Milky was the center of the Universe and the only galaxy in our universe. The latter statement was disproven when Hubble observed Andromeda which was another galaxy outside our own in 1923, work which he later presented at the American Astronomical Society (Hubble, 1925) and corroborated the earlier result of Ernst Öpik in 1922 (who measured its distance at 450 kPc compared to Hubble’s measurement of 285 kPc (Opik, 1922)). It was not until 1952 Walter Baade was able to present his work done with the 200 inch Hale telescope at the International Astronomical Union to show the Milky Way was a fairly typical galaxy, not especially dissimilar from others (Baade, 1948, 1952, 1956), indicating even our own galaxy was not special in the Universe. After this, very few proponents arguing against the Cosmological Principle were left.

There remained a debate over the preferred cosmological model, with the two main competing models being the present day Big Bang Theory of Cosmology and a Steady State Cosmology. The Steady State cosmology put forward by Hoyle (1948); Bondi & Gold (1948) in 1948 proposed that density of the universe remains unchanged throughout time, and sought to explain the expansion of the Universe through the continuous creation of matter as the Universe expanded. This was in contrast to the Big Bang theory of cosmology which stated the Universe had a highly dense and hot beginning relative to today - notions of which go as far back as to Lemaître who spoke of a “Primeval Atom” in 1931 (Kragh, 2015). This debate would not be settled until the discovery of the Cosmic Microwave Background (CMB) which we discuss later in Section 2.4.1.

We focus however, on a contemporary of Shapley, Edwin Hubble, who was working on measuring the velocities and distances of distant objects in 1929 (Hubble, 1929). It was during this period, Hubble published a paper describing the relation between the radial distance of extragalactic ‘nebulae’ and their velocity. It would turn out later these nebulae were galaxies themselves. The figure he created can be seen in Fig. 2.1. This provided a key piece of observational evidence that the Universe was expanding. Though, it would be improper to solely attribute this discovery to Hubble as the work was built on research that came before. Slipher (1913, 1915, 1917) had already measured the radial velocity of a large number of nebulae and noted the majority of them were moving away from us and these measurements were also used by

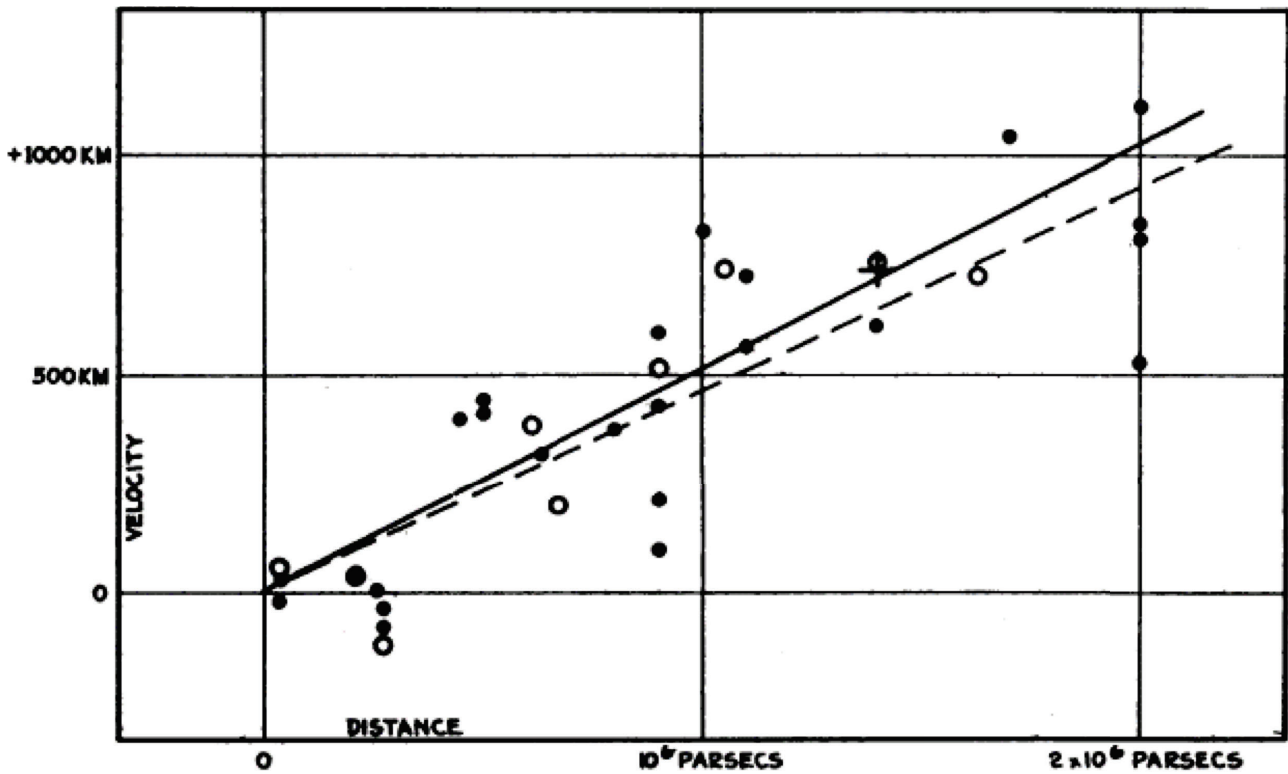


Figure 2.1: Figure 1 from [Hubble \(1929\)](#) showing the linear relationship between distance and velocity for extragalactic ‘nebulae’.

Hubble, [Einstein \(1917\)](#) and [de Sitter \(1917\)](#) suggested that the Universe could be described by a set of relativistic field equations proposed by [Einstein \(1915\)](#). [Friedmann \(1922\)](#) presented solutions of the field equations from [Einstein \(1915\)](#), showing that the solutions supported a dynamical model. Lemaître later showed these solutions supported an expanding Universe in 1927 (his work was translated to English from French later in 1931 ([Lemaître, 1931](#))). This was in opposition to the then favoured static model of the Universe by Einstein.

Hubble’s observations were able to confirm the theory of an expanding universe made by Lemaître, though he himself passed no judgement, preferring to leave the interpretation of his observations to those “competent to discuss the matter with authority” as he wrote in a letter to de Sitter [Nussbaumer & Bieri \(2011\)](#). His observations involved measuring distances using Cepheid variables wherein he noted that everything further than 300 kPc was moving away from us. Moreso, the further away an object was, the greater the recession velocity (inferred using its redshift).

The relation between radial distance and recession velocity can be simply stated as

$$v_{obs} = H_0 r + v_{pec} \quad (2.1)$$

where v_{obs} is the recession velocity, r is the radial distance and H_0 is the present day value of Hubble-Lemaître constant. From Fig. 2.1, it can be seen the linear relationship is not exact as the objects have a separate velocity component, independent of that implied by the expansion of the Universe, so a ‘peculiar’ velocity term, v_{pec} is added to account for this, which is independent from the dispersion that exists due to measurement noise.

To measure the distances, Hubble was able take advantage of the periodicity of the luminosities of Cepheid variables, a relation which was first noticed by Leavitt (Leavitt, 1908; Leavitt & Pickering, 1912), making them the first standardisable candles in use to measure distances. Hubble could then make use of their apparent observed magnitude to infer their distances (more on distances in Section 2.3). The range of objects Hubble looked at had a distance of less than 2Mpc, but it can be shown this linear relationship holds up 100Mpc (Mukhanov, 2005). At further distances than this, peculiar velocities are negligible, however the linear relationship also starts to break down. It should be noted that H_0 is the value of the Hubble-Lemaître constant today and it is often written as $H_0 = 100h\text{kms}^{-1}\text{Mpc}^{-1}$ where $h \sim 0.7$.

2.2 The Friedmann Equations

One of the cornerstones of modern Cosmology are the Friedmann equations from which a cosmological model that supports an expanding Universe can be derived. In this section, I will briefly outline the general formalism and physics required to derive the Friedmann equations.

At the core of it, most of the modern underpinning of Cosmology is based around Einstein’s field equations from General Relativity (Einstein, 1915) which connects the spatial curvature of our universe with the energy content and can be stated as

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = 8\pi GT_{\mu\nu} + \Lambda g_{\mu\nu} \quad (2.2)$$

Here $R_{\mu\nu}$ is the Ricci curvature tensor, $g_{\mu\nu}$ is the metric tensor, R is the Ricci scalar tensor, G is Newton's gravitational constant and $T_{\mu\nu}$ is the stress-energy tensor. This equation (and all following equations in this chapter) is given in natural units where the speed of light $c = 1$. The left hand side in Eq. (2.2), often called the Einstein tensor, $G_{\mu\nu}$, represents the spatial curvature of our Universe. The right hand side represents the energy content of the Universe. The Ricci curvature tensor and Ricci scalar tensor defining the Einstein tensor can also both be calculated as a function of the metric $g_{\mu\nu}$. Therefore, $G_{\mu\nu}$ can be specified completely through the metric. The stress-energy tensor for a perfect fluid of pressure, p and density, ρ , with 4-velocity u^μ may be written as

$$T_{\mu\nu} = (\rho + p)u^\mu u^\nu + pg^{\mu\nu}. \quad (2.3)$$

A note here on Λ which is the cosmological constant and was added by Einstein to allow for a static solution by altering the dynamics to counteract the effects of gravity. After the advent of Hubble's observation that showed the Universe was expanding, this usage was no longer common. However, after the discovery that the expansion of the Universe is accelerating (Riess et al., 1998; Perlmutter et al., 1999), which could not be accounted for by current known densities, Λ was reinstated as a means of explaining this acceleration as a positive value of Λ corresponds to an acceleration in the expansion. Its inclusion is done usually with an extra energy density within the stress-energy tensor in Eq. (2.3).

To solve Einstein's field equation, one has to specify the metric tensor $g_{\mu\nu}$ which describes the local geometric structure of space-time. For an unperturbed expanding isotropic Friedmann-Lemaître-Robertson-Walker (FLRW) universe, it is defined under cartesian coordinates as

$$g_{\mu\nu} = \text{diag}(-1, a(t), a(t), a(t)) \quad (2.4)$$

for which we adopt the sign convention $(-, +, +, +)$ in $g_{\mu\nu}$. Here, $a(t)$ is the scale factor of expansion defined such that we can define the physical radial distance $r(t)$ and the comoving

distance, χ , that is independent of the expansion of the universe with the formula

$$r(t) = a(t)\chi. \quad (2.5)$$

For an expanding Universe, $a(t)$ increases with time. For use further on, alongside the comoving distance we also define the comoving time, η as

$$t = a(t)\eta. \quad (2.6)$$

The Hubble parameter can also be expressed in terms of this scale factor using

$$H(t) = \frac{\dot{a}(t)}{a(t)}. \quad (2.7)$$

With the simplifying assumptions of isotropy and homogeneity over large distances, the FLRW distance metric ([Friedmann, 1922, 1924](#); [Lemaître, 1931, 1933](#); [Robertson, 1935](#); [Walker, 1937](#)) can then be defined in cartesian coordinates as

$$ds^2 = g_{\mu\nu}dx^\mu dx^\nu \quad (2.8)$$

where ds is a spacetime interval. Under spherical polars, Eq. (2.8) and Eq. (2.4) reduces to the line element

$$ds^2 = -dt^2 + a(t)^2 \left(\frac{dr^2}{1 - kr^2} + r^2 d\theta^2 + r^2 \sin^2\theta d\phi^2 \right). \quad (2.9)$$

This is the FLRW metric where $a(t)$ is the scale factor of expansion and k is the spatial curvature defined as $+\frac{|a|^2}{a^2}$, $-\frac{|a|^2}{a^2}$, or 0 meaning it can take values +1, -1 or 0 which correspond to closed, open and flat Universes respectively. Depending on the curvature this causes the comoving

distance to be rewritten as

$$r_\chi = S_k(\chi) = \begin{cases} \sin \chi, & k = +1 \\ \chi, & k = 0 \\ \sinh \chi, & k = -1 \end{cases} \quad (2.10)$$

Solutions to Einstein's field equation are based around the first and second derivatives of the scale factor, known as the Friedmann equations, and are given by

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{k}{a^2} + \frac{\Lambda}{3} \quad (2.11)$$

,

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3p) + \frac{\Lambda}{3} \quad (2.12)$$

To arrive at this solution we make the assumption that the energy content of the Universe can be modelled as perfect fluid so that the $T_{\mu\nu}$ takes the form given in Eq. (2.3). ρ has two primary contributors, the radiation density from relativistic particles, ρ_γ , and the matter density from pressureless non-relativistic matter, ρ_m . However, it is common to include the term related to the cosmological constant Λ as a vacuum energy contribution, ρ_Λ , as stated before and is often called the dark energy density. The curvature term is also included as a density parameter with ρ_k . Therefore we can define the total energy density of the Universe as

$$\rho(t) = \rho_\gamma(t) + \rho_m(t) + \rho_\Lambda(t) + \rho_k(t). \quad (2.13)$$

These are all related to the pressure p via the equation of state

$$p_i = w\rho_i \quad (2.14)$$

where i is a subscript denoting the relevant component.

2.2.1 Density Evolution

We can derive the evolution of the density contributions in Eq. (2.13) by first deriving the conservation equation from the two Friedmann equations Eq. (2.11) and Eq. (2.12) as

$$\dot{\rho}_i + 3H(\rho_i + p_i) = 0. \quad (2.15)$$

By substituting Eq. (2.14) into this, we arrive at the relativistic continuity equation (Dodelson, 2003), sometimes called the fluid equation, as

$$\dot{\rho}_i + 3\frac{\dot{a}}{a}(1 + w)\rho_i = 0. \quad (2.16)$$

Solving this equation leads to the solution

$$\rho_i = \rho_{i,0}a(t)^{-3(1+w)} \quad (2.17)$$

where $\rho_{i,0}$ is the value of the energy density today. We can therefore determine the evolution of the density parameters given the equation of state parameter w . The functional form of these evolutions is given in Table. 2.1.

Table 2.1: The evolution of the different density terms in our universe with time and the different values of w in the equation of state for each density contribution in ρ

Density Term	$\rho(t)$	w
$\rho_m(t)$	$\propto a^{-3}$	0
$\rho_\gamma(t)$	$\propto a^{-4}$	1/3
$\rho_\Lambda(t)$	$\propto a$	-1
$\rho_k(t)$	$\propto a^{-2}$	-1/3

The evolution of the non-relativistic matter density term ρ_m as an inverse cubic is easily understood from the classical Newtonian viewpoint on how density evolves with volume. For the radiation term ρ_γ however, there is an extra factor of a^{-1} to account for redshifting as spacetime expands. A common assumption for density contribution of ρ_Λ is that it is constant in time with $w = -1$. It is easy to see from the first Friedmann equation why the curvature density ρ_k

has an a^{-2} dependence meaning it is defined by the equation of state parameter $w = -1/3$ and scales at a rate slower than matter and radiation.

From Eq. (2.11), if we assume a flat Universe ($k=0$) and a Universe with no cosmological constant ($\Lambda = 0$) we can derive a critical density, ρ_c , such that

$$\rho_c = \frac{3H^2}{8\pi G}. \quad (2.18)$$

This value can be numerically evaluated as $1.8788h^2 \times 10^{-30} \text{kg/m}^3$ and denotes the value that makes our universe spatially flat. The uniqueness of this number is more easily seen if we recast the units in terms of solar masses, M_\odot , and Mpc so that it becomes

$$\rho_c = 2.78h^2 \times 10^{11} M_\odot/\text{Mpc}^3 \quad (2.19)$$

which is close to the current density of galaxies and our universe when considering most galaxies are typically separated by around $\sim 1\text{Mpc}$ and usually contain of the order $10^{11} - 10^{12}$ stars which have a mass of roughly $1M_\odot$ per star. We can then define a density parameter, Ω_i , using the different density components in ρ as a ratio of this critical density such that

$$\Omega_i = \frac{\rho_i}{\rho_c} \quad (2.20)$$

and

$$\Omega_{i,0} = \frac{\rho_{i,0}}{\rho_c} \quad (2.21)$$

being the value of this density parameter today at time t_0

This naturally leads to the condition that

$$\Omega_m + \Omega_\gamma + \Omega_\Lambda + \Omega_k = 1. \quad (2.22)$$

From here we can now write the first Friedmann equation in terms of these critical density so

that

$$H^2(t) = H_0^2 [\Omega_{m,0}a(t)^{-3} + \Omega_{\gamma,0}a^{-4} + \Omega_{\Lambda,0} + \Omega_{k,0}a^{-2}]. \quad (2.23)$$

2.3 Distance Determinations

One critical aspect of this thesis, is the ability to measure distances accurately in order to investigate the late-time acceleration the Universe is undergoing. While in the 'local' Universe, the method of parallaxes and Cepheid variables are used, for objects very far away (> 50 Mpc), this is not possible. There are multiple ways to measure the distance to an object at these distances. We discuss two methods below that are instrumental in the generation of 'standard candles' and 'standard rulers' that have theoretical underpinnings dependent on our cosmological parameters and therefore allow us to constrain cosmological parameters using observations. To define these methods, we must first understand how photons propagate through spacetime.

In curved spacetime, freely falling photons travel along null geodesics which are defined by the line element

$$ds^2 = 0 \quad (2.24)$$

The assumption of isotropy and homogeneity means we can assume light travels only radially which implies

$$dr^2 = \frac{dt^2}{a^2(t)}. \quad (2.25)$$

We can therefore define the comoving distance χ from equation Eq. (2.5) as

$$\chi = \int_{t_e}^{t_o} \frac{dt}{a(t)} \quad (2.26)$$

where t_e is the time the photon was emitted and t_o is the time the photon was observed. However, the wavelength of the light is stretched as the Universe expands. This effect is encapsulated in the definition of the redshift z where

$$z = \frac{\lambda_o}{\lambda_e} - 1. \quad (2.27)$$

Here λ_e is the wavelength of the photon when its emitted at the source and λ_o the wavelength when it is observed. The expansion means that objects further away are reddened, hence the term redshift. Redshift is related to the scale factor $a(t)$ by the relation

$$1 + z = \frac{a(t_0)}{a(t)} \quad (2.28)$$

where t_0 is the time today. Given that we usually define the scale factor such that $a(t_0) = 1$ this reduces the equation to

$$a(t) = \frac{1}{1 + z}. \quad (2.29)$$

We can therefore write the comoving distance as

$$\chi = \frac{1}{a_0} \int_0^z \frac{dz'}{H(z')} \quad (2.30)$$

where

$$H(z)^2 = H_0^2 E^2(z) \quad (2.31)$$

and

$$E^2(z) = [\Omega_{m,0}(1+z)^3 + \Omega_{\gamma,0}(1+z)^4 + \Omega_{\Lambda,0} + \Omega_{k,0}(1+z)^2] \quad (2.32)$$

2.3.1 Luminosity Distance

One of the most importance distance measures that I use in this thesis is the luminosity distance, D_L , which is defined as

$$D_L^2 = \frac{L}{4\pi F} \quad (2.33)$$

where D_L is the luminosity distance measured in Mpc, L is the absolute luminosity of the source object and F is the *observed* flux. This can also be written as a function of our cosmological parameters with the equation

$$D_L = \frac{1+z}{H_0 \sqrt{|\Omega_k|}} \text{sinn} \left\{ \sqrt{|\Omega_k|} \int_0^z \frac{dz}{E(z)} \right\} \quad (2.34)$$

where

$$\text{sinn}(x) = \begin{cases} \sin(x), & \Omega_k < 0 \\ x, & \Omega_k = 0 \\ \sinh(x), & \Omega_k > 0 \end{cases} \quad (2.35)$$

This can then be transformed into the distance modulus measure μ using the equation

$$\mu(z) = 5 \log \left[\frac{D_L}{1\text{Mpc}} \right] + 25 \quad (2.36)$$

where the divisor of 1 Mpc ensures the argument is also a dimensionless quantity. Observationally, the distance modulus is given by

$$\mu = m - M \quad (2.37)$$

where m is the apparent magnitude of the observed object and M is the absolute magnitude of the object (the apparent magnitude it would have if it was located at a distance of 10pc and there were no dimming caused by interstellar matter or dust).

For standardisable candles, the above is important in probing cosmological parameters as predictions of the distance modulus can be made using an assumed cosmology and then compared with the observed distance modulus. A key feature of standardisable candles such as Type Ia Supernovae is that their absolute magnitude can be determined to a high degree of precision and accuracy (discussed more in Chapter 3).

2.3.2 Angular Diameter Distance

One final distance measure we define is the angular diameter distance d_A which is defined as

$$D_A = \frac{l}{\theta} \quad (2.38)$$

which is when an object of known length l is subtended by an angle θ along the perpendicular line of sight. This is related to the comoving distance by

$$D_A(z) = \frac{S_k(\chi)}{1+z} \quad (2.39)$$

and the luminosity distance d_L by

$$D_L(z) = (1+z)^2 D_A(z). \quad (2.40)$$

Here it is clear to see that, unlike in Euclidean geometry, the angular diameter distance and luminosity distance are not the same, with the relationship between the two referred to as the distance duality relation or the Etherington reciprocity theorem ([Etherington, 1933](#)). As a complement to the luminosity distance which are used to define a ‘standard candle’, the angular diameter distance can be used to define ‘standard rulers’ with which we can constrain cosmological parameters (discussed further in [Section 2.4.1](#) and [2.4.2](#)).

2.4 Observational Probes

Constraining the density parameters in [Eq. \(2.23\)](#) is one of the primary focuses of modern day cosmology. For example, it is easy to show that a matter dominated Universe (Einstein de Sitter model) is unfeasible as this would then lead to an estimate of the age of the Universe as

$$t_0 = \int_0^\infty \frac{dz}{H(z)(1+z)} = \frac{2}{3} H_0^{-1}. \quad (2.41)$$

For an approximate value of $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$, this would imply the age of the Universe to be around $\sim 9 \text{ Gyrs}$. In the time where this model was popular, H_0 was believed to be $H_0 \sim 100 \text{ km s}^{-1} \text{ Mpc}^{-1}$ leading to an age of $\sim 7 \text{ Gyrs}$. Both are contradictory to the age implied by globular clusters though which are of the order 11 Gyrs ([VandenBerg et al., 2013](#)).

Considering the Universe is close to flat (and confirmed to a high degree by Planck ([Planck](#)

([Collaboration et al., 2018](#))) and we are in an era where the radiation density is negligible (it is only dominant at early times), this would imply a significant contribution from the dark energy density, Ω_Λ , when considering the cosmic sum rule in Eq. (2.22). In this section, we discuss the primary probes that allow us to constrain cosmological parameters which are the Cosmic Microwave Background (CMB) and Baryon Acoustic Oscillations (BAO). Complementary to these probes, SNIa were also instrumental in clarifying the status of the cosmological parameters and the non-zero cosmological constant value of Λ , but discussion of this is deferred to chapter 3 and we limit our discussion here to the CMB and BAO.

2.4.1 The Cosmic Microwave Background

While the debate between Big Bang cosmology and Steady State cosmology continued in the 20th century, as early as 1961, questions about the validity of the Steady State cosmology had been proposed when it was noticed counts of bright radio sources such as quasars or radio galaxies could only be observed in the distant past ([Ryle & Clarke, 1961](#)) which was at odds with Steady State cosmology that would presume some could be found in recent history too.

It was not until 1965 that the Steady State theory was definitely ruled out, when the Cosmic Microwave Background (CMB) predicted by the Big Bang was first detected. The CMB is composed of the first freely streaming photons to appear when they decoupled from the primordial plasma of the Universe. It represents the oldest part of the sky and was first detected accidentally by Penzias and Wilson ([Penzias & Wilson, 1965](#)). This discovery was crucial in favouring Big Bang Cosmology over the Steady State Cosmology.

From this discovery, Cosmologists were able to infer that the earth is bathed in a background microwave radiation that is approximately the same in all directions. The uniformity of this temperature in all directions is an indication of how causally connected everything was in the past, providing near decisive evidence in favour of the Big Bang theory over the Steady State theory.

In the following decades, these results were further confirmed by the ‘Cosmic Background

Explorer' (COBE) (Mather et al., 1994). From this mission, Cosmologists were able to accurately measure the temperature of the CMB background at $T \sim 2.752$ and saw that it was an almost perfect blackbody spectrum, but also saw small temperature anisotropies existed. Later missions attempted to measure these anisotropies such as 'Balloon Observations Of Millimetric Extragalactic Radiation and Geophysics' (BOOMERANG) (de Bernardis et al., 2000), the 'Millimeter Anisotropy Experiment Imaging Array' (MAXIMA) (Balbi et al., 2000), the 'Wilkinson Microwave Anisotropy Probe' (WMAP) (Hinshaw et al., 2009) and most recently Planck (Planck Collaboration et al., 2018). From these experiments, it was clearly evident that temperature anisotropies of the order 1 in 100,000 existed. These fluctuations are key to placing constraints on our current cosmology as their nature can be predicted from assumptions about early Universe physics and late time evolution to define 'standard rulers' that can be related to the angular diameter distance to constrain cosmological parameters.

To see this how this works, first let us assume the temperature of the CMB conforms to an isotropically radiating black body with temperature T . We can consider a perturbation of ΔT that creates a fluctuation $\Theta(\theta, \phi)$ corresponding to a direction (θ, ϕ) in the sky in spherical coordinates. We can expand this fluctuation in terms of spherical harmonics to write

$$\Theta(\theta, \phi) = \sum_{\ell=1}^{\infty} \sum_{m=-\ell}^{\ell} a_{\ell m} Y_{\ell}^m(\theta, \phi) \quad (2.42)$$

where $a_{\ell m}$ controls the scale of temperature anisotropies at different scales. What we are interested in to constrain cosmological parameters however, is not the value of $a_{\ell m}$, but the statistical properties of it, the radiation angular power spectrum which is defined as

$$C_{\ell} = \langle |a_{\ell m}|^2 \rangle. \quad (2.43)$$

Due to rotational invariance, this quantity therefore does not depend on the position (θ, ϕ) and thus cannot depend on the m index either, meaning it is purely a function of ℓ . The index ℓ defines the angular scale of the anisotropy with small ℓ corresponding to large angular scales and large ℓ for small scales. Typically the size of the anisotropy corresponds to $\sim \frac{180^\circ}{\ell}$

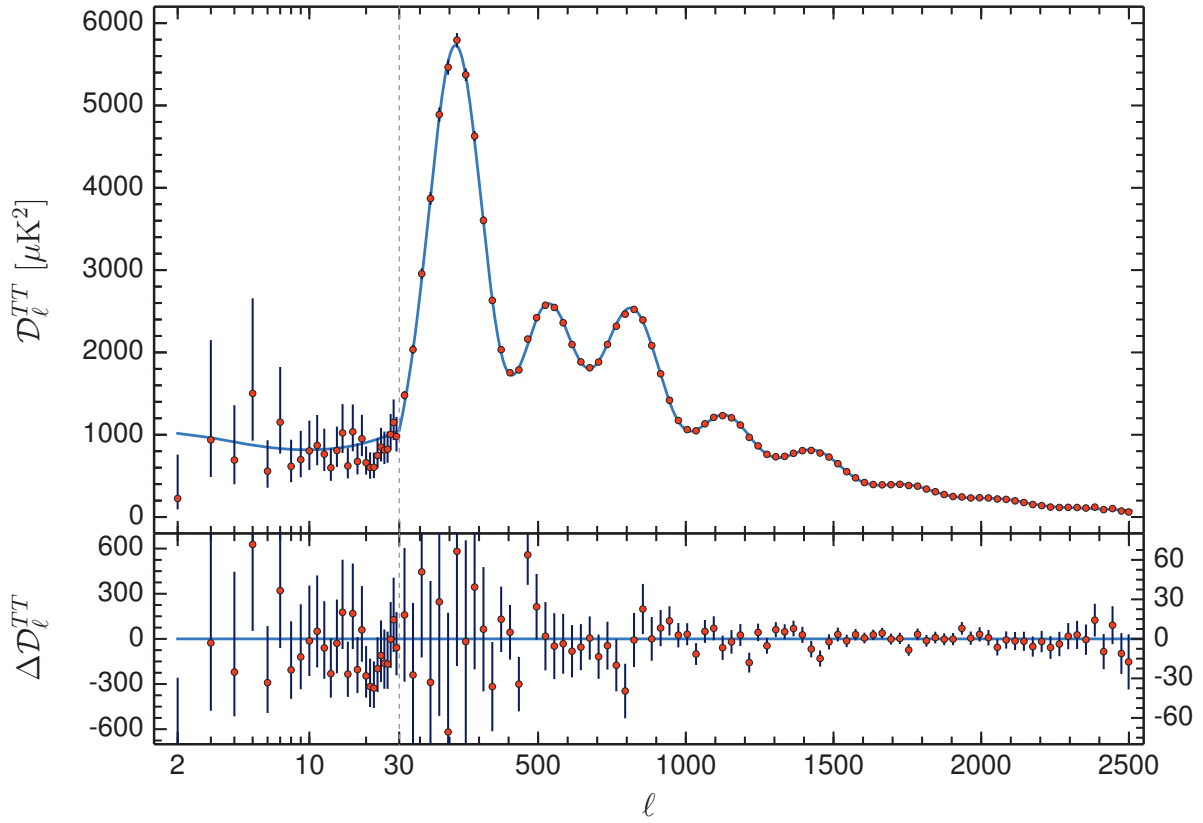


Figure 2.2: The CMB Temperature-Temperature angular power spectrum. The dipole associated with $\ell = 1$ is not plotted here. Theoretical predictions are in blue and observed data points in red. The bottom plot represents the residual error of the prediction with the observed. The unique x-axis scale here is such that the observed data points are intended to be roughly evenly distributed. Figure taken from [Planck Collaboration et al. \(2018\)](#)

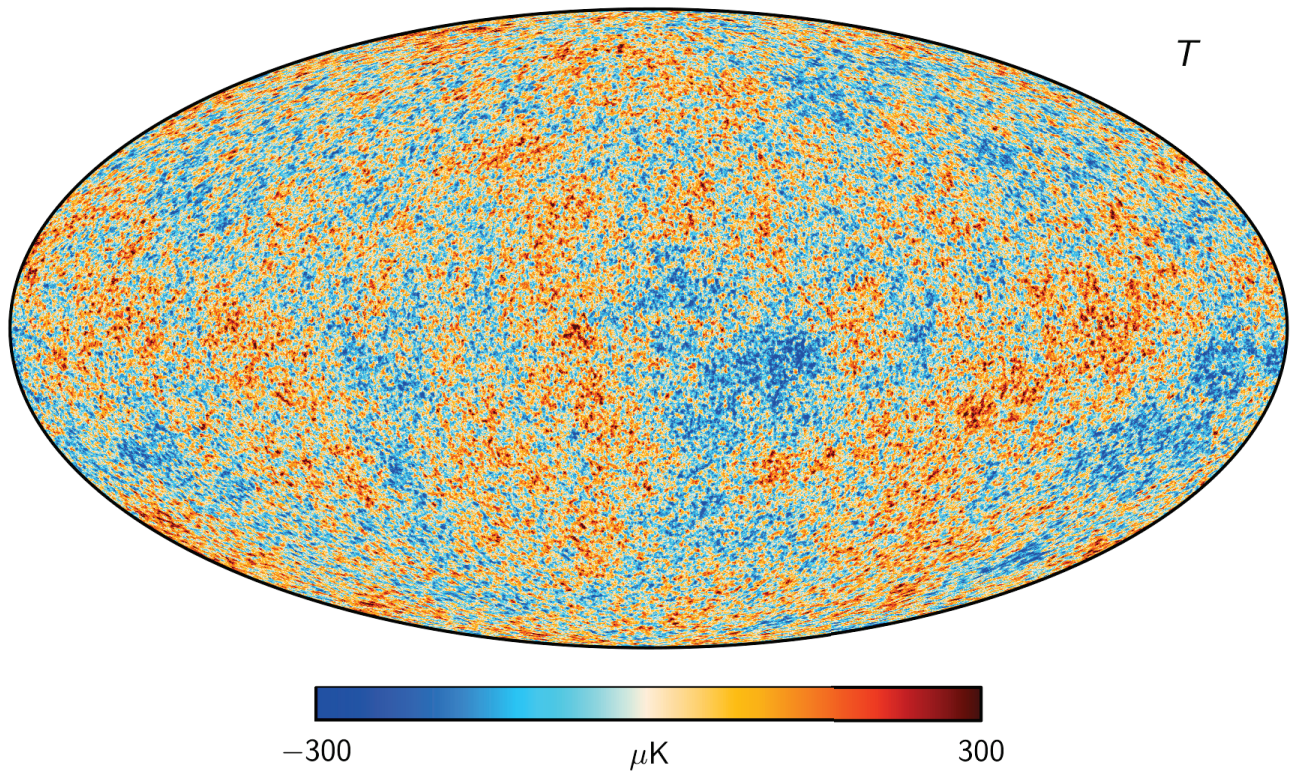


Figure 2.3: The CMB temperature map taken from ESA/Planck 2018 ([Planck Collaboration et al., 2018](#))

The resultant power spectrum determined by [Planck Collaboration et al. \(2018\)](#) can be seen in Figure 2.2. We can see on the very largest scales where $\ell < 15$, there is a non-zero value that is roughly constant across this range called the Sachs-Wolfe plateau and is caused by the variation in gravitational potential between regions.

The cosmic fluid would have constantly been undergoing periods of rarefaction and compression due to the combined effects of the photon pressure and gravitation that would either dampen or increase the anisotropies and thereby creating the oscillating peaks seen in the CMB power spectrum. The first peak at $\sim \ell = 200$ corresponds to an angular scale of approximately 1° . The anisotropies associated with these and higher modes is easily seen in the planck CMB map in Figure 2.3

These peaks in the power spectrum can be used to constrain cosmological parameters from their relation to the angular diameter distance which depends on cosmological parameters. For example, the baryon density can be derived from the ratio of the second and third peaks. Analysis by the Planck collaboration was able to determine that the Universe is nearly completely flat

with $\Omega_k = 0.0007 \pm 0.0019$ and that the matter density was $\Omega_m = 0.3147 \pm 0.0074$. The baryon density on the other hand was determined to have a value $\Omega_b = 0.0500 \pm 0.0003$, implying a significant fraction of the matter density cannot be explained by the baryon density alone. This unseen matter is commonly called ‘dark matter’ and is the subject of ongoing research.

2.4.2 Baryon Acoustic Oscillations

The distribution of the aforementioned unseen dark matter can be traced by looking at the large scale structure of the Universe. In other words, the distribution of galaxies and their clustering properties. One of the ways this is done is by defining the power spectrum from the galaxy two point correlation function.

The oscillation caused by gravitation and radiation pressure that caused the CMB anisotropies also gave rise to a second ‘standard ruler’ that could be used to constrain cosmological parameters. This oscillation created spherical acoustic waves that travelled at relativistic speeds ([Hu & White, 2004](#)). At the point of decoupling when the Universe had a temperature of approximately $T \sim 3000K$, these wave speeds were reduced to the point where the propagation of the spherical waves become ‘frozen in’, meaning their evolution from that point onwards was predominantly dictated by the expansion of the Universe and their effect can be seen in the large scale structure of the Universe ([Eisenstein & Hu, 1998](#)). Measuring the effect of these frozen in waves on the galaxy power spectrum provides a probe of the growth of their structure and these oscillations are referred to as Baryon Acoustic Oscillations (BAO). By relating the angular separation in the sky of these structures to the power spectrum and the angular diameter distance (which depends on our cosmology), cosmological parameters can be constrained. Therefore, analysing the galaxy distribution not only provides insight into the dark matter distribution but also provides an avenue with which to probe dark energy and constrain cosmological parameters.

The Sloan Digital Sky Survey (SDSS) is a multi-spectral imaging survey that has many phases with the most recent that continues to this day being SDSS-V. [Eisenstein et al. \(2005\)](#) was able to measure the BAO (with SDSS-I) using the measured spectroscopic redshifts of 46,748

luminous red galaxies (LRGs) over 3,816 square-degrees of sky and out to a redshift of approximately $z = 0.47$. They measured the BAO at having a size of approximately 150Mpc which was further confirmed by SDSS-III's Baryon Oscillation Spectroscopic Survey (BOSS) survey ([Anderson et al., 2012](#)) and the extended BOSS (eBOSS) survey of SDSS-IV ([Bautista et al., 2020](#)). The comparison of the results derived from the SDSS LRGs and BOSS CMASS galaxies can be seen in Fig. [2.4](#).

The standard ruler constraints from BAO's are not capable of breaking the degeneracy of the results derived from the CMB to constrain the cosmological parameters. Conclusive determination of the late-time acceleration of our universe required a third probe in the form of Type Ia Supernovae that came earlier in 1998. Their use in breaking the degeneracy further is discussed in Chapter [3](#).

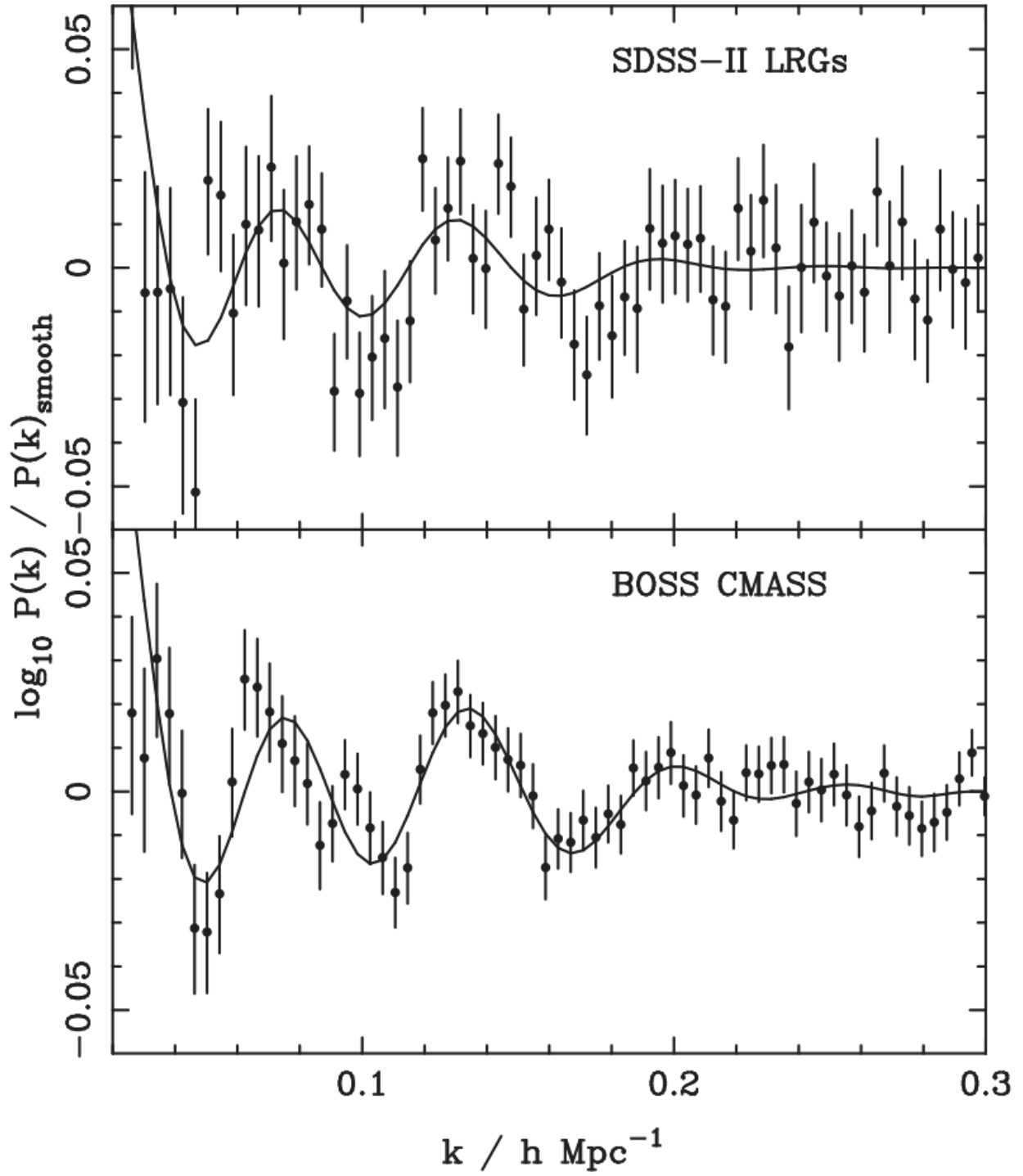


Figure 2.4: Here we show the power spectrum derived from the SDSS-II LRGs results (top) and compare them with the BOSS DR9 CMASS galaxies results (bottom). Solid lines show the best-fit models.

Chapter 3

Type Ia Supernovae

In this chapter, I discuss the third major probe of our Cosmology which are Type Ia Supernovae (SNIa) which were used to show the late-time acceleration of the Universe ([Riess et al., 1998](#); [Perlmutter et al., 1999](#)). This groundbreaking result which won the 2011 Nobel Prize was because of how SNIa can be standardised so that they only show small variation in their peak absolute magnitudes ($\sim 0.1\text{mag}$) meaning they can be used as ‘standard candles’. An overview of what Supernovae are, their origins and formation mechanisms, the standardisation procedure used to make them standard candles, and the constraints we place on our Cosmology with them are discussed in this chapter.

3.1 A Historical Overview of Supernovae

Supernovae (SNe) of all types are formed during the end of a stars life. Their high luminosity means they are easily detectable even when residing in galaxies outside our own with an instantaneous power output of the order of 10^{53}erg s^{-1} ([Woosley & Janka, 2005](#)). SNe are responsible for the creation of the heavy elements that compose our planet and our bodies. These transient objects in the sky last for up to several months before fading away leading to them often being called ‘Guest Stars’. [Joglekar et al. \(2011\)](#) claims they were first observed potentially as early as 4600BC based on carvings found in rock walls in the Burzhama region of Kashmir, but this

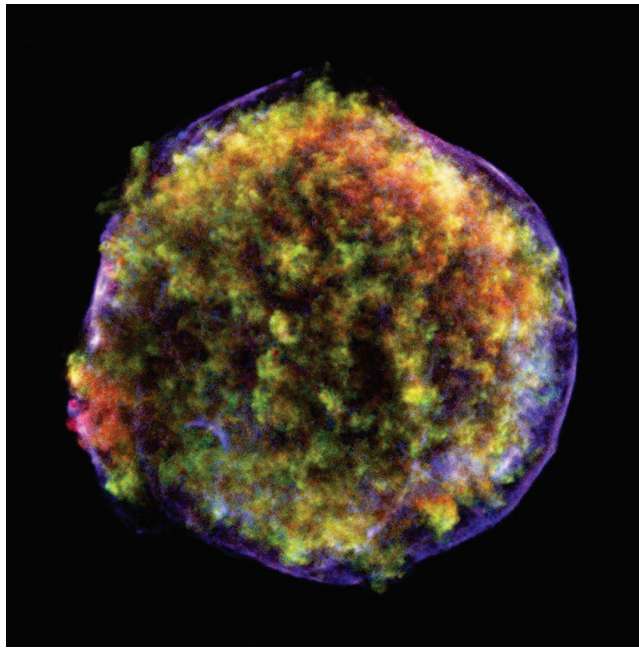


Figure 3.1: The remnants of Tycho's 'De Stella Nova' or SN1572 as seen in X-ray light. This was captured using the Chandra X-Ray Observatory Telescope (Image Credit: NASA/CXC/Rutgers/J.Warren & J.Hughes et al).

is still disputed. The first official recording that Astronomers agree on however, was not until 185AD with SN185 which was observed by Chinese Astronomers ([Zhao et al., 2006](#)). The name 'Nova' is taken from the Latin meaning 'new' and its usage in astronomical context can be traced as far back to the Danish Astronomer Tycho Brache in 1572 who used it to describe a 'new star' or 'De Stella Nova' SN1572 which appeared in Cassiopeia ([Decourchelle, 2017](#)). Despite having occurred a few centuries ago, its remnants can still be observed by modern day instruments as shown in Figure. 3.1. In modern astronomy however, the word 'nova' is attributed to the explosive expulsion of the outer hydrogen layers of a white dwarf causing the hydrogen to fuse into helium. Supernovae however, cause elements heavier than iron and nickel to be produced. Notably, after a nova, the white dwarf remains intact compared to a supernovae which leaves behind a neutron star remnant or a black hole. A third class of nova known as 'Helium Nova' have also been proposed and are characterised by the lack of hydrogen in the white dwarf ejecta ([Kato & Hachisu, 2003](#)).

Supernovae were first recognised as distinct objects by [Baade & Zwicky \(1934\)](#) who saw there was a more rare class of nova, which appeared to be on average 10 magnitudes (10000 times) greater than normal i.e., a 'Supernova' as they came to be known, where they believed the

cause was the transition of a normal size star to a neutron star. These objects were observed to be less frequent than standard novae.

Aside from SN185 and SN1572 mentioned previously, there exists only a handful of other historical examples of Supernovae within our galaxy, of which their location made them visible to the naked eye. The remnants of these supernovae have been instrumental into gaining insight into the various astrophysical processes that create supernovae and their impact on the surrounding environment. The brightest SNe observed, SN1006, was seen in the constellation of Lupus which was observed by Astrologers and Astronomers all over the world ([Winkler et al., 2003](#)) and reached an estimated V-band magnitude of -7.5 ([Katsuda, 2017](#)); The remnant of SN1054 is now known as the Crab Nebula ([Blandford & Bühler, 2017](#)); SN1181 also appeared in the constellation of Cassiopeia and it is thought the radio and x-ray pulsar ‘3C 58’ is a remnant of it ([Kothes, 2017](#)); 32 years after Tycho’s SN1572, SN1604 i.e., Kepler’s Supernovae, was observed and is the most recent galactic supernova according to historical records ([Vink, 2017](#)).

The last supernovae to be observed by the naked eye was SN1987A ([Ken’ichi & Shigeyama, 1988](#)), however, this event did not occur within the Milky Way, but in a satellite galaxy close by called the Large Magellanic Cloud located approximately 51.4 kPc from earth. Its uniqueness lay in the fact that aside from being observable by eye, it was also observed with modern instrumentation in all parts of the electromagnetic spectrum. Notably, a few hours before the visible light reached earth, neutrino emissions were detected at several observatories and this event was significant into gaining insight into the core-collapse mechanism that leaves behind a neutron star as the neutrino emission occurs before the generation of light in the visible spectrum ([Alexeyev et al., 1988](#)).

3.2 Supernovae Origins and Classification

Supernovae are thought to occur broadly via two main mechanisms, either via the Core-Collapse (CC) mechanism ([Janka, 2012](#)) or via the thermonuclear runaway reaction of an accreting white dwarf ([Nomoto et al., 1996](#)). A few short years after their discovery, [Minkowski \(1941\)](#)

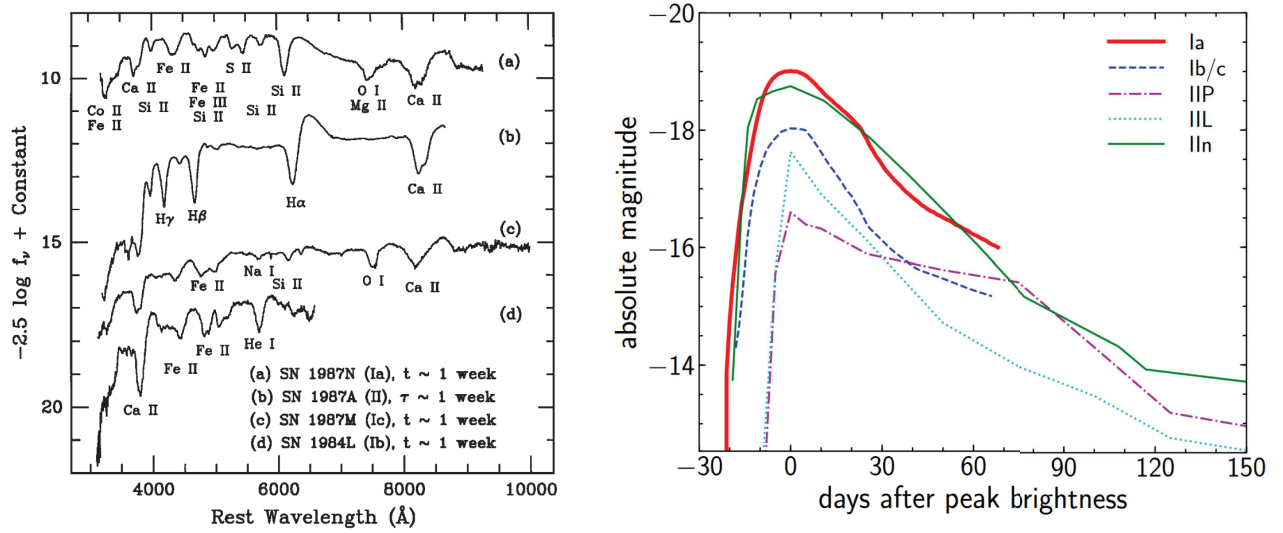


Figure 3.2: Left: Figure taken from [Filippenko \(1997\)](#) showing spectra of SNe, displaying distinctions between the four major types and subtypes. Here the variables t and τ represent time after observed B-band maximum and time after core collapse, respectively. The y axes here is essentially “AB magnitudes” as defined by [Oke & Gunn \(1983\)](#). Right: Peak absolute V-band magnitudes. Figure from [Oguri \(2019\)](#)

broadly classified them into two groups, Type I and Type II SNe based on whether they showed Hydrogen or not in their spectra. Over time, as supernova discovery and data quality improved, this classification was refined to include numerous sub types and their classification became based on their spectral appearance near maximum light ([Filippenko, 1997](#)). An example figure showing the various spectral features from different SNe subtypes as well as how their peak magnitude evolves is shown in Fig. 3.2. With the advent of new surveys such as from the Vera Rubin Observatory Legacy Survey of Space and Time (LSST) survey ([LSST Dark Energy Science Collaboration, 2012](#)), this classification scheme may become subject to change as richer data becomes available to better differentiate SNe.

A summary of the main types and the spectral features that define them are also provided in Table. 3.1.

It should be noted here that Type II SNe can sometimes be further subdivided into other classes based on qualitative features related to their light curve decay after peak. Type IIP SNe exhibit a plateau shortly after the peak that lasts for approximately 100 days. These are also the most common type of Type II SNe. Their progenitors are thought to be red supergiants with masses in the range $10 - 17M_\odot$ ([Arcavi, 2017](#)). Type IIL exhibit a very linear-like decline

Table 3.1: Summary of supernovae classifications based on their spectral absorption features as well as their primary formation scenario. Further subdivisions are sometimes defined based on qualitative features related to the light curve evolution after the peak.

Type	Spectral Absorption Features	Formation Scenario
Ia	No H. SiII present.	Thermonuclear Runaway Reaction
Ib	No H, SiII. HeI present.	Core-Collapse
Ic	No H, HeI, SiII.	Core-Collapse
II	H	Core-Collapse

after the peak. The progenitors for Type IIL however are less clear (see [Arcavi \(2017\)](#) again). Type IIn contain very narrow hydrogen emission lines in their spectra. It is presumed this is an artifact of how the progenitor interacts with a sufficiently thick circumstellar medium before the explosion ([Chugai, 1991](#)). For Type Ib/c, the progenitors are assumed to be large Wolf-Rayet Stars that have shed their hydrogen and helium envelope through strong stellar winds ([Gaskell et al., 1986](#)) or that it is a low-mass star that has had its envelope stripped by a close binary companion ([Nomoto et al., 1995](#)).

Since the spectral features of a SNIa are important for its classification, the photometry alone is not enough to confirm their typing. This requires spectroscopy to confirm the typing (discussed more in Section 3.7.2).

3.2.1 Superluminous Supernovae

In recent times, a new class of supernova have been proposed and are referred to as ‘Superluminous Supernovae’. These can occur as both Type I and Type II supernovae ([Howell, 2017](#)) and occur around a thousand times less often than regular SNe ([Prajs et al., 2016](#)). With peak B-band magnitudes of < -21 mag and occurring at redshifts $z > 2$, they represent a potential new way to probe our cosmology at higher redshifts ([Cooke et al., 2012](#); [Quimby et al., 2011](#); [Curtin et al., 2019](#)). There are competing theories that discuss the cause of their extreme luminosities that include the spin-down from compact magnetic objects know as magnetars and ^{56}Ni interaction powered models ([Moriya et al., 2018](#)). They are relatively rare in nature, but their high luminosity means they can often be seen at more further distances than regular SNe.

For a more complete overview of Superluminous Supernovae see [Gal-Yam \(2012\)](#); [Moriya et al. \(2018\)](#).

3.3 Core-Collapse Mechanism

While the primary focus of this thesis on SNIa which explode via a thermonuclear runaway reaction we provide first a brief description of the core-collapse mechanism SNe that are not of Type Ia can undergo. A more complete description can be found in [Foglizzo \(2016\)](#); [Yamaoka \(2016\)](#).

For large stars with masses in the range of $> 8M_{\odot}$ their life generally ends as core-collapse SNe once the core has gone through the stages of hydrogen, helium, carbon, neon, oxygen and silicon fusion until the degenerate core of the progenitor star is typically composed primarily of iron. The electron degeneracy pressure from the core cannot support the surrounding non-degenerate matter, resulting in catastrophic collapse of the core at a speed that is almost a quarter of the speed of light ([Woosley & Janka, 2005](#)) until all that remains is a dense neutron star approximately 10km in radius. Depending on the size of the star, the remnant may also become a black hole ([Begelman, 2003](#)).

For Progenitor stars with masses in the range $130-250M_{\odot}$ and low metallicity (most commonly believed to be Population III stars, the first stars that formed), they often undergo a core-collapse explosion because of electron-positron production in the core and are referred to as ‘pair-instability’ supernovae that are powerful enough to not leave a remnant behind at all. The lack of remnant means they are difficult to detect and they have yet to be confirmed by observation, though some currently observed SNIa have been theorised to be the product of pair-instability explosions ([Moriya et al., 2010](#); [Rakavy & Shaviv, 1967](#)).

3.4 Thermonuclear Runaway Reaction and Type Ia Supernova

In comparison to the supernovae that undergo core-collapse, SNIa are believed to be formed when main sequence stars of mass less than $8M_{\odot}$ become carbon-oxygen White Dwarfs (WDs). The white dwarf is in stable equilibrium as further gravitational collapse is prevented by electron degeneracy pressure. However, if a mechanism can be provided by which the white dwarf can continue to accrete further material it can explode as a supernovae as they approach (and in certain scenarios discussed later, exceed) the Chandrasekhar limit of $\sim 1.4M_{\odot}$. The Chandrasekhar limit is when the electron degeneracy pressure of the white dwarf is no longer capable of supporting the star against gravitational collapse and therefore defines the maximum allowable mass of a white dwarf.

The cause of the thermonuclear runaway reaction that creates this explosion is the burning of carbon and occurs shortly before the star reaches the Chandrasekhar limit. This kickstarts a ‘simmering’ that lasts for a few hundred years. As this burning phase progresses, large instabilities are created which eventually creates the runaway reaction (see [Hillebrandt & Niemeyer \(2000\)](#); [Piersanti et al. \(2022\)](#) for more complete details). The fusion of iron-group elements (Cr, Mn, Fe, Co, Ni) via this process is enough to disintegrate this star within fractions of a second. The primary driver of the light we see from SNIa are driven by radioactive decay of the material synthesized in the explosion, primarily ^{56}Ni .

The burning can progress via two modes, either subsonic speed (deflagration) or supersonic speed (detonation). Neither on their own is enough to account for the observed production of elements. Pure detonation methods mainly only produce the iron group elements and not enough of the intermediate mass elements such as silicon, calcium and sulphur that are seen in Ia spectra (see left plot of [Fig. 3.2](#)). As an alternative to this, a deflagration model was proposed ([Nomoto et al., 1976](#)), however this had the issue the SNIa brightness was too low because of the slow(er) rate of burning in comparison to detonation models. Today, the most favoured models are ones where initially there is a deflagration component with a later transition to a

detonation regime, but the exact mechanism via which the transition happens is still unclear (see [Gamezo et al. \(2004\)](#); [Mazzali et al. \(2007\)](#); [Poludnenko et al. \(2019\)](#)).

The iron group elements generated from these burning scenarios are responsible for the decay chain powering the light curve after the initial explosion ([Diehl & Timmes, 1998](#)), with the primary decay being



The initial decay of ^{56}Ni into ^{56}Co is most responsible for powering the light curve up to the peak brightness. The late time evolution (~ 30 days post-peak) is driven mostly by the decay of ^{56}Co to ^{56}Fe . The peak brightness of a SNIa is directly correlated and proportional to the amount of ^{56}Ni produced and this is known as ‘Arnett’s rule’ ([Arnett, 1982](#)). One of the key signatures of SNIa that also differentiate them from other Type I SNe is the presence of SiII doublet absorption lines at $\lambda = 6347\text{\AA}$ and $\lambda = 6371\text{\AA}$ (see [Fig. 3.2](#)).

3.4.1 Progenitor Scenarios

In both burning scenarios mentioned in the previous section, mass accretion of the WD is the crucial factor. WDs are generally formed well below the Chandrasekhar limit ($< 1M_{\odot}$) so a mechanism by which they can accrete mass is required. Despite their use as standardisable candles, exact understanding of the progenitor scenarios that generate SNIa is an open problem.

The WD progenitor can come in three different forms, the lower-mass Helium-WD, the intermediate mass Carbon-Oxygen WD and the higher mass Oxygen-Neon WD. The lower mass Helium WDs and higher mass Neon-Oxygen WDs can be ruled out as their decay products inferred from simulations are inconsistent with what is observed ([Woosley et al., 1986](#); [Nomoto & Sugimoto, 1977](#)). Furthermore, the population of Neon-Oxygen WDs are too low to replicate observed SNIa rates ([Saio & Nomoto, 1985](#); [Saio & Nomoto, 2004](#); [Nomoto & Kondo, 1991](#); [Gutierrez et al., 1996](#)). Therefore, this leaves CO WDs as the only viable channel for creating SNIa.

For the accretion phase, two principle models have been proposed. The first is a *Single Degen-*

erate scenario where a single WD accretes mass from larger companion star (typically a red giant or main sequence star) from Roche lobe overflow or strong stellar winds (Maguire, 2017) until shortly before it reaches the Chandrasekhar limit and explodes (Li & van den Heuvel, 1997).

The second scenario is the *Double Degenerate* case where the companion star is another WD. The pair inspiral towards each emitting gravitational waves as they do so (Webbink, 1984; Maguire, 2017). When the pair get close enough usually the smaller WD has its Roche lobe (the region around a star in a binary system within which orbiting material is gravitationally bound to that star) filled and is tidally disrupted by the larger WD, resulting in the formation of an accretion disk around the primary (Pakmor et al., 2012). The rapid accretion of matter in this scenario means that there is a possibility of the explosion happening with the exploding WD exceeding the Chandrasekhar limit (Webbink, 1994; Kamiya et al., 2012).

It was initially believed that the single degenerate case is how the majority of SNIa occurred, which would explain the apparent standardisation of SNIa, however, there is now growing body of evidence that the double degenerate case is more common than initially believed with Hernández et al. (2012).

3.4.2 Peculiar Type Ia Supernovae

The variation in progenitor systems means that subgroups of ‘peculiar’ SNIa can be identified, these can both be under- and over-luminous with respect to the average SNIa. For example, one of the largest groups of peculiar SNIa are under-luminous and are identified as SNIax or ‘SN2002cx-like’ supernovae. These are believed to occur primarily in young/late-type galaxies and have been shown to have lower ejecta velocities. The current leading theory on them is that they are formed from pure deflagration models (Foley et al., 2013; Jha, 2017).

Another large group of peculiar SNIa are ‘Super-Chandrasekhar’ SNe that are noted to be over-luminous. They are a sub-class of the aforementioned ‘Superluminous Supernovae’ and are believed to almost certainly come from double degenerate progenitors (Branch, 2006; Howell

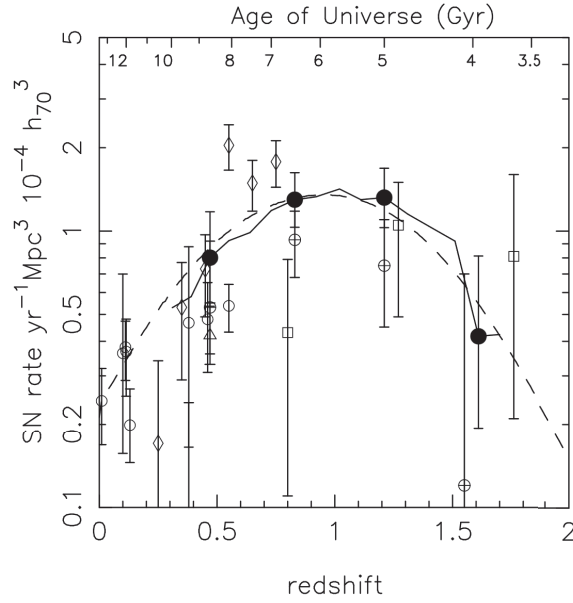


Figure 3.3: Type Ia supernova rates. The dashed line is the best fit model (taken from [Dahlen et al. \(2008\)](#), see therein for more details on the model and the data points present).

[et al., 2006](#); [Scalzo et al., 2010](#); [Taubenberger et al., 2011](#)). The lack of companion stars in this scenario makes this hard to verify however.

3.4.3 Type Ia Supernova Rates

SN Ia have been noted to be present in both, ‘young’ galaxies undergoing periods of active formation and ‘old’ galaxies with little to no star formation present. Core-collapse SNe in contrast, only occur in ‘young’ galaxies ([Li et al., 2011](#)). The SN Ia rate itself is a function of the star-formation rate in these galaxies (see [Mannucci et al. \(2005\)](#); [Sullivan et al. \(2006\)](#)). Because the progenitors of SN Ia are difficult to determine, the evolution of the star-formation rate as a function of redshift provides important information about their origin. An example of the SN Ia formation rate based on data by SDSS is shown in Fig. 3.3. It can be seen there is a steep decline in SN Ia formation rates after $z = 2$.

3.5 Type Ia Supernovae Cosmology

Based on the physics discussed previously, it can be seen SNIa are objects that can sufficiently be standardised to use as ‘standard candles’ with a fixed peak absolute magnitude, M . We can also define the apparent observed magnitude m_B which, as the subscript denotes, is the apparent magnitude as measured in the B-band. This means we can write the observed distance modulus as

$$\mu = m_B - M. \quad (3.2)$$

This can directly be compared with what is predicted by theory (see Section 2.3.1) where the distance modulus is given by

$$\mu(z, \ell) = 5 \log \left[\frac{c}{H_0} \frac{d_L(z, \ell)}{1 \text{Mpc}} \right] + 25. \quad (3.3)$$

where $\ell = \{\Omega_{M,0}, \Omega_{\Lambda,0}, \Omega_{K,0}, H_0, w\}$ is the cosmological parameters of interest.

SNIa are not usable as standard candles straight away because there is still large variability around their peak absolute magnitudes. To reduce this residual variability sufficiently a light-curve fitting procedure is applied.

3.6 Light Curve Fitting

For most SNIa there is a rise time to peak-brightness of ~ 20 days and then a decline of roughly 1-3 magnitudes month in the B-band. A series of corrections can be applied to account for correlations of absolute peak magnitude with their light curve decline rate and colour. This correlation with light curve decline rate is often known as the ‘width-luminosity’ relation. The relation is such that brighter supernovae have their light curve decay more slowly and dimmer supernovae light curves decay faster. This correlation feature was first observed by Rust (1974) and further corroborated later by Pskovskii (1977, 1984) and Phillips (1993);

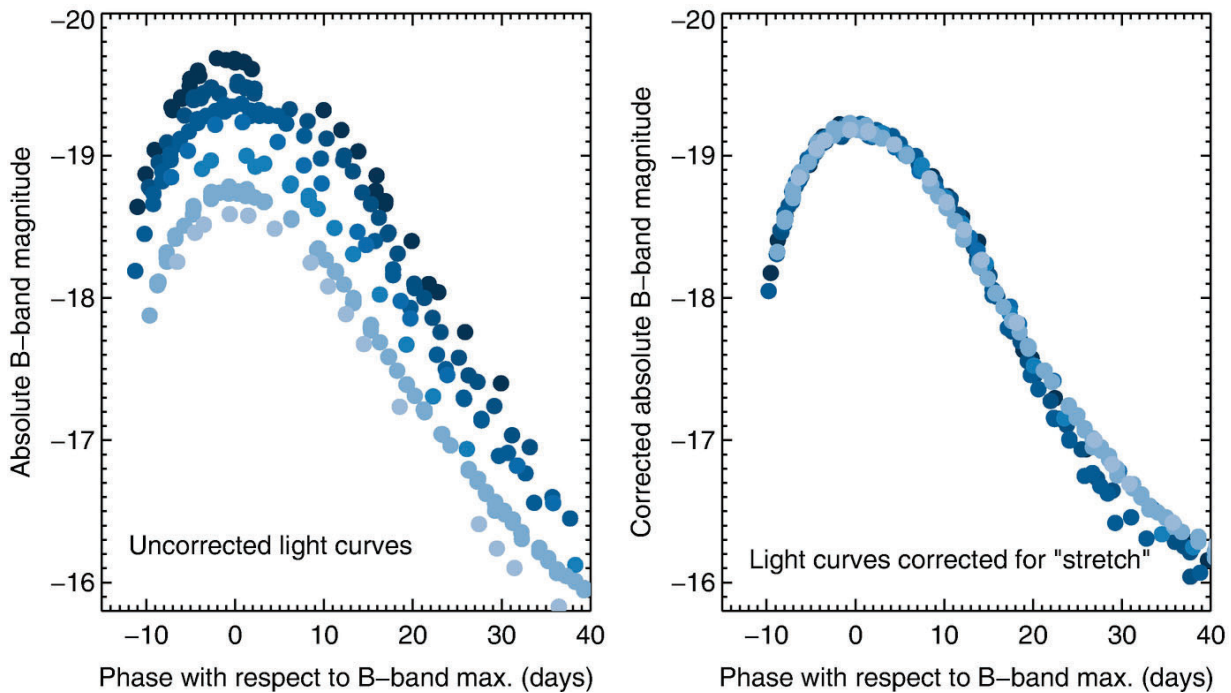


Figure 3.4: Here a set of example light curves in the B-band are displayed. The light curves were from data by [Hicken et al. \(2009a\)](#) and [Stritzinger et al. \(2011\)](#). We can see an example of the width-luminosity relation here and how after applying linear Phillips corrections the corrected SNIa show a small residual scatter in their peak absolute B-band magnitude. Figure taken from [Maguire \(2017\)](#)

[Phillips et al. \(1999\)](#) using larger samples of SNe. Later a correlation with color, the ‘colour-luminosity’ relation, was also observed where it was noticed that brighter SNIa were ‘bluer’ ([Hamuy et al., 1995](#); [Riess et al., 1996](#); [Perlmutter et al., 1997](#)). Empirical corrections that account for correlations of absolute peak magnitude with their decline time and colour can then be applied to standardise the SNIa data ([Tripp, 1998](#); [Phillips et al., 1999](#)). Light curves of SNIa can be standardised so that the residual scatter of their peak *B*-band magnitude is sufficiently small (~ 0.1 mag) to infer cosmological parameters, as was first demonstrated by [Riess et al. \(1998\)](#) and [Perlmutter et al. \(1999\)](#). An example of the reduction in scatter after correcting for these correlations can be seen in Fig. 3.4.

To reduce the scatter in the peak magnitude, often a light curve fitter is employed. Many light curve fitters exist, that work in various ways, but the data analysed in this thesis was analysed using the Spectral Adaptive Light-Curve Template 2 (SALT2) light curve fitter ([Guy et al., 2007](#)) which is currently the most popular fitter in use and is an improved version of

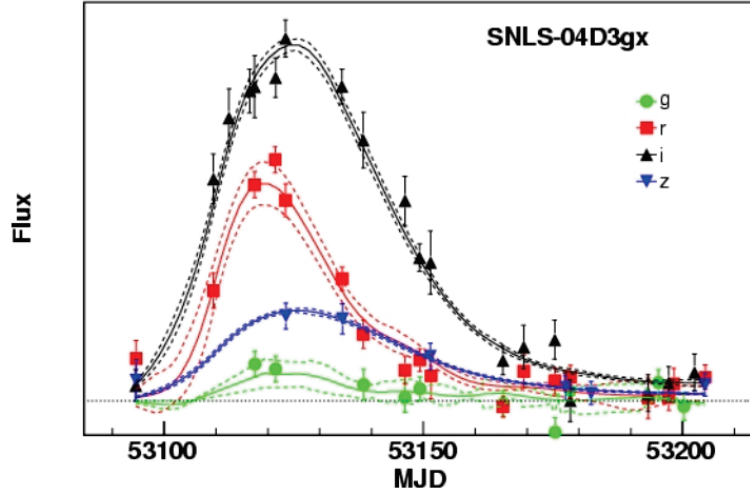


Figure 3.5: Observed light-curve points of SNIa SNLS-04D3gx along with the light-curves derived from the model fit. The solid lines indicate the fit when the SN is not included in the training sample and the dashed lines represent the 1σ uncertainty of the fit. Taken from [Guy et al. \(2007\)](#)

the original SALT fitter ([Guy et al., 2005](#)) which uses the methodology followed by [Perlmutter et al. \(1999\)](#). Recently however, a new version SALT3 has been announced ([Kenworthy et al., 2021](#)). We limit our discussion to the process followed by SALT2 though which is used for the analysis in this thesis.

In SALT2, the data points for a SNIa are fitted to a light curve template (see Fig. 3.5 for an example). This light curve template is ‘trained’ using both nearby and distant SNIa. As part of this fitting, summary statistics that can be used to apply empirical corrections are derived and can be defined as

- m_B – the rest frame B-band maximum apparent magnitude of the SNIa.
- x_1 – the stretch parameter which uses the width luminosity relation and the decline rate Δm_{15} (the decline in apparent magnitude after 15 days from the peak) and were related together in [Guy et al. \(2007\)](#) from a polynomial regression as

$$\Delta m_{15} = 1.09 - 0.161x_1 + 0.013x_1^2 - 0.00130x_1^3 \quad (3.4)$$

- c – the colour excess, which uses the colour-luminosity relation and is given by

$$c = E(B - V) + 0.057 \quad (3.5)$$

where $E(B - V)$ is the colour excess between the intrinsic and observed apparent magnitudes in the B band compared to the V band at maximum luminosity, i.e, $E(B - V) = (B - V) - (B_0 - V_0) = (m_B - m_V) - (M_B - M_V)$. The extra constant is required as -0.057 is the chosen reference color of a SNIa in Vega magnitudes (Guy et al., 2005). It should be noted here that colour can be positive or negative. Shifts in either direction can be caused by multiple effects such as intrinsic variation in color or the effects of dust reddening. SALT2's methodology is unable to differentiate between the different sources and measures the combined effect.

These estimates for the best fit values of m_B , x_1 and c are provided with a covariance matrix of their uncertainty from the light curve fit for each SNIa as

$$C = \begin{pmatrix} \sigma_{m_B}^2 & \sigma_{m_B, x_1} & \sigma_{m_B, c} \\ \sigma_{m_B, x_1} & \sigma_{x_1}^2 & \sigma_{x_1, c} \\ \sigma_{m_B, c} & \sigma_{x_1, c} & \sigma_c^2 \end{pmatrix} \quad (3.6)$$

It is stressed again that these are purely empirical corrections with the astrophysical origins for them not fully understood as of yet. The summary statistics are then applied linearly to the supernovae peak B-band absolute magnitude. It is at this step the ‘width-luminosity’ and ‘brighter-bluer’ relationship is encoded via a linear regression formula. This empirically corrected absolute magnitude, M^ϵ is related to original peak magnitude, M , via the linear Tripp relation (Tripp, 1998).

$$M^\epsilon = M + \alpha x_1 - \beta c, \quad (3.7)$$

where α and β are nuisance parameters that control the slope of the stretch and colour correction respectively. This means the distance modulus can now be written as

$$\mu = m_B - \alpha x_1 + \beta c - M \quad (3.8)$$

which is used to compare with the theoretical distance modulus when constraining our cosmology. The fitting in the second step is not done by SALT2 itself, however. Constraining the best fit values of the nuisance parameters and our cosmology is done with the **BAHAMAS** methodology presented in Section 4.3.

As mentioned in the definition of colour, SALT2 cannot differentiate between the variation in colour due to intrinsic variation in the population of supernovae or from reddening due to interstellar dust. At lower wavelengths, light is absorbed and scattered to a greater degree by interstellar dust. The net effect is to cause reddening of the colour. It is relatively simple to correct for dust owing to the Milky Way as it is well measured along different lines of sight (Johansson, 2015). Dust from other components however, such as host galaxy dust, are more difficult to correct. The dust-size distribution is often captured via the total-to-selective extinction ratio R_V where R_V is defined as

$$R_V = \frac{A_V}{E(B - V)} \quad (3.9)$$

where $E(B - V)$ is the colour excess as before and A_V is the extinction in V-band. The Milky Way has the value $R_V \sim 3.1$ and relatively little scatter of $\sigma_{R_V} = 0.18$ (Schlafly et al., 2016). However values between 1.5 and 6 and have also been inferred by some authors (Draine, 2003; Krisciunas et al., 2006) with a lot of the lower end of these readings from SNIa with $E(B - V) > 0.3$ (Amanullah et al., 2015; Hoang, 2017). Many users of SNIa data mitigate the effects of dust systematics by restricting the data to ‘normal’ SNIa by applying colour cuts of $E(B - V) \lesssim 0.3$ as was done with the JLA data (Betoule et al., 2014) that is used in this thesis.

Many other light curve fitters exist in Supernovae Cosmology that attempt to standardise based off the ‘width-luminosity’ and ‘colour-luminosity’ relationship which include (but not limited to) MLCS (Riess et al., 1998) and its successor MLCS2k2 (Jha et al., 2007). MLCS, unlike SALT2, explicitly attempts to account for dust reddening by separating it from the intrinsic variation

in colour. BayeSN (Mandel et al., 2009, 2011) is a Hierarchical Bayesian method for modelling the light curves of Near IR SNIa where the effects of dust are significantly mitigated relative to optical data. SNooPy (Burns et al., 2010), similar to MLCS also attempts to separate dust-based reddening of colour from intrinsic variation of colour. PISCOLA (Müller-Bravo et al., 2021) uses a Gaussian Process based method to standardise light curves without the need for a base template to fit the curves to. SUGAR (Léget et al., 2020) includes two extra parameters to account for variability in addition to the colour and stretch parameters of SALT2. These are associated with the supernovae ejecta velocities and their calcium lines and leads to a colour law that is more consistent with Milky Way dust extinction which SALT2 does not provide.

3.7 Concordance Cosmology

Before we discuss our current accepted cosmology, I introduce here the two most common models that allow for dark energy, Λ CDM and the closely related w CDM model. They are defined from writing the Friedmann equation as

$$H(z)^2 = H_0^2(\Omega_M(1+z)^3 + \Omega_\gamma(1+z)^4 + \Omega_K(1+z)^2 + \Omega_\Lambda \exp \left[\int_0^z \frac{1+w(z)}{1+z} dz \right]). \quad (3.10)$$

When $w = -1$ and constant, this is referred to as a Λ CDM model, otherwise it is a w CDM model. Note, for dark energy to be repulsive and cause acceleration in the Universe, $w < -1/3$ is a strict requirement.

The largest compilation of SNIa used for cosmology purposes is the Pantheon sample (Scolnic et al., 2017) which contains 1048 spectroscopically confirmed ‘normal’ SNe. It should be noted however a ‘Pantheon+’ dataset is due for imminent release with over 1500 spectroscopically confirmed SNe (see Brout et al. (2022) preprint). The constraints it places on our Cosmology can be seen in Fig. 3.6. The left panel of Fig. 3.6 shows the constraints on a w CDM model. Here we can see the complementary nature of SNIa with respect to fits provided by CMB and BAO

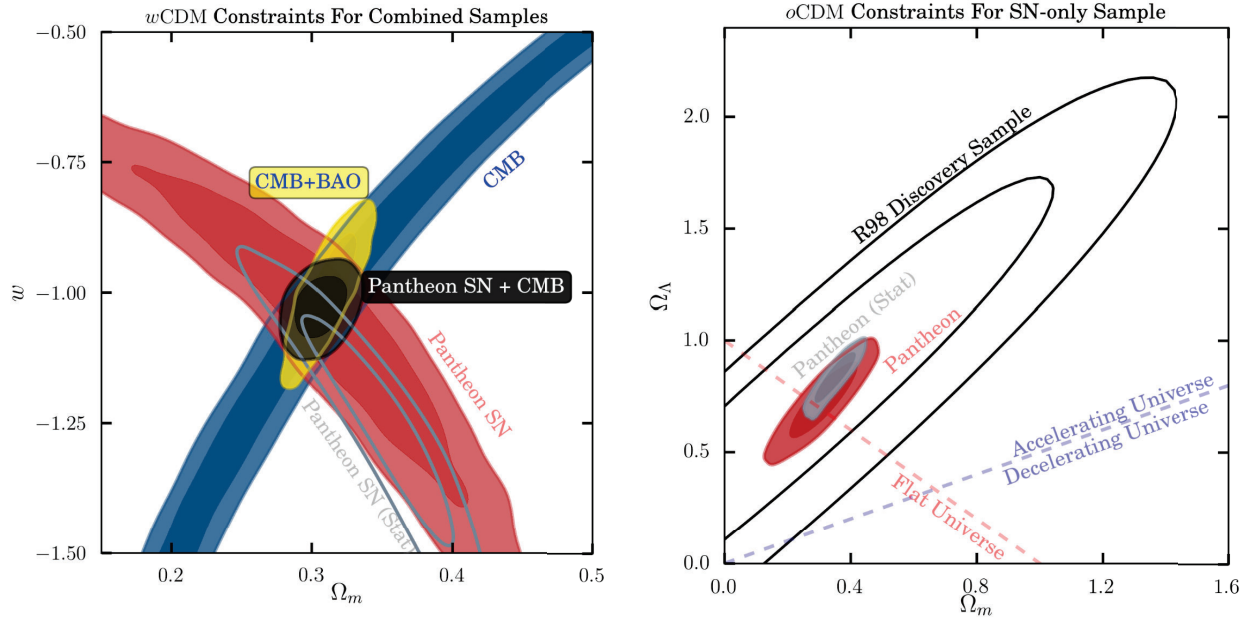


Figure 3.6: Cosmological constraints from the Pantheon SNIa sample. The left figure shows the constraints for a w CDM model and the combined constraints using BAO and CMB measurements. The right shows constraints from SNIa only on a Λ CDM model which does assume strict flatness ($\Omega_k = 0$). Figures taken from Scolnic et al. (2017).

probes where they serve to break the degeneracy in allowed parameters significantly because of the orientation of the distribution in the $w - \Omega_m$ plane. The inferred results have best fit values of $w = -1.014 \pm 0.040$, $\Omega_m = 0.307 \pm 0.008$ which are consistent with a model that has a cosmological constant Λ with $w = -1$. Furthermore within an o CDM model (where flatness is not assumed), the right panel of Fig. 3.6 indicates that the universe is currently experience a period of late-time acceleration in the expansion rate of the universe. A result that won the 2011 Nobel Prize when it was displayed by Riess et al. (1998); Perlmutter et al. (1999). This has lead to currently accepted concordance Λ CDM cosmology, the naming of which implies three major components to our universe - The cosmological constant, cold dark matter and ordinary matter.

Since the initial discovery of the accelerated expansion of the universe, the field has since moved on to attempting to characterising and measuring the equation of state parameter $w = p/\rho$ and its time evolution with groups such as the *Dark Energy Task Force* (Albrecht et al., 2006) deciding the appropriate experiments to conduct for this purpose. The characterisation of w by the pantheon sample of SNIa can be seen in Fig. 3.7 by using a model parameterisation from

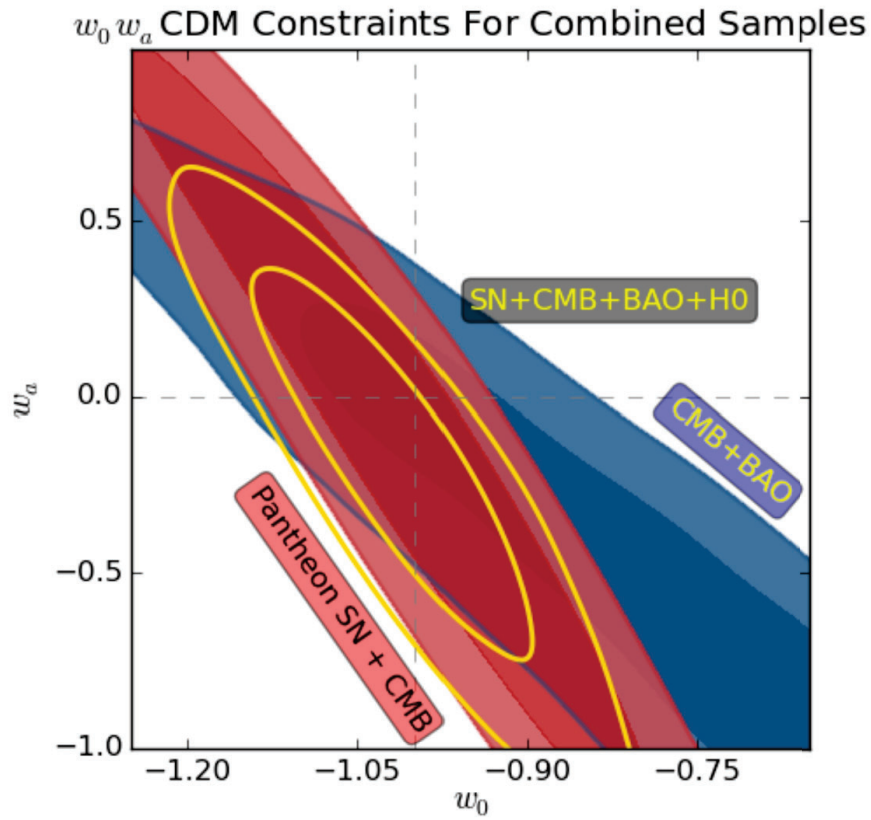


Figure 3.7: Results from the pantheon SNIa sample from trying to constrain an evolving w CDM model where the model is defined according to Eq. (3.11). Figure taken from Scolnic et al. (2017)

Chevallier & Polarski (2001) where

$$w(z) = w_0 + w_a z / (1 + z). \quad (3.11)$$

The results are consistent with a model that is not time evolving and that the data is not able to determine an evolving value of w based on the sensitivity than can be probed by current SNIa data.

Some tensions exist in the results derived from SNIa however, with one of the most notable being the ‘Hubble Tension’. While the Hubble tension is not the central focus of this thesis, it is worth discussing to motivate the need to consider deviations from Λ CDM in Chapters 5 and 6. SNIa top out the distance ladder with the first two rungs given by parallax and Cepheid variable measurements, the latter of which are also used anchor and calibrate SNIa. SNIa provide a measurement of the Hubble-Lemaître constant today of $H_0 = 74.03 \pm 1.42 \text{ km s}^{-1} \text{ Mpc}$ from the SH_0 ES collaboration (Riess et al., 2019). This is in 4.4σ tension with the Planck 2018 measurements of $H_0 = 67.4 \pm 0.5 \text{ km s}^{-1} \text{ Mpc}$ (Aghanim et al., 2020), potentially calling in to questions whether our model of flat Λ CDM is correct or whether an alternate more exotic model is needed. Measurements of other probes of H_0 are indicated in Fig. 3.8 wherein it can also be seen that there is a general tension between measurements of H_0 at early and late times.

3.7.1 Precision Cosmology and Systematic Uncertainties

The state of the field as it currently stands, is that we are at point where we can probe our cosmology with less than 1% uncertainty when all probes are combined. Fig. 3.9 shows the residual error to the best fit cosmology on the Hubble diagram is also approximately 1%. Reducing these errors even further is one of the major goals of current cosmological analysis. Indeed, we are currently in the era of what is colloquially called ‘Precision Cosmology’ where uncertainties are on the scale of a percent and increasingly systematic uncertainties dominate with respect to the statistical uncertainties.

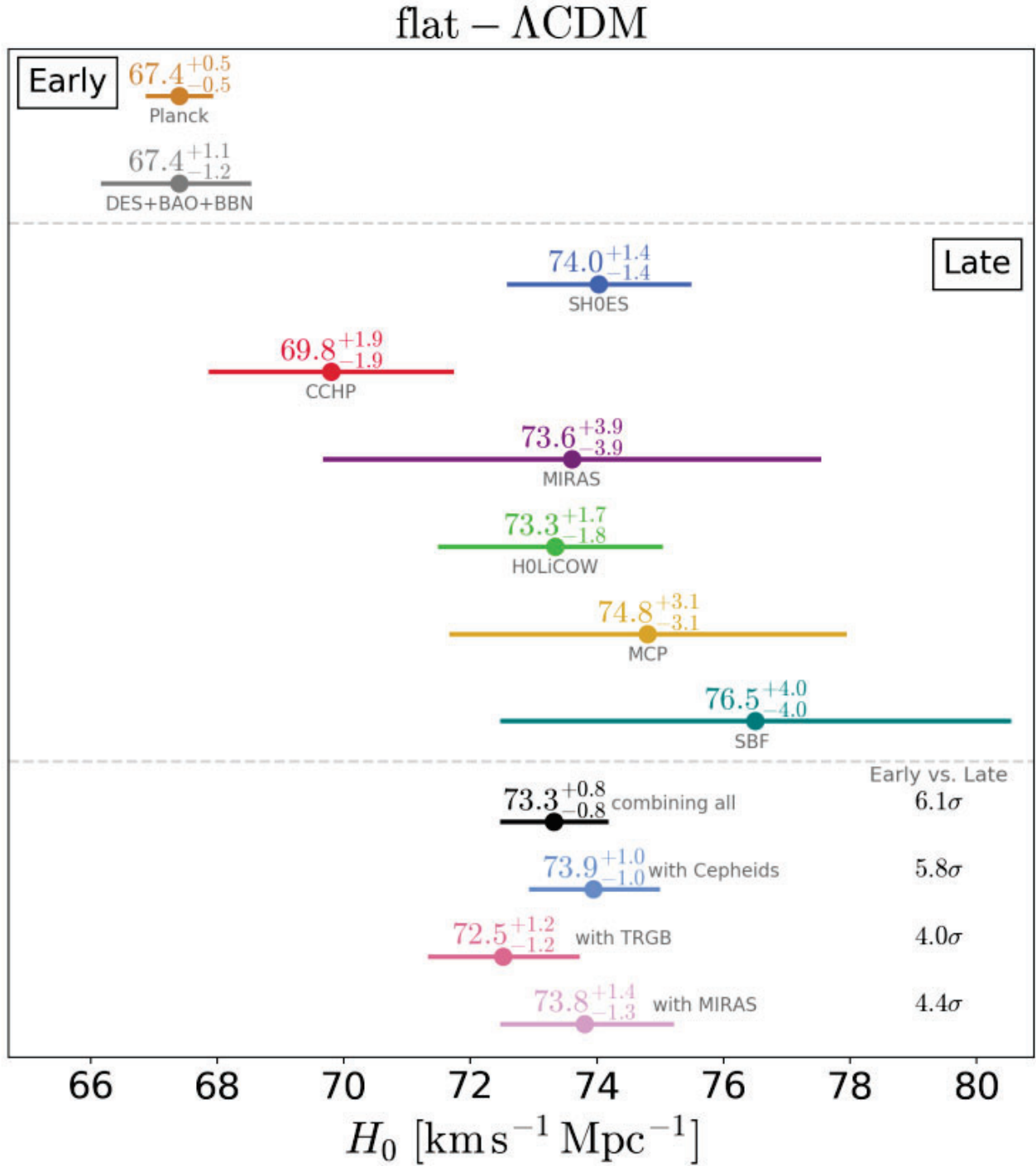


Figure 3.8: Compilation of Hubble Constant predictions and measurements that was presented or discussed at the workshop “Tensions between the Early and the Late Universe” in 2019. Figure taken from a summary of the workshop by Verde et al. (2019).

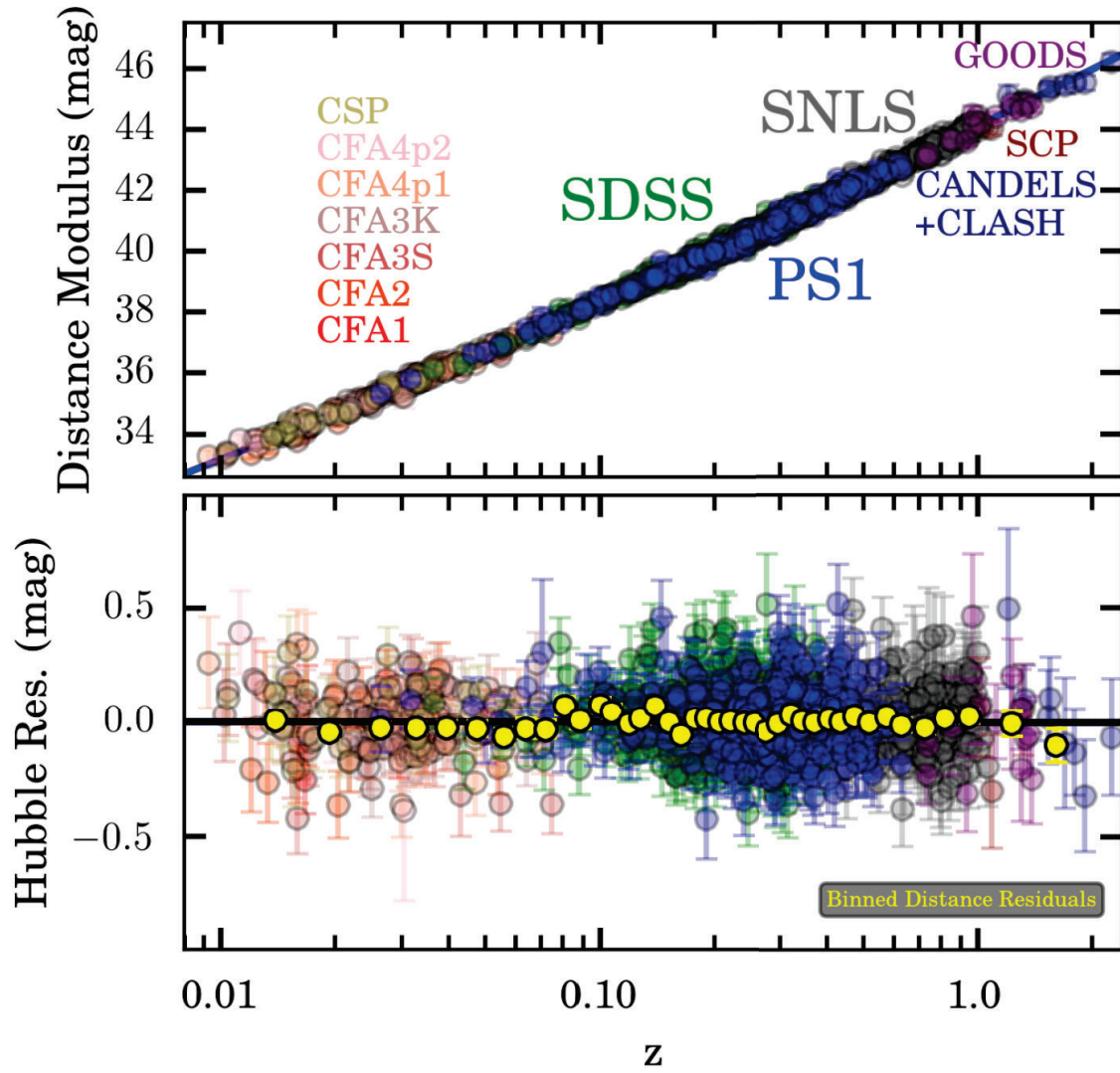


Figure 3.9: Hubble diagram from the pantheon SNIa sample. The top shows the measured distance modulus with data points coloured according to the survey the data was measured by. A line based on the best fit cosmology is plotted on top in blue. The bottom shows the residual error between the best fit cosmology and the data. Figure taken from [Scolnic et al. \(2017\)](#)

Handling systematic uncertainties in SNe datasets is key to reducing the uncertainty in our best-fit cosmology further. I briefly go through here some of the most discussed sources of systematic uncertainty in SNIa analysis.

- Calibration - Cosmological constraints depend on measuring the relative brightness of SNIa at different redshifts. Flux calibration errors associated with this are typically the largest source of uncertainty (Conley et al., 2011). Furthermore, as calibration is done early in the data pipeline, systematics associated with this can propagate to each stage of the analysis. The flux error effect on the measurement of the SNIa propagates to the K-corrections, extinction corrections and the final magnitude measurements which are used to constrain cosmological parameters. In addition, while most future surveys are likely to measure most if not all of their SNe with a single, well calibrated system such as LSST (LSST Dark Energy Science Collaboration, 2012) and DES (DES Collaboration et al. (2018)), current cosmological bounds are derived from a compilation of surveys such as the Joint Light-Curve Analysis (JLA) data (Betoule et al., 2014) and the Pantheon data. Calibration offsets between the data sets can also potentially result in a bias when fitting a cosmology. In the case of the Pantheon data (Scolnic et al., 2017) which is a multi-survey compilation, calibration uncertainty of 2-6 mmag (depending on the sample) accounts for approximately half (0.02) of the total uncertainty on the measurement of w .
- Dust - To accurately use SNIa as standard candles, one needs to account for the interaction of light with dust which may be from the circumstellar material around the exploding star, the interstellar medium of the host galaxy, the intergalactic medium and the dust in the Milky Way itself. Dust is important to account for as it can scatter and absorb light, thereby dimming the SNe. Space-based observatories such as Nancy Roman Grace Telescope (previously called WFIRST) and Euclid will have Near-Infrared (NIR) capabilities. Measurements of SNIa in NIR have been noted to have better standardisation properties than those in the optical and are noted to suffer less from systematics due to dust related effects (Mandel et al., 2009, 2011; Wood-Vasey et al., 2007; Wood-Vasey et al., 2008; Krisciunas et al., 2004). In addition, one can consider looking at SNe in the

outskirts of galaxies where there is less dust (Hill et al., 2018). These NIR data can also be combined in conjunction with optical data with the NIR data used to properly model dust extinction of the host galaxy that affects the optical data as in Mandel et al. (2021).

- **Light Curve Fitters** - The light curve fitters used to standardise SNIa can also be a source of the systematic bias. For example, the widely used SALT2 and MLCS2k2 fitter can each introduce different systematic biases. The reference samples for training MLCS2k2 are low- z SNe so the training can be biased. Furthermore, MLCS2k2 assumes colour shifts separate from intrinsic variation are in one direction only by assuming all shifts are caused by dust based reddening. SALT2 on the other hand uses data from both high and low redshift, but biases can be caused by the intrinsic changes in redshift that are not picked up such as systematic biases for a calibration filter in a given redshift range that ends up biasing the whole sample across over the entire redshift range during training or not accounting for the evolution of the light curve parameters with redshift (see Kessler et al. (2009) for a greater discussion of MLCS2k2 and SALT2).
- **Selection Bias** - Because of the nature of the SNIa measurements and general flux limited measurements, they inherently suffer from Malmquist bias (Malmquist, 1922), wherein brighter objects are easier to detect (discussed in much greater detail in Chapter 7). Many attempts at constraining cosmology with SNIa involve the use of large scale Monte-Carlo simulations to sample the fraction of SNIa that can be measured at a provided redshift given your instrument detection requirements and then correct for this bias. This procedure is used in deriving the results from the Pantheon compilation of SNIa (Scolnic et al., 2017). Handling selection bias is one of the goals of this thesis and is discussed further in Chapter 7.
- **Gravitational Lensing** - The derivation of the luminosity distance in Section 2.3.1 assumed isotropy and homogeneity which is a valid approximation at large scales based on CMB data, however at small scales, inhomogeneities in local structure can have an affect on inferred distances. The foreground of galaxies have a lensing effect from gravitational interactions with points of higher density. The effect is to create a scatter in the observed

SNIa fluxes that increases with redshift (Bergström et al., 2000; Holz & Linder, 2005; Mörtzell et al., 2001). These are expected to statistically average out for a sufficiently large and complete sample, but for small samples or magnitude limited samples this can be source of bias (Jönsson et al., 2010).

- Peculiar Velocities - Similar to gravitational lensing effects, local matter inhomogeneities can induce a systematic bias at small redshifts. The density fluctuations can induce peculiar velocities of $\sim 300 \text{ km s}^{-1}$ which are observationally not possible to separate from the recessional velocity when interpreting the redshift. Aside from the peculiar velocity of the galaxy, the SNIa itself can have a peculiar velocity with respect to the galaxy motion. At smaller redshifts, where this peculiar velocity is relatively large compared to the Hubble flow, this can cause an error in the distance modulus approximately given by

$$\sigma_\mu \approx \frac{v}{c} \frac{5}{\ln 10} \left[\frac{1+z}{z(1+z/2)} \right]. \quad (3.12)$$

Out to a redshift of 0.015 this can be approximately 0.1 mag, greater than the residual error that is often aimed for when standardising the SNIa peak brightness. Most authors as a precaution impose a lower redshift cutoff to mitigate this effect (typically $z > 0.02$) (Riess et al., 1998; Perlmutter et al., 1997; Davis et al., 2011). The effect of peculiar velocities and how we account for it at low redshift are discussed further in Chapter 5.

- Host Galaxy Mass - In general, several properties of the SNIa host galaxy can be sources of systematic bias. Sullivan et al. (2010) finds that SNIa found in more massive galaxies are brighter after standardisation. It has therefore become common practice to split the SNe into two groups, based on their host-galaxy stellar mass (Sullivan et al., 2010; Lampeitl et al., 2010; Roman et al., 2018), with the mass threshold between the two groups being around value of $\log_{10}(M_g) = 10$, where M_g is the host-galaxy mass measured in solar masses. Smith et al. (2020) found a difference of up to 0.04 mag in the average intrinsic magnitude of the two groups, and application of the Tripp relation to the two groups separately may further reduce the residual dispersion in SNIa absolute magnitude. The

extent the mass split reduces the residual dispersion by is not noted in [Smith et al. \(2020\)](#), however, the implementation of this mass split by [Shariff et al. \(2016\)](#) saw that the higher mass group in the JLA compilation of SNe ([Betoule et al., 2014](#)) had a residual dispersion that was smaller by ~ 0.007 , with the lower mass group having a residual dispersion that is conversely larger by $\sim .0.006$. [Brout & Scolnic \(2021\)](#) has recently cast doubt on the robustness of this mass-step correction, which they ascribed instead to incorrect dust modeling and that the mass step can be explained away by the variability in the dust parameter R_V between massive and small galaxies. [Thorp et al. \(2021\)](#), however, finds the dust parameter across a low and high mass galaxy samples consistent with the global value and does not account for the mass step within their Bayesian heirachical model. Furthermore, [Uddin et al. \(2020\)](#) finds a significant mass-step in both optical and near IR SNIa data where, in the latter, the effect of dust extinction is relatively small. Another follow up paper by [Johansson et al. \(2021\)](#) claims that any significant mass-step in optical data disappears when fitting an R_V value to each individual supernovae as opposed to two groups of high and low mass galaxies. [Thorp & Mandel \(2022\)](#), however, details how modelling values of R_V for individual SNIa could lead to an overestimate of the sample variance of R_V . The ultimate origin of the mass step remains unclear, and it might relate to stellar population age ([Childress et al., 2014](#)) and metallicity ([Sullivan et al., 2010](#); [Gallagher, 2008](#)).

- Host Star Formation Rate - Another oft discussed property for standardising SNIa is the host local Star Formation Rate (SFR) as a useful covariate which is related to their age. [Rigault et al. \(2020\)](#) was able to find that SNIa in predominantly younger environments with greater star formation rates were typically dimmer by about 0.163 mag (5.7σ) than those found in older environments because of the dust and gas rich environment, in which stars form, attenuates the brightness. This is comparable to the difference seen based on the mass-step correction and corroborated by different authors on different samples ([Kim et al., 2018](#); [Uddin et al., 2017](#)), though some claim the effects is smaller at around 0.051 mag (2.1σ) ([Jones et al., 2018](#)). This has a knock on consequence on the determination of the value of the Hubble constant as the extent of the star formation

is correlated to brightness. This can bias H_0 measurements, especially as the calibrator Cepheids tend to be found in such active star forming environments, though the full extent of such effects is yet to be resolved (Rigault et al., 2015; Jones et al., 2015; Riess et al., 2016).

- **Host Galaxy Metallicity** - The host galaxy metallicity is also implied to have a significant effect on the standardisation of supernova. Most theoretical reasoning leads to the conclusion that different metallicities lead to different amounts of ^{56}Ni produced during explosion which powers the SNIa optical light curve (Foley & Kirshner, 2013), however the effect is not as strong as the other two host galaxy covariates. D'Andrea et al. (2011) found also that higher-metallicity hosts had SNIa that were brighter by 0.1 mag on average, however, later Childress et al. (2013) found that the effect is smaller at ~ 0.077 mag on a separate sample. In addition, they noticed a small correlation with the hubble residuals of a SNIa and the hosts with higher metallicity had higher stretch and redder colour. Campbell et al. (2016) measured a weak correlation at 1.4σ with the hubble residual and the SNIa host metallicity.

3.7.2 Future Surveys

The next generations of surveys are expected to target larger and richer amounts of data than what we currently have by measuring an increased number of SNe and in a larger number of wavelength bands. It is expected in the next decade that $\sim 10^7$ SNIa will be detected which is many times more than then current ~ 1000 available in current SNIa datasets and can be used to reduce sources of uncertainty further as well as provide novel new ways of probing our cosmology. The source of this increase in data is being driven by surveys such as the Vera Rubin Observatory Legacy Survey of Space and Time (LSST) (LSST Science Collaboration et al., 2009), Nancy Roman Grace Telescope (Spergel et al., 2015) and the James Webb Space Telescope (JWST) (Gardner et al., 2006).

Surveys like LSST will measure hundreds of thousands SNIa *per year*. Because of the increased size of the data from future surveys, it is not always possible to get the spectroscopy of the

data. As can be seen from the discussion in Section 3.2, the spectral features of a SNe are important for classifying their typing. Therefore, spectroscopic confirmation is important in analysis to confirm SNIa datasets do not suffer from contamination of non SNIa. Learning to classify SNe from just their photometry is one of the major challenges of the next generation of surveys. To this end, the Astrophysics community created a public challenge on the Kaggle platform¹ where a sample (from simulation of the kind of data LSST is expected to observe) is provided. The goal of the challenge to create an algorithm capable of accurately classifying observables from just photometry alone (The PLAsTiCC team et al., 2018; Malz et al., 2019). The winner of the challenge was the *AVOCADO* algorithm (Boone, 2019) which follows a similar procedure to the gaussian process based data augmentation method of *STACCATO* by Revsbech et al. (2018). These algorithms, and future ones, are unlikely to be 100% accurate in their classification and means of the handling the probability an object is not a SNIa is required. One methodology of doing so is the Bayesian Estimation of Multiple Species (BEAMS) formalism of Kunz et al. (2007) which is designed to reduce biased parameter estimates made from contaminant supernovae in data that are not of Type Ia.

Some independent probes of our cosmology with SNIa that differ from the method discussed in this chapter, but will supported by this new richer dataset include (but not limited to)

- Gravity with Peculiar Velocities - the peculiar velocities of the galaxy motions for the lowest redshift objects are correlated with the inhomogeneous structure of the universe at small scales. A peculiar velocity power spectrum can be constructed from such measures that are useful for constraining cosmological parameters such as the product of the growth rate and amplitude of mass fluctuation, $f\sigma_8$. The current best estimate of this parameter from SNIa studies is $f\sigma_8 = 0.428^{+0.046}_{-0.045}$ (Huterer et al., 2017). Howlett et al. (2017) discusses how LSST can be used to significantly reduce the uncertainty on these results to just a few percent.
- Weak Lensing signal - SNIa are subject to lensing effects that can dim or brighten them based on whether they pass through overdense or underdense regions. This lensing signal

¹<https://www.kaggle.com/c/PLAsTiCC-2018>

can be used to infer cosmological parameters (Wang, 1999, 2005; Macaulay et al., 2020; Shah et al., 2022) or using them to assess dark matter halo properties (Metcalf, 1999; Goliath & Mörtzell, 2000; Jönsson et al., 2008). To do this, the lensing distribution $p(\mu, z)$ needs to be reconstructed. Some efforts to this end have been made so far such as by Smith et al. (2013) where they look at a sample of SDSS SNIa, however, it is estimated that ~ 600 SNIa are required per redshift bin to have sufficiently strong statistics. It is expected that Nancy Roman Grace Telescope will provide enough SNIa for multiple redshift bins in the range $1 < z < 2$ that will provide an independent probe of our cosmology for that range.

- Strong Lensing of SNIa - Strong Lensing of SNIa is expected to be one of the sources of data that will help to constrain our cosmology. The process to do this involves a method known as *Time-Delay Cosmography* which measures the delay in arrival time for multiple images of a strongly lenses transient. Via this procedure one can derive a ratio of cosmological distances to the source and the lens which is inversely proportional H_0 as well as being sensitive to other cosmological parameters. This method has already been used successfully with strongly lensed quasar data (Suyu et al., 2018; Birrer et al., 2020). In terms of SNIa, the first lens was detected by Kelly et al. (2015), and was named ‘SN Refsdal’ after Sjur Refsdal who first proposed using such lenses for Cosmological inference (Refsdal, 1964). The first such lens that was strongly resolved and multiply imaged was iPTF16geu (Goobar et al., 2017) which Dhawan et al. (2019) was able to determine magnification, extinction, and time-delay estimates for.

A summary of the sources of data and their redshift distribution as well the kind of cosmological probes that can be performed is shown in Fig. 3.10 and is taken from the Astro2020 decadal survey white paper (Scolnic et al., 2019).

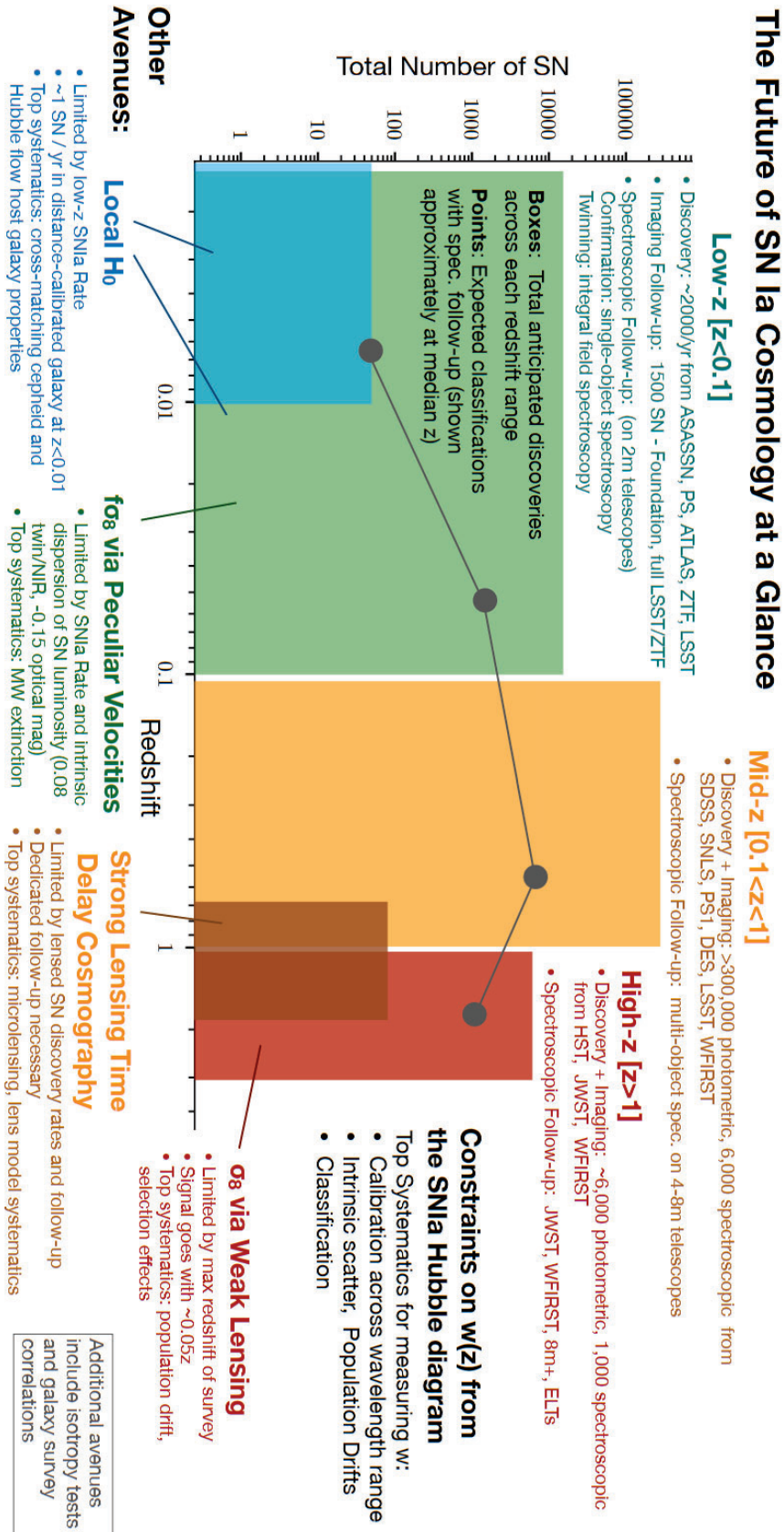


Figure 3.10: A summary of the sources of SNIa data in the 2020's and the probes of cosmology that can be performed with them. Figure taken from (Scolnic et al., 2019)

Chapter 4

Bayesian Statistics

One of the foundational pillars of the scientific method is the use of statistical analysis. This is a big departure from the days of old when Ernest Rutherford supposedly said *“If your experiment needs statistics, you ought to have done a better experiment”*. The process of empiricism leads us to make observations about the world which we can then use in order to draw conclusions and update our beliefs. This sort of reasoning process is inevitably a probabilistic exercise. However, observations and experiments are never perfect, as measurements are subject to stochastic errors which can have many sources. These statistical, or ‘aleatoric’, uncertainties are a limiting factor in determining how informative the data can be. In addition, the models, measuring apparatus and assumptions we use to analyse the data are also rarely perfect in practice which then become a source of systematic, or ‘epistemic’, error. Incorporating these various sources of uncertainties in a rigorous and consistent manner is of utmost importance to ensure the conclusions we derive from observed data are accurate reflections of the physics of the universe.

The seminal results of [Riess et al. \(1998\)](#); [Perlmutter et al. \(1999\)](#) that indicated the late-time acceleration of the universe used only ~ 50 SNIa which is a far cry from the volumes of data we currently have and are expected to have access to within the next decade as mentioned in Chapter 3. With the increasing volume and quality of data there is a corresponding reduction in the statistical uncertainties, giving rise to the era of precision cosmology. The existence of new physics is often inferred via detections at the $3 - 4\sigma$ level necessitating the need for statistical

analyses that are rigorous in their analysis of uncertainties, since the increased significance of any resulting biases could mistakenly point to the existence or lack of new physics. This increased influx of data presents a significant computational challenge for statistical inference, as methods must scale efficiently to these very large datasets. The higher precision constraints that can be placed with these larger datasets also mean that previously neglected systematic uncertainty effects must now be accounted for by providing more detailed and complex models that can account for previously more subtle effects.

In this chapter, I will provide a brief introduction to the fundamentals of Bayesian statistics and then move on to discussing the necessary numerical tools that allow us to fully utilise Bayes' Theorem. I will then introduce the Bayesian Hierarchical Model **BAHAMAS** ([March et al., 2011](#); [Shariff et al., 2016](#)) which we employ for much of the analysis in this thesis. Comparisons of the **BAHAMAS** algorithm with classical frequentist methods is also given. For a more in-depth overview of the topic of Bayesian analysis in the context of Astrophysics, I refer the reader to ([Trotta, 2008](#); [Loredo, 1992](#); [Hobson et al., 2009](#); [Liddle et al., 2009](#)).

4.1 Bayesian Analysis

The history of Bayesian analysis goes all the way back to the 18th century when it was first introduced by the Reverend Thomas Bayes, whose work was published posthumously after his death by his friend Richard Price ([Bayes & Price, 1763](#)) in a document titled 'An Essay towards solving a Problem in the Doctrine of Chances'. This essay included the theorems of conditional probability which now are part of the cornerstones Bayes' Theorem.

The way we think about Bayes' Theorem today is perhaps best attributed to Pierre-Simon Laplace instead who independently around the same time, in 1814, formalised the mathematics used for Bayes' Theorem as it is used today. He used this to realise the skewed birth ratio of males to females was a 'general law of the human race' as opposed to statistical anomaly and measured the masses of Jupiter and Saturn to within 1% of the correct value (based on today's measurements). However, his work on the Central Limit Theorem late in life meant that he

rarely used Bayesian statistics upon his death, using frequentist statistics instead (discussed further in Section 4.1.1). These statistics were in general much easier to compute than their Bayesian counterpart at the time and usage of Bayes' theorem gradually fell out of favour.

Though Bayes' Theorem did not find popularity again until the late 20th Century where our computational power increased enough to make Bayesian problems more tractable, before then, there were still pockets of scientists using it great aplomb. Alan Turing famously used a Bayesian approach to reduce the number of possible dials in his enigma machine he had to consider as new information came in and Arthur Bailey used it to determine financial rates data (see, [McGrayne \(2011\)](#) for details on both usages). In Astronomy and Cosmology itself, Bayesian methods have seen a significant uptick of usage in recent decades (see Fig. 1 of [Trotta \(2008\)](#)). For a more complete historical context, I refer the reader again to [McGrayne \(2011\)](#).

4.1.1 Interpretations of Probability

Before I can extend our discussion further, it is important first to denote what a probability is. There are competing schools of thought here and even today the correct interpretation is the subject of much debate. For much of history, the dominant interpretation of probability was that of the 'frequentist' perspective. Within the frequentist framework, probability is defined as the relative frequency of an event compared to the number of trials performed in a sequence of repeated trials, in the limit where the number of trials tends to infinity. The typical example used to exemplify this interpretation is that of flipping a coin. For the frequentist interpretation, the probability of the coin toss being tails is given by the fraction of tosses that are tails in an infinite sequence of tosses.

This differs from the Bayesian interpretation of probability which defines probability as the 'degree of belief' in an outcome. This interpretation of probability is noted to bypass some of the pitfalls of the frequentist interpretation. One immediate consequence of this formulation is that it does not require repeated trials, useful for situations where a trial cannot be repeated. Perhaps the most relevant example of this is our Universe. There is only one universe so it is difficult to answer, in a principled sense, questions of the kind "*what is the probability our*

Universe is flat?”. Furthermore, [Trotta \(2008\)](#) argues the frequentist definition can be circular as it assumes a fixed probability in the repeated trials, which is difficult to justify as it is also the same quantity we are trying to estimate. It also not clear often how many trials is sufficient to get a valid estimate of the probability, since infinite trials is clearly not feasible.

4.1.2 Bayes’ Theorem

The Bayesian interpretation of probability views probability as a degree of belief about the value of a parameter given our current state of information. In this context, parameters are random variables that are conditional on our current state of knowledge and have their own probability distributions which encode our uncertainty of the parameter’s value.

The Bayesian interpretation of probability can be shown to come about naturally while following the axiomatic rules of probability related to conditional probability. Consider first, the event A which has probability $p(A)$ of occurring and the event B which has probability $p(B)$ of occurring. Here the notation writes $p(x)$ as shorthand for $p(X = x)$ which is the probability that a random variable X has the value x . Where x is discrete (continuous), the function $p(x)$ is the probability mass function (probability density function), with the requirement that the summation (integration) over the whole range of allowed values of x , $\sum_i p(x_i)$ ($\int p(x)dx$), is 1.

If you wanted to define the probability $p(A|B)$, which is the probability of event A occurring given the event B has happened, the multiplication law of probability allows you to define the conditional probability as a function of the joint probability $p(A, B)$ to give you

$$p(A|B) = \frac{p(A, B)}{p(B)}. \quad (4.1)$$

It can be seen this is consistent with the Bayesian definition of probability wherein our knowledge of A is conditionally dependent on information given by B and A is a random variable given by a probability distribution

Similarly $p(B|A)$ can be defined as

$$p(B|A) = \frac{p(A, B)}{p(A)}. \quad (4.2)$$

By noting the common term $p(A, B)$ (which happens because of the transitivity rule $p(A, B) = p(B, A)$) we can then combine both equations to get

$$p(A, B) = \frac{p(A, B)}{p(B)} = \frac{p(A, B)}{p(B)}. \quad (4.3)$$

This is easily rearranged to give the final Bayes' Theorem equation

$$p(A|B) = \frac{p(B|A)p(A)}{p(B)}. \quad (4.4)$$

The power in using Eq. (4.4) becomes apparent via a simple remapping of labels. Suppose you had observed some data, d , and a hypothesis for a model, M , that takes in n parameters, $\theta = \{\theta_1, \dots, \theta_n\}$. A natural question you may wish to ask is “what is the probability of the parameters θ given the observed data d ?”. In other words, if one wishes to know $p(\theta|d)$. Via the remapping $A \rightarrow \theta$ and $B \rightarrow d$, we can rewrite Bayes' theorem as

$$p(\theta|d) = \frac{p(d|\theta)p(\theta)}{p(d)}. \quad (4.5)$$

which immediately gives us the answer we require provided we can define the three terms on the right. It is also often common to make explicit here that our hypothesis on the parameters θ depends on our model M by adding in this extra conditional to write

$$p(\theta|d, M) = \frac{p(d|\theta, M)p(\theta|M)}{p(d|M)}. \quad (4.6)$$

The term $p(\theta|d, M)$ is called the joint ‘posterior probability’ of the parameters θ and is the main quantity of interest in the Bayesian approach. $p(\theta|M)$ is known as the ‘prior’ which denotes the

prior belief one has about the parameters of the model before the data is observed (discussed further in Section 4.1.3). $p(d|\theta, M)$ is known as the ‘Likelihood’ and is the probability of the data occurring conditioned on the set of parameters θ and our model M . For the frequentist approach, maximisation of this quantity is the main goal. It should be noted that the knowledge provided from this likelihood quantity is different from what is typically the main quantity of interest in scientific enquiries where inferring that a dataset would be unlikely under a particular hypothesis is not the same as inferring how likely that hypothesis is, the latter of which lines up more intuitively with what one would want as the result of a scientific enquiry and is given by the Bayesian posterior density.

One feature of this Bayesian formulation is the ease by which one can marginalise out nuisance parameters. Nuisance parameters are often labelled as such due to their ability to influence data but being of no interest for inference. One of the most frequent examples in astrophysics is the problem of dealing with background source counts, b , that makes the true source count uncertain, s . Dealing with them within a Bayesian context is a simple process which involves inferring the joint probability distribution of s and b , $p(s, b)$, and then b can be removed from the joint density via marginalisation.

The rules of marginalisation can be derived from the simple axioms of probability theory. The marginalisation rule states

$$p(A) = \sum_i p(A, B_i) \quad (4.7)$$

where in the continuous limit this is

$$p(A) = \int p(A, B)dB. \quad (4.8)$$

Here B can be considered the nuisance parameter and A the main parameter of interests. The effect of B on the joint density can be removed by summing over all possible values of B as described in Eq. (4.7) or Eq. (4.8) to get a probability density that only depends on A . This integration is not always easy to perform over a large number of dimensions, though, as will be discussed later, this is easy to do numerically when one has samples proportional to the

joint density. One benefit of this process of marginalisation is that the uncertainty associated with this unknown nuisance parameter is propagated onto the parameters of interest in the posterior. Fixing the nuisance parameter to a reasonable value, as is common in some analyses, will incorrectly account for the uncertainty by generally underestimating it and will lead to a posterior that is narrow and an inaccurate representation of the state of knowledge of the system.

One further feature of the Bayesian formulation is that in many cases it can be shown that the frequentist results are a special case of the Bayesian case under some implicit assumed uniform prior. The Bayesian formulation simply makes this implicit prior explicit.

4.1.3 Priors

As can be seen from equation Eq. (4.6), one of the vital ingredients for using Bayes' formula is $p(\theta|M)$. This quantity is called the 'prior' and denotes our prior belief about the parameter of interest before the data has been seen and may even be informed by the results of a previous experiment which is the cause of the famous quote by Lindley "yesterdays posterior is today's prior" (Lindley, 2000). The prior is the subject of much criticism of the Bayesian approach as it implies the posterior is a consequence of a subjective state of knowledge as opposed to an objective one and is therefore subjective itself. However, priors also allow us to restrict the support for the likelihood to 'sensible' regions when there is relevant external information e.g., in an experiment where you are trying to determine the mass of an object, relevant external information is that measured masses are positive. Furthermore, it can be shown under different subjective prior assumptions, repeated application of Bayes' theorem will cause asymptotic convergence to a common posterior density which is objective as required (provided the initial prior is non-zero where the likelihood is large). Furthermore, under this asymptotic limit, Bernstein-von Mises theorem states the posterior converges to a multivariate normal distribution centered on the Maximum Likelihood Estimate (MLE) (Vaart, 1998). One importance of this is that, due to the asymptotic correctness, it links Bayesian and frequentist inference together and given sufficient data, one can make valid frequentist statements about estimation

and uncertainty on parameters from the Bayesian posterior distribution.

Another notion that is important for priors is the idea of ‘proper priors’. A valid probability distribution is one where $\int p(x)dx = 1$. Priors that obey this rule are known as ‘proper priors’ and those that do not are ‘improper priors’ (Sivia & Skilling, 2006). For the purposes of model comparison (discussed later in section 4.1.5), properly normalised priors are a requirement and is also a prerequisite to having a properly normalised posterior density in Bayes’ formula.

Despite the subjectivity that can be involved in choosing priors, an attempt can be made however to construct priors mathematically to make them minimally informative and remove this subjectivity. Care needs to be taken here when constructing these uninformative priors. While on an initial glance, a uniform prior may make for a suitably uninformative prior where $p(\theta) \propto 1$, without upper and lower bounds on the prior, it does not have a finite integral and therefore cannot be a proper prior. Imposing these lower and upper bounds however, makes them no longer uninformative as the range of allowed values is now restricted. The uniform prior is also parameterisation-dependent, so under a change of variables, say from θ to ϕ , the prior on $p(\phi)$ can be informative if the transformation is non-linear. Finally, in a larger number of dimensions, D , having a uniform prior on all the parameters will cause most of the samples of the prior to concentrate in a D -dimensional shell with constant variance, causing the ‘concentration-of-measure’ phenomenon. On the other hand, the volume is concentrated in the corners of the hypercube meaning a high dimensional uniform prior can also end up highly informative (see Trotta (2008) for more details).

One of the approaches to constructing priors that are uninformative was given by Jeffreys (1939). These are sometimes called ‘Jeffrey’s priors’ where the prior on θ can be given by

$$p(\theta) \propto \sqrt{I_F(\theta)} \quad (4.9)$$

where $I_F(\theta)$ is the Fisher Information and is given by

$$I_F(\theta) = - \int \left(\frac{d^2}{d^2\theta} \log p(d|\theta, M) \right) p(d|\theta, M) dd \quad (4.10)$$

The Jeffrey's prior has the property that it is invariant under reparameterisation which makes them a common feature in Bayesian analysis. However, they are not always possible to normalise so care needs to be taken when using them.

Another alternative set of priors called 'reference priors' have also been proposed ([Bernardo, 1979](#)) which is based on information theory. The method of constructing reference priors is such that they maximise the gain in information gain as you update your prior from your posterior. This information gain is frequently measured using the Kullback-Leibler (KL) divergence, D_{KL} ([Kullback & Leibler, 1951](#)), which is defined as

$$D_{KL}[p(\theta|M)||p(\theta|d, M)] = \int p(\theta|M) \log \frac{p(\theta|M)}{p(\theta|d, M)} d\theta. \quad (4.11)$$

and is sometimes known as the 'relative entropy'. However since the prior represents your belief before you observe the data and the true posterior, one usually maximises instead the expectation of the KL divergence between the posterior and prior when averaged over the data distribution instead of using the data directly to maximises the information gain. In other words,

$$I(p(\theta|M)) = \int D_{KL}[p(\theta|M)||p(\theta|d, M)] p(d) dd. \quad (4.12)$$

The prior that maximises this quantity is the reference prior. Typically this can be difficult to compute however due to the high-dimensionality of the integral, though numerical algorithms exist that can reduce the dimensionality of the problem for efficient computation (see section 3 of [Dey \(2005\)](#)).

Certain combinations of likelihoods and priors that are 'conjugate distributions' can yield analytical posterior distributions. This involves defining a 'conjugate prior'. When combining a conjugate prior with an appropriate likelihood, the posteriors of these models have the same probability distribution as the prior, but with different hyperparameters. In this scenario, the posterior and prior are referred to as conjugate distributions. This means we can analytically write the posterior distributions in a closed form. Furthermore, for these distributions it is usually easy to calculate summary statistics of interest such as the mean and variance which

makes them useful. The downside is that one is restricted in their choice of likelihood and prior which may not be suitable for modelling the problem of interest.

4.1.4 Parameter Inference

One of the most common uses of Bayes' theorem is for parameter inference for a given model. This could be for example, inferring the parameters that constrain a Λ CDM model. In this scenario, if one is not interested in model comparison, one can usually drop the normalising constant of Bayes' Theorem and it is common to write it in the form

$$p(\theta|d) \propto p(d|\theta)p(\theta) \quad (4.13)$$

where I have dropped the explicit conditioning on the model M for clarity and it is implicit instead since for parameter inference purposes the model is fixed. A consequence of this is that the right hand side of Eq. (4.13) may no longer integrate to unity and be a valid probability density. For the purposes of parameter inference though, this is sufficient. Another consequence of the fixed model is that the denominator $p(d|M)$ is a constant which is why it can be neglected in this form of Bayes' theorem.

One can also combine independent probes of Λ CDM model, such as SNIa and weak lensing, by taking the product of the likelihoods to create a joint likelihood, providing they are independent probes and uncorrelated, or this correlation will need to be taken into account in the likelihood (see [Krause et al. \(2017\)](#)).

With the likelihood and priors defined, parameter inference can be done. The process for doing so usually involves drawing samples from the posterior or something proportional to the posterior in the case of Eq. (4.13). The sampling used for such purposes are typically Markov Chain Monte Carlo (MCMC) or Nested sampling schemes (discussed further in section 4.2).

4.1.5 Bayesian Evidence and Model Comparison

One feature of Bayes' theorem is that a simple route for model comparison is provided, where model comparison is defined as computing the relative probability between two models. Unlike in the case of parameter inference, the normalisation of the posterior needs to be computed for model comparison. This requires the computation of the denominator in Eq. (4.6), $p(d|M)$. This term is usually referred to as the 'Bayesian evidence' or the 'model likelihood' but can also be known as the 'marginal likelihood' because it can be computed by taking the numerator of Eq. (4.6), $p(d|\theta, M)p(\theta|M)$, and marginalising out the dependence on θ such that

$$p(d|M) = \int p(d|\theta, M)p(\theta|M)d\theta. \quad (4.14)$$

This expression makes clear why the Bayesian evidence is the normalising constant to create a valid posterior density, because it is the integral over the numerator. This allows the Bayesian evidence to be thought of as the average of the likelihood over the prior space for the model of choice. It is clear here the importance of the priors and how we choose them as it can have a significant impact on the final value of the computed evidence. The integral in Eq. (4.14) is in general difficult to compute numerically when the dimensionality of θ is high. Methods for computing this quantity are discussed further in section 4.2.

The importance of this quantity in model comparison is the reason the normalising constant can no longer be ignored like in parameter inference and the normalisation of the posterior is required. To understand why, consider the probability of M_i being the true model conditioned on the observed data d . This conditional probability can be given by application of the Bayes' rule as

$$p(M_i|d) = \frac{p(d|M_i)p(M_i)}{p(d)}. \quad (4.15)$$

As before, we can rewrite the denominator by using the marginalisation rule to get

$$p(M_i|d) = \frac{p(d|M_i)p(M_i)}{\sum_j p(d|M_j)p(M_j)} \quad (4.16)$$

where to compute the denominator in Eq. (4.16) we summed over all possible models M_j to compute the normalising term $p(d)$. However, the possible models to consider can be infinite so computing this quantity is not practical in most scenarios. Instead, within the Bayesian formulation the preferred route is to compare the ratio between the probability of two models M_1 and M_2 conditioned on the observed data such that

$$\frac{p(M_1|d)}{p(M_2|d)} = \frac{p(d|M_1)p(M_1)}{p(d|M_2)p(M_2)}. \quad (4.17)$$

In this scenario, the difficult to compute denominator from Eq. (4.16) has now cancelled, leaving behind only the ratio of the model evidences (the normalising quantity in Eq. (4.6)) and the ratio of the model priors. It is common to call the ratio of the model evidences the ‘Bayes factor’, B_{12} , where

$$B_{12} = \frac{p(d|M_1)}{p(d|M_2)}. \quad (4.18)$$

In the scenario where each model is equally probably, the model prior ratio becomes unity. Therefore, the relative probability of the of the models is given by this Bayes factor. The quantity B_{12} then states the degree of support for M_1 over M_2 . Often this is measured on a qualitative scale known as the ‘Jeffrey’s Scale’ and is shown in Table 4.1

Table 4.1: Summary of the Jeffrey’s scale for evaluating the Bayes’ Factor. The table and thresholds have taken from Trotta (2008). The odds indicate how favoured model M_1 is over M_2 .

$ \ln B_{12} $	Odds	Strength of Evidence
< 1.0	$\lesssim 3 : 1$	Inconclusive
1.0	$\sim 3 : 1$	Weak evidence
2.5	$\sim 12 : 1$	Moderate evidence
5	$\sim 150 : 1$	Strong evidence

One feature of this form of model comparison is its ability to quantify the quality of the data fit, but also penalise the model complexity based on the number of free parameters to reduce the chance of overfitting. Models with fewer parameters are preferred because more parameters leads to a larger prior volume which reduces the average of the likelihood over the prior space and therefore reduces value of $p(d|M)$ when evaluating it. However, this only applies

to parameters that can be constrained by the model. Parameters that are not constrained by the model are not penalised in the Bayes' Factor. This therefore requires the goodness-of-fit of the data to be better if this more complex model is to be preferred over a simpler model with fewer parameters. This is in accord with 'Occams razor' or the 'principle of parsimony' which states that a simpler model should be preferred if it explains the data well (Sivia & Skilling, 2006; Ariew, 1976).

To remove the dependence on the model when inferring the posterior, $p(\theta|d, M)$, one can also consider the idea of 'model averaging' (see Fragoso et al. (2017)) where the marginalisation rule is used such that

$$p(\theta|d) = \sum_i p(\theta|d, M_i)p(M_i|d). \quad (4.19)$$

This implicitly assumes you have exhaustively summed over all possible models M_i and ensures the uncertainty in the model is incorporated in the parameters of interest.

4.2 MCMC Algorithms and Nested Sampling

As discussed in Section 4.1.2, unless the posterior has a closed form solution, for example, by using conjugate priors and likelihoods, then the posterior is usually inferred through some sampling scheme. The literature on sampling schemes is extensive, but for the purposes of this thesis I restrict the discussion to Markov Chain Monte Carlo (MCMC) schemes and Nested sampling schemes which are the most common forms of sampling in Bayesian contexts and the latter is of the type used in the research exhibited in this thesis. Many variants of MCMC exists, each designed with different properties in mind. I outline below the salient properties of MCMC schemes that are common to all.

In MCMC methods, the goal is to produce a series of samples that form what is called a 'Markov chain' that converges to an equilibrium distribution that matches that of our target distribution, in this case, the posterior density $p(\theta|d, M)$.

A defining feature of a Markov chain is that it satisfies the Markov property. This

property states that in a sequence of random variables X^i for $i = \{1, \dots, N\}$, the next value in the sequence only depends on the current value. In other words

$$p(X^i | X^1, \dots, X^{i-1}) = p(X^i | X^{i-1}). \quad (4.20)$$

In the context of our Bayesian formulation, the MCMC chains takes steps in the parameter space defined by θ whilst following this property. The step itself defined by the transition probability $T(\theta^i, \theta^{i+1})$ which gives the probability of your next step θ^{i+1} conditioned on the current value θ^i .

To know when the Markov chain has converged and become stationary one necessary (but not sufficient) condition is that the chain should obey the ‘detailed balance condition’ property which states

$$p(\theta^i)T(\theta^i, \theta^{i+1}) = p(\theta^{i+1})T(\theta^{i+1}, \theta^i). \quad (4.21)$$

One of the properties of Markov chains is the ‘ergodic’ property which ensures that given a sufficient number of samples, the chain should eventually converge to the target distribution.

For MCMC in practice, there are some numerical considerations that need to be heeded:

- While ergodicity of the chains guarantees that it will eventually converge to the target distribution, if a poor starting point is chosen, the final chain can be unrepresentative of the target distribution. As a remedy to this, some fractional percentage of the first samples are discarded ([South et al., 2021](#)). This is frequently referred to as the ‘burn-in’ period.
- Successive chains in the sample are generally correlated because of the Markov property. If independent samples are required, then a process known as ‘thinning’ can be used where a sample is taken every n points ([Gelman et al., 2013](#))
- If the posterior being sampled can have multiple peaks the chain can get stuck in a local maxima. Different MCMC algorithms handle this in different ways, but in general, proper exploration of the parameter space and experimentation with MCMC algorithms

is required to get around this.

- Checking the chain has converged can in general be difficult to assess but several procedures and tests exist (see [Gelman et al. \(2013\)](#) for more information).

Whereas usually the marginal distribution for a single parameter $p(\theta_i|d)$ is quite difficult to obtain, having access to a set of samples from an MCMC chain makes this significantly easier. Consider, if we wanted the marginal distribution $p(\theta_1|d)$. One would need to perform the following integral

$$p(\theta_1|d) = \int p(\theta|d) d\theta_2 d\theta_3 \dots d\theta_n \quad (4.22)$$

This is numerically difficult for large n . As the Markov chain contains samples from the full posterior, their density reflects the density of the full posterior. Therefore, the marginal distribution for the parameters of interest can easily be obtained via numerical marginalisation by binning θ_1 from the chain and counting the number of samples for the relevant parameter that fall in each bin to construct the density of $p(\theta_1|d)$. A similar procedure can be used to construct any m -dimensional marginal distribution where $m < n$.

Many implementations of MCMC schemes exist such as Metropolis-Hasting ([Metropolis et al., 1953](#); [Hastings, 1970](#)), Gibbs ([Geman & Geman, 1984](#)), Hamiltonian Monte Carlo ([Duane et al., 1987](#)) and more.

4.2.1 Nested Sampling

While MCMC algorithms are prevalent in Bayesian analysis, for this thesis we use a sampling algorithm that is not MCMC. Therefore, I limit our discussion here to the one primarily employed in the research presented in this thesis which is ‘Nested’ sampling ([Skilling, 2006](#)) and one of its public implementations and variations of it, `MultiNest` ([Feroz & Hobson, 2008](#); [Feroz et al., 2009](#)). The MCMC implementations mentioned, are primarily targeted at parameter inference as opposed to model comparison and are not suited to calculation of the Bayesian evidence for model comparison and often extra computationally expensive steps are taken to do so.

Examples of these steps include Thermodynamic integration with annealing, using the Laplace approximation if the likelihood is unimodal, or the Savage-Dickey Density Ratio (SDDR) for the case of nested models (see [Annis et al. \(2019\)](#); [Heck \(2018\)](#); [Trotta \(2008\)](#) for more detail).

In contrast to these methods, nested sampling follows an ‘evidence first’ approach whose primary goal is to compute the Bayesian evidence, but as a byproduct generates chains that can also be used for parameter inference. I provide here a high level overview that is the core of most nested sampling implementations and follows the procedure outlined by [Skilling \(2004, 2006\)](#).

Consider first a likelihood function $\mathcal{L}(\theta)$, a prior distribution, $\pi(\theta)$, and Bayesian evidence, Z which is the normalising constant of our posterior. This can be used to define the posterior density function

$$p(\theta|d) = \frac{\mathcal{L}(\theta)\pi(\theta)}{Z}. \quad (4.23)$$

The method behind nested sampling involves reparameterising the evidence calculation in terms of a new parameter \mathcal{X} which denotes the fractional prior volume above some likelihood threshold λ . For this purpose and ease of calculation, the likelihood is sorted to be monotonically decreasing as a function of \mathcal{X} which varies from 0 to 1. We can define this fractional volume as

$$\mathcal{X}(\lambda) = \int_{\mathcal{L}(\theta) > \lambda} \pi(\theta) d\theta. \quad (4.24)$$

The likelihood can then be rewritten as

$$\mathcal{L}(\mathcal{X}(\lambda)) \equiv \lambda. \quad (4.25)$$

The result of this reparameterisation is that we can restate the evidence integral as

$$Z = \int_0^1 \mathcal{L}(\mathcal{X}) d\mathcal{X} \quad (4.26)$$

which is now a much simpler 1-dimensional integral as opposed to the potentially difficult to compute multidimensional one in θ . The task here is now to approximate $\mathcal{X}(\lambda)$ in a computa-

tionally efficient manner.

In nested sampling a set number of random points (sometimes known as the ‘live points’) are drawn from the prior for which a corresponding value of \mathcal{X} is calculated and the likelihood for each point is evaluated. Then, the lowest likelihood point is discarded (but the value of \mathcal{X} is kept recorded) and a new point is sampled from the prior with the condition that its likelihood has a greater value than the point discarded. Via this manner, \mathcal{X} is probabilistically and geometrically compressed from 1 to 0 and the sampling concentrates around regions with the greatest likelihood. After some stopping criterion is reached (dependent on the implementation), the sampling is stopped and the integral is easily evaluated numerically, such as by using a trapezoidal rule as initially done by [Skilling \(2004, 2006\)](#).

The main difficulty from the above procedure is sampling points from the prior with hard likelihood constraint that the new likelihood is greater than a likelihood boundary (called an iso-likelihood contour) set by the previously discarded point. As the scheme progresses, this becomes exponentially harder if one samples naively from the prior as more samples from the prior are discarded before a valid prior sample can be found. Efficient sampling to get around this issue is what differentiates the different implementations of nested sampling.

We discuss here the specific of the algorithm associated with **MultiNest** ([Feroz & Hobson, 2008](#); [Feroz et al., 2009](#)) which is employed for much of the analysis within this thesis. The **MultiNest** sampler works by using ellipsoidal rejection sampling. The procedure for this works by sampling from a restricted space formed by an ellipse around the current samples rather than the prior itself. To not bias the sampling, a collection of ellipses approximating the likelihood boundary are generated. This collection of ellipses is generated by using a k-means clustering analysis on a set of the previous samples. New sample are drawn from these ellipses. Samples are then rejected until the likelihood boundary is respected, replacing the lowest likelihood point which is removed and stored as before in the general algorithm outline. Because of the restricted sampling space created by these ellipsoids, this rejection occurs less frequently than when compared to naively sampling the prior. An extra benefit of using multiple ellipsoids is that the algorithm can naturally handle multimodal posteriors as each mode can be assigned

different ellipsoids to sample from. This is a feature that traditional MCMC algorithms struggle with as going from one mode to another requires the samples crossing a low-probability regions between modes which is difficult to do.

Other popular implementations of nested sampling used in cosmological contexts include **PolyChord** (Handley et al., 2015), **DyNesty** (Speagle, 2020), **UltraNest** (Buchner, 2021).

4.3 Bayesian Hierarchical Modelling of SNIa

The mechanics of how Bayesian models work mean that it is easy to propagate uncertainty and correlations between parameters in a principled manner to the final inference and posterior distribution. This opens up the avenue to do modelling at multiple levels, where the probability distributions can be assigned at each level and uncertainty associated with parameters at each level can be propagated forward to intermediate levels (which may not be of interest) and the final level. In essence, a ‘hierarchy’ is built where each level has conditional distributions dependent on the values of the previous level and this sort of model is known as a Bayesian Hierarchical Model (BHM). The method takes advantage of the fact that hyperparameters of our priors can have probability distributions assigned to them with hyperpriors to build the ‘hierarchy’.

One advantage of this hierarchical approach is that by using information across groups of observations at different levels, inference about individual samples, or lower-level parameters, can be informed by one another causing the marginal posteriors to tighten. This phenomenon is often referred to as ‘shrinkage’ or ‘borrowing strength’ (Gelman et al., 2013) makes BHM’s suitable for modelling populations. March et al. (2011) also shows that when the errors on independent variables are large, and the sample size is relatively small, then BHM’s lead to smaller biases in the inferred parameters compared to the frequentist methods.

However, one benefit of frequentist methods over the Bayesian method is their smaller computational requirements. Bayesian models are also typically more sensitive to the model description and can be more biased if incorrect assumptions are used. One set of assumptions here is the

priors which must be properly justified, especially when sample sizes are small owing to the potentially large impact these have on the final posterior.

Bayesian hierarchical models are often presented as ‘Directed Acyclic Graphs’ (DAGs), which are graphical diagrams that show the conditional dependence of parameters in the model. The edges connecting two nodes (parameters) of the DAG show the two nodes are conditionally dependent. As the graph is ‘directed’ this also shows which way the conditional dependence is. This means DAG’s efficiently provide clarity on a model and its dependencies in visual manner while also providing a logical structure that is easily interpreted by machines and programming frameworks.

4.3.1 The χ^2 Statistic

Before we discuss the BHM, BAHAMAS, that is used for SNIa analysis, a brief detour is taken to discuss the χ^2 based method that is also commonly used for inference as a comparison point.

As stated before, the likelihood is the probability of the data occurring conditioned on the set of parameters θ and our model M . One of the most common forms of the likelihood that is constructed in frequentist frameworks is the Gaussian likelihood. Consider a collection of N measurements $\hat{d} = \{\hat{d}_1, \dots, \hat{d}_n\}$, with an associated error matrix Σ which can be assumed to be known and constant. Here in this thesis, variables with a hat symbol indicate it is a ‘measured’ value. Given that we have a model M and some model parameters $\theta = \{\theta_1, \dots, \theta_N\}$ that it takes as input, we can define the probability of the data occurring under this model and parameters i.e, the likelihood, constructed as a Gaussian as

$$p(\hat{d}|\theta) = \frac{1}{|2\pi\Sigma|^{-1/2}} \exp\left(-\frac{1}{2}\chi^2\right) \quad (4.27)$$

where the term is χ^2 is

$$\chi^2 = (\hat{d} - d(\theta))^T \Sigma^{-1} (\hat{d} - d(\theta)) \quad (4.28)$$

and $d(\theta)$ is the theoretical value of d expected for a given set of θ . Under frequentist frameworks,

the general goal is to find the θ that maximises the likelihood, $p(\hat{d}|\theta)$. For a constant error matrix Σ , this is equivalent to minimising the value of χ^2 . The term χ^2 follows the chi-square distribution with ν degrees of freedom (*d.o.f*) (Wilks, 1938; Casella & Berger, 2001). ν is defined as $\nu = n - N$. It can be shown the mean of this chi-square distribution is ν . Values of θ that are derived from fitting Eq. (4.28) is expected to fall close to the mean such that $\chi^2/\nu \approx 1$.

4.3.2 BAHAMAS

In this thesis, we adopt a BHM for the statistical analysis of SNIa data, called BAHAMAS (March et al., 2011; Shariff et al., 2016). The model in this case is ‘hierarchical’ because it refers to a layer of unobserved (so-called ‘latent’) variables for each SNIa, corresponding to the true value of their light-curve-derived properties (as opposed to the noisy measured value). The latent variables are marginalised over in the inference, and are constrained in virtue of the fact that they are all generated from the same underlying population distribution, which is modelled with a set of hyperparameters, themselves determined from the data.

This Bayesian methodology for SNIa analysis pioneered in Mandel et al. (2009); March et al. (2011); Mandel et al. (2011) has been adopted and extended in several other papers, including e.g., UNITY (Rubin et al., 2015), STEVE Hinton et al. (2018) and Simple-BayeSN (Mandel et al., 2017) (see also Nielsen et al. (2016) for a profile likelihood interpretation). Here, we briefly summarize BAHAMAS.

We denote with a hat symbol observed quantities, in order to distinguish them from the latent (i.e., unobserved) variables in our model. For each SNIa i , the data d_i can be summarised by a vector

$$d_i = \{\hat{z}_i, \hat{c}_i, \hat{x}_{1_i}, \hat{m}_{B_i}, \hat{C}_i\}, \quad (4.29)$$

where \hat{z}_i is the observed redshift, \hat{m}_{B_i} is the observed peak B-band apparent magnitude, \hat{x}_{1_i} and \hat{c}_i are observed ‘stretch’ and ‘colour’ corrections, which are summary statistics of the lightcurve of the SNIa obtained with the lightcurve fitter SALT2 (Guy et al., 2005; Guy et al., 2007) during the standardization procedure. Furthermore, $\hat{C}_i = \text{Cov}(\hat{c}_i, \hat{x}_{1_i}, \hat{m}_{B_i})$ is a 3×3

variance-covariance matrix that describes the measurement error on the observables.

In BAHAMAS, latent variables for each SNIa are introduced, in order to model each source of uncertainty according to its origin: measurement error, population scatter and intrinsic (residual) variability. A probabilistic hierarchical model is built as follows: each SNIa has latent variables M_i^ϵ , x_{1_i} and c_i , representing the objects' 'true' (i.e., noiseless) absolute magnitude, stretch correction and colour correction, respectively. These latent variables follow normal distributions, representing population variability of the SNe and parameterised by their means and variances:

$$x_{1_i} \sim \mathcal{N}(x_{1\star}, R_{x_1}^2), \quad (4.30)$$

$$c_i \sim \mathcal{N}(c_\star, R_c^2), \quad (4.31)$$

$$M_i^\epsilon \sim \mathcal{N}(M_0^\epsilon, \sigma_{\text{res}}^2) \quad (4.32)$$

where $x_{1\star}$, c_\star and M_0^ϵ are the population means and $R_{x_1}^2$, R_c^2 and σ_{res}^2 are the population variances, all of which are also estimated from the data. We collect the population-level parameters in a vector of variables $\vartheta \equiv \{x_{1\star}, c_\star, M_0^\epsilon, R_{x_1}^2, R_c^2, \sigma_{\text{res}}^2\}$. The intrinsic magnitude of each SNIa M_i , is modified by applying the linear 'Tripp relation' (Tripp, 1998), so that $M_i \rightarrow M_i^\epsilon \equiv M_i + \alpha x_{1_i} - \beta c_i$, where the set of regression coefficients $\mathcal{B} = \{\alpha, \beta\}$ are nuisance parameters that control the slope of the stretch and colour correction, respectively. Therefore, M_i^ϵ is a linear function of M_i that features a lower population variance, represented by σ_{res}^2 . This procedure is what allows us to standardise the SNIa's as mentioned in section 3.6 and reduce sufficiently the residual standard deviation of their peak absolute magnitude so that they can be used as luminosity distance indicators. Shariff et al. (2016) was able to apply BAHAMAS to determine their SNIa data had a mean residual standard deviation around the peak absolute magnitude of $\sigma_{\text{res}} \sim 0.1$, similar to the results obtained by Riess et al. (1998); Perlmutter et al. (1999). At the latent level, the apparent peak magnitude m_{Bi} is related to the standardised intrinsic magnitude M_i^ϵ via the distance modulus of Eq. (3.3):

$$m_{Bi} = \mu_i(\hat{z}_i, \ell) - \alpha x_{1_i} + \beta c_i + M_i^\epsilon. \quad (4.33)$$

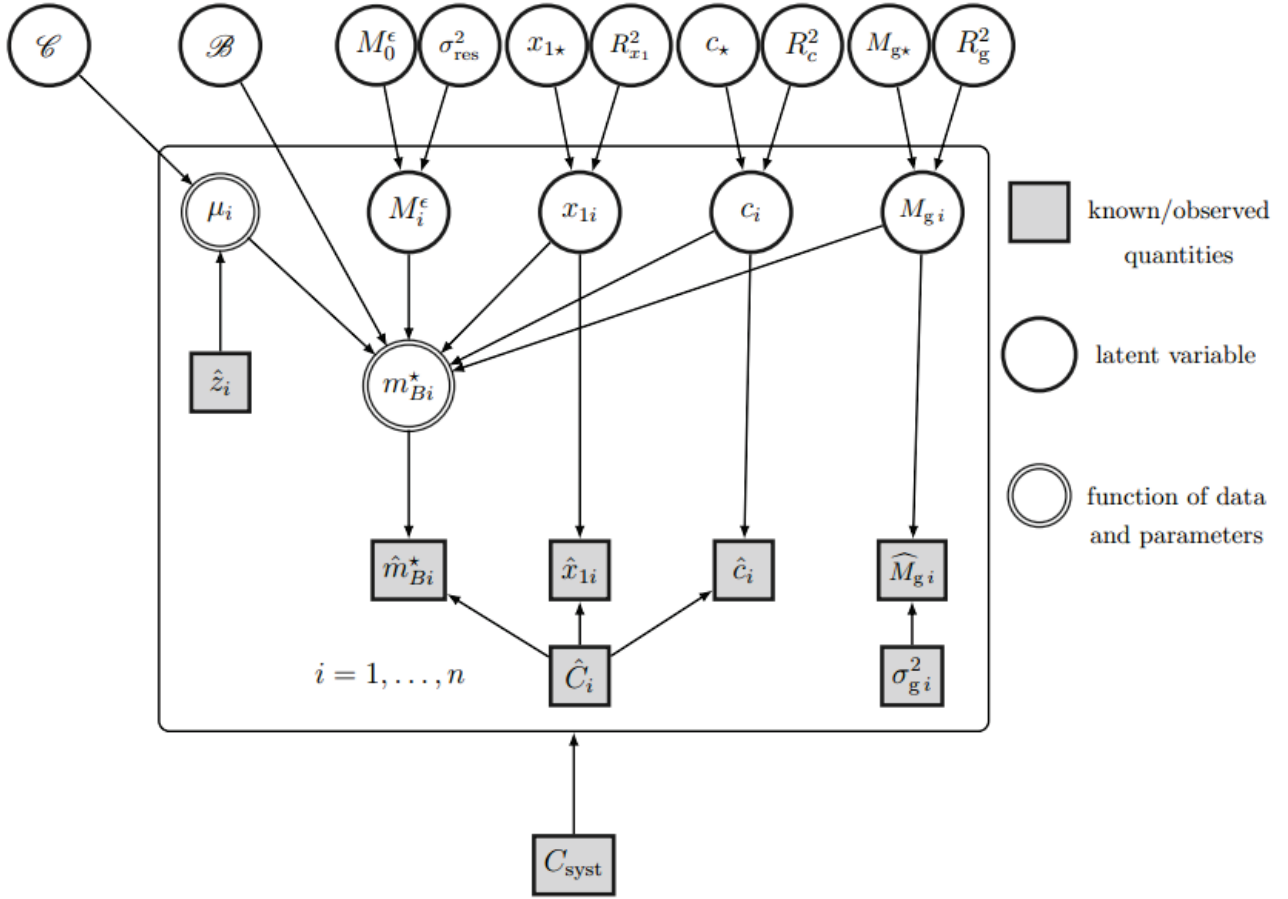


Figure 4.1: The DAG showing the BAHAMAS model. Figure adapted from Shariff et al. (2016)

Finally, the observed values of $\{\hat{m}_{Bi}, \hat{x}_{1i}, \hat{c}_i\}$ are modelled as normally distributed around their latent values, with variance-covariance matrix given by \hat{C}_i . Additionally, a systematic errors covariance matrix, C_{syst} , that correlates different SNIa (for example, because of calibration uncertainties common between SNIa within the same survey) is included when available. Inference is based on the marginal distribution of the quantities of interest, ℓ , which includes uncertainty at all levels of the hierarchy.

The DAG used by (Shariff et al., 2016) to depict the model is shown in Fig. 4.1. This DAG also shows how BAHAMAS has a module for including a host galaxy mass-step correction with the terms: $\hat{M}_{g,i}$, the host galaxy mass measurement, $\sigma_{g,i}$ the measurement error, $M_{g,i}$ the latent unobserved mass. The population parameters defining the normal distribution the latent galaxy mass is drawn from are $M_{g,\star}$ and R_g . Shariff et al. (2016) showed that adding a mass-step or a more general linear covariate as a function of host-galaxy mass has little impact on the

ensuing cosmological parameters inference. Therefore, in this thesis we do not adopt a mass-step correction.

We refer the reader to [March et al. \(2011\)](#) and [Shariff et al. \(2016\)](#) for the full mathematical detail of the BAHAMAS model, marginalization procedure and algorithms used for sampling the resulting posterior distribution. It is emphasised here that BAHAMAS is not a fully complete BHM as it still lacks features that could account for effects in the data such as those due to metallicity, host galaxy properties and selection effects (an issue which is addressed in Chapter 7) and thus requires traditional ‘ad-hoc’ corrections to the data still to account for these effects.

4.3.3 Comparison with the χ^2 method

In terms of the traditional chi-squared approach that has been followed by many authors ([Astier et al., 2006](#); [Kowalski et al., 2008](#); [Amanullah et al., 2010](#); [Conley et al., 2011](#)), the quantity to minimise is

$$\chi^2 = \sum_{i=1}^N \frac{\mu_i(\hat{z}_i, \ell) - \hat{\mu}_i}{\sigma_{\mu_i}^2} \quad (4.34)$$

Here $\sigma_{\mu_i}^2$ has three sources with

$$\sigma_{\mu,i}^z = (\sigma_{\mu,i}^{fit})^2 + (\sigma_{\mu,i}^z)^2 + \sigma_{res}^2. \quad (4.35)$$

These three sources are:

- $\sigma_{\mu,i}^{fit}$ which is the fitting error from the light curve fits. Typically computed as $\Psi^T C_i \Psi$ with $\Psi = \{-\beta, \alpha, 1\}$. It is easy to see from this definition that the global fit parameters α and β therefore appear in both the numerator and denominator of Eq. (4.34).
- The redshift error, $\sigma_{\mu,i}^z$, which is the measurement error of the redshift given by the host galaxy and is usually comprised of error due to peculiar velocity uncertainty in spectroscopic measurements.
- The residual dispersion, σ_{res} , that remains after corrections for stretch and colour. As discussed in chapter 3.7.1, this residual dispersion can have many physical sources. It

is important to note that this term is not known ‘a priori’ and is inferred during the parameter estimation process (see [Vishwakarma & Narlikar \(2010\)](#)).

Inference in the χ^2 approach involves minimising by sampling over the parameters space and simultaneously attempting to fit the cosmological parameters of interest ℓ and the global population parameters $\{\alpha, \beta, M_o\}$.

This traditional approach however has been shown to have certain downsides which using a BHM mitigates. In addition BAHAMAS exhibits extra advantageous properties. These downsides of the traditional approach and the BHM advantages include

- The parameter estimation is usually iterative and involves trying to achieve $\chi^2/d.o.f \approx 1$. Authors such as [Astier et al. \(2006\)](#); [Wang et al. \(2006\)](#); [Kowalski et al. \(2008\)](#) realised however this can lead to large biases in the recovered value of some parameters such as β with the bias being as large as 6σ according to simulations ([March et al., 2011](#)). The source of this bias has been traced to the observation that the error on the colour correction parameter c_i is as large as or larger than the width of the distribution of values of c_i especially for high-redshift SNe. This is solved by BAHAMAS because of its hierarchical model that propagates the errors properly at each level.
- The assumption of gaussianity behind the χ^2 method is not always statistically justified. A Bayesian model allows an easy breakaway from this assumption by simply modifying the likelihood as required.
- The Tripp relation regression coefficients, α and β , appear in both the numerator and denominator of the χ^2 expression. Therefore there is a statistical problem with trying to estimate both the location and variance of the population parameters for stretch and colour.
- This approach has been shown to have better coverage statistics overall than the χ^2 method traditionally employed, and leads to a reduction of mean squared errors for the recovered cosmological parameters by a factor of $\sim 2 - 3$ when deployed on simulated data ([March et al., 2011](#)).

- As part of the χ^2 procedure σ_{res} is adjusted to obtain $\chi^2/d.o.f \approx 1$. Adjusting σ_{res} to obtain the goodness-of-fit makes it difficult to carry out further goodness-of-fit tests on the model itself. The Bayesian version on the other hand provides an easy route for model comparison via computation of the Bayesian evidence.
- The interpretation of the 95% credible interval in Bayesian methods is more statistically justified compared to the 95% confidence interval quoted in frequentist limits which is only valid in the limit where the number of times an experiment is repeated goes to infinity, something that is not possible in cosmological contexts.
- The frequentist results can be recovered by Bayesian methods when using uniform priors as in this case the posterior is proportional to the likelihood and therefore maximising the the posterior is equivalent to maximising the likelihood as done in frequentist contexts.

It should be noted however, that despite extra model complexity involved in frequentist methods to solve issues that have easy solutions in Bayesian contexts, computation is typically very quick in a frequentist framework which is why it has been historically favoured. In Bayesian methods the trade off is often the opposite where specifying a model in a principled manner is easier, but the computation is harder. With recent advances in computation power this is less of an issue these days however. Because of this, recently in the SNIa cosmology space, there has been a great increase of Bayesian Hierarchical Models for SNIa cosmology. To name a few, STEVE ([Hinton et al., 2018](#)), UNITY ([Rubin et al., 2015](#)), BEAMS ([Kunz et al., 2007, 2013](#)), zBEAMS ([Roberts et al., 2017](#)), BayesSN ([Mandel et al., 2011](#)), Simple-BayeSN ([Mandel et al., 2017](#)), BayeSN-SED ([Mandel et al., 2021](#)) and many more, each attempting to solve different aspects of the pipeline such as cosmological fits, types of SNIa data, light curve fitting or handling of non-SNIa contaminants.

An extra point of note here is that BAHAMAS in its current form neglects the effects of missing data i.e., selection effects, caused by Malmquist bias ([Malmquist, 1922](#)). It relies on being given data has that already corrected for this effect. A more complete description of selection effects and accounting for it within BAHAMAS is provided in Chapter 7.

Chapter 5

New Constraints on Anisotropic Expansion from SNIa

In this chapter, I show how **BAHAMAS** can be used to constrain potential anisotropies in our universe. I consider two possible forms this anisotropy can take, one which restricts the anisotropy to a local redshift scale and one which is constant in redshift. In addition we consider constraining these anisotropies within two cosmological framework, the Λ CDM model as well the model-independent ‘cosmographic expansion’ ([Visser, 2004](#)). The majority of the work presented here is based on the published work of [Rahman et al. \(2021\)](#).

5.1 Introduction

A fundamental assumption underpinning the cosmological concordance model is the cosmological principle, namely that the universe is homogeneous and isotropic on sufficiently large scales. Given the ubiquity of the cosmological principle, an observational test of this assumption is an important step towards validating our best description of the large scale universe. Testing of homogeneity is hampered by the need of surveying extremely large scales (see [Maartens \(2011\)](#)), although recent studies have found the transition to homogeneity at high ($z \sim 2$) redshift consistent with expectations from the Λ CDM cosmological concordance model ([Gonçalves et al.,](#)

2018, 2020).

The assumption of isotropy has been tested over a range of redshifts and with many different probes, from the relatively local universe out to the redshift of recombination. Analyses of Cosmic Microwave Background (CMB) anisotropies data obtained by the Wilkinson Microwave Anisotropy Probe (WMAP) and the Planck satellite found up to $\sim 3\sigma$ evidence of breaches of statistical isotropy in the form of power asymmetry between hemispheres, multipole alignments, anomalous clustering of directions, although the significance of these results is difficult to assess, partially because of issues of *a posteriori* testing (Bennett et al., 2013; Akrami et al., 2014; Schwarz et al., 2016; Planck Collaboration et al., 2016). Quasar polarization directions also appear to be aligned along anomalous directions in the CMB (Hutsemékers et al., 2005) and with coherence scales in excess of 500 Mpc (Hutsemékers, D. et al., 2014; Friday et al., 2022), in potential disagreement with the cosmological principle. Investigating the distribution of galaxies on large scales, Sarkar et al. (2019) found however good agreement between the predictions of Λ CDM and the Sloan Digital Sky Survey data, with a transition to isotropy observed beyond a length scale of $200h^{-1}$ Mpc (where h is the dimensionless Hubble-Lemaître parameter). More recently, Secrest et al. (2020) reported a one-sided $4.9\text{-}\sigma$ rejection of the hypothesis that the dipole in a sample of 1.3 million quasars is purely due to our motion with respect to the CMB.

Supernovae Type Ia can be used to test the second expression of the cosmological principle, namely that the expansion of the universe is isotropic. SNe type Ia observations can also be used to test the hypothesis of isotropy in the expansion of the universe underpinning the Friedmann-Lemaître-Robertson-Walker (FLRW) metric of the concordance cosmological model, which exhibits an isotropic scale factor $a(t)$. To this end, various authors have analyzed increasingly large SNIa compilations with different statistical approaches, often with sharply discordant results.

After early works (Kolatt & Lahav, 2001; Schwarz & Weinhorst, 2007; Gupta et al., 2008), Cooke & Lynden-Bell (2010) analysed a subset of 250 SNIa from the Union compilation (Kowalski et al., 2008) with $z > 0.2$ with a maximum likelihood approach to constrain a dipolar mod-

ulation to the luminosity distance, finding no significant deviation from isotropy. However, [Cai & Tuo \(2012\)](#) claimed that the deceleration parameter shows a preferred direction in the Union2 ([Amanullah et al., 2010](#)) compilation of 557 SNIa, a result corroborated by the analysis of [Antoniou & Perivolaropoulos \(2010\)](#), who combined SNIa data with other cosmological probes. [Jiménez et al. \(2015\)](#) analysed the same Union2 data, additionally including the SNLS3 data, and showed that previous claims of anisotropy disappear if one accounts for correlations among the observations by including the full data covariance matrix in the analysis. Other null results of anisotropic expansion include [Heneka et al. \(2014\)](#) and [Lin et al. \(2016a\)](#), who investigated the Joint Light-curve Analysis (JLA) compilation ([Betoule et al., 2014](#)) of 740 SNIa (see also [Sun & Wang \(2019\)](#), who obtain discrepant results from three different compilations of SNIa namely Union2.1, JLA and Constitution ([Hicken et al., 2009b](#))). Similarly [Andrade et al. \(2018a\)](#), find the JLA data prefers isotropy, with the results being inconclusive on the Union2.1 data. [Javanmardi et al. \(2015\)](#) however, finds that taking into account the inferred alignment of the dipole on the Union2.1 data with the Cosmic Microwave Background (CMB) temperature anisotropy, then the null hypothesis of isotropy cannot be rejected at 95-99% level. An important distinction between these data sets is that Union2/2.1 and Constitution have no corrections for the peculiar velocities of the SNe host galaxies, whereas JLA does. [Bernal et al. \(2017\)](#) found from Union2 and LOSS data potential differences between hemispheres in the isotropy of the deceleration parameter. The situation becomes more confused when considering the largest SNIa compilation to date, Pantheon, encompassing 1048 objects in the redshift range $0.01 < z < 2.3$ ([Scolnic et al., 2018b](#)). A major hurdle to any re-analysis that uses Pantheon is the lack of a publicly available full correlation matrix for its SNIa light curve standardisation coefficients, which hampers a principled statistical approach. Nevertheless, several papers have attempted to use Pantheon-derived measurements of the distance modulus as a function of redshift to investigate potential deviations from an isotropic expansion, finding isotropy is still favoured ([Sun & Wang, 2018b](#); [Soltis et al., 2019](#); [Zhao et al., 2019](#); [Andrade et al., 2018b](#)).

Despite the existence of substantially larger compilations like Pantheon, the JLA remains a useful data set for analyses of this kind, because all the necessary statistical and systematics covariance matrices are publicly available, unlike e.g. the Pantheon set. Recently, [Colin et al.](#)

(2019b) (henceforth C19) claimed 3.9σ evidence for a dipole in the deceleration parameter from a maximum likelihood analysis of JLA data, leading to a lack of statistical evidence for acceleration in the expansion. The claim was disputed by Rubin & Heitlauf (2020) (henceforth, RH20), who pointed out the incorrect use of heliocentric redshifts in C19 and other technical assumptions about selection effects which, when corrected, remove the preference for a dipole and restore the high significance for an accelerated expansion. (See also the discussion in Rubin & Hayden (2016), itself a rebuttal of Nielsen et al. (2016).) A further reply by Colin et al. (2019a) appears to concede some technical points, but not the overall conclusion on the actual lack of statistical significance for an accelerated expansion.

The aim of this work is to clarify the status of claims for a statistically significant dipole in the accelerated expansion of the universe, especially in light of the ongoing controversy. In so doing, we also revisit the important question of the level of statistical evidence in favour of the accelerated expansion in an isotropic universe from SNIa data alone. We address the criticisms of the published JLA data made by C19, who claimed that the peculiar velocity corrections made to the JLA SNIa data based on local bulk flows are incorrect. In this chapter, we introduce a state-of-the-art treatment of peculiar velocities, which are independently constrained using the 2M++ galaxy catalogue (Lavaux & Hudson, 2011; Boruah et al., 2020a; Boruah et al., 2020b), re-derive correlated peculiar velocity uncertainties (both statistical and systematic) from a fully consistent flow model, and upgrade the Bayesian hierarchical model BAHAMAS Shariff et al. (2016) to include an new treatment of residual colour-based selection effects in SNIa data.

The remaining of this chapter is structured as follows: section 5.2 introduces the cosmological model, the anisotropy model, our Bayesian framework, the data used, our new peculiar velocities treatment and our new colour-based selection effects correction. Section 5.3 demonstrates the performance of our method on simulated data. Our results from the JLA data, both in terms of parameter inference and Bayesian model comparison, are presented in section 5.3.2. Our conclusions are given in section 5.4.

5.2 Methodology and Data

5.2.1 Cosmological model and dipole modulation

We investigate the isotropy of the expansion in both a model-specific and a model-independent way: firstly, we consider the Λ CDM model for the underlying cosmology; secondly, we use the so-called ‘Cosmographic expansion’ (i.e., a Taylor expansion of the scale factor as a function in time) as a model-independent description of the underlying matter-energy density of the universe.

For the purposes of this chapter, I modify our notation which is different from that introduced in earlier chapters. I now define the Λ CDM model to have cosmological parameters $\mathcal{P}_1 = \{\Omega_m, \Omega_\Lambda, H_0\}$, where Ω_m and Ω_Λ have the same definitions as before; H_0 is fixed to $H_0 = 72$ km/s/Mpc, as it is exactly degenerate with the SNIa intrinsic magnitude. The curvature parameter Ω_κ then is given by

$$\Omega_\kappa = 1 - \Omega_m - \Omega_\Lambda \quad (5.1)$$

and we assume a universe with constant dark energy equation of state, $w(z) = -1$. We denote by \bar{z} the redshift of a comoving galaxy seen by an observer who is also at rest w.r.t. to the CMB restframe (i.e., the ‘cosmological’ redshift, with no peculiar velocities from either the source or the observer) and by z_{hel} the redshift for an observer in the Sun’s frame of reference¹. The measured redshift in our heliocentric frame of reference is given by \hat{z}_{hel} , and it differs from z_{hel} by measurement noise. The redshift in our heliocentric frame of reference, z_{hel} , differs from the redshift of a comoving observer, \bar{z} , by virtue of peculiar velocities of the source and the observer, and gravitational red/blueshifts due to the local gravitational potential at the location of the source and observer. In the following, we neglect gravitational effects, which are subdominant (see however [Calcino & Davis \(2017\)](#)) and focus instead on the impact of peculiar velocities.

The relationship between heliocentric redshift, z_{hel} , and the redshift of a comoving galaxy as

¹We neglect the distinction between geocentric and heliocentric frames of references (the difference due to the ~ 30 km/s orbital speed of the Earth is of order $\Delta z \sim 10^{-5}$), since redshift measurements are routinely reported in the heliocentric frame and also already corrected for atmospheric refraction.

seen by an observer at rest w.r.t. the CMB, \bar{z} is given by:

$$(1 + z_{\text{hel}}) = (1 + z_{\text{CMB}})(1 + z_{\text{pec}}^{\odot}) \quad (5.2)$$

$$(1 + z_{\text{CMB}}) = (1 + \bar{z})(1 + z_{\text{pec}}^{\text{SN}}) \quad (5.3)$$

where z_{pec}^{\odot} is the redshift induced by the peculiar velocity of the Solar System w.r.t. the CMB restframe, while $z_{\text{pec}}^{\text{SN}}$ is the redshift caused by the peculiar velocity of the SNIa w.r.t. the CMB frame. The second equality introduces the redshift in the CMB restframe, z_{CMB} , i.e., the frame in which our motion w.r.t. the CMB has been removed². With the above definitions, we can write the luminosity distance to redshift \bar{z} , as (Davis et al., 2011)

$$d_L(\bar{z}, z_{\text{pec}}^{\odot}, z_{\text{pec}}^{\text{SN}}, \mathcal{P}_1) = \frac{c}{H_0} \frac{(1 + \bar{z})(1 + z_{\text{pec}}^{\odot})(1 + z_{\text{pec}}^{\text{SN}})^2}{\sqrt{|\Omega_{\kappa}|}} \times \text{sinn} \left\{ \sqrt{|\Omega_{\kappa}|} \int_0^{\bar{z}} \frac{dz}{E(z)} \right\}. \quad (5.4)$$

$E(z)$ depends on our choice of cosmology and for the Λ CDM universe, is given by

$$E^2(z) = \Omega_M(1 + z)^3 + \Omega_{\Lambda} + \Omega_{\kappa}(1 + z)^2. \quad (5.5)$$

The $\text{sinn}(x)$ function is defined as

$$\text{sinn}(x) = \begin{cases} x & \text{if } \Omega_{\kappa} = 0 \\ \sin(x) & \text{if } \Omega_{\kappa} < 0 \\ \sinh(x) & \text{if } \Omega_{\kappa} > 0 \end{cases} \quad (5.6)$$

In our model-independent approach, we follow Visser (2004) and Taylor-expand the scale factor

²A source of confusion in the literature is the widespread use of the term ‘CMB restframe’ to denote what we call \bar{z} (i.e., the cosmological redshift, with no peculiar motions from either source nor observer). This misleading nomenclature is for example used by Betoule et al. (2014), as well as in the data products of the JLA data release.

of the FLRW metric up to third order in time around t_0 (today), as:

$$a(t) = a_0 \{1 + H_0(t - t_0) - \frac{1}{2}q_0 H_0^2(t - t_0)^2 + \frac{1}{3!}j_0 H_0^3(t - t_0)^3 + O([t - t_0]^4)\} \quad (5.7)$$

where q_0 is the dimensionless deceleration parameter, defined as

$$q_0 = -\frac{1}{a} \frac{d^2 a}{dt^2} \left[\frac{1}{a} \frac{da}{dt} \right]_{t=t_0}^{-2} \quad (5.8)$$

and j_0 is the so-called ‘jerk’,

$$j_0 = +\frac{1}{a} \frac{d^3 a}{dt^3} \left[\frac{1}{a} \frac{da}{dt} \right]_{t=t_0}^{-3}, \quad (5.9)$$

which is also dimensionless. This model-independent expansion only relies on the FLRW metric but makes no assumption about the underlying matter-energy density, and leads to the following form of the luminosity distance at redshift \bar{z} :

$$d_L(\bar{z}, z_{\text{pec}}^{\odot}, z_{\text{pec}}^{\text{SN}}, \mathcal{P}_2) = \frac{(1 + z_{\text{pec}}^{\odot})(1 + z_{\text{pec}}^{\text{SN}})^2}{1 + \bar{z}} \frac{c\bar{z}}{H_0} \times \left[1 + \frac{1}{2}(1 - q_0)\bar{z} - \frac{1}{6}(1 - q_0 - 3q_0^2 + j_0 - \Omega_{\kappa})\bar{z}^2 + O(\bar{z}^3) \right], \quad (5.10)$$

where c is the speed of light and the model-independent parameters are $\mathcal{P}_2 = \{H_0, q_0, j_0, \Omega_{\kappa}\}$. From expressions (5.4) or (5.10), we obtain the isotropic distance modulus, μ_I , using the standard formula

$$\mu_I(\bar{z}, z_{\text{pec}}^{\odot}, z_{\text{pec}}^{\text{SN}}, \mathcal{P}_a) = 25 + 5 \log_{10} \frac{d_L(\bar{z}, z_{\text{pec}}^{\odot}, z_{\text{pec}}^{\text{SN}}, \mathcal{P}_a)}{1 \text{ Mpc}}, \quad (5.11)$$

where $a = 1, 2$ depending on the chosen parameterization.

There are several different ways one can parameterize the possibility of anisotropic expansion, depending on the underlying physical origin for the effect. A spherical harmonics expansion introduces, to lowest order, a dipolar modulation in the direction of a SNIa situated at redshift \bar{z} and in direction \mathbf{n}_{SN} in the sky, with \mathbf{n}_{SN} a unit vector pointing from the centre of the coordinate system (the Earth) to the location of the SNIa on the celestial sphere. Different authors have taken different approaches in the literature, with no consensus as to which quantity should be modulated: one could expand the scale factor $a(t)$, the luminosity distance, the comoving

distance, the Hubble parameter, the matter density, the cosmological constant density, or the distance modulus. Each of these possibilities leads to a different anisotropic imprint onto the Hubble-Lemaître law. A dipole moment that is constant with distance, r , in the peculiar velocity field (i.e., a bulk flow) leads to

$$c\bar{z} \approx H_0 r + D_v(\mathbf{n}_{\text{dip}} \cdot \mathbf{n}_{\text{SN}}), \quad (5.12)$$

By contrast, a constant dipole in either H_0 or r leads to

$$c\bar{z} \approx H_0 r + H_0 D_H r (\mathbf{n}_{\text{dip}} \cdot \mathbf{n}_{\text{SN}}), \quad (5.13)$$

which increases linearly with distance. Another possibility is to modulate the distance modulus directly:

$$\mu = \mu_I(\bar{z}, z_{\text{hel}}, \mathcal{P}_a) (1 + D_\mu F(\bar{z})(\mathbf{n}_{\text{dip}} \cdot \mathbf{n}_{\text{SN}})), \quad (5.14)$$

where $F(z)$ is a function of redshift alone which can be used to localize the dipole at a given length scale. Yet another approach, adopted by C19, is to model the dipole on the deceleration parameter, q_0 , in a Cosmographic expansion:

$$q_0(z) = q_m + D_{q_0} F(\bar{z})(\mathbf{n}_{\text{dip}} \cdot \mathbf{n}_{\text{SN}}). \quad (5.15)$$

In this work, we add the dipole to either the distance modulus, Eq. (5.14), or to the deceleration parameter, Eq. (5.15), and consider both $F(z) = 1$ and, following C19, an exponentially decaying function of redshift with characteristic scale given by the free parameter S , namely $F(z) = \exp(-z/S)$. These two forms have the effect of either creating a dipole that is constant in redshift or constrained to a local scale which could arise for reasons such as existing within a cosmic void.

As noted in previous works, a phenomenological approach as the one taken here that perturbs an underlying FLRW metric may not be entirely consistent. An alternative route would require specifying a physical model for the anisotropy, and then derive the ensuing predictions for the

distance modulus and compare those with observations, as done for example in the context of an ellipsoidal universe from Bianchi type I models (Campanelli et al., 2011). However, the advantage of a purely phenomenological approach is that it remains agnostic about the underlying cause of any anisotropy, and it provides constraints on the level of anisotropy that can then be applied to other models.

5.2.2 Bayesian Hierarchical Model

In this work, we improve on previous works constraining anisotropy from SNe data by adopting the fully Bayesian hierarchical model BAHAMAS for the statistical analysis of SNIa data. See section 4.3 for a full description of BAHAMAS.

To use BAHAMAS within the context of this work on anisotropy we redefine some key parameters. As before, we still denote with a hat symbol observed quantities, in order to distinguish them from the latent unobserved variables in our model.

For each SNIa i , the data d_i can be summarised by a vector

$$d_i = \{\hat{z}_{\text{hel},i}, \hat{c}_i, \hat{x}_{1,i}, \hat{m}_{B_i}, \hat{C}_i\}, \quad (5.16)$$

where \hat{C}_i , $\hat{m}_{B_i}, \hat{x}_{1,i}$ and \hat{c}_i have their usual definitions as defined in section 4.3. $\hat{z}_{\text{hel},i}$ is the observed heliocentric redshift. On the standard deviation scale, the measurement error for redshift for SNIa with spectroscopic follow-up is $\sigma_z^{\text{SPSN}} \sim 5 \times 10^{-3}$ when the redshift is determined from the SNIa spectrum alone, and $\sigma_z^{\text{SPhost}} \sim 5 \times 10^{-4}$ when it is obtained from host-galaxy spectra (Zheng et al., 2008). The redshift measurements are independent from each other and from all other observables. We discuss this uncertainty further in section 5.2.6. (See also Calcino & Davis (2017) for the potentially important impact of systematic redshift errors as small as $\Delta z \sim 10^{-4}$.)

The hierarchical modelling then proceeds as before and we can say at the latent level, the apparent peak magnitude m_{B_i} is related to the standardised intrinsic magnitude M_i^ϵ via the

isotropic distance modulus of Eq. (5.11):

$$m_{Bi} = \mu_I(z_{\text{hel},i}, \bar{z}_{,i}, \mathcal{P}_a) - \alpha x_{1_i} + \beta c_i + M_i^\epsilon. \quad (5.17)$$

This is then related to the observed values of $\{\hat{m}_{Bi}, \hat{x}_{1_i}, \hat{c}_i\}$ and are modelled as normally distributed around their latent values, with variance-covariance matrix given by \hat{C}_i . The systematics matrix C_{syst} is included as before as well. Inference is based on the marginal distribution of the quantities of interest, \mathcal{P}_a , which includes uncertainty at all levels of the hierarchy. An extra note is added here to emphasise that our method is not fully Bayesian, as there are aspects of the data (discussed in the next section) which are not captured within our model such as the selection effects on magnitude which are instead handled with ‘corrections’ that have been applied to the data. This corrected data has **BAHAMAS** applied to it. Increasing the scope of **BAHAMAS** to handle more data effects so these ad hoc corrections to the data are not required is the subject of future work.

5.2.3 Data

The largest SNIa compilation to date is the ‘Pantheon sample’ (Scolnic et al., 2018a), which contains 1048 spectroscopically confirmed SNIa. This compilation includes 279 new SNIa discovered by the Pan-STARRS1 (PS1) Medium Deep Survey (Chambers et al., 2016) in addition to the previous SNe discovered by previous catalogues to create the total.

The Pantheon sample only provides estimates (and associated uncertainties) of the distance modulus for each SNIa but does not include the covariance matrices of either the measurement error for each SNIa nor the systematic covariance matrix across the whole data set. Because our Bayesian hierarchical model also requires these covariances over the light curve fit parameters as opposed to the covariance over distance modulus provided by the Pantheon data, we instead use the smaller ‘Joint Light-Curve Analysis’ (JLA) compilation (Betoule et al., 2014). The JLA data contains 740 SNIa including 374 SNIa from the SDSS-II survey (Frieman et al., 2008; Sako et al., 2018), 239 from SNLS (Conley et al., 2011; Sullivan et al., 2011), a low- z sample of 118

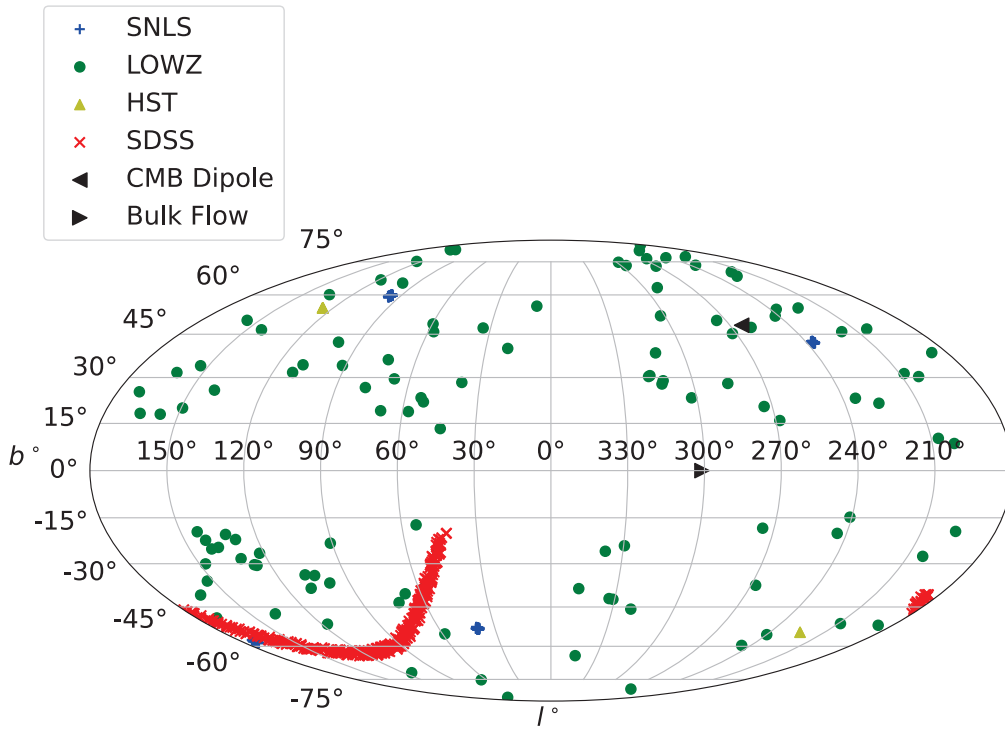


Figure 5.1: A plot of the JLA SNe showing their location in the sky in galactic coordinates, as well as the direction of the CMB dipole (left-pointing triangle) and bulk flow (right-pointing triangle), as determined by [Boruah et al. \(2020a\)](#).

SNIa at $z < 0.1$ which comprises numerous smaller surveys, and nine Hubble Space Telescope SNIa. These have been fit and standardised using the SALT2 Light-Curve Fitter ([Guy et al., 2007](#)).

An overview of the distribution of JLA objects in the sky is provided in Fig. 5.1. The long stripe in the lower left hemisphere is from the SDSS objects. It is clear to see that the distribution of the JLA objects in the sky is highly anisotropic.

One way the sky coverage limits this project is that there has to be sufficient SNIa data in a region around where the dipole points. This particularly applies to any dipole pointing along the galactic plane as there is a dearth of SNIa along the region $-15^\circ < b < 15^\circ$ owing to the obscuring effect of the galactic plane.

5.2.4 Accounting for Colour-dependent Selection Effects

Towards the high end of the redshift range of a given survey, SNe that are intrinsically brighter (smaller M_i) or bluer (smaller c_i) are more likely to be observed and to be followed up spectroscopically to confirm their type. This selection bias must be accounted for to avoid bias in the estimates of the cosmological parameters. Ignoring magnitude-based selection effects leads to estimates of the distance modulus that are biased low: at a given redshift, the average observed magnitude is smaller (i.e., the observed peak flux brighter) than the population mean, which leads to an estimated distance modulus that is biased low. This effect reduces or even obliterates the preference for a non-zero cosmological constant. Traditionally, this has been addressed by ‘correcting back’ the estimates of the distance modulus by the average bias in each redshift bin, established with forward simulations of data subject to selection effects. This approach is adopted e.g., by [Betoule et al. \(2014\)](#). More recently, this method has been extended and refined with the so-called ‘BEAMS with Bias Corrections’ (BBC) method ([Kessler & Scolnic, 2017](#)). The JLA analysis only corrected for magnitude-based selection effects, concluding that no additional correction were necessary for colour ([Betoule et al. \(2014\)](#), Fig. 11), despite observing a downward trend in observed colour with redshift for SNSL and SDSS.

In the context of a Bayesian analysis, however, selection effects are treated differently: the posterior is conditional on the observed data ([Kelly, 2007](#)), which leads to a re-weighting factor increasing the statistical weight of SNe that are less likely to be observed (see Eq. (5.20) below). [Rubin et al. \(2015\)](#) introduced a general formalism for the selection function that was further developed by [Hinton et al. \(2018\)](#). In practice, however, this formalism typically requires several simplifying assumptions that may be difficult to justify (e.g., a well-sampled SNIa redshift distribution, a selection function that is described by a normal cumulative distribution function (CDF), independence of the selection probability from the underlying cosmology). An improved selection effects treatment is presented in Chapter 7. For now, we take the ad-hoc correction applied to B-band apparent magnitude measurement in the JLA data as is and focus instead on the potential residual corrections that need to be accounted for in colour.

[Rubin & Hayden \(2016\)](#) argue that uncorrected-for colour-dependent selection effects remaining

in the JLA data (after bias correction of the data) should be addressed by introducing a population colour mean that is both redshift- and survey-dependent. However, at the top level of the BHM we would like a *physically-meaningful* population mean that describes the underlying, population-level latent mean colour, itself a reflection of the physical properties of the SNe. Such a colour mean can be a function of redshift, to reflect evolution in the physical properties of SNe with lookback time, but it cannot be survey-dependent, for clearly the latter dependency is caused by survey-specific selection effects and, in virtue of being survey-specific, *cannot* be the consequence of changing underlying physical properties of the SNe being observed. Thus, by using population-level variables to address a survey-induced selection effect, the method advocated by Rubin & Hayden (2016) goes against the physical interpretability of the BHM; we prefer the population-level variables to be tied to the physics of the SNIa explosion mechanism rather than the survey-induced selection effects. While genuine population level changes with redshift can be thought to occur owing to redshift related changes associated with metallicity, star formation rate, and other galaxy covariates, to properly account for any supposed drift in the population means with redshift and make it distinct from any selection effects, the implemented model would need the drift parameters to be independent of any selection effects/telescope parameters if it is to be principled. Something which is not done in the case of Rubin & Hayden (2016). Furthermore, as pointed out by Dam et al. (2017), modeling residual colour drift with redshift as advocated by Rubin & Hayden (2016) introduces undesirable degeneracies with cosmological parameters in the cosmographic expansion.

Here, we account for residual colour-based selection effects with an approximate method that captures the spirit of the correct Bayesian procedure (with the full treatment outline in Chapter 7). In general, we aim to base the likelihood function on the distribution of the data d_i for SNIa i conditional on it having been observed, which is denoted by an indicator variable $I_i = 1$. ($I_i = 0$ would indicate that SNIa i is not observed.) Denote by $\hat{D} = \{\hat{d}_1, \dots, \hat{d}_n\}$ a random sample of SNe where all SNe have equal probability of being observed (i.e., there are no selection effects in \hat{D}). In the presence of selection effects, we observe a non-representative sample of $n_{\text{obs}} < n$ SNe from \hat{D} , whereby a SNIa i is observed with probability given by the selection function $p(I_i = 1|\hat{d}_i, \Psi)$, where \hat{d}_i are the observed (noisy) data, and Ψ are parameters describing the

selection function (assumed known). The distribution of the observed \hat{d}_i , conditional on SNIa i being observed, is given by

$$p(\hat{d}_i|I_i = 1, \Psi, \Theta) = \frac{p(I_i = 1|\hat{d}_i, \Psi)p(\hat{d}_i|\Theta)}{p(I_i = 1|\Psi, \Theta)}, \quad (5.18)$$

where $\Theta = \{\mathcal{P}_a, \vartheta\}$ are the parameters of the hierarchical model (including the parameterization of the distance modulus, \mathcal{P}_a , and the population-level distribution parameters, ϑ). We assume that the selection probability conditional on the observed data, \hat{d}_i , i.e., $p(I_i = 1|\hat{d}_i, \Psi)$ in the numerator of Eq. (5.18), does not depend on Θ and thus is an ignorable constant if (as we assume here) the selection function and its parameter Ψ are known. The quantity $p(\hat{d}_i|\Theta)$ is the likelihood in the absence of selection effects, and the denominator gives the probability of observing a SNIa, irrespective of the value of the data:

$$p(I_i = 1|\Psi, \Theta) = \int d\hat{d}_i p(I_i = 1|\hat{d}_i, \Psi)p(\hat{d}_i|\Theta). \quad (5.19)$$

Omitting the ignorable multiplicative constants in the numerator of Eq. (5.18) yields the likelihood function of Θ including selection effects for a sample of n_{obs} observed SNe, $\hat{d}_{\text{obs}} \equiv \{\hat{d}_1, \dots, \hat{d}_{n_{\text{obs}}}\}$,

$$p(\hat{d}_{\text{obs}}|\{I_i = 1\}_{i=1}^{n_{\text{obs}}}, \Psi, \Theta) \propto \frac{p(\hat{d}_{\text{obs}}|\Theta)}{p(I = 1|\Psi, \Theta)^{n_{\text{obs}}}}, \quad (5.20)$$

where we have dropped the dummy index i in the denominator. Therefore, selection effects are accounted for by dividing the likelihood function of the observed data, $p(\hat{d}_{\text{obs}}|\Theta)$, by a ‘correction factor’ that gives the probability of making n_{obs} observations.

Thus far, we have been entirely general. Next we specify the form of the selection function, $p(I_i = 1|\hat{d}_i, \Psi)$, entering in Eq. (5.19). The probability of a SNIa being selected, and spectroscopically followed-up to determine its type, depends primarily on its magnitude and colour. (SNe with larger stretch parameter x_1 are slower declining and thus remain visible and potentially detectable for longer, but this effect is subdominant.) The data correction procedure in [Betoule et al. \(2014\)](#) should in principle account for both magnitude- and colour-based selection, but their discussion makes it clear that there are large uncertainties in the determination

of the selection probability that enters their forward simulation of the data. For example, [Betoule et al. \(2014\)](#) mention that the SDSS spectroscopic follow-up target selection favours intrinsically bluer SNe, introducing complex colour-dependency in the selection function. In light of such difficult-to-simulate selection effects, we advocate a method that estimates any residual selection effect (after the data correction procedure of [Betoule et al. \(2014\)](#)) directly from the observed data.

We wish to account for residual colour-based selection effects that may remain in the data. Therefore, we assume that $p(I_i = 1|\hat{d}_i, \Psi)$ depends only on \hat{c}_i and $\hat{z}_{\text{hel},i}$, with the N_s redshift bins for survey s chosen as discussed in section 5.2.3, and factorize both the selection function and the likelihood in a product over redshift and survey bins, assumed independent of each other. Within each redshift, and survey bin, we allow a different selection function, which is derived below. With these assumptions, the probability of observing n_{obs} SNe just depends on the color data as (with the shorthand notation $I_{n_{\text{obs}}} = 1$ denoting $\{I_1 = 1, \dots, I_{n_{\text{obs}}} = 1\}$):

$$p(I_{n_{\text{obs}}} = 1|\Psi, \Theta) = \prod_{s=1}^4 \prod_{j=1}^{N_s} \left(\int d\hat{c}_i p_{sj}(I_i = 1|\hat{c}_i, \Psi) p_{sj}(\hat{c}_i|\Theta) \right)^{N_{sj}}, \quad (5.21)$$

where N_s is the number of bins for survey s and N_{sj} the number of observed SNe in bin sj . Within each redshift bin for survey s , we parameterize the selection function as a normal cumulative distribution function (CDF), and assume that we observe a SNIa i with colour \hat{c}_i with probability:

$$p_{sj}(I_i = 1|\hat{c}_i, \Psi) = \Phi \left(\frac{c_{sj}^{\text{obs}} - \hat{c}_i}{\sigma_{sj}^{\text{obs}}} \right), \quad (5.22)$$

where

$$\Phi(x) = \int_{-\infty}^x \mathcal{N}_y(0, 1) dy \quad (5.23)$$

is the CDF of a standard normal, and $\mathcal{N}_y(\mu, \sigma^2)$ is a Gaussian distribution in y with mean μ and variance σ^2 . In Eq. (5.22), c_{sj}^{obs} is the colour value at which there is a 50% probability of observing a SNIa in redshift bin j and for survey s ; σ_{sj}^{obs} denotes the width of the transition from the regime where all objects are observed, i.e. for $(c_{sj}^{\text{obs}} - \hat{c}_i)/\sigma_{sj}^{\text{obs}} \ll 0$, to the regime where no objects are observed, where $(c_{sj}^{\text{obs}} - \hat{c}_i)/\sigma_{sj}^{\text{obs}} \gg 0$, for the bin sj being considered. In Eqs. (5.28)-

(5.29) below we show how to estimate the selection function parameters $\Psi = \{(c_{sj}^{\text{obs}}, \sigma_{sj}^{\text{obs}})\}$ ($s = 1, \dots, 4, j = 1, \dots, N_s$). As an approximation, we ignore uncertainty in the resulting estimates and assume the parameters are known exactly³.

In principle, we would like to use the likelihood function of [March et al. \(2011\)](#) and [Shariff et al. \(2016\)](#) as the second term of the integrand in Eq. (5.19). Since we only wish to account for residual colour-based selection effects, however, we ignore the part of the BAHAMAS likelihood that relates colour to magnitude via the Tripp linear relation (Eq. (C2) in [March et al. \(2011\)](#)), and instead only consider the distribution of colour values that one would obtain when integrating out the latent colour variables conditional on all other variables in the BHM, leading to the simple expression for the likelihood entering into Eq. (5.21):

$$p_{sj}(\hat{c}_i|\Theta) = \mathcal{N}_{\hat{c}_i}(c_\star, R_c^2 + \bar{\sigma}_{c,sj}^2), \quad (5.24)$$

where $\bar{\sigma}_{c,sj}$ is the average colour measurement error for the n_{sj} SNe in bin sj (for simplicity, we assume all n_{sj} SNe in bin sj have the same colour measurement error, given by $\bar{\sigma}_c$; we also ignore correlation between colour and stretch and magnitude) Eq. (5.24) features a redshift-independent conditional expectation of colour, described by c_\star . This formalizes the assumption of [Rubin & Hayden \(2016\)](#) within BAHAMAS that the observed drift to bluer SNe with redshift within a survey is a consequence of selection effects and not of a change in the underlying population colour distribution with redshift.

With the above elements, we can compute the probability of observing SNIa i in redshift bin j for survey s by integrating over its colour, \hat{c}_i , obtaining ⁴

$$p_{sj}(I_i = 1|\Psi, \Theta) = \Phi \left(\frac{c_{sj}^{\text{obs}} - c_\star}{\sqrt{(\sigma_{sj}^{\text{obs}})^2 + R_c^2 + \bar{\sigma}_{c,sj}^2}} \right). \quad (5.25)$$

For a given sj bin, when $c_\star \ll c_{sj}^{\text{obs}}$ selection effects are irrelevant, because the survey is seeing the entire colour population, and accordingly $p(I_i = 1|\Psi, \Theta) \rightarrow 1$ from Eq. (5.25). However,

³We ignore uncertainty in our estimates of Ψ , and we defer the evaluation of the impact of this approximation to Chapter 7

⁴It is useful to recall that $\int_{-\infty}^{\infty} \Phi\left(\frac{\mu-x}{\sigma}\right) \mathcal{N}_x(\nu, \tau^2) dx = \Phi\left(\frac{\mu-\nu}{\sqrt{\sigma^2+\tau^2}}\right)$.

$p(I_i = 1|\Psi, \Theta)$ becomes smaller for values of $c_\star > c_{sj}^{\text{obs}}$, with the difference measured in units of the total standard deviation, i.e., when the survey is preferentially seeing the bluer part of the population because of colour-based selection bias. In this case, the observed distribution of the N_{sj} objects in the sj bin deviates from the latent distribution, and the likelihood that ignores selection effects would incorrectly penalize this value of c_\star . According to Eq. (5.20), the correction factor in the denominator of Eq. (5.20) increases the weight given to observed SNe with $c_\star > c_{sj}^{\text{obs}}$.

Finally, there remains the issue of determining the value of Ψ , the selection function parameters. Ideally, one would do so from forward simulation of surveys, but this is unpractical for our purposes, and unfeasible for the low- z sample, which is obtained from a collection of telescopes with poorly understood selection functions. Furthermore, as noted earlier, colour-based selection effects might escape *ab initio* modeling of this kind, as indicated in e.g. Betoule et al. (2014). As an alternative, we estimate the value of Ψ in each redshift- and survey-bin sj , by matching the first and second moment of the empirical colour distribution within the bin to the marginal distribution of colour based on the right-hand-side of Eq. (5.20), understood as a distribution over observed colour values \hat{c}_i within each bin. Let $m_C(t)$ denote the moment generating function of the random variable C (the observed colour), defined as

$$m_C(t) \equiv \int_{-\infty}^{\infty} e^{ty} f_C(y) dy, \quad (5.26)$$

where $f_C(c)$ in our case is given by (in the bin under consideration)

$$f_C(c) = \frac{\mathcal{N}_c(c_\star, R_c^2 + \bar{\sigma}_{c,sj}^2) \Phi\left(\frac{c_{sj}^{\text{obs}} - c}{\sigma_{sj}^{\text{obs}}}\right)}{p_{sj}(I_i = 1|\Psi, \Theta)}, \quad (5.27)$$

with denominator given by Eq. (5.25). Our strategy, known as the *method of moments*, is to analytically compute the first two moments of the distribution, set them equal to the empirical moments, and solve the resulting system of equations to obtain estimates of the selection function parameters in each survey and redshift bin, $\{c_{sj}^{\text{obs}}, \sigma_{sj}^{\text{obs}}\}$. Details of the calculation are presented in Appendix A, where we show that the moment generating function is given by Eq. (A.5),

its first moment, $dM_C(t)/dt|_{t=0}$, by Eq. (A.18), and its second moment, $d^2M_C(t)/dt^2|_{t=0}$, by Eq. (A.20). Specifically, in each bin, we set the first moment equal to the empirical mean, and set the second central moment equal to the empirical variance:

$$\left. \frac{dM_C(t)}{dt} \right|_{t=0} = \langle \hat{c}_i \rangle, \quad (5.28)$$

$$\left. \frac{d^2M_C(t)}{dt^2} \right|_{t=0} - \left. \frac{dM_C(t)}{dt} \right|_{t=0}^2 = \frac{1}{1 - N_{sj}} \sum_{i=1}^{N_{sj}} (\hat{c}_i - \langle \hat{c}_i \rangle)^2, \quad (5.29)$$

where $\langle \hat{c}_i \rangle = \frac{1}{N_{sj}} \sum_{i=1}^{N_{sj}} \hat{c}_i$ is the empirical mean. We set the population mean and standard deviation to $c_\star = -0.0022$ and $R_c = 0.0758$, the empirical mean and standard deviation from the lowest two bins of the SDSS, SNLS and Low- z surveys, where colour-based selection effects are expected to be negligible. We then solve the resulting coupled equations to determine $\{c_{sj}^{\text{obs}}, \sigma_{sj}^{\text{obs}}\}$. As an alternative one can also consider the method of maximum likelihood which was found to give similar results to our moments method, albeit with a slightly slower runtime, therefore we stick with the method of moments for this work.

In cases where N_{sj} is small (~ 10) and the sample variance is small, the method of moments yields an estimate of σ_{sj}^{obs} near or equal to zero. This is similar to a well-known pathology in the estimation of the shape parameter in the skew-normal distribution (see e.g. [Azzalini & Arellano-Valle \(2012\)](#)). A simple solution is to impose a lower cutoff to the value of σ_{sj}^{obs} . We chose a cutoff value $\sigma_{sj}^{\text{obs}} > 0.01$, since any smaller value leads to a selection function indistinguishable from a step function. Tests of our method of moments on simulated data are provided in Appendix A.

Because our selection effect model assumes in Eq. (5.24) that the colour observations are independent, we set the corresponding covariances in the variance-covariance matrix for the systematic effects to zero. This has a minor effect on our estimates of the cosmological parameters, as those terms are subdominant with respect to other off-diagonal terms.

We apply the method of moments to estimate Ψ to each of the four sub-surveys separately, and we verify its accuracy by simulating colour observations from the model of Eq. (5.24), assuming a Normal constant-in-redshift latent colour distribution with mean $c_\star = 0$ and standard deviation

$R_c = 0.1$. We then apply the reconstructed selection function in each redshift bin, and compare the resulting distribution of simulated SNe with the observed sample within that bin. This comparison is shown in Fig. 5.2. For the simulation study and the real JLA data in Section 5.3.2, we divided each sub-survey in 5 approximately equal-spaced redshift bins, with the following exceptions: the HST data is treated in one single bin owing to the small number of SNIa in HST. For the same reason, in the Low- z survey data, the two highest-redshift bins are combined into a single bin. For the SDSS survey, the right edge of the 4th bin has its boundary shifted 0.015 in redshift space to the right, in order to account for a discontinuity in the data, where the population of objects have a lower colour and are separated by a gap in redshift. The agreement between real data and simulation is excellent, thus validating our approach.

5.2.5 A New Derivation of Peculiar Velocity Corrections

Our motion w.r.t. the CMB frame is measured precisely by the temperature dipole observed in CMB anisotropies. The most accurate result is from [Planck Collaboration et al. \(2018\)](#), giving a velocity $v_{\odot-\text{CMB}} = 369.82 \pm 0.11$ km/s in the direction $l = 264.021^\circ \pm 0.011^\circ$, $b = 48.253^\circ \pm 0.005^\circ$. This induces in the non-relativistic limit a redshift correction ([Davis et al., 2011](#))

$$z_{\text{pec}}^{\odot} \approx -\frac{v_{\odot-\text{CMB}}}{c}(\mathbf{n}_{\text{CMB}} \cdot \mathbf{n}_{\text{SN}}) \lesssim 10^{-3}, \quad (5.30)$$

where \mathbf{n}_{CMB} is a unit vector in the direction of the CMB dipole and \mathbf{n}_{SN} is a unit vector in the direction of the SNIa. Given the small uncertainties in the measurement for $v_{\odot-\text{CMB}}$ and the CMB dipole direction, we can consider z_{pec}^{\odot} as known exactly, and thus neglect measurement error on this quantity (as it is $\sim 3 \times 10^{-7}$). Estimating $z_{\text{pec}}^{\text{SN}} = v_{\text{pec}}^{\text{SN}}/c$ requires knowledge of the peculiar velocity of the SNIa in the CMB frame, $v_{\text{pec}}^{\text{SN}}$. This can be measured either from a peculiar velocity field survey (for example, by using the Fundamental Plane (FP) relation or the Tully-Fisher (TF) relation to measure the distance to a galaxy, and then subtracting from the observed velocity the expansion component obtained from the Hubble-Lemaître law) or derived from linear perturbation theory applied to a smoothed density field. The latter approach has a long history, originally having been used to predict the peculiar velocities of FP

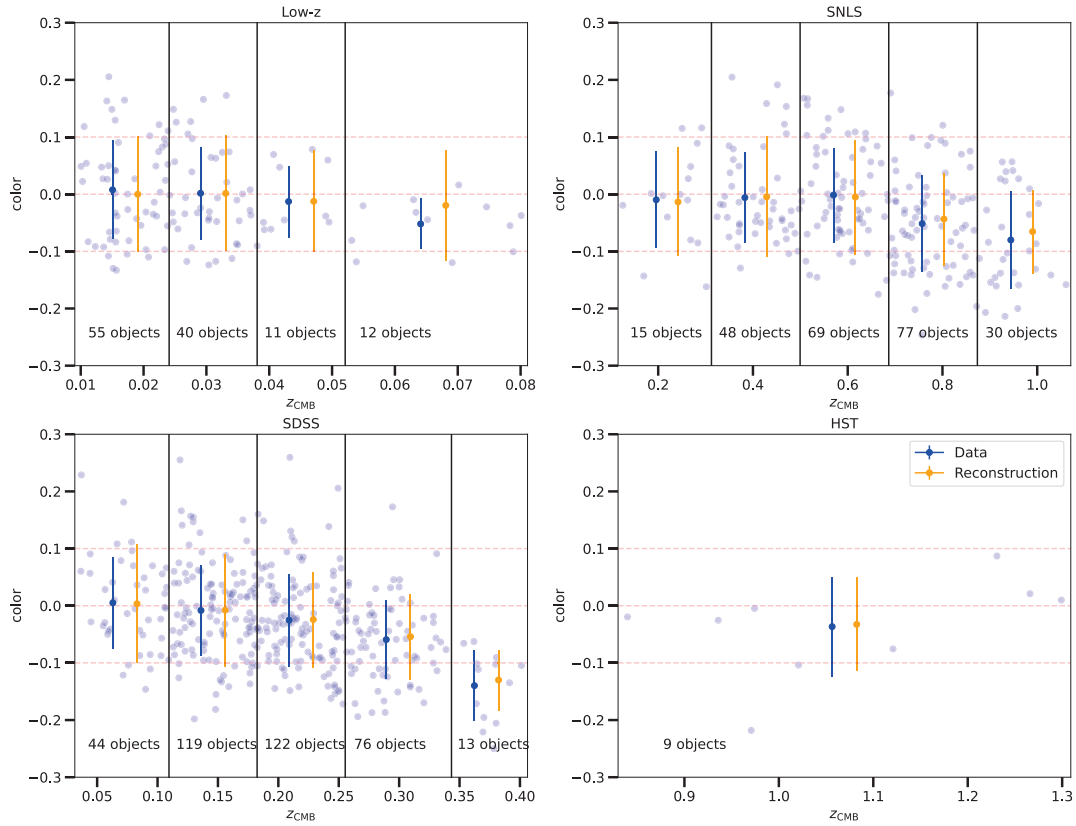


Figure 5.2: Colour-based selection function in binned JLA data, with redshift bins boundaries indicated by the vertical black lines. The blue circles are the individual SNe, the blue errorbars represent the data mean and standard deviation within each top-hat bin, while the orange errorbars give the mean and standard deviation of simulated data from the model using the reconstructed selection function in that bin (shifted horizontally for ease of comparison).

and TF samples (Hudson, 1993; Strauss & Willick, 1995; Davis et al., 1996) but more recently applied to peculiar velocity data that include SNe type Ia (Riess et al., 1997; Radburn-Smith et al., 2004; Pike & Hudson, 2005; Neill et al., 2007; Turnbull et al., 2012; Carrick et al., 2015; Boruah et al., 2020b; Lilow & Nusser, 2021; Stahl et al., 2021).

While the peculiar velocity of the SNIa becomes rapidly negligible for $z \gtrsim 0.1$, it is important for local objects ($z \ll 0.1$), where it can be significant w.r.t. the expansion velocity (up to $\sim 30\%$) and where it leads to much larger changes in the apparent magnitude, due to the steeper gradient of the distance modulus at low redshift. For example, at $z = 0.01$ an uncorrected SNIa peculiar velocity $v_{\text{pec}}^{\text{SN}}$ induces a redshift systematic error $\delta z = v_{\text{pec}}^{\text{SN}}/c$, which corresponds to a significant change in the theoretical distance modulus $\delta\mu \approx \frac{d\mu}{dz}\delta z \approx 5/\ln(10)(\delta v/(cz)) = 0.14 \text{ mag}$ for $v_{\text{pec}}^{\text{SN}} = 200 \text{ km/s}$. To avoid difficulties with peculiar velocities, earlier SNIa cosmological analyses routinely adopted a lower redshift cutoff z_{cut} , removing SNe below z_{cut} ; for example, Kessler et al. (2009) used $z_{\text{cut}} = 0.02$; Riess et al. (2007) used $z_{\text{cut}} = 0.023$. Recently, Huterer (2020) estimated the impact of uncorrected peculiar velocities on the Pantheon sample from numerical N-body simulations, and recommended a cutoff $z_{\text{cut}} = 0.02$ to protect against significant bias to cosmological parameters. However, a better way that does not discard useful data at low redshift is to assign uncertainties that scale with distance, as we do here.

The JLA sample contains 37 SNe with $z_{\text{hel}} < 0.02$, and 110 with $z_{\text{hel}} < 0.05$, for which an appropriate treatment of peculiar velocities is required if they are to be used in the cosmological analysis – particularly in our case, where we wish to use them to constrain a local dipole in the expansion. To first order in redshift Eq. (5.2) gives

$$\bar{z} = z_{\text{hel}} - z_{\text{pec}}^{\odot} - z_{\text{pec}}^{\text{SN}}, \quad (5.31)$$

meaning that the redshift of a comoving observer, \bar{z} , is obtained from the measured heliocentric redshift by subtracting our local dipole (z_{pec}^{\odot}) and the redshift due to the SNIa peculiar velocity, $z_{\text{pec}}^{\text{SN}}$.

The model used in Betoule et al. (2014) to estimate $z_{\text{pec}}^{\text{SN}}$ has been criticised by C19, who

highlighted potential bulk flow velocity discontinuities at $z = 0.04$, pointed out that peculiar velocity corrections arbitrarily disappear beyond $200/h$ Mpc ($z \sim 0.067$, the limit of the galaxy density field measurements from which the peculiar velocities were derived) and that the residual uncorrelated velocity dispersion of $\sigma_v = 150$ km/s might be underestimated. While RH20 pointed out technical flaws with the analysis of C19, it is important in the light of this valid criticism to revisit the issue of low-redshift peculiar velocity corrections here.

To this end, in this work we replace the peculiar velocity corrections used by [Betoule et al. \(2014\)](#) – which rested on the IRAS PSCz catalogue from [Branchini et al. \(1999\)](#) – with the more recent ones obtained by [Carrick et al. \(2015\)](#). We follow [Boruah et al. \(2020a\)](#); [Boruah et al. \(2020b\)](#), who carried out a thorough comparison between density reconstruction from galaxy redshift surveys and kernel smoothing of peculiar velocity data methods. We adopt here their peculiar velocity field inferred from 69,160 galaxies from the 2M++ galaxy redshift catalogue ([Lavaux & Hudson, 2011](#)). The catalogue covers almost the entire sky (with the notable exception of the plane of the galaxy), is highly complete out to $200/h$ Mpc ($z \sim 0.067$) in the region covered by 6dF and SDSS, and out to $125/h$ Mpc ($z \sim 0.041$) in the region covered by 2MRS. We thus remove⁵ the SNIa peculiar velocity corrections for the low- z JLA sample that are in common with the A2 sample of [Boruah et al. \(2020a\)](#) (107 objects), and replace them with new values obtained as follows.

The radial peculiar velocity for a SNIa at comoving distance r and direction \mathbf{n}_{SN} is obtained from the luminosity-weighted density field \mathbf{v}_{gal} as

$$v_{\text{pec}}^{\text{SN}}(r, \mathbf{n}_{\text{SN}}, \theta) = \mathbf{n}_{\text{SN}} \cdot (\beta_v \mathbf{v}_{\text{gal}}(r, \mathbf{n}_{\text{SN}}) + \mathbf{V}_{\text{ext}}) \quad (5.32)$$

where $\theta = \{\beta_v, \mathbf{V}_{\text{ext}}\}$, with a rescaling parameter $\beta_v = 0.411 \pm 0.020$ and external residual bulk flow velocity (in Galactic Cartesian coordinates) $\mathbf{V}_{\text{ext}} = [52 \pm 20, -163 \pm 21, 49 \pm 16]$ km/s (how we treat and propagate the uncertainties in these values is addressed in section 5.2.6). We follow the methodology of [Boruah et al. \(2020a\)](#), with the difference that we only use the SFI++

⁵Differently from C19, we do not remove the magnitude bias corrections made to the JLA SNe, as they are important to account for selection effects, nor do we neglect the contribution of peculiar velocities uncertainty to the covariance matrix, which we re-derive for our case.

peculiar velocity sample (therefore not including A2 SNe data) in order to avoid circularity (i.e. using SNe data to predict the peculiar velocity correction for the same SNe data).

We do not wish to use the distance modulus information from a SNIa at this stage of the analysis, only its observed redshift in the CMB restframe, $\hat{z}_{\text{CMB}} \equiv \hat{z}_{\text{hel}} - z_{\text{pec}}^{\odot}$. Firstly, the observed redshift in the CMB restframe is corrected to the average redshift of the group to which the host galaxy belongs. Differently from JLA, we correct the CMB restframe redshift for *all* host galaxies, including those in clusters and poorer groups. This is necessary to suppress the highly non-linear velocity contribution to the observed redshift, and it leads to deviations of a few percent in \hat{z}_{CMB} in most low- z SNe, compared with the value used by JLA, see Fig. 5.3. However, there are 6 SNe that show much larger changes in their CMB frame redshift, up to $\sim 30\%$ (highlighted in green in Fig. 5.3); two of them (**sn2007ci** and **sn2001cz**) are in common with the outliers in peculiar velocity, identified in Fig. 5.4 and discussed below in this section. It is emphasised here that no special treatment of these 6 outliers occurs in this work and they are highlighted here for demonstration purposes.

Then, we compute the expected peculiar velocity by marginalizing over the unknown comoving distance of the SNIa, r :

$$\langle v_{\text{pec}}^{\text{SN}} \rangle = \int dr p(r | c\hat{z}_{\text{CMB}}) v_{\text{pec}}^{\text{SN}}(r, \mathbf{n}_{\text{SN}}), \quad (5.33)$$

where $v_{\text{pec}}^{\text{SN}}$ is computed self-consistently from the flow model, Eq. (5.32), and $p(r | \hat{z}_{\text{CMB}})$ is the probability density function (pdf) for r given the observed redshift. This can be linked via a variable transformation to the pdf for the true (latent) CMB redshift of the SNIa, $z_{\text{CMB}}(r)$, via:

$$p(r | c\hat{z}_{\text{CMB}}) = p(cz_{\text{CMB}}(r) | c\hat{z}_{\text{CMB}}) \left| \frac{\partial cz_{\text{CMB}}(r)}{\partial r} \right|, \quad (5.34)$$

where from Eq. (5.3)

$$z_{\text{CMB}}(r) = \bar{z}(r) + (1 + \bar{z}(r)) v_{\text{pec}}^{\text{SN}}(r, \mathbf{n}_{\text{SN}}) / c, \quad (5.35)$$

and the cosmological redshift at comoving distance r , $\bar{z}(r)$, is computed for the Λ CDM concordance model, with parameters as in Table 5.1. Note that the transformation between r and z_{CMB} may not be unique due to the existence of triple-valued regions from r being a cubic

function of z_{CMB} . However, in practice we checked that the reconstructed velocity field indeed leads to unique transformation between r and z_{CMB} for the parameters under consideration. Nevertheless, it can lead to ‘flat’ regions in redshift space (i.e., where $\frac{\partial z_{\text{CMB}}}{\partial r} \approx 0$), which result in large uncertainties in the expected peculiar velocity.

The 2M++ reconstruction employs linear perturbation theory to predict the velocities. As shown in Carrick et al. (2015), this leads to an uncertainty due to non-linearities of $\sigma_v^{\text{NL}} = 150$ km/s. Therefore, assuming a Gaussian uncertainty, we can write the probability of the predicted redshift in the CMB frame for a SNIa at comoving distance r given its observed redshift transformed in the CMB frame in Eq. (5.34) as:

$$p(cz_{\text{CMB}}(r)|c\hat{z}_{\text{CMB}}) = \mathcal{N}(c\hat{z}_{\text{CMB}}, (\sigma_v^{\text{NL}})^2) \quad (5.36)$$

As a check for the robustness of our method, we also estimated the predicted peculiar velocity using an iterative prescription. In this alternative method, we start from the observed CMB restframe redshift as an initial, rough approximation for the comoving distance (under the fiducial Λ CDM assumption), taking $v_{\text{pec}}^{\text{SN}} = 0$. In subsequent iterations, an updated estimate for the comoving distance is obtained using Eq. (5.35) and the velocity estimate we get from the reconstruction at the given comoving distance. This step is repeated until convergence. The iterative method gives broadly consistent results as the marginalization-based method described above. However, the iterative prescription may underestimate the uncertainty in the vicinity of triple-valued regions, and therefore we elected to use the marginalization method instead.

Our new peculiar velocity corrections obtained from Eq. (5.33) are compared against those used in Betoule et al. (2014) in Fig. 5.4, which are obtained from the JLA data products via⁶

$$v_{\text{corr,JLA}} = c \left(\frac{1 + z_{\text{hel}}}{(1 + \bar{z})(1 + z_{\text{pec}}^{\odot})} - 1 \right). \quad (5.37)$$

The correlation coefficient between our peculiar velocity corrections and that used in Betoule

⁶In Eq. (5.37) we use the notation introduced in this paper, but we notice that our \bar{z} is (confusingly) denoted ‘ z_{CMB} ’ in the Betoule et al. (2014) and associated data products.

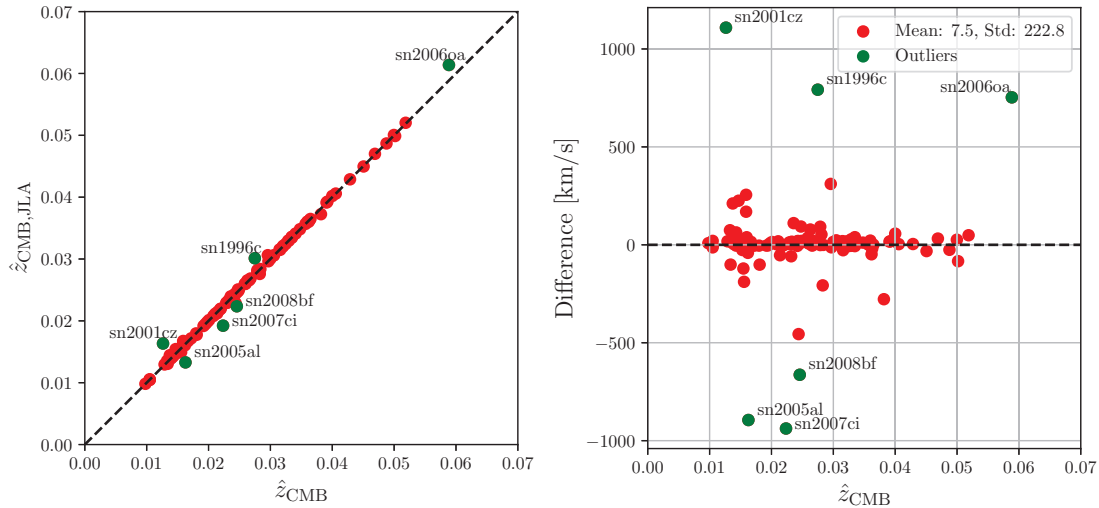


Figure 5.3: Left panel: comparison of the CMB restframe redshift used in this work for the low-redshift sample (horizontal axis, 107 SNe) with that of the JLA analysis (vertical axis). There are 6 SNe (highlighted in green) with a difference exceeding 3 standard deviations around the identity line (dashed). Right panel: the same comparison but showing the fractional differences between redshifts on the vertical axis.

et al. (2014) is only ~ 0.60 . There are also 6 SNe (labeled in Fig. 5.4) that are more than 3 standard deviations of the sample away from the identity line. Several reasons can explain the differences between our peculiar velocities and those used in JLA: firstly, the density field used to predict peculiar velocities in JLA was based on the IRAS PSCz survey, which is likely to be noisier than 2M++ everywhere. Secondly, in the JLA analysis only galaxies in rich clusters are corrected to the mean redshift of the cluster, whereas here we correct the CMB restframe redshift for all host galaxies, including those in clusters and poorer groups.

For SNe beyond $z = 0.067$ (in the SNLS, SDSS and high- z samples), the relative importance of the peculiar velocity corrections diminishes as the recession velocity of the Hubble flow increases, so the detailed treatment becomes less important. In the JLA analysis, peculiar corrections have been set to 0 beyond redshift $z \sim 0.067$, the limit of the galaxy survey from which said corrections were derived. We do the same here, noting that beyond redshift $z \sim 0.067$ the exact treatment of peculiar velocity correction becomes less crucial since their relative importance diminished quickly.

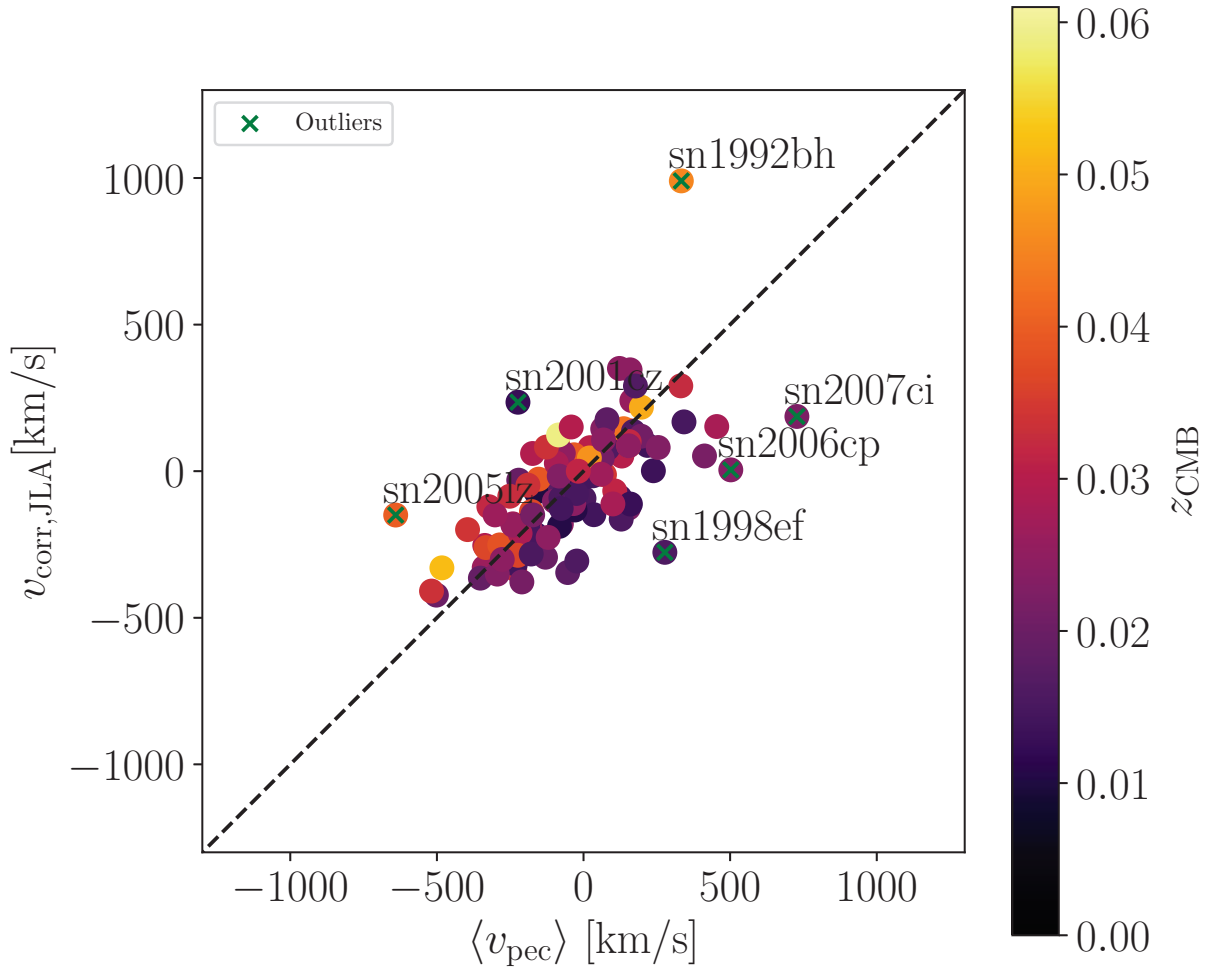


Figure 5.4: Comparison of the peculiar velocity corrections used in this work (horizontal axis) with those adopted in the JLA analysis (vertical axis) for the 107 SNe in the A2 low-redshift sample common to the JLA data. The colour coding gives the CMB restframe redshift used in this work. Outliers (defined as more than 3 standard deviations away from the dashed identity line) are labeled.

5.2.6 Peculiar Velocity Correction Uncertainties

As we have replaced the peculiar velocities of JLA with our own, new estimates, we also update their error analysis associated with the peculiar velocities treatment. First, we remove the diagonal term from the JLA statistical covariance matrix representing the uncertainty from peculiar velocity corrections. We also remove the ‘pecvel’ contribution to the systematic covariance matrix. We replace them with the following terms, flowing from our updated treatment of peculiar velocity corrections.

From Eq. (5.31), the covariance of the cosmological redshift value for SNe i, j is given by (without including the negligible error in z_{pec}^{\odot}) is given by:

$$\Xi_{z,ij}^2 = \delta_{ij}^K \sigma_{z,i}^2 + \Sigma_{z,ij}^2, \quad (5.38)$$

$$\Sigma_{z,ij}^2 = \delta_{ij}^K (\sigma_{\text{NL}}^2 + \sigma_{2\text{M}++}^2(z_{\text{CMB}}) + \sigma_{v,i}^2) / c^2 + C_{ij}^{\text{flow}} / c^2, \quad (5.39)$$

where $\sigma_{z,i}$ is the spectroscopic redshift measurement uncertainty, $\sigma_{\text{NL}} = 150$ km/s is the uncertainty in the peculiar velocity due to non-linearities, $\sigma_{2\text{M}++}$ is the redshift-dependent uncertainty due to survey incompleteness, $\sigma_{v,i}$ is the standard deviation of the average peculiar velocity prediction, Eq. (5.33), and C_{ij}^{flow} is the correlated covariance coming from uncertainty in the flow model (δ_{ij}^K is the Kronecker delta). We address each term in turn.

For the 107 low- z SNe in our analysis, the largest reported statistical uncertainty in their redshift measurement in the JLA data release is $\max_i \sigma_{z,i}^{\text{JLA}} = 0.0014$, corresponding to a velocity uncertainty of 420 km/s. The JLA data release also has $\sigma_{z,i}^{\text{JLA}} = 0$ for 10 of the 107 low- z SNe and $\sigma_{z,i}^{\text{JLA}} < 0$ for 42 SNIa. In order to resolve the issue of 0 or negative redshift uncertainties and to be conservative, we adopt the following prescription for the standard deviation of the spectroscopic uncertainty:

$$\sigma_{z,i} = \max(\sigma_{z,i}^{\text{JLA}}, 5 \cdot 10^{-4}). \quad (5.40)$$

where the floor value of $5 \cdot 10^{-4}$ represents a conservative estimate of the spectroscopic redshift determination from host galaxies spectra. For these 42 SNIa, the floor here translates to a value of 150 km/s, similar to the value of σ_{NL} . This conservative estimate of the uncertainty is used

in the absence of better data, though it should be noted for some low- z objects the uncertainty can be of the order 10's of km/s. The conservative estimate can make it more difficult to detect weak effects from having the uncertainties inflated. However, this impact is expected to be minimal as the velocities from the heliocentric corrections dominate, since they are typically greater than 150km/s and as high as 1000km/s (see Figure 2 of C19). Furthermore, this floor only applies to the smaller subset of SN with incorrect errors.

The term $\sigma_{\text{NL}} = 150$ km/s represents uncertainty in the linear velocity prediction due to unaccounted-for non-linearities, which we fix at the value recommended in Carrick et al. (2015). However, the uncertainty in our reconstructed peculiar velocity increases with redshift, an effect that was ignored in previous work: firstly, the predicted peculiar velocities for tracers near the outer edge of the 2M++ catalogue (200 Mpc/ h) have larger uncertainty because of unaccounted-for structures outside of the survey limits, as well as because of lack of coverage beyond 125 Mpc/ h for part of the sky (Hollinger & Hudson, 2021); secondly, the noise increases at larger distances due to the smaller number of galaxies with larger weights that are used to represent the density field (Lilow & Nusser, 2021). We capture these effects via the redshift-dependent term:

$$\sigma_{2\text{M}++}(z_{\text{CMB}}) = \begin{cases} \sigma_1 z_{\text{CMB}} & \text{for } z_{\text{CMB}} < z_{400}, \\ \sigma_1 z_{400} & \text{for } z_{\text{CMB}} \geq z_{400}, \end{cases} \quad (5.41)$$

where $z_{400} = 0.138$ is the redshift corresponding to a radial comoving distance of 400/ h Mpc, and σ_1 is chosen so that the total peculiar velocity rms beyond z_{400} , i.e. $(\sigma_{\text{NL}}^2 + \sigma_{2\text{M}++}^2(z_{400}))^{1/2}$, equals 380 km/s. This prescription also approximately matches the Λ CDM prediction at the 2M++ boundary, $z = 0.067$, where $(\sigma_{\text{NL}}^2 + \sigma_{2\text{M}++}^2(0.067))^{1/2} = 227$ (km/s). This is in contrast with the original JLA analysis which uses a redshift-independent 150 km/s uncertainty throughout the redshift range.

The term $\sigma_{v,i}^2$ is the variance of $v_{\text{pec}}^{\text{SN}}$ under the distribution given by Eq. (5.34), i.e.

$$\sigma_{v,i}^2 = \langle (v_{\text{pec}}^{\text{SN}})^2 \rangle - \langle v_{\text{pec}}^{\text{SN}} \rangle^2. \quad (5.42)$$

Finally, we translate the redshift covariance of Eq. (5.38) into a magnitude covariance via linear propagation of errors using the isotropic distance modulus of Eq. (5.11), i.e.

$$\sigma_{m,ij}^2 = \Sigma_{z,ij}^2 \left. \frac{\partial \mu_I}{\partial \bar{z}} \right|_{\substack{z_{\text{hel}}=\hat{z}_{\text{hel},i} \\ \bar{z}=\bar{z}_i}} \left. \frac{\partial \mu_I}{\partial \bar{z}} \right|_{\substack{z_{\text{hel}}=\hat{z}_{\text{hel},j} \\ \bar{z}=\bar{z}_j}} + \delta_{ij}^K \sigma_{z,i}^2 \left(\left. \frac{\partial \mu_I}{\partial z_{\text{hel}}} \right|_{\substack{z_{\text{hel}}=\hat{z}_{\text{hel},i} \\ \bar{z}=\bar{z}_i}} \right)^2 \quad (5.43)$$

where \bar{z}_i is computed from Eq. (5.31), and the distance modulus derivatives are evaluated at the fiducial cosmological parameter values given in Table 5.1.

So far, we have considered a fixed value for the flow parameters, $\theta = \{\beta_v, \mathbf{V}_{\text{ext}}\}$ entering in Eq. (5.32). The uncertainties in the inferred flow parameters lead to correlated uncertainties in the peculiar velocities which needs to be accounted for, and that in previous work are usually considered a source of systematic error. Our parameterised flow model allows us to translate them into a statistical error, as follows. In order to estimate the covariance coming from uncertainty in the flow model parameters, we draw 10^4 posterior samples of the flow parameters θ_k ($k = 1, \dots, 10^4$) from the fitted flow model, using the method of [Boruah et al. \(2020b\)](#), and we calculate the average peculiar velocity, $\langle v_{\text{pec},i}^{\text{SN}} \rangle$ for all the 107 SNe in our low- z sample from those samples. We then estimate the covariance of the average peculiar velocity between SNIa i and j as $C_{ij}^{\text{flow}} = \text{Cov}(\langle v_{\text{pec},i}^{\text{SN}} \rangle, \langle v_{\text{pec},j}^{\text{SN}} \rangle)$, where the covariance matrix is computed from the k samples. Since the value of $\sigma_{v,i}$ above varies among the k samples (although the variation is small, $\lesssim 10\%$), we use the average of $\sigma_{v,i}$ from the 10^4 posterior samples. In accord with terminology used in the literature, we call this term the ‘systematic uncertainty’, although as noted above we have actually translated it into a statistical uncertainty. There are no changes to the x_1 and c terms of the covariance and these are left unchanged from the original JLA analysis.

We show in the top panel of Fig. 5.5 the square root of the diagonal entries of the peculiar velocities covariance matrix, translated into magnitude covariance, i.e. $\sigma_{m,ii}$ in Eq. (5.43)). The bottom panel shows the square root of the diagonal entries of the total magnitude covariance matrix (including all other magnitude uncertainties). Our values are compared with the original JLA values on the same figure. The largest difference in the total value of σ_m appears in the

low redshift range, where our re-analysis modifies the associated peculiar velocities, which are dominant in this redshift range. In general, the net effect is to increase the statistical uncertainty while decreasing the systematic uncertainty with respect to the JLA analysis: at the median redshift of the low- z sample, $z = 0.0243$, the average diagonal σ_m due to statistical uncertainty in the peculiar velocities is 0.076 in our analysis (vs 0.045 JLA), while the average systematic diagonal error is 0.020 in our work (vs 0.039 JLA). Overall, the total magnitude uncertainty due to peculiar velocities is increased by $\sim 30\%$ in our analysis (at the median redshift) compared to JLA. In the bottom panel of Fig. 5.5, we compare the total uncertainties on the apparent magnitude (including all statistical and systematic uncertainties) between this work and the JLA analysis, showing that our magnitude uncertainties are generally larger, especially at low redshifts where the new peculiar velocity uncertainties dominate the error budget.

5.2.7 Choice of Priors

As always in a Bayesian analysis, particular attention must be paid to priors, especially in the present case where we are interested in performing not only parameter inference but also model comparison (for an overview of the issue, see e.g. [Trotta \(2008\)](#)).

The priors for the cosmological parameters \mathcal{P}_1 and for the other parameters in the hierarchical model are chosen as in [Shariff et al. \(2016\)](#), to which we refer for fuller details. A summary is provided in Table 5.1. For the cosmographic expansion parameters \mathcal{P}_2 , the Uniform(0, 2) priors in Ω_m, Ω_Λ translate into the following non-uniform prior for the deceleration parameter $q_0 = \Omega_m/2 - \Omega_\Lambda$:

$$p(q_0) = \begin{cases} \frac{2}{3}(q_0 + 2) & \text{for } -2 \leq q_0 < -1, \\ 2/3 & \text{for } -1 \leq q_0 < 0, \\ \frac{2}{3}(1 - q_0) & \text{for } 0 \leq q_0 \leq 1, \\ 0 & \text{otherwise.} \end{cases} \quad (5.44)$$

Given that the likelihood's support is almost entirely within the region $-1.0 \lesssim q_0 \lesssim 0$, we choose the slightly simplified, uniform prior $q_0 \sim \text{Uniform}(-2, 1)$. Since the jerk j_0 and spatial

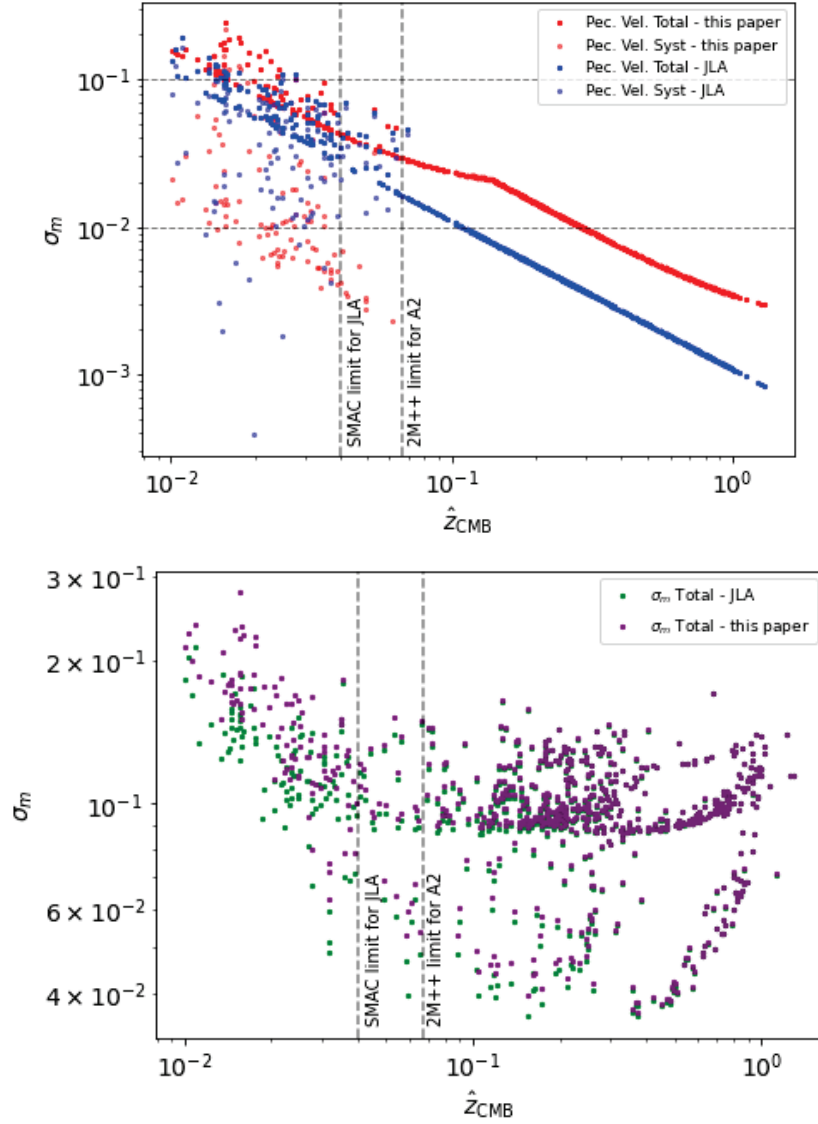


Figure 5.5: Top panel: statistical and systematic (diagonal entries of the covariance matrix only) uncertainties in the apparent magnitude induced by peculiar velocity corrections for our re-analysis compared with JLA. Bottom panel: total diagonal σ_m for our reanalysed and original JLA data. These components include the statistical error from the SALT2 fits and peculiar velocities as well systematics from several other components that are outlined in [Betoule et al. \(2014\)](#).

curvature Ω_κ appear in the degenerate combination $j_0 - \Omega_\kappa$ in the second-order term in the cosmographic expansion, Eq. (5.10), we adopt a uniform prior on this combination, with ranges motivated by the range of physically plausible curvature values. This includes the fact that some authors such as C19 report a recovery of $j_0 - \Omega_K$ that is less than 1 as expected in Λ CDM and closer to zero so we center our prior around zero and set a uniform prior on this combination at $j_0 - \Omega_K \sim \text{Uniform}(-2, 2)$.

The dipole amplitude parameter, D_μ , is a positive quantity for which the obvious choices of priors are a uniform or log-uniform prior. The lower boundary of the former is naturally 0, while for the latter a lower cutoff must be imposed in order for the prior to be proper (i.e., normalizable). This however is problematic for the model selection outcome: as the likelihood becomes flat (i.e., insensitive to the value of D_μ) once D_μ drops below a certain threshold, the marginal posterior becomes equal to the marginal prior for arbitrarily lower values of D_μ with a log-uniform prior. This has an influence on the Bayes factor, as a larger prior range (i.e., a lower D_μ cut-off in a log-uniform prior) leads to a less favourable model selection outcome for the anisotropic model. In order to avoid this difficulty, we choose a uniform prior on D_μ itself. The upper prior cutoff is set by the characteristic scale expected by a dipole signal. This could be gleaned from a theoretical model, or, in a phenomenological approach such as ours, guided by the order of magnitude of previous upper limits on the parameter, which is of order $\sim 10^{-3}$ (see Table 5.4). Such upper limits can however be considerably relaxed in the case of a dipole that is decaying with redshift, leading to upper limits of order $\sim 10^{-1}$ even when no dipole is present (see our simulated case of Fig. 5.8). In order to accommodate such a scenario, we choose a uniform prior $D_\mu \sim \text{Uniform}(0, 0.2)$.

We choose to sample the area of the sky in a uninformative manner since we do not have any prior belief of the directions a dipole might be pointing to. Requiring rotational invariance on the surface of the 2-sphere leads to a uniform distribution on the Galactic longitude of the dipole vector, $l_d \sim \text{Uniform}(0, 2\pi)$ (in radians), and a uniform distribution on the cosine of the latitude of the dipole vector, $\cos(b_d) \sim \text{Uniform}[0, 1]$, with $b_d \in [0, \pi/2]$. Flipping the sign of b_d is equivalent to the transformation $l_d \rightarrow l_d + \pi \bmod 2\pi$ and $D_\mu \rightarrow -D_\mu$. Hence in order to cover the possibility of a dipole pointing in a direction in the southern Galactic hemisphere,

we extend the dipole amplitude to negative values, and therefore our prior is modified to $D_\mu \sim \text{Uniform}(-0.2, 0.2)$. Similar considerations lead to a prior for the dipole amplitude on the deceleration parameter $D_{q_0} \sim \text{Uniform}(-30, 30)$.

For the prior on the exponential scale parameter, S , we need to select a lower boundary (lest D_μ becomes unidentifiable and to stop pathologies associated with $S = 0$), which we take to be the scale of the lowest redshift SNe in our data, namely $S = 0.01$; for the upper boundary, we take $S = 0.1$ as it is known that the bulk flow does not disappear at least out to $z \sim 0.067$. In summary, our prior is thus $S \sim \text{Uniform}[0.01, 0.1]$.

5.3 Simulations and Tests of Methodology

5.3.1 Simulated Data

To test the ability of our setup to recover the anisotropy parameters within the BAHAMAS framework, we forward simulate data from our model in a similar manner to [March et al. \(2011\)](#), adding however JLA-like Galactic coordinates to our simulated SNe in order to carry out inference on a potential dipole. The fiducial values for the parameters are listed in Table 5.1. We consider a case with no dipole ($D_\mu = 0$) and a case with a large dipole ($D_\mu = 0.02$), pointing in the approximate direction of the CMB dipole and with an exponential scale, $S = 0.026$, matching the value preferred by the results of C19. The steps to generate the simulated data are as follows:

1. Draw a value for the latent CMB restframe redshift, $z_{\text{CMB},i}$, for each SNIa. z_{hel} is assumed to equal this CMB redshift. We draw 740 objects at the same redshifts as the real JLA data to ensure the binning of the data is the same when applying selection effects correction on simulated data as it is on real data. Notice that our simulations do not include the issue of peculiar velocity corrections, which are assumed to have already been performed.

Table 5.1: Supernova population parameters, cosmological parameters and dipole model parameters adopted in this work, together with the prior choices and fiducial values for simulation studies. ‘SD’ stands for standard deviation

SN Ia population distributions and covariates	Symbol	Fiducial value	Prior distribution
Mean absolute magnitude of SNIa	M_0	-19.3	$M_0 \sim \text{Normal}(-19.3, 2^2)$
Residual scatter of SNIa magnitude after corrections	σ_{res}	0.1	$\sigma_{\text{res}} \sim \text{InvGamma}(0.003, 0.003)$
Coefficient of stretch covariate	α	0.14	$\alpha \sim \text{Uniform}(0, 1)$
Coefficient of colour covariate	β	3.2	$\beta \sim \text{Uniform}(0, 4)$
Mean of stretch	x_*	0.0	$x_* \sim \text{Normal}(0, 10^2)$
Mean of colour	c_*	0.0	$c_* \sim \text{Normal}(0, 1^2)$
SD of stretch distribution	R_x	1.0	$R_x \sim \text{LogUniform}(-5, 2)$
SD of colour distribution	R_c	0.1	$R_c \sim \text{LogUniform}(-5, 2)$
Parameters controlling the expansion history			
Matter energy density	Ω_M	0.3	$\Omega_M \sim \text{Uniform}(0, 2)$
Dark energy density	Ω_Λ	0.7	$\Omega_\Lambda \sim \text{Uniform}(0, 2)$
Deceleration Parameter	q_0	-0.55	$q_0 \sim \text{Uniform}(-2, 1)$
Jerk and spatial curvature	$j_0 - \Omega_K$	1	$j_0 - \Omega_K \sim \text{Uniform}(-2, 2)$
Hubble-Lemaître constant [km/s/Mpc]	H_0	72	Fixed
Anisotropic expansion parameters			
Galactic longitude of dipole [rad]	l_d	$\{-, 4.60\}$	$l_d \sim \text{Uniform}(0, 2\pi)$
Galactic latitude of dipole [rad]	b_d	$\{-, 0.84\}$	$\cos(b_d) \sim \text{Uniform}(0, 1), b_d \in [0, \pi/2]$
Amplitude of dipole on μ	D_μ	$\{0, 0.02\}$	$D_\mu \sim \text{Uniform}(-0.2, 0.2)$
Amplitude of dipole on q_0	D_{q_0}	$\{0, 10\}$	$D_{q_0} \sim \text{Uniform}(-30, 30)$
Exponential scale of dipole	S	$\{-, 0.026\}$	$S \sim \text{Uniform}(0.01, 0.10)$

2. Compute $\mu_i(\hat{z}, \Theta)$ using the fiducial values for our chosen cosmology, whether Λ CDM or the Cosmographic expansion.
3. Apply the dipolar modulation to the distance modulus using Eq. (5.14).
4. Draw the latent parameters x_{1i} , c_i and M_i from the normal distributions, $x_{1i} \sim N(x_*, R_x^2)$, $c_i \sim N(c_*, R_c^2)$ and $M_i \sim N(M_0, \sigma_{\text{res}}^2)$ respectively.
5. Compute m_{Bi} using x_{1i} , c_i , M_i and the Phillip's relation equation in Eq. (5.11).
6. Draw the value of the standard deviations $\sigma_{x_{1i}}$, σ_{c_i} , and σ_{m_i} , from the appropriate normal distributions fitted to the errors in the JLA data and use them to construct the 3×3 covariance matrix for each SNe as $C = \text{diag}(\sigma_{c_i}^2, \sigma_{x_{1i}}^2, \sigma_{m_i}^2)$.
7. Draw the observed SALT2 parameters from $\hat{x}_{1i} \sim N(x_{1i}, \sigma_{x_{1i}}^2)$, $\hat{c}_i \sim N(c_i, \sigma_{c_i}^2)$ and $\hat{m}_{Bi} \sim N(m_{Bi}, \sigma_{m_i}^2)$.
8. Apply the selection function on the colour values drawn in the previous step. We use the values of σ_{sj}^{obs} and c_{sj}^{obs} inferred from the real JLA data in section 5.2.4 for the redshift bins the SNIa falls in. If a given SNIa is not selected, we cycle back to step (iv) and redraw that SNIa. The process ends when all 740 objects are selected.
9. Generate positions for SNIa in the sky which match the positions of the JLA data. The non-isotropic distribution of the data has an effect on our ability to constrain a dipole, so it is important to match the real JLA SNIa positions for a realistic simulation.

An example of the simulated data, compared with the JLA dataset, is shown in Fig. 5.6 and 5.7, for a Λ CDM realisation. Some banding of the data can be seen of the plots for σ_{x_1} , σ_c , and σ_{m_B} . This results from the different survey components comprising the JLA data. We can see that our simulated data does not capture this banding in the errors because we draw our errors from Gaussian distributions, but this approximation has little quantitative effect on our simulations and our recovered biases are similar to that of March et al. (2011). The well defined 'stripe' in the plot of the galactic coordinates of the data is from the SDSS component of the JLA data.

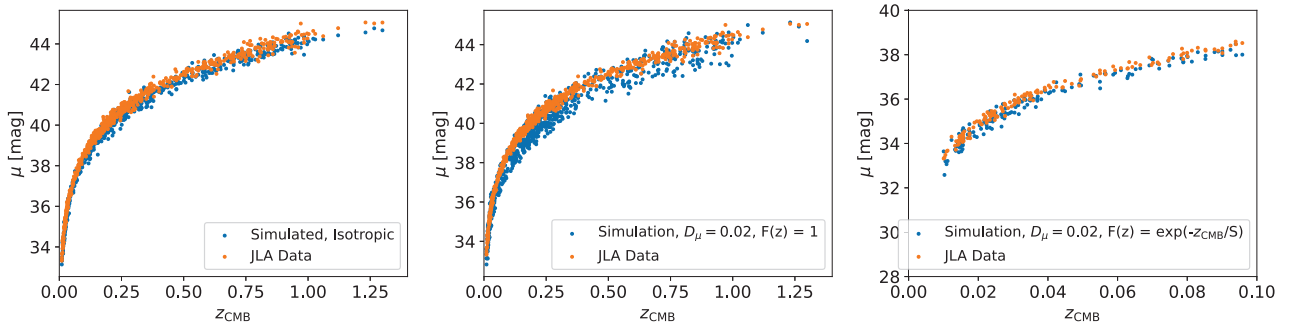


Figure 5.6: Simulated SNIa data generated using the JLA dataset as a reference point. The simulations assume a Λ CDM model and no anisotropy in the distance modulus for the left most model. A dipole of value $D_\mu = 0.02$ is present for the second plot. The third plot also has this value of the dipole, but restricted to a local scale ($z \sim 0.1$) by multiplying the dipole term by the function $F(z) = \exp(-z/0.026)$. For this third plot the redshift has been truncated to only show the redshift range where the dipole is noticeable.

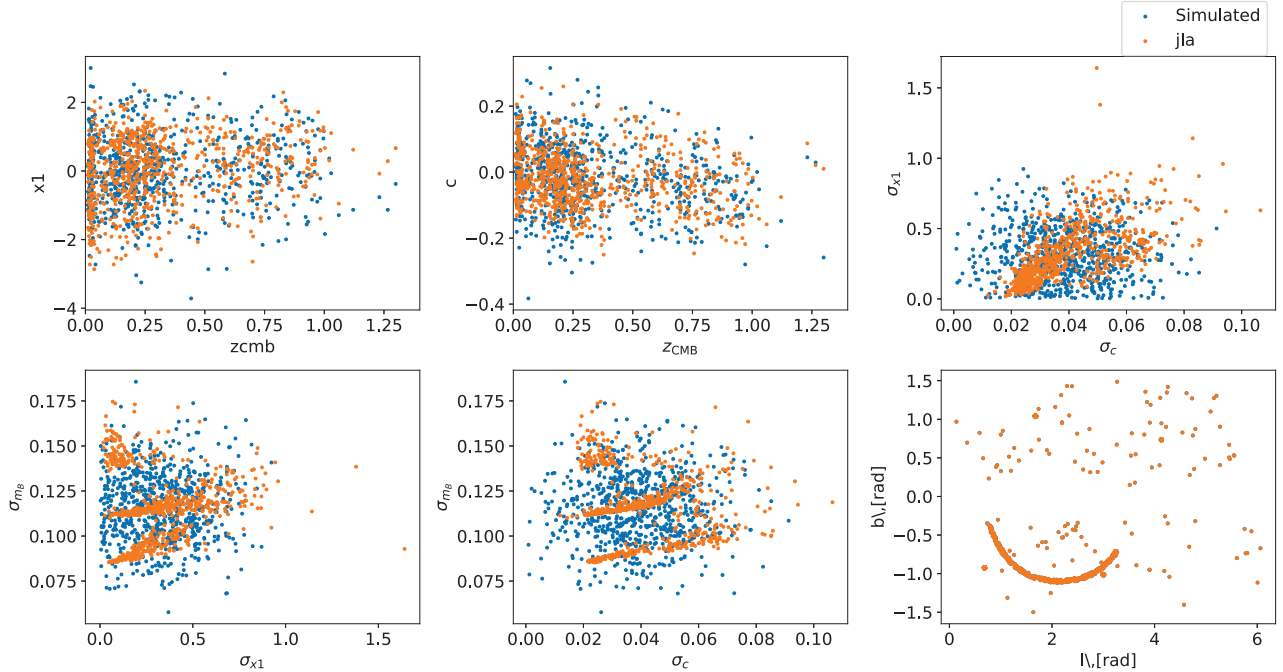


Figure 5.7: Simulated SNIa data generated using the JLA dataset as a reference point, assuming Λ CDM as a fiducial model: distribution of colour (c), stretch (x_1) and the respective standard deviation entering the observational error covariance matrix, as well as the SNIa's sky coordinates.

We test the ability of our method to retrieve the true fiducial values listed in Table 5.1 from the simulated data under the 7 different scenarios below:

1. **Λ CDM-Isotropic** simulates an isotropic universe ($D_\mu = 0$) from Λ CDM and the inference also assumes a Λ CDM cosmology.
2. **Λ CDM-D** simulates a dipole ($D_\mu = 0.02$) with $F(z) = 1$ from Λ CDM and the inference also assumes a Λ CDM cosmology.
3. **Λ CDM-D-exp** simulates a dipole ($D_\mu = 0.02$) with $F(z) = \exp(-z/S)$ from Λ CDM and inference also assumes a Λ CDM cosmology, with the additional free parameter S .
4. **Cosmographic-Isotropic** simulates an isotropic universe ($D_\mu = 0$) from a cosmographic expansion and inference also assumes a cosmographic expansion.
5. **Cosmographic-D** simulates a dipole with $F(z) = 1$ from a cosmographic expansions and inference also assumes a cosmographic expansion.
6. **Cosmographic-D-exp** simulates a dipole ($D_\mu = 0.02$) with $F(z) = \exp(-z/S)$ and inference also assumes a cosmographic expansion, with the additional free parameter S .
7. **Cosmographic-D-exp*** simulates data as in Λ CDM-D-exp but the reconstruction adopts the cosmographic expansion instead. This serves to assess the bias in parameter reconstruction from the cosmographic expansion when the reconstruction model is misspecified.

For each parameterisation we generate 10 realisations of the data; each realisation contains 740 SNIa objects to be similar in size to the JLA data. The posterior results are averaged over the 10 realisations for each parameterisation when we reconstruct the parameters in section 5.3.2. We do not use the Gibbs sampler of Shariff et al. (2016) for posterior sampling, but rather adopt PyMultiNest (Buchner et al., 2014), an implementation of the Nested Sampling algorithm Multinest (Feroz & Hobson, 2008; Feroz et al., 2009; Feroz et al., 2013). The benefit is that we can also compute the Bayesian evidence which we will use for Bayesian model comparison in section 5.3.4.

5.3.2 Parameter Reconstruction from Simulations

We use BAHAMAS to construct posterior distributions (averaged over 10 data realizations) for the set of cosmological parameters \mathcal{P}_1 (i.e., Λ CDM) or \mathcal{P}_2 (i.e., cosmographic expansion) and anisotropy parameters $\{l_d, b_d, D_\mu, S\}$. Although they are sampled over during reconstruction, SNe population parameters and SALT2 coefficients are numerically marginalised over in corner plots and not visualised as they are not the focus of this chapter. The difference between the 1D marginal posterior mean (averaged over realizations) and the true value of each parameter is displayed in Table 5.2. We observe that in all cases except for the scenario Cosmographic-D-exp* the difference is a fraction of a standard deviation, hence entirely within realization and sampling noise. The model misspecification of Cosmographic-D-exp*, however, does lead to shifts of up to $\sim 1.4\sigma$ in the reconstructed cosmographic parameters, a reflection of the fact that the data have been generated under a different model, namely Λ CDM, than has been assumed in the reconstruction. However, the difference for the anisotropy parameters remains below 0.5σ .

It is instructive to investigate the expected constraints on the dipole amplitude when the simulated data are from a isotropic universe (scenario Λ CDM-Isotropic), shown in Fig. 5.8. The 1-sided 95% upper limit we can place on D_μ for this simulated data is $D_\mu \leq 8.08 \times 10^{-4}$ (top panel). When however we introduce the additional freedom of an exponential scale parameter S , the constraints in D_μ degrade by two orders of magnitude, as a small value of S confines any anisotropy to very small redshifts where the statistical power of our data is small and therefore degenerate with many values of D_μ which leads to the entire prior space on D_μ being well explored right up to the prior edge (bottom panels of Fig. 5.8.). The 95% upper bound becomes $D_\mu < 4.36 \times 10^{-2}$ and $S < 4.56 \times 10^{-2}$. Qualitatively similar results apply for the Cosmographic expansion.

The above simulations do not include colour-based selection effects for simplicity. A simulation study including colour-based selection appears in Appendix A (see Fig. A.2) and shows that averaging the posterior distributions across replicates yields 2-dimensional and 1-dimensional marginal distributions that are centered on the true values of the parameters.

Table 5.2: Difference between the 1D marginal posterior mean and the true value used to simulated JLA-data, averaged over 10 data realizations each, $\Delta x \equiv x_{\text{true}} - \bar{x}$ for each parameter x . In parenthesis, the difference is expressed in units of the 1D posterior standard deviation. The top (middle) section shows the case where Λ CDM (Cosmographic expansion) is assumed both for the simulation and the inference; the bottom row assumes Λ CDM for the simulation and the the Cosmographic expansion for the inference – a case of model miss-specification.

Model	Δq_0	$\Delta(j_0 - \Omega_\kappa)$	$\Delta\Omega_{\nu_f}$	$\Delta\Omega_\Lambda$	$\Delta l(\text{rad})$	$\Delta b(\text{rad})$	$ \Delta D_\mu (\times 10^{-2})$	$\Delta S(\times 10^{-2})$
Simulation: Λ CDM; reconstruction: Λ CDM								
(ia) Λ CDM-Isotropic, $F(z) = 0$	-	-	0.03(0.51 σ)	0.05(0.36 σ)	-	-	$< 1.00 \times 10^{-1}(2\sigma)$	-
(ib) Λ CDM-Isotropic, $F(z) = \exp(-z/S)$	-	-	0.03(0.51 σ)	0.05(0.36 σ)	-	-	$< 4.96 \times 10^{-1}(2\sigma)$	$< 9.29(2\sigma)$
(ii) Λ CDM-D	-	-	-0.02(0.19 σ)	0.02(0.10 σ)	-0.003(0.20 σ)	-0.003(0.25 σ)	0.036(0.13 σ)	-
(iii) Λ CDM-D-exp	-	-	0.03(0.91 σ)	0.07(0.72 σ)	0.02(0.20 σ)	-0.02(0.17 σ)	-0.16(0.28 σ)	-0.10(0.21 σ)
Simulation: Cosmographic; reconstruction: Cosmographic								
(iv) Cosmographic-Isotropic, $F(z) = 0$	-0.02(0.20 σ)	0.03(0.01 σ)	-	-	-	-	-	-
(v) Cosmographic-D	-0.06(0.56 σ)	0.27(0.58 σ)	-	-	-0.0116(0.37 σ)	-0.003(0.26 σ)	-0.01(0.35 σ)	-
(vi) Cosmographic-D-exp	-0.01(0.10 σ)	-0.05(0.14 σ)	-	-	-0.04(0.20 σ)	0.01(0.08 σ)	-0.13(0.30 σ)	0.16(0.37 σ)
Simulation: Λ CDM; reconstruction: Cosmographic								
(vii) Cosmographic-D-exp	-0.11(1.05 σ)	-0.63(1.37 σ)	-	-	0.11(0.60 σ)	0.038(0.37 σ)	-0.128(0.44 σ)	-0.165(0.24 σ)

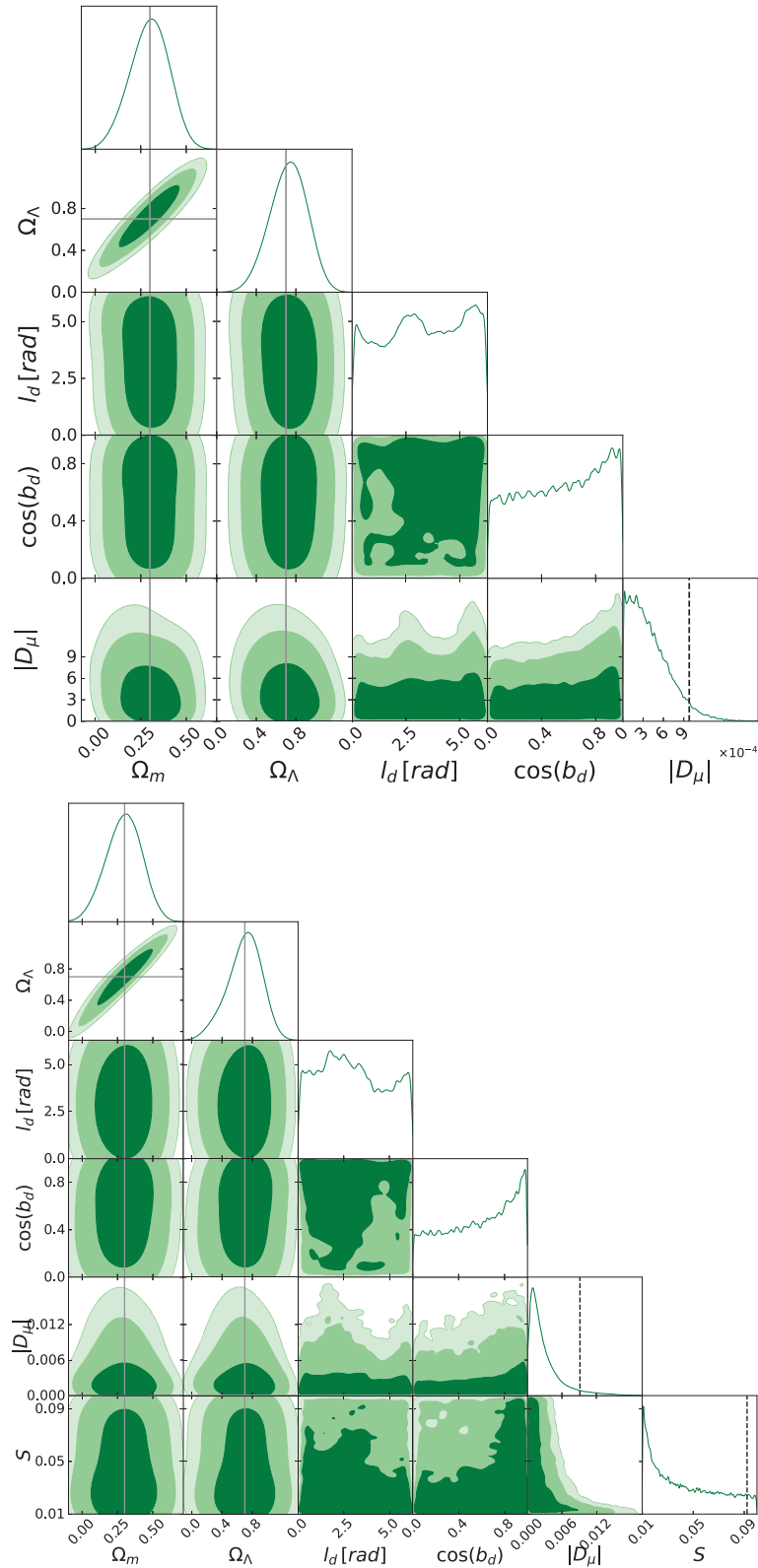


Figure 5.8: Posterior distributions from simulated JLA-like data, averaged over 10 data realizations. The simulations assume a Λ CDM isotropic universe ($D_\mu = 0$); the reconstructions allow for a dipole in the distance modulus with $F(z) = 1$ (left) and $F(z) = \exp(-z/S)$ (right). Contours in 2D plots enclose joint 1, 2 and 3 σ highest posterior density (HPD) credible regions; the vertical dashed line in the 1D marginals for D_μ and S delimits the 2σ upper limit. Notice the different scales in the axis of D_μ between the two cases.

5.3.3 Parameter Inference

We begin by presenting the impact of our new treatment of peculiar velocities and colour selection effects correction on the constraints on Ω_m, Ω_Λ for the case of an isotropic universe, as shown in Fig. 5.9. In the top panel, we compare the constraints using the old JLA peculiar velocities and z_{CMB} as in [Betoule et al. \(2014\)](#) (green) to the ones obtained using our new, group-corrected values of the CMB restframe redshifts (blue) of the source SNIa with no correction or peculiar velocities added for the observer, and additionally replacing the JLA observer peculiar velocity corrections and uncertainties with our newly derived ones (orange). The constraints on the parameters of our models from JLA data are summarised in Table 5.3, for both the Λ CDM model and the Cosmographic expansion. In the top section we also investigate the impact of our newly derived values for z_{CMB} , peculiar velocity corrections and colour-based selection effects on Ω_m, Ω_Λ constraints in an isotropic universe. Starting from the same treatment as [Betoule et al. \(2014\)](#), we find $\Omega_m = 0.306 \pm 0.087$, in good agreement with the value in [Betoule et al. \(2014\)](#), $\Omega_m = 0.295 \pm 0.034$ but with significantly larger uncertainty, perhaps on account of the different statistical approach. When replacing the CMB restframe redshifts used in [Betoule et al. \(2014\)](#) with the ones presented here, we find $\Omega_m = 0.253 \pm 0.089$, a shift of about half a standard deviation according to our uncertainty, but of 1.5σ in units of the standard deviation quoted by [Betoule et al. \(2014\)](#). The effect of using the new peculiar velocity corrections (with their newly derived associated covariance matrix) while maintaining the value of CMB restframe redshift from [Betoule et al. \(2014\)](#) results in a more modest shift, $\Omega_m = 0.297 \pm 0.089$ (case 2M++, old z_{CMB}). When using both the new redshift values and our newly derived peculiar velocity corrections in combination, we obtain $\Omega_m = 0.273 \pm 0.090$. All these results do not use our new colour-based selection effects corrections; once those are included, the constraint on the matter density shifts back to a value close to the original JLA analysis, namely $\Omega_m = 0.290 \pm 0.091$ (but notice the larger uncertainty on our result).

In the bottom panel of Fig. 5.9, we observe a shift in the posterior towards lower Ω_Λ and larger Ω_m when adding the systematic covariance matrix (including our new peculiar velocity covariances) to the statistical covariance matrix, as already noticed by [Shariff et al. \(2016\)](#).

Further adding the correction for colour selection effects shifts the posterior only slightly.

The second section of Table 5.3 presents our constraints on the distance modulus dipole parameters, also comparing the impact of using the colour selection effect correction (as indicated in the third column), which is found to be quite minor on all constraints. Posterior 1D and 2D distributions are shown in Fig. 5.10 for the Λ CDM model and in Fig. 5.11 for the Cosmographic expansion. The posterior distribution for the dipole $|D_\mu|$ in the top panel of Fig. 5.10 for the Λ CDM model with $F(z) = 1$ peaks at 0, and we set a 1-tailed 95% upper bound of $|D_\mu| < 5.93 \times 10^{-4}$ (95.45% probability) – a factor of ~ 2 tighter than the limits derived by Lin et al. (2016a) from the same data, namely $|D_\mu| < 1.98 \times 10^{-3}$. The dipole direction is correspondingly unconstrained.

In the second case, shown in the bottom panel of Fig. 5.10, a scale function of the form $F(z) = \exp(-z/S)$ is used to constrain the dipole to local region. As expected from our simulations, we find a degeneracy between the dipole parameter and its scale, whose effect is to degrade the upper limits on the dipole amplitude to $|D_\mu| < 1.05 \times 10^{-2}$ for the Λ CDM case. We can see $|D_\mu|$ is well explored again, right up to its prior edge because of this degeneracy, which contributes to these weaker constraints. The limits on the dipole scale S are also weak, with the 1D marginal distribution stretching all the way to the prior boundary ($S = 0.10$), but peaking near the lower prior boundary. The very weak preference for a non-zero dipole (seen in the peak away from 0 in the 1D marginal distribution) could be an indication of a residual effects of the bulk flow, which points broadly in the same direction as the more prominent peak in the l_d posterior distribution. Such departures from perfect isotropy are weak, and not dissimilar from what we observed in our isotropic universe simulations (Fig. 5.8, bottom panel) – hence they can be easily ascribed to the result of random noise. With regards to constraints on cosmological parameters, with respect to the Λ CDM-Isotropic case, we observe only a very mild shift in their value as a consequence of the introduction of a potential dipole in the model, $|\Delta\Omega_M| = 0.005$ and $|\Delta\Omega_\Lambda| = 0.006$. Both are shifts of less than 0.1 standard deviations of the posterior, and are similar in scale for the case with $F(z) = 1$. A similar result is seen in the Cosmographic case with $|\Delta q_0| = 0.009$ which is also quite minor relative to the posterior standard deviation.

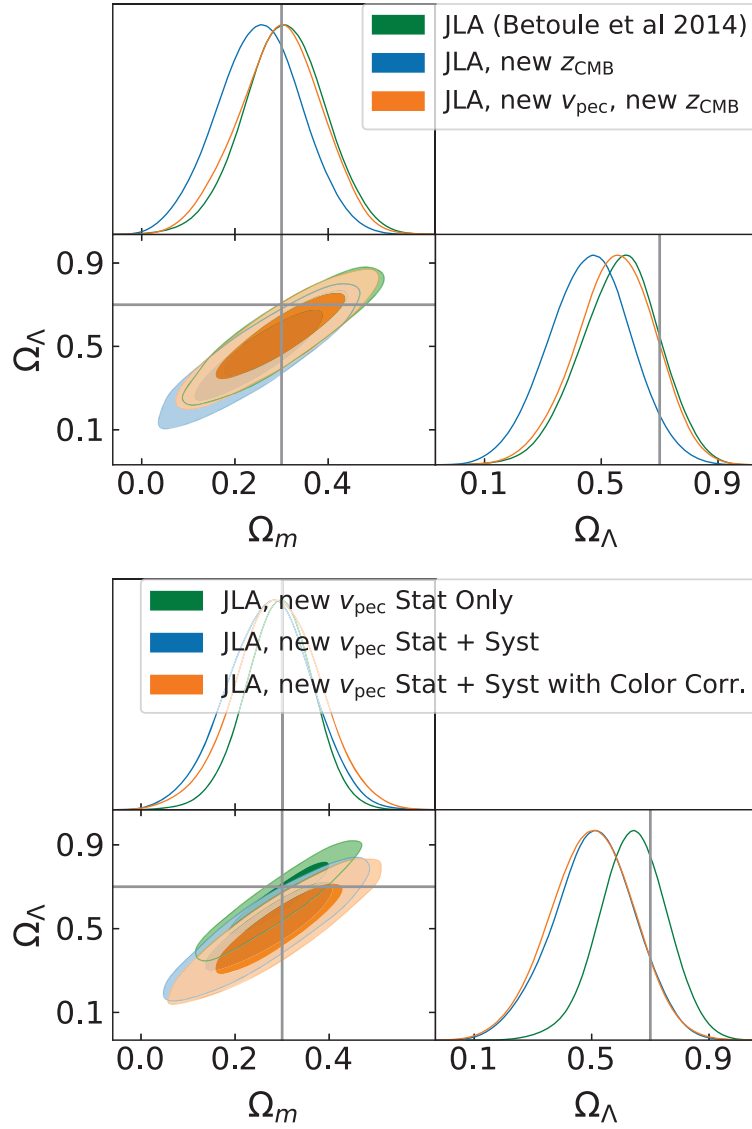


Figure 5.9: Impact of our new treatment of peculiar velocities and colour selection effects correction on the constraints on Ω_m, Ω_Λ for the case of an isotropic universe. Top panel: comparison of marginal posteriors Ω_m, Ω_Λ for isotropic Λ CDM using the old JLA peculiar velocity uncertainties and z_{CMB} (green) to the ones obtained using our new, group-corrected values of the CMB restframe redshifts (blue) of the source SNIa with no peculiar velocity corrections added for the observer, and additionally replacing the JLA observer peculiar velocity corrections and uncertainties with our newly derived ones (orange). Bottom panel: posterior using only the statistical covariance matrix (green), adding the systematic covariance matrix (including our new peculiar velocity covariances, blue) and further adding colour selection effects corrections (orange). In this panel, we adopt our newly derived, group-corrected CMB restframe redshifts.

Table 5.3: Parameter constraints from 1D marginal posterior distributions from JLA data, and outcome of Bayesian model comparison (last column) for the models and data considered. The second column indicates the peculiar velocity corrections adopted: ‘PSCz’ is the same treatment as in Betoule et al 2014; ‘2M++’ indicates the new treatment in this paper; ‘PSCz, new z_{CMB} ’ uses the peculiar velocity corrections of Betoule et al 2014 but our newly derived CMB restframe redshift, while ‘2M++, old z_{CMB} ’ uses our newly derived peculiar velocity corrections but the same z_{CMB} as Betoule et al 2014. The last column is the difference between the model’s log-evidence and log-evidence of the ΛCDM -Isotropic or Cosmographic-Isotropic case (with or without colour correction, as appropriate), i.e., the log of the Bayes factor. A negative value indicates that the model under consideration is disfavoured w.r.t. the isotropic universe. The \star on some of the data for $\Delta \ln(Z)$ are to indicate that the calculation involved a comparison for a Cosmographic expansion with a restricted prior on the deceleration parameter of $q_0 \sim [-2, 0]$ which limits it to an accelerating universe only.

Model	$v_{\text{pec}}^{\text{SN}}$ correction	Colour selection effects correction	Ω_m	Ω_Λ	l_d [rad]	b_d [rad]	$ D_\mu $	Dipole scale, S	Quantity modulated	$\Delta \ln(Z)$
ΛCDM -Isotropic	PSCz	No	0.306 ± 0.087	0.562 ± 0.135	-	-	-	-	-	-
ΛCDM -Isotropic	PSCz, new z_{CMB}	No	0.253 ± 0.089	0.460 ± 0.139	-	-	-	-	-	-
ΛCDM -Isotropic	2M++, old z_{CMB}	No	0.297 ± 0.089	0.547 ± 0.137	-	-	-	-	-	-
ΛCDM -Isotropic	2M++	No	0.273 ± 0.090	0.508 ± 0.140	-	-	-	-	-	0.0
ΛCDM -Isotropic	2M++	Yes	0.290 ± 0.091	0.500 ± 0.141	-	-	-	-	-	0.0
ΛCDM only-Isotropic	2M++	No	0.037 ± 0.028	0	-	-	-	-	-	-5.65
ΛCDM only-Isotropic	2M++	Yes	0.037 ± 0.028	0	-	-	-	-	-	-4.85
ΛCDM -D	2M++	No	0.263 ± 0.090	0.493 ± 0.140	Unconstrained	Unconstrained	$< 5.75 \times 10^{-4}$ (95%)	-	μ	-8.68 ± 0.01
ΛCDM -D	2M++	Yes	0.285 ± 0.094	0.490 ± 0.143	Unconstrained	Unconstrained	$< 5.93 \times 10^{-4}$ (95%)	-	μ	-6.81 ± 0.01
ΛCDM -D-Exp	2M++	No	0.269 ± 0.087	0.499 ± 0.137	Unconstrained	Unconstrained	$< 9.81 \times 10^{-3}$ (95%)	$< 8.75 \times 10^{-2}$ (95%)	μ	-5.27 ± 0.08
ΛCDM -D-Exp	2M++	Yes	0.286 ± 0.091	0.487 ± 0.142	Unconstrained	Unconstrained	$< 1.05 \times 10^{-2}$ (95%)	$< 8.71 \times 10^{-2}$ (95%)	μ	-3.49 ± 0.01
\downarrow	\downarrow	\downarrow	q_0	$j_0 - \Omega_\kappa$	\downarrow	\downarrow	\downarrow	\downarrow	\downarrow	\downarrow
Cosmographic-Isotropic	PSCz	No	-0.352 ± 0.092	0.114 ± 0.391	-	-	-	-	-	-
Cosmographic-Isotropic	PSCz, new z_{CMB}	No	-0.288 ± 0.095	-0.132 ± 0.370	-	-	-	-	-	-
Cosmographic-Isotropic	2M++, old z_{CMB}	No	-0.342 ± 0.093	0.071 ± 0.390	-	-	-	-	-	-
Cosmographic-Isotropic	2M++	No	-0.320 ± 0.094	-0.254 ± 0.383	-	-	-	-	-	0.0
Cosmographic-Isotropic	2M++	Yes	-0.302 ± 0.090	-0.023 ± 0.357	-	-	-	-	-	0.0
Cosmographic- $(q_0 = 0)$	2M++	No	0	-0.953 ± 0.140	-	-	-	-	-	$-3.90 \pm 0.06^*$
Cosmographic- $(q_0 = 0)$	2M++	Yes	0	-0.881 ± 0.134	-	-	-	-	-	$-3.28 \pm 0.09^*$
Cosmographic- $(q_0 \geq 0)$	2M++	No	$< 0.067 \times 10^{-2}$ (95%)	-0.922 ± 0.141	-	-	-	-	-	$-7.39 \pm 0.05^*$
Cosmographic- $(q_0 \geq 0)$	2M++	Yes	$< 0.074 \times 10^{-2}$ (95%)	-0.922 ± 0.144	-	-	-	-	-	$-7.00 \pm 0.09^*$
Cosmographic-D	2M++	No	-0.307 ± 0.093	-0.085 ± 0.375	Unconstrained	Unconstrained	$< 6.43 \times 10^{-4}$ (95%)	-	μ	-6.36 ± 0.02
Cosmographic-D	2M++	Yes	-0.291 ± 0.095	-0.070 ± 0.371	Unconstrained	Unconstrained	$< 6.32 \times 10^{-4}$ (95%)	-	μ	-6.47 ± 0.02
Cosmographic-D-Exp	2M++	No	-0.307 ± 0.089	-0.069 ± 0.353	Unconstrained	Unconstrained	$< 1.00 \times 10^{-2}$ (95%)	$< 8.68 \times 10^{-2}$ (95%)	μ	-2.96 ± 0.05
Cosmographic-D-Exp	2M++	Yes	-0.293 ± 0.094	-0.048 ± 0.362	Unconstrained	Unconstrained	$< 9.63 \times 10^{-3}$ (95%)	$< 8.94 \times 10^{-2}$ (95%)	μ	-3.26 ± 0.01
\downarrow	\downarrow	\downarrow	q_0	$j_0 - \Omega_\kappa$	\downarrow	\downarrow	$ D_\mu $	\downarrow	\downarrow	\downarrow
Cosmographic-D	2M++	No	-0.305 ± 0.094	-0.083 ± 0.376	Unconstrained	Unconstrained	$< 6.06 \times 10^{-2}$ (95%)	-	q_0	-7.12 ± 0.02
Cosmographic-D	2M++	Yes	-0.286 ± 0.094	-0.084 ± 0.367	Unconstrained	Unconstrained	$< 6.29 \times 10^{-2}$ (95%)	-	q_0	-7.45 ± 0.08
Cosmographic-D-Exp	2M++	No	-0.316 ± 0.097	-0.040 ± 0.370	Unconstrained	Unconstrained	< 22.46 (95%)	$< 5.59 \times 10^{-2}$ (95%)	q_0	-1.22 ± 0.02
Cosmographic-D-Exp	2M++	Yes	-0.296 ± 0.093	-0.043 ± 0.359	Unconstrained	Unconstrained	< 23.45 (95%)	$< 6.00 \times 10^{-2}$ (95%)	q_0	-1.42 ± 0.02

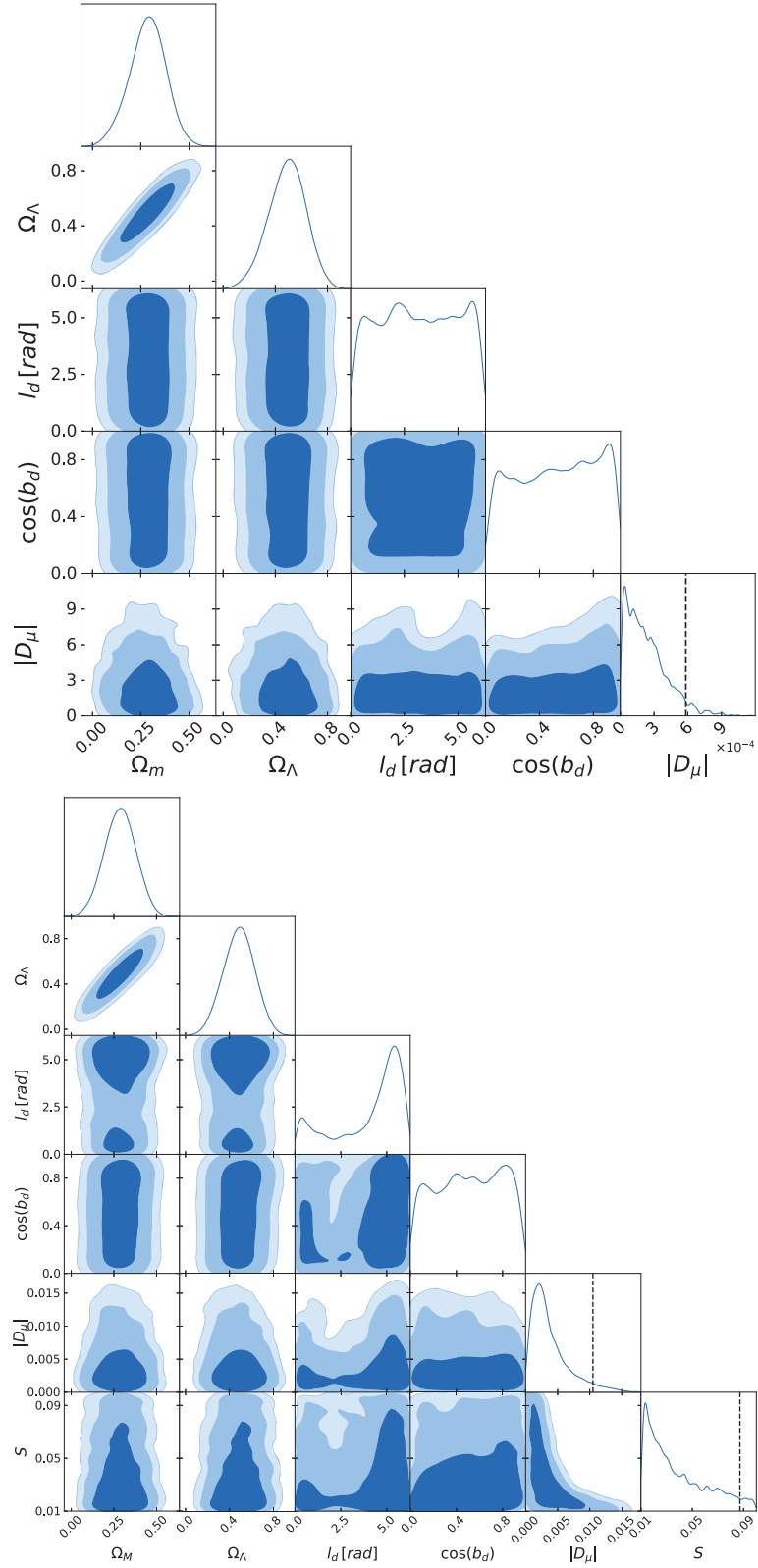


Figure 5.10: Posterior inference allowing for a distance modulus dipole from JLA data, assuming the Λ CDM model with $F(z) = 1$ (left) and $F(z) = \exp(-z/S)$ (right), including colour-based selection effects correction. Contours in 2D plots enclose joint 1, 2 and 3 σ HPD credible regions; the vertical dashed line in the 1D marginals for D_μ and S delimits the 2σ upper limit.

For the isotropic Cosmographic expansion, the posterior mean for the deceleration parameter increases from $q_0 = -0.352 \pm 0.092$ when using the old PSCz velocity corrections and z_{CMB} to $q_0 = -0.302 \pm 0.090$ when updating both velocity corrections and CMB restframe redshifts to the new values we derived here (the error indicates the standard deviation of the posterior, not the uncertainty on the mean). When introducing the possibility of a dipole, the posterior mean for q_0 hovers around -0.30 , depending on the dipole model and whether we adopt the colour selection effects corrections. These values are quite a bit larger than the expectation under ΛCDM , namely $q_0 = -0.55$, but not as large as the results reported in Table 2 of [Colin et al. \(2019b\)](#) ($q_0 = -0.157$ in our notation), who ascribed the shift of the deceleration parameter towards 0 to evidence for an anisotropic universe. Our analysis (with its improved treatments of peculiar velocities and colour selection effects) disagrees with those conclusions: the marginal posterior probability (obtained from our posterior samples, not a Gaussian approximation to the posterior) for $q_0 \geq 0$ (i.e., no acceleration) for the $F(z) = 1$ case with selection effects corrections and a distance modulus dipole is 9.3×10^{-4} . We return on this question from the point of view of Bayesian model comparison in section 5.3.4.

The constraints on the dipole parameters for the Cosmographic expansion model are qualitatively similar to those presented for the ΛCDM model, as shown in Fig. 5.11 and detailed in the central two sections of Table 5.3. For the $F(z) = 1$ case, the posterior dipole amplitude peaks at 0 and we set a 95% upper limit $|D_{q_0}| < 6.32 \times 10^{-4}$. There is no significant evidence for a dipole moment in the Cosmographic expansion case under our data and models. A qualitatively similar picture holds for the $F(z) = \exp(-z/S)$ case, albeit with weaker limits on the dipole amplitude owing to the degeneracy with the scale parameter S .

For a more direct comparison with the results of C19, we have also investigated the same model as C19, where the dipole modulation is applied to the deceleration parameter q_0 , as in Eq. (5.15), rather than to the distance modulus μ . Differently from C19, we did *not* remove the bias corrections to the magnitude, we kept the direction of the dipole free (as opposed to being fixed in the CMB dipole direction), used our new peculiar velocity corrections and CMB restframe redshifts (as opposed to heliocentric redshifts) and applied our new colour selection effects. The resulting posteriors are shown in Fig. 5.12 and constraints presented in the bottom

section of Table 5.3.

We see qualitatively similar results to the case where we model the dipole on q_0 in the Cosmographic expansion. For the $F(z) = 1$ case, the posterior dipole amplitude again peaks at 0 and we set a 95% upper limit $|D_{q_0}| < 6.29 \times 10^{-2}$. A qualitatively similar picture holds for the $F(z) = \exp(-z/S)$ case with a preference for a small S value. This is in contrast with [Colin et al. \(2019b\)](#), who saw a preference of a scale value $S = 0.0262$, a likely consequence of their removal of peculiar velocity correction and the use of heliocentric redshifts.

It should be noted that the posteriors indicated in Figure 5.12 do cover a range that includes the inferred value of the dipole amplitude detected by C19 at $D = 8.03$ for the case of the exponential scale decay model. However, this should not be interpreted as evidence for their dipole model as we will see by model comparison in section 5.3.4. Due to the nature of random sampling and our prior constraints, there is a non zero probability that some part of the posterior includes their value given enough samples, but also, from the shape of our posteriors on the JLA data, it is clear the preference for a dipole amplitude of zero as it is not strongly peaked at a value away from zero. Such behaviour can be seen in our plot of a simulation with a diople present where the posterior is strongly peaked at the simulated dipole value in Fig. A.2

5.3.4 Bayesian Model Comparison

We compare the isotropic expansion model to the alternatives featuring a dipole via Bayesian model comparison, and report the difference in the log of the Bayesian evidence (i.e., the log of the Bayes factor) in Table 5.3:

$$\Delta \ln(Z) = \ln B_D - \ln B_I, \quad (5.45)$$

where B_I is the Bayesian evidence for the isotropic model (either Λ CDM or Cosmographic expansion) and B_D is the evidence for a model featuring a dipole, with priors as in. Table 5.1. A value of $\Delta \ln(Z) < 0$ indicates a preference for the isotropic model. According to the Jaynes' scale for the strength of evidence, values of $|\Delta \ln(Z)| = 2.5(5.0)$ correspond to moderate

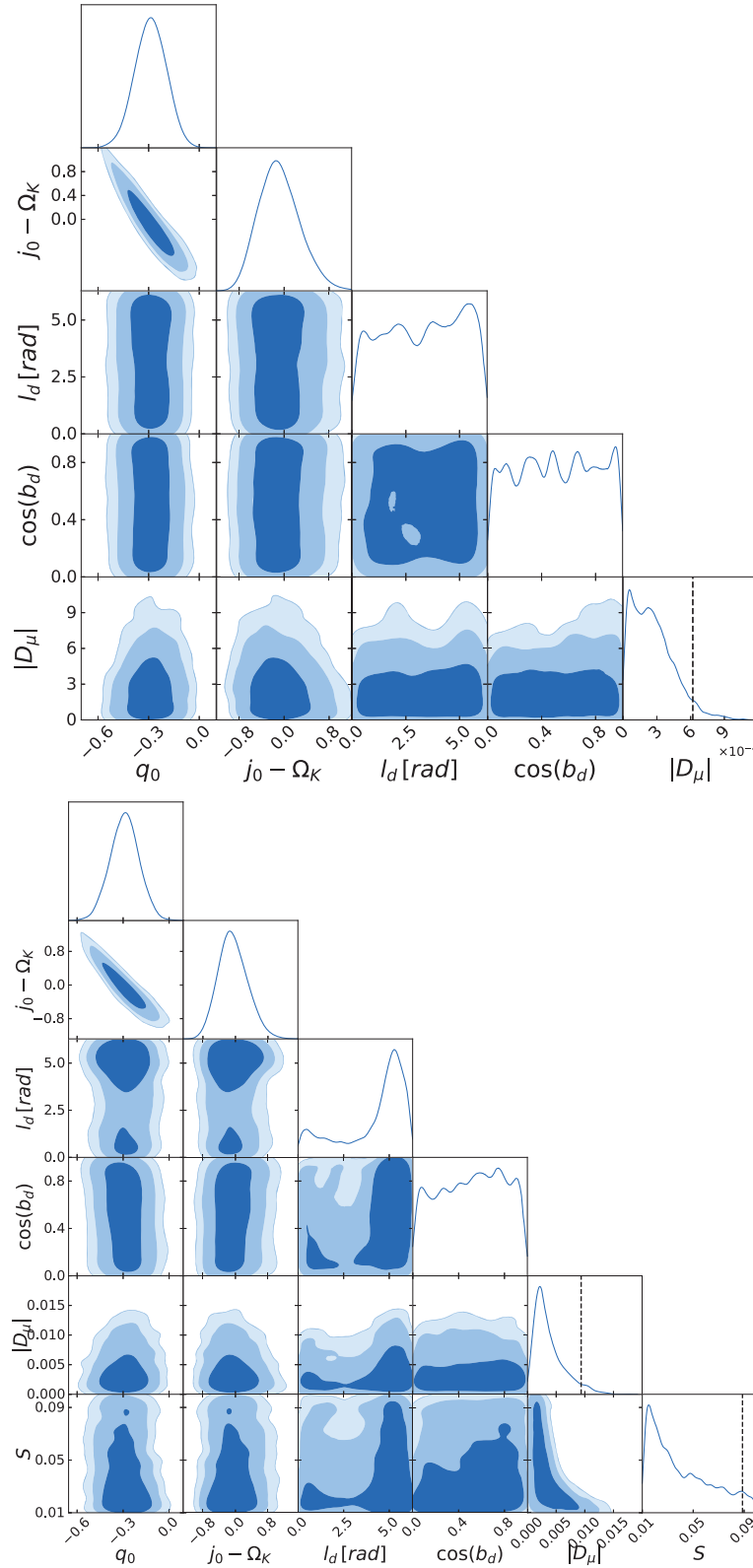


Figure 5.11: Posterior inference allowing for a distance modulus dipole from JLA data, assuming the Cosmographic expansion model with $F(z) = 1$ (left) and $F(z) = \exp(-z/S)$ (right), including colour-based selection effects correction. Contours in 2D plots enclose joint 1, 2 and 3 σ HPD credible regions; the vertical dashed line in the 1D marginals for D_μ and S delimits the 2σ upper limit.

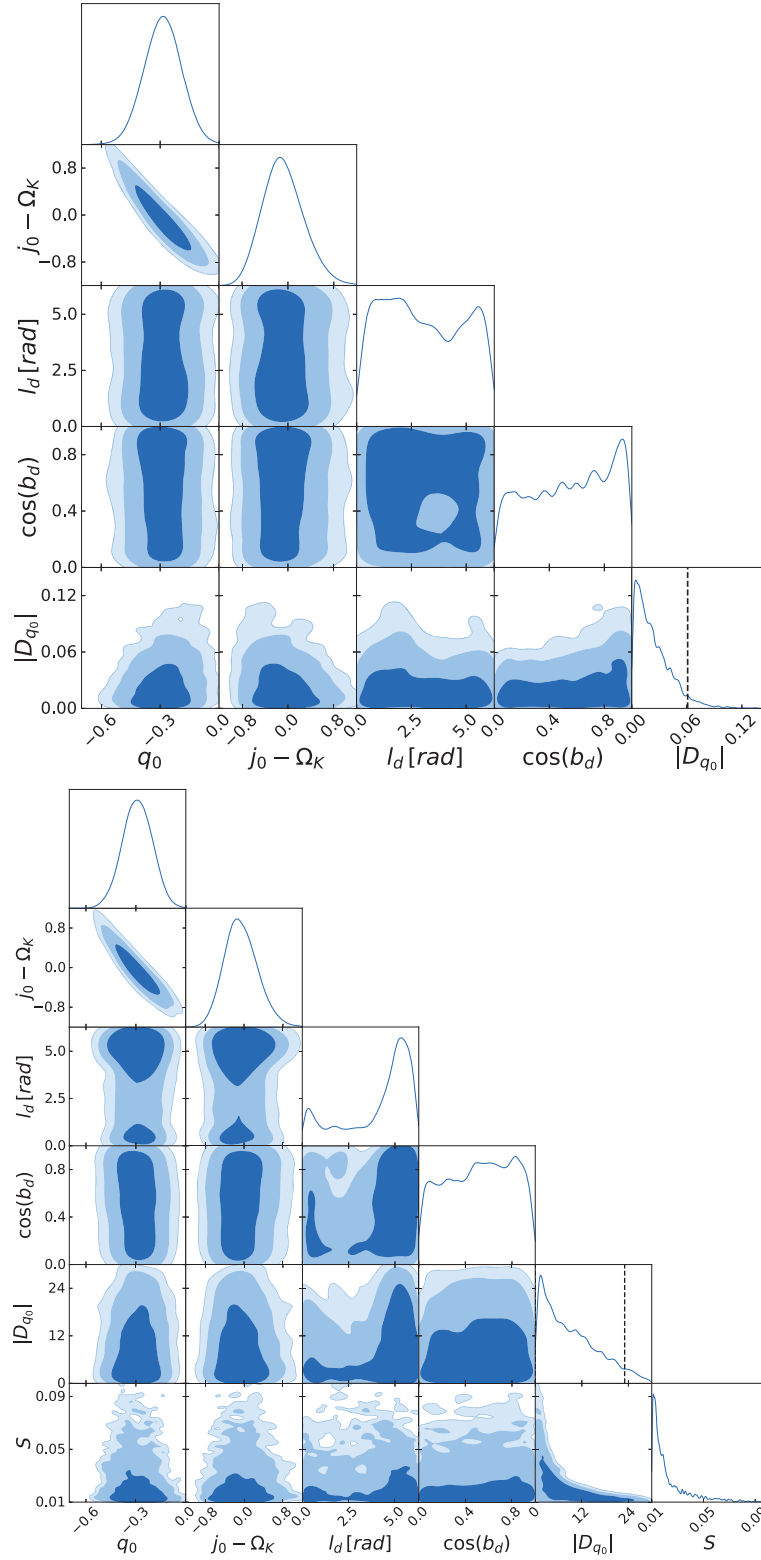


Figure 5.12: Posterior inference allowing for a distance modulus dipole from JLA data, assuming the Cosmographic expansion model with $F(z) = 1$ (left) and $F(z) = \exp(-z/S)$ (right), including colour-based selection effects correction. Contours in 2D plots enclose joint 1, 2 and 3 σ HPD credible regions; the vertical dashed line in the 1D marginals for D_{q_0} and S delimits the 2σ upper limit. The dipole here is modelled directly on the deceleration parameter rather than the distance modulus, similar to C19.

Table 5.4: Comparison of our results with previous searches for a dipolar modulation. We only include comparable results that utilize the ‘dipole fitting’ approach (as opposed to the ‘hemisphere comparison’ method). Reported detections (at higher than 95% significance) are highlighted in boldface. The third column ‘ v_{pec} corr’ states whether peculiar velocity corrections were adopted, and if so from which data set they were derived. It should be noted here, previous significant detections involved data that had no peculiar velocity corrections.

Reference	Data	v_{pec} corr	l_d (deg)	b_d (deg)	Dipole amplitude	Quantity modulated
Cooke & Lynden-Bell (2010)	Union	None	309	43	0.14 ± 0.12	d_L
Lin et al. (2016b)	Union2.1	None	$171.8^{+42.0}_{-42.0}$	$9.9^{+20.3}_{-20.3}$	0.160 ± 0.115	Ω_m
Mariano & Perivolaropoulos (2012)	Union2	None	309^{+18}_{-18}	$-15.1^{+11.5}_{-11.5}$	$(1.3 \pm 0.6) \times 10^{-3}$	μ
Yang et al. (2014)	Union2.1	None	$307.1^{+16.2}_{-16.2}$	$-14.3^{+10.1}_{-10.1}$	$(1.2 \pm 0.5) \times 10^{-3}$	μ
Wang & Wang (2014)	Union2.1+GRB	None	$309.2^{+15.8}_{-15.8}$	$-8.6^{+10.5}_{-10.5}$	$(1.37 \pm 0.57) \times 10^{-3}$	μ
Lin et al. (2016a)	JLA	PSCz	316^{+107}_{-110}	-5^{+41}_{-60}	$< 1.98 \times 10^{-3}$ (95%)	μ
Sun & Wang (2018a)	Union2.1	None	$309^{+15.5}_{-15.7}$	$-8.9^{+11.2}_{-9.8}$	$(1.46 \pm 0.56) \times 10^{-3}$	μ
Sun & Wang (2018a)	Constitution	None	$67.0^{+66.5}_{-66.2}$	$-0.6^{+25.2}_{-26.3}$	$(4.4 \pm 5.0) \times 10^{-4}$	μ
Sun & Wang (2018a)	JLA	PSCz	Unconstrained	Unconstrained	Unconstrained	μ
Sun & Wang (2018c)	Pantheon	2M++	329^{+101}_{-28}	37^{+52}_{-21}	$(3.7^{+2.5}_{-3.7}) \times 10^{-4}$ (95%)	μ
Zhao et al. (2019)	Pantheon	2M++	$306.00^{+82.95}_{-125.01}$	$-34.20^{+16.82}_{-54.93}$	$< 1.16 \times 10^{-4}$ (95%)	μ
This work	JLA	2M++	Unconstrained	Unconstrained	$< 5.93 \times 10^{-4}$ (95%)	$\mu, F(z) = 1$
This work	JLA	2M++	Unconstrained	Unconstrained	$< 1.05 \times 10^{-2}$ (95%)	$\mu, F(z) = \exp(-z/S)$
Colin et al. (2019b)	JLA	None	264.021 (fixed)	48.253 (fixed)	$-8.03^{+2.05}_{-2.05}$	$q_0, F(z) = \exp(-z/S)$
Rubin & Heitlauf (2020)	JLA	None	264.021 (fixed)	48.253 (fixed)	$-8.65^{+2.2}_{-2.6}$	$q_0, F(z) = \exp(-z/S)$
Rubin & Heitlauf (2020)	JLA	PCSz	264.021 (fixed)	48.253 (fixed)	$-1.1^{+3.2}_{-3.4}$	$q_0, F(z) = \exp(-z/S)$
This work	JLA	2M++	Unconstrained	Unconstrained	$< 6.32 \times 10^{-2}$ (95%)	$q_0, F(z) = 1$
This work	JLA	2M++	Unconstrained	Unconstrained	< 13.68 (95%)	$q_0, F(z) = \exp(-z/S)$

(strong) evidence for one of the models being compared. The Jaynes' scale is only used here as a qualitative description of the strength of evidence for a given model, but it is up to the practitioner to decide what is sufficient evidence for making a decision based on the implied relative probability between the models from the Bayes Factor. For equal prior probability for the models, the quantity $\exp(\Delta \ln(Z)) = B_D/B_I$ gives the posterior odds between the isotropic and the dipole expansion models, which are approximately 12:1 (150:1) for moderate (strong) evidence (see [Trotta \(2008\)](#) and chapter 4 for details on Bayesian model comparison). When considering the numerical odds derived from the Bayesian evidence, it is important to bear in mind that these can be very sensitive to the choice of prior distribution, particularly the prior on the parameters that only appears in the more complicated model (i.e., the model featuring the dipole). While we believe that our choices of prior distributions are well justified in Section 5.2.7, researchers that make other choices for their prior distributions may compute odds ratios that differ from those that we report.

In Table 5.1 we only carry out pair-wise model comparisons between models that use the same data and same treatment of colour selection effects, for comparing models with different data and/or assumptions about the data generating process would be meaningless. The evidences (and associated uncertainties) are estimated with `PyMultinest`, which was run with 400 live points and an evidence tolerance of 0.5. We observe that the isotropic model is favoured over all others: in the Λ CDM scenario, the constant dipole model is disfavoured with odds ranging from 900:1 to almost 6000:1, depending on the adoption of colour selection effects. The model with a dipole falling off with redshift is also disfavoured, albeit with smaller odds ranging between 32:1 and 194:1.

A similar picture holds for comparison between models in the Cosmographic expansion case. Here, the odds against the anisotropic models are generally smaller than in Λ CDM, owing to the smaller parameter space volume ratio between posterior and prior, which control the strength of the Occam's razor effect in favour of the isotropic model. We also note that when introducing a dipole scale parameter S , despite the larger number of free parameters in this model w.r.t. the case where the dipole is constant in redshift, the Bayes factor against it is actually smaller than the constant-in-redshift dipole case. This can be explained by noting that the introduction of

S as a free parameter leads to much less stringent limits on the dipole amplitude because of the degeneracy explained above. Since S itself is only weakly constrained, the Occam's razor effect for these two parameters is weakened, leading to a weaker preference for the isotropic model. As a consequence, the Cosmographic-D-Exp model where the dipole is on the deceleration parameter is only very mildly disfavoured (odds smaller than 5:1) w.r.t. the isotropic model – a consequence of the fact that the additional parameters for this model cannot be strongly constrained, and hence the posterior odds remain approximately equal to the prior odds.

As another case, we reproduce the setup used in C19, namely removing peculiar velocity corrections entirely, using heliocentric redshifts (i.e., z_{hel} instead of \bar{z} in Eq. (5.31), a choice that imprints the dipole due to the solar system's motion onto the data, as pointed out by Rubin & Heitlauf (2020)) and removing from the covariance matrix all uncertainties associated with peculiar velocity corrections. In this setup, we compare the evidence for an isotropic Cosmographic expansion with that of a dipolar modulation of the form $F(z) = \exp(-z/S)$ either on the deceleration parameter, Eq. (5.15) (as in C19), or on the distance modulus, Eq. (5.14). We adopt a Gaussian prior with a standard deviation of 10° on the dipole direction, centred on the bulk flow direction from Boruah et al. (2020b), namely $l_{bf} = 301^\circ \pm 4^\circ$, $b_{bf} = 0^\circ \pm 3^\circ$ (in excellent agreement with the results of Said et al. (2020), obtained using the Fundamental Plane relation, namely $l_{bf} = 304^\circ \pm 4^\circ$, $b_{bf} = 1^\circ \pm 4^\circ$). When the dipole is modelled on the deceleration parameter, as in C19, the Cosmographic anisotropic model is still disfavoured with respect to the isotropic one, with odds of approximately 17:1 ($\Delta \ln(Z) = -2.84 \pm 0.08$). Although the anisotropic model achieves a better quality of fit by absorbing the dipole in the data, from an Occam's razor perspective it remains disfavoured due to its additional, unwarranted model complexity. We observe a similar effect (if stronger) when the dipole is modelled instead on distance modulus μ , with odds of approximately 150:1 ($\Delta \ln(Z) = 5.03 \pm 0.03$) in favour of the isotropic model. We can repeat this comparison in the isotropic Λ CDM case and compare that with an anisotropic Λ CDM model with $F(z) = \exp(-z/S)$, finding $\Delta \ln(Z) = -4.92 \pm 0.15$, which again favours the isotropic model with odds of 136:1.

It is interesting to view here the Bayes factors related to the simulated data. For this, we simulate data sets with no dipole in both the Cosmographic and Λ CDM case and calculate the

Bayes factors from comparing the evidences for an isotropic model and an anisotropic model with the exponential scale decay. The Bayes factor in this scenario is computed as in 5.45 so that a negative value of the Bayes factor favours the isotropic model and a positive value the dipole model. We also simulate data that has a genuine dipole present in the Cosmographic and Λ CDM case and compute the Bayes factors again as before. Each set of simulations and inferences are run for 100 iterations using the steps and parameter values outlined in section 5.3.1.

The top of Figure 5.13 shows all Bayes factors computed prefer the isotropic model when no dipole is present in the simulated data with a spread of values that is similar to the spread of values seen in Table 5.3. The standard deviation of $\Delta \ln Z$ in this case is 0.79 (Λ CDM) and 0.69 (Cosmographic). On the other hand, while the bottom of Figure 5.13 shows all computed Bayes factors prefer the dipole model for data with a dipole present, the strength of the preference is much higher with the computed Bayes factors being larger in magnitude but also variance, with both sets of $\Delta \ln Z$ having standard deviations of 7.5 (Λ CDM and Cosmographic).

The above further lends evidence to the isotropic model as the distribution of Bayes factors unanimously favour the correct model in each set of simulations.

Another interesting question is the strength of evidence in favour of an accelerating universe in the isotropic expansion case. The Bayes factor between the isotropic Λ CDM model and an isotropic model with no dark energy ($\Omega_\Lambda = 0$) disfavors the latter with odds in excess of 120:1 (including selection effects corrections). For completeness (and to compare the above Bayesian model comparison results with a hypothesis testing approach), we have also computed the log-likelihood difference for the best-fit parameter values:

$$\Delta \ln(L) = \ln \hat{L}_{\Omega_\Lambda=0} - \ln \hat{L}_{\Lambda\text{CDM}} \quad (5.46)$$

where $\hat{L}_{\Lambda\text{CDM}}$ is the maximum likelihood value for the Λ CDM model and $\hat{L}_{\Omega_\Lambda=0}$ is the maximum likelihood value for a universe with no dark energy. We find $\Delta \ln(L) = -6.35$ (with selection effects correction), which we translate into a p -value using Chernoff theorem (as the hypothesis

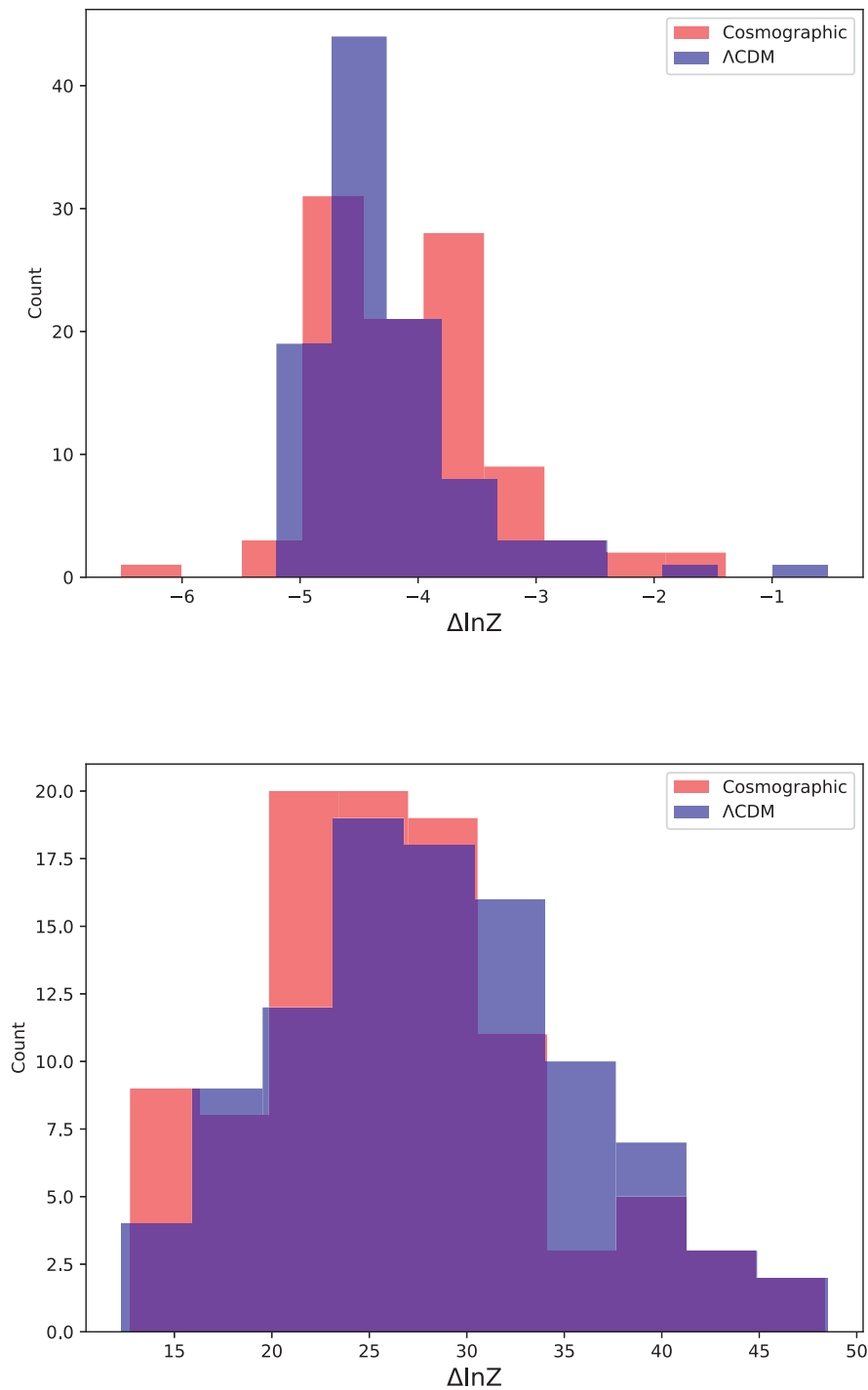


Figure 5.13: Top: The distribution of Bayes factors for simulated data with no dipole present. Bottom: The distribution of Bayes factors for simulated data with an exponentially decaying dipole present. In both of the above, plots the Bayes factor is computed such that negative values prefer isotropic models and positive values favour the exponentially decaying dipole model.

being tested, $\Omega_\Lambda = 0$, lies on the boundary of the allowed parameter space). For the null hypothesis of a $\Omega_\Lambda = 0$ universe which contains no dark energy, we obtain a p -value of 5.9×10^{-3} meaning such a universe as a null hypothesis is easily rejected under the common significance threshold of 5%.

In the phenomenological Cosmographic expansion setting, the accelerating isotropic model, with a uniform prior on $q_0 \in [-2, 0)$ is favoured with odds of approximately 26:1 when compared to a coasting universe, i.e., $q_0 = 0$. The accelerating universe is preferred with odds of almost $\sim 1100:1$ when compared with a decelerating model, i.e. one with a uniform prior $q_0 \in (0, 1]$.

5.4 Summary and Conclusions

We have revisited the question of a dipolar anisotropy in the expansion of the universe, and derived new constraints on a possible dipolar modulation from SNIa data. Our approach builds on the Bayesian hierarchical model **BAHAMAS**, which has been extended to include a new approximate correction for residual colour-based selection effects. We have also upgraded the treatment of peculiar velocities and host galaxy redshifts from the original JLA paper by [Betoule et al. \(2014\)](#), by adopting state-of-the-art flow models constrained using the 2M++ galaxy catalogue. Finally, we have improved the treatment of both statistical and systematic uncertainties pertaining to peculiar velocity corrections – the dominant source of error for $z \lesssim 0.1$ SNIa, which are all-important for a robust, accurate and precise measurement of anisotropy in the local expansion.

We did not find any evidence for a deviation from isotropy, either in the framework of Λ CDM or in phenomenological Cosmographic expansion. We placed tight constraints on the amplitude of a possible dipole both in the distance modulus and on the deceleration parameter. Our upper bounds are more stringent by a factor of ~ 2 than the results previously obtained from the same data sets with a comparable approach. We note that all previous searches that have claimed a significant detection of anisotropy have neglected peculiar velocity corrections.

We have used the framework of Bayesian model comparison to evaluate the Bayes factor between

models featuring a dipole and an isotropically expanding universe (both in Λ CDM and in the Cosmographic expansion). We found moderate to strong Bayesian evidence against an anisotropic expansion. We have also evaluated the evidence in favour of acceleration, finding that a non-zero cosmological constant is preferred, using JLA SNe alone, by odds of 120:1, a result corroborated by a more traditional p -value approach based on a frequentist hypothesis test, which rejects $\Omega_\Lambda = 0$ with a p -value of 5.9×10^{-3} . In the Cosmographic expansion, a decelerating universe is disfavoured with odds of almost 1100 : 1 w.r.t. an accelerating one. We conclude that the preferred model remains the Λ CDM isotropically expanding model.

The effect of our peculiar velocity measurements on H_0 measurements is difficult to quantify within our Bayesian model as the parameter is exactly degenerate with the SNIa intrinsic magnitude and hence is fixed $H_0 = 72 \text{ km/s/Mpc}$ in our analysis. However, we provide our corrections to the JLA data publicly so others wishing to use it for studies relating to H_0 can do so.

In this work we adopted the JLA compilation since we were unable to use the most recent Pantheon sample, owing to the full covariance matrix (including systematics) not being publicly available. We plan to apply our new framework, including up-to-date peculiar velocity corrections, to the Foundation sample (Foley et al., 2018; Jones et al., 2019), together with the recent data release of the Dark Energy Survey (DES Collaboration et al., 2018). In the near future, we expect to be able to obtain even tighter constraints on possible anisotropy in the expansion from upcoming, larger SNIa surveys like the one that will be delivered by the Vera Rubin Observatory Legacy Survey of Space and Time (LSST) (LSST Science Collaboration et al., 2009). The data products generated from our analysis in this chapter is available from <https://zenodo.org/record/5854639>.

Chapter 6

A Cosmological Model for Anisotropic Expansion

Most studies looking into observational tests for isotropy do so with a parameterized model that typically introduces a preferred direction in the expansion via a dipole term in either the luminosity distance, the distance modulus or the deceleration parameter as was done in Chapter 5. One of the key questions from a physical point of view, however, is how to link any potential deviation from isotropy at late times with the initially isotropic universe suggested by the CMB. An attempt was made to do this also in Chapter 5 by restricting the modelled dipole to a local scale. However, that chapter sought to do this under a standard Λ CDM context by simply including a dipole term in the distance modulus and deceleration parameter without modifying the underlying Friedmann-Lemaître-Robertson-Walker (FLRW) metric which assumed isotropy. Here we instead take a more phenomenologically consistent route here by considering a rescaling of the expansion in the FLRW metric that is time-dependent and then deriving the corresponding luminosity distance modulus formulas that would arise. The simplest way to do that is to suppose a regime change that occurs at some time before the current era, when both homogeneity and isotropy are broken. One possible method of doing this is to associate this regime change with the properties of dark energy, which might be something more complicated than a cosmological constant, and hence might be inhomogeneous at later times leading to an

anisotropic expansion. Therefore, I shall consider here a simple model, which can be thought of as an extension of the Λ CDM model, where dark energy undergoes a phase transition around the time of dark matter – dark energy equality, that is at approximately $z \sim 0.3$, which leads to isotropy-breaking and hence to a deviation from the Cosmological Principle at late times. The aim of this chapter is to compare such a model with the Λ CDM isotropic model, in order to constrain any deviations from isotropy and compare this result with the analysis presented in the previous chapter.

The structure of this chapter is the following. In Section 6.1 I present a simple model where the universe undergoes a phase transition around the time of dark matter-dark energy equality, leading to an anisotropic expansion rate, and compute the luminosity distance and distance modulus in this model. In Section 6.2 we use the BAHAMAS Bayesian Hierarchical Model and JLA SNIa data to constrain the parameters of our anisotropic model, and carry out Bayesian model comparison with the Λ CDM model. Finally, I present my conclusions in Section 6.3.

Throughout the chapter we use the standard convention for the signature of the metric $(-1, 1, 1, 1)$ and work in units in which the speed of light is unity: $c = 1$.

6.1 A Simple Model for Anisotropic Expansion

With the aim of reconciling the strong constraints on anisotropies at early times from the CMB while allowing for possible departures at late times, I conjecture here that dark energy could source an anisotropic expansion. For the sake of simplicity, I consider a simple, one-parameter model, where the universe expands at a different rate along one direction. I start by introducing a Cartesian system of coordinates centred on the Sun, with the three axes labeled as u, v, w as opposed to the standard x, y and z labels, in order to avoid confusion with redshift, which will be denoted by z as usual. For now, we shall assume that the u, v directions expand isotropically, while the w direction expands at a different rate (later, we generalize this to an arbitrary direction in space). We work in a spatially flat FLRW metric, and introduce a time-dependent

function $\epsilon(t)$ that rescales the expansion of the $u - v$ plane::

$$\begin{aligned}
 ds^2 &= -dt^2 + a^2(t)[\epsilon^2(t)(du^2 + dv^2) + dw^2] \\
 &= a^2(\eta)[-d\eta^2 + \epsilon^2(\eta)(du^2 + dv^2) + dw^2] \\
 &= a^2(\eta)d\tilde{s}^2,
 \end{aligned} \tag{6.1}$$

where

$$\begin{aligned}
 d\tilde{s}^2 &= -d\eta^2 + \epsilon^2(\eta)(du^2 + dv^2) + dw^2 \\
 &= \epsilon^2(\tau)[-d\tau^2 + du^2 + dv^2] + dw^2 \\
 &= \epsilon^2(\tau)d\hat{s}^2 + dw^2,
 \end{aligned} \tag{6.2}$$

and

$$\begin{aligned}
 d\hat{s}^2 &= -d\tau^2 + du^2 + dv^2 \\
 &= -d\tau^2 + dr^2.
 \end{aligned} \tag{6.3}$$

Thus we have three metrics ds^2 , $d\tilde{s}^2$ and $d\hat{s}^2$, related by conformal transformations. Let us recall now that under a conformal transformation, $ds^2 = f^2(x)ds'^2$, the luminosity distance transforms as ([Ivanov et al., 2018](#))

$$d_L = \frac{f_o^2}{f_s} d'_L, \tag{6.4}$$

where f_o is the value of the conformal factor at the observer, and f_s is the value of the conformal factor at the source. Also, the luminosity distance in $d+1$ dimensions is related to the luminosity distance in d dimensions by

$$d_L^{(d+1)} = \sqrt{d_L^{(d)2} + w^2}. \tag{6.5}$$

Now, using Eq. (6.4) and Eq. (6.5), we can derive the following expressions for the luminosity distance in each of the three metrics:

$$\hat{d}_L = r, \tag{6.6}$$

$$\tilde{d}_L = \sqrt{\frac{\epsilon_o^4}{\epsilon_s^2} r^2 + w^2}, \quad (6.7)$$

$$d_L = \frac{a_o^2}{a_s} \sqrt{\frac{\epsilon_o^4}{\epsilon_s^2} r^2 + w^2}, \quad (6.8)$$

where r is the radial distance in the $u - v$ -plane with $r^2 = u^2 + v^2$. Going now in spherical coordinates, we write $w = R \cos \theta$ and $r = R \sin \theta$, where R is radial comoving distance and θ measures the angle from the North Galactic pole (the w axis), Eq. (6.8) simplifies to

$$\begin{aligned} d_L(R, \theta) &= \frac{a_o^2}{a_s} R \sqrt{\frac{\epsilon_o^4}{\epsilon_s^2} \sin^2 \theta + \cos^2 \theta} \\ &= d_{L,FRW} \sqrt{\frac{\epsilon_o^4}{\epsilon_s^2} \sin^2 \theta + \cos^2 \theta}, \end{aligned} \quad (6.9)$$

where $d_{L,FRW} = \frac{a_o^2}{a_s} R$ is the luminosity distance in a FLRW spacetime.

Eq. (6.9) assumed the anisotropy points along the w axis, which we can say is in the coordinate frame S . To generalize it to the case where the anisotropic direction is $\mathbf{n}_{an} = (\theta_a, \phi_a)$ in spherical polars where θ_a and ϕ_a are measured in S , we can consider a rotation about the v axis of angle θ_a with rotation matrix $R_v(\theta_a)$ followed by a rotation about the w axis of angle ϕ_a with rotation matrix $R_w(\phi_a)$ to get the composite rotation

$$T(\theta_a, \phi_a) = R_w(\phi_a) R_v(\theta_a) = \begin{bmatrix} \cos \theta_a \cos \phi_a & -\sin \phi_a & \sin \theta_a \cos \phi_a \\ \cos \theta_a \sin \phi_a & \cos \phi_a & \sin \theta_a \sin \phi_a \\ -\sin \theta_a & 0 & \cos \theta_a \end{bmatrix}. \quad (6.10)$$

Under this rotation, the original coordinate frame, S , can then be transformed to another coordinate frame S' where the symmetry of Eq. (6.9) holds. In doing so we can obtain the following expression for the luminosity distance to an object in direction $\mathbf{n}_{obj} = \{\theta, \phi\}$ by first

defining the coordinates u', v', w' in this new S' frame as

$$\begin{bmatrix} u' \\ v' \\ w' \end{bmatrix} = T(\theta_a, \phi_a) \begin{bmatrix} R \sin \theta \cos \phi \\ R \sin \theta \sin \phi \\ R \cos \phi \end{bmatrix}. \quad (6.11)$$

We can therefore use Eq. (6.8) but with the substitution $r \rightarrow r'$ and $w \rightarrow w'$ where $r'^2 = u'^2 + v'^2$. Following the derivation as normal then, the luminosity distance becomes

$$\begin{aligned} d_L(\mathbf{n}_{\text{obj}}, R|\mathbf{n}_{\text{an}}) = d_{L,FRW} \left[\frac{\epsilon_o^4}{\epsilon_s^2} (\cos^2 \theta \sin^2 \theta_a + 2 \cos \theta_a \cos \theta \cos \phi \sin \theta_a \sin \theta \right. \\ \left. + \sin^2 \theta \cos^2 \theta_a \cos^2 \phi + \sin^2 \theta \sin^2 \phi) + \cos^2 \theta_a \cos^2 \theta \right. \\ \left. - 2 \sin \theta_a \cos \theta_a \sin \theta \cos \theta \cos \phi + \sin^2 \theta_a \sin^2 \theta \cos^2 \phi \right]^{1/2}. \end{aligned} \quad (6.12)$$

Notice that the luminosity distance depends only on θ_a , not on ϕ_a . This is expected since isotropy is broken along a single axis in the new frame S' and the rotational symmetry we see in Eq. (6.9) along the axis being rescaled is still preserved. Furthermore, a property of rotation matrices is that the distance to a point from the origin obtained via the pythagorean sum of the coordinates is preserved so we can ignore having to transform back to frame S as the physical quantity we are interested in, the luminosity distance we will later use for parameter inference, remains the same in either frame. Therefore, the anisotropy direction can be fully specified by two parameters, θ_a and ϵ_0 and we ignore entirely any dependence ϕ_a . Galactic coordinates are used in this this work as they are a spherical coordinate system and supernovae objects often have their positions given by galactic coordinates. However, if we wish to use galactic coordinates such that $\mathbf{n}_{\text{obj}} = (b_{sn}, l_{sn})$ and $\mathbf{n}_{\text{an}} = (b_a, l_a)$, we must note galactic coordinates measure the polar angle from the plane of the galaxy as so that the range of b is $-\pi/2 < b < \pi/2$, compared to θ_a and θ that fall in the range $0 < \theta_a, \theta < \pi$. To apply Eq. (6.12) we must shift the galactic latitude coordinates such that it is in the correct range using the relation $\theta = \pi/2 - b_{sn}$ and $\theta_a = \pi/2 - b_a$ (see Fig. 6.1 for details). The azimuth angles in standard spherical coordinates are measured the same way as galactic longitude and therefore

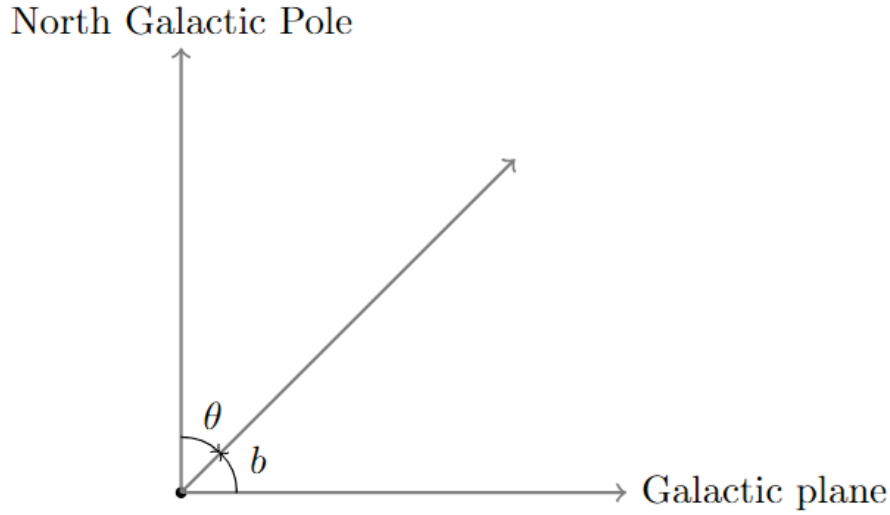


Figure 6.1: Diagram showing the difference between how b is measured in galactic coordinates compared to θ in spherical coordinates.

$\phi = l_{sn}$. The luminosity distance can then be used to calculate a distance modulus as before in previous chapters.

We can now design the function $\epsilon(t)$ so as to capture the behaviour described before: at early times the universe is homogeneous and isotropic, and therefore $\epsilon \rightarrow 1$, while at late times the isotropy may be broken and ϵ settles down to some value different from unity. Since we conjecture this isotropy-breaking behaviour to be associated with dark energy, we model the transition to occur roughly at the time of dark matter–dark energy equality, t_{eq} , and so to have a typical time-scale equal to the time scale for the transition from matter domination to dark energy domination, T_{eq} . A simple form of $\epsilon(t)$ implementing such a behaviour is

$$\epsilon(t) = -\left(\frac{1}{2} - \frac{\epsilon_0}{2}\right) \tanh\left(\frac{t - t_{eq}}{T_{eq}}\right) + \frac{1}{2} + \frac{\epsilon_0}{2}, \quad (6.13)$$

where $0 \leq \epsilon_0 \leq 1$ is a free parameter determining the amount of isotropy breaking. Fig. 6.2 shows the evolution of ϵ as a function of time for $\epsilon_0 = 0.60$, $\epsilon_0 = 0.85$, and $\epsilon_0 = 0.95$. The constant ϵ_0 represents the maximal amount of isotropy breaking – it is equal to the value of ϵ at future timelike infinity. The constant t_{eq} is specified by the physics at matter–dark energy equality, as we will see later. The bottom plot of Fig. 6.2 also shows the evolution of the ratio $\epsilon_o^4/\epsilon^2(t)$ that is used to compute the luminosity distance in Eq. (6.9) where ϵ_o is evaluated at

$t = 0$ for an observer located at our current location in time.

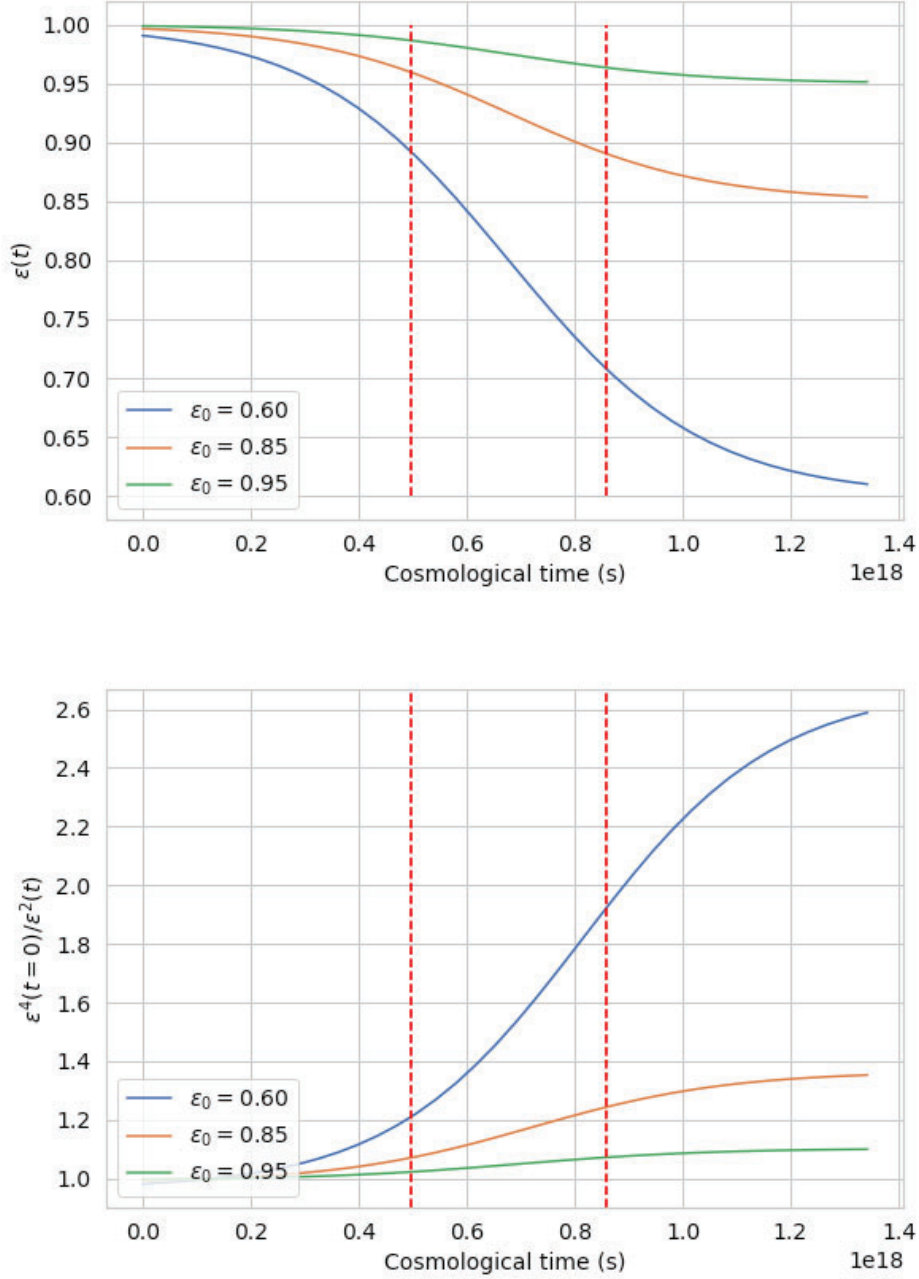


Figure 6.2: The function $\epsilon(t)$ and the ratio $\epsilon^4(t=0)/\epsilon^2(t)$ for $\epsilon_0 = 0.60, 0.85, 0.95$. The time interval marked by the red dashed lines corresponds to T_{eq} .

In order to rewrite Eq. (6.9) and Eq. (6.12) as functions solely of the redshift z and the polar angles θ and ϕ , we need to express Eq. (6.13) as a function of redshift, z . For that we need to solve for the evolution of the scale factor $a(t)$ during the time when both matter and dark energy are important, i.e. we need to solve the Friedmann equations with both matter and

dark energy sources (we set $k = 0$)¹. It should be noted, the set up here is similar to the Bianchi I class of anisotropic models which are allowed to vary at different rates in different directions. However, we are unable to use ‘off the shelf’ pre-derived cosmologies associated with this, as our cosmology has a time varying component that controls when the anisotropy appears (around dark matter-dark energy equality). This necessitates the need to derive our cosmological equations again using the derivation as follows:

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho, \quad (6.14)$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3P), \quad (6.15)$$

where $\rho = \rho_m + \rho_\Lambda$ with

$$\rho_m = \frac{1}{2}\rho_{eq}\left(\frac{a_{eq}}{a}\right)^3, \quad \rho_\Lambda = \frac{1}{2}\rho_{eq}, \quad (6.16)$$

where ρ_{eq} is the total density at the time of the equality and a_{eq} is the scale factor at equality. While dark energy is not really a cosmological constant in the anisotropic model, it will be approximately so, and therefore it is justified to treat it as a constant as a first-order approximation. Since our goal is to solve for $t(z)$ in order to substitute it in Eq. (6.13), any deviation from FLRW and a cosmological constant would appear as second-order in ϵ . Thus we are allowed to estimate $t(z)$ by assuming that the metric is FLRW. Using Eq. (6.16), the first Friedmann equation (the constraint equation) becomes

$$\dot{a}^2 = \frac{4\pi G}{3}\rho_{eq}a_{eq}^3\frac{1}{a} + \frac{4\pi G}{3}\rho_{eq}a^2, \quad (6.17)$$

while the second Friedmann equation (the dynamical equation) is not independent anymore - it can be derived from the first by differentiating with respect to t .

¹If we wanted to be pedantic, we would have to solve Einstein’s Field equation with the metric ansatz Eq. (6.1) and use the resulting equations instead of the Friedmann equations. However, this is unnecessary since the Friedmann equations would still hold approximately if we assume that the effect of ϵ is very small, which will certainly be true at late times.

The solution of Eq. (6.17) with the initial condition $a(t=0) = 0$ is

$$a = a_{eq} \left[\sinh \left(\frac{3}{2} \sqrt{\frac{4\pi G}{3} \rho_{eq} t} \right) \right]^{2/3}, \quad (6.18)$$

which can be easily verified by differentiating with respect to time. Inverting that, we get

$$\begin{aligned} t(a) &= \frac{2}{3} \left(\frac{4\pi G}{3} \rho_{eq} \right)^{-1/2} \operatorname{arcsinh} \left[\left(\frac{a}{a_{eq}} \right)^{3/2} \right] \\ &= \frac{2}{3} \Omega_{\Lambda,0}^{-1/2} H_0^{-1} \ln \left[\left(\frac{a}{a_{eq}} \right)^{3/2} + \sqrt{1 + \left(\frac{a}{a_{eq}} \right)^3} \right], \end{aligned} \quad (6.19)$$

where to get to the second line we used that $\rho_{\Lambda,0} = \rho_{\Lambda,eq} = \frac{1}{2}\rho_{eq}$ and that $\operatorname{arcsinh}(x) = \ln(x + \sqrt{x^2 + 1})$. Using $a = (1+z)^{-1}$ (we set $a_0 = 1$) this can be rewritten as

$$\begin{aligned} t(z) &= \frac{2}{3} \left(\frac{4\pi G}{3} \rho_{eq} \right)^{-1/2} \operatorname{arcsinh} \left[\left(\frac{1+z_{eq}}{1+z} \right)^{3/2} \right] \\ &= \frac{2}{3} \Omega_{\Lambda,0}^{-1/2} H_0^{-1} \ln \left[\left(\frac{1+z_{eq}}{1+z} \right)^{3/2} + \sqrt{1 + \left(\frac{1+z_{eq}}{1+z} \right)^3} \right]. \end{aligned} \quad (6.20)$$

The time of equality is given by

$$t_{eq} = \frac{2}{3} \Omega_{\Lambda,0}^{-1/2} H_0^{-1} \operatorname{arcsinh}[1] \approx 7 \times 10^{17} s. \quad (6.21)$$

In addition, we can easily determine the redshift at equality z_{eq} and the scale factor at equality a_{eq} , finding:

$$a_{eq} = (1+z_{eq})^{-1} = \left(\frac{\Omega_{m,0}}{\Omega_{\Lambda,0}} \right)^{1/3}. \quad (6.22)$$

The time scale for the dark matter–dark energy transition T_{eq} is given by the inverse of the Hubble scale at the time of equality

$$T_{eq} = \left(\frac{a}{\dot{a}} \right)_{eq} = \frac{1}{\sqrt{2}} \Omega_{\Lambda,0}^{-1/2} H_0^{-1} \approx 3.8 \times 10^{17} s \quad (6.23)$$

where the above values have been calculated for $\Omega_M = 0.3$ and $\Omega_\Lambda = 0.7$ to indicate their representative values. By substituting $t(z)$ from Eq. (6.20) into $\epsilon(t)$ from Eq. (6.13) we obtain

$\epsilon(z)$, shown in Fig. 6.3. For larger redshifts (earlier times), $\epsilon(z)$ gets closer to the isotropic case of $\epsilon_0 = 1$ as desired. We also in the same figure plot the evolution of the ratio $\epsilon^4(z=0)/\epsilon^2(z)$ which used to compute the luminosity distance.

Substituting $\epsilon(z)$ in Eq. (6.9) leads to the luminosity distance function in the anisotropic model $d_{L,anis}(z, \theta)$ in the case where the w -direction is the direction of anisotropy.

$$d_{L,anis}(z, \mathbf{n}_{obj} | \mathbf{n}_{an}) = d_{L,FRW}(z) \sqrt{\frac{\epsilon(z=0)^4}{\epsilon^2(z)} \sin^2 \theta + \cos^2 \theta}. \quad (6.24)$$

For the case of an arbitrary anisotropic direction, one obtains the luminosity distance by substituting $\epsilon(z)$ in Eq. (6.12). Eq.(6.24) describes a ‘squashed universe’ where the luminosity distance is larger in the w direction compared to directions in the $u - v$ plane, where the deviation from isotropy is largest. Fig. 6.4 shows a plot of Eq. (6.24) in the case of a test SNIa object at $\theta = \pi/2$ for different values of ϵ_0 including for the Λ CDM model which we can see corresponds exactly to anisotropic model with $\epsilon_0 = 1$ as expected. The test object being at $\theta = \pi/2$, means it lies along the $u - v$ plane which is expanding at a slower rate than the w direction so for smaller ϵ_0 the luminosity distance is correspondingly smaller. In addition, in the bottom plot of Fig. 6.4 we fix the test SNIa at redshift $z = 0.5$ and then we compute the luminosity distance as the angle θ varies. As expected the the luminosity distance is minimised as θ is increased until it reaches a minimum at an angle of $\pi/2$.

6.2 Constraints from JLA SNIa data

6.2.1 Overview of Statistical Model

In this section, we take the model derived in the previous section and now we use Type Ia SNe to constrain anisotropies arising from Eq. (6.12) by performing a Bayesian parameter estimation of all the parameters simultaneously, including standardization coefficients, cosmological parameters and anisotropy parameters ϵ_0, b_a . We carry this parameter estimation out with the Bayesian Hierarchical Model BAHAMAS.

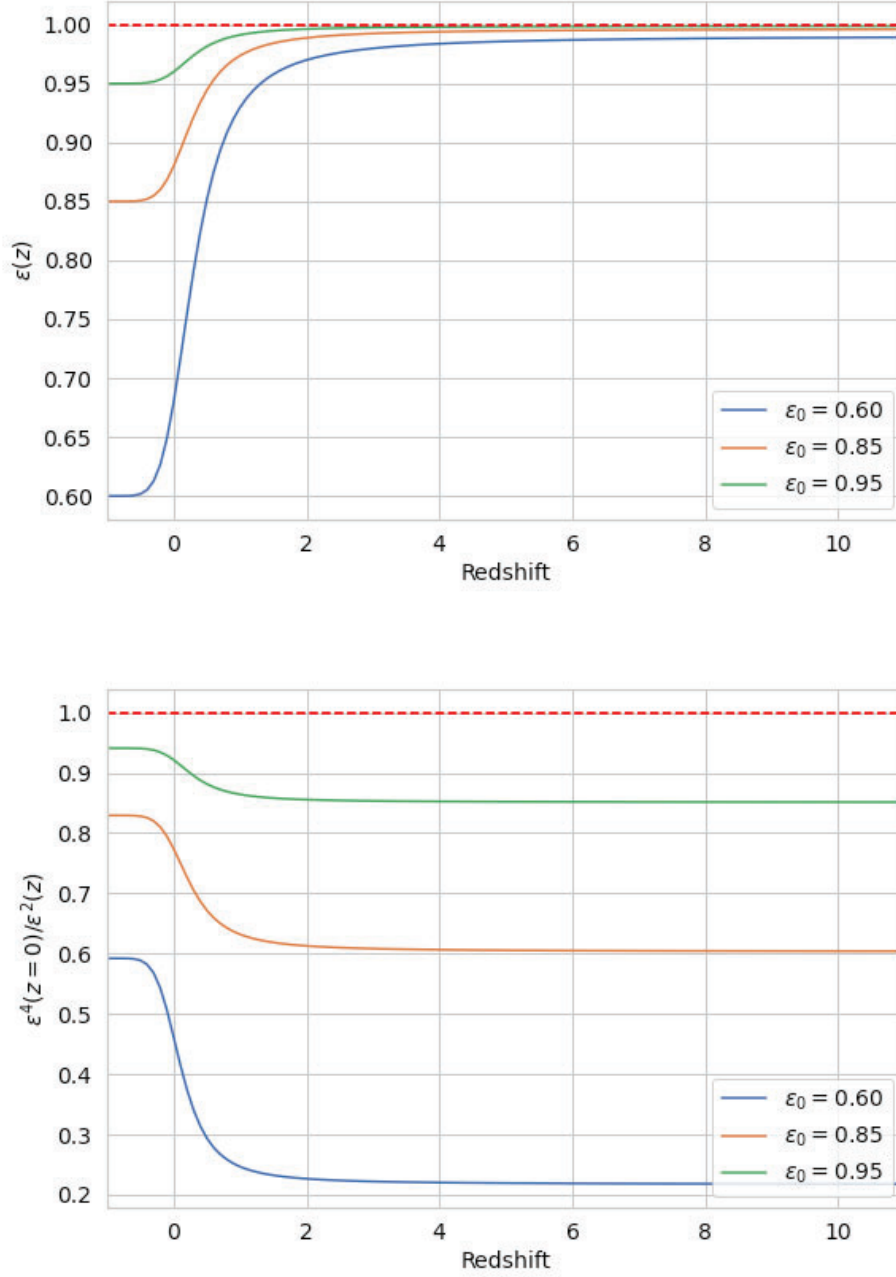


Figure 6.3: The function c and the ratio $\epsilon^4(z=0)/\epsilon^2(z)$ for $\epsilon_0 = 0.60, 0.85, 0.95$. The red dashed line corresponds to $\epsilon(z) = 1$ which is the case of perfectly isotropic cosmology as in Λ CDM.

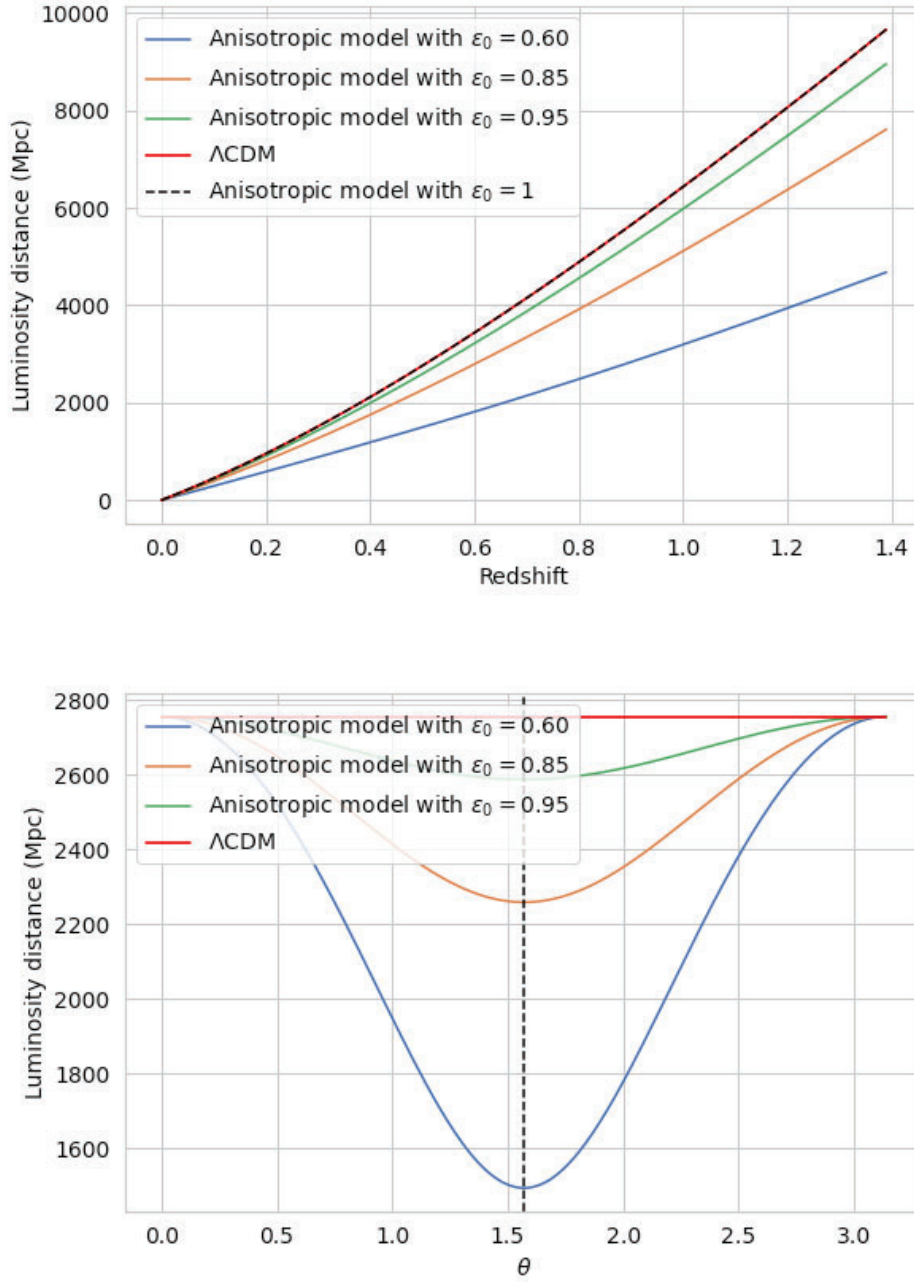


Figure 6.4: The top plot shows the luminosity distance as a function of redshift for the Λ CDM model (red), and for the anisotropic model for a test SNIa object at $\theta = \pi/2$ for $\epsilon_0 = 0.60$ (blue), $\epsilon_0 = 0.85$ (orange), $\epsilon_0 = 0.95$ (green), and $\epsilon_0 = 1$ (dashed black). We can see the anisotropic model overlaps exactly with the Λ CDM model when $\epsilon_0 = 1$ as expected. In addition, the bottom plot shows how the luminosity distance evolves with θ for a test SNIa object at redshift $z = 0.5$ when the anisotropy is aligned along the w axis. As expected, the greatest effect (where the luminosity distance is minimised) is felt at an angle $\pi/2$ to the w axis which is marked in a black dashed line.

For this work, we modify **BAHAMAS** so that the luminosity distance is now of the form Eq. (6.12).

Since our model considers a flat universe, for which $\Omega_\Lambda = 1 - \Omega_M$, the inference is effectively

only over Ω_M and we can remove Ω_Λ . Therefore, the final list of parameters **BAHAMAS** performs parameter inference over encompasses the population hyperparameters, the SALT2 regression coefficients, the cosmological parameter Ω_M , the anisotropy direction parameter b_a and the isotropy breaking parameter ϵ_0 :

$$\Theta = \{c_*, R_c, x_*, R_x, M_0^\epsilon, \sigma_{\text{res}}, \alpha, \beta, \Omega_M, b_a, \epsilon_0\}. \quad (6.25)$$

We follow the new treatment presented in the previous chapter and [Rahman et al. \(2021\)](#), where we updated the peculiar velocity treatment of the low- z SNIa in the JLA catalogue and introduced a new procedure to account for residual colour-based selection effects.

For our prior selection on the parameters of the **BAHAMAS** model, we use the same priors used by the previous chapter and [Shariff et al. \(2016\)](#). The only new parameters to apply priors on are on the galactic coordinate of the anisotropy direction, b_a , and the transition parameter ϵ_0 for which we use the uniform priors $b_a \sim \text{Uniform}[-\pi/2, \pi/2]$ and $\epsilon_0 \sim \text{Uniform}[0, 1]$.

6.2.2 Tests on Simulated Data

To test the capability of **BAHAMAS** to perform parameter inference in the scenario considered in this chapter, we forward simulate from the hierarchical model that **BAHAMAS** is based off using a Λ CDM universe with $\Omega_M = 0.3$ and $\Omega_\Lambda = 0.7$. The population parameters used for these simulations, such as for the SALT2 covariates are simulated using the population distributions in [4.30](#), [4.31](#) and [4.32](#) with parameters as listed in [Table 6.1](#) along with the priors used for them during the inference step. We consider two simulated data sets. One where there is anisotropy in the data with $\epsilon_0 = 0.85$ in the direction of $b_a = \pi/4$ and one where there is no anisotropy that corresponds to $\epsilon_0 = 1$. 740 objects were simulated in the same position in redshift and galactic coordinates as the JLA objects. For each data set we construct 10 realisations of the data.

In this chapter, the parameter inference is done using **PyMultiNest** ([Buchner et al., 2014](#)) as before. Using this we can compute the Bayesian Evidence which we will use for model selection

Table 6.1: Supernova population, cosmological and anisotropy model parameters adopted in this work, together with the prior choices and fiducial values for simulation studies. ‘SD’ stands for standard deviation.

SNIa population distributions and covariates	Symbol	Simulation Value	Prior Distribution
Mean absolute magnitude of SNIa	M_0	-19.3	$M_0 \sim \text{Normal}(-19.3, 2^2)$
Residual scatter of SNIa magnitude after corrections	σ_{res}	0.1	$\sigma_{\text{res}} \sim \text{InvGamma}(0.003, 0.003)$
Coefficient of stretch covariate	α	0.14	$\alpha \sim \text{Uniform}(0, 1)$
Coefficient of colour covariate	β	3.2	$\beta \sim \text{Uniform}(0, 4)$
Mean of stretch	x_*	0.0	$x^* \sim \text{Normal}(0, 10^2)$
Mean of colour	c_*	0.0	$c^* \sim \text{Normal}(0, 1^2)$
SD of stretch distribution	R_x	1.0	$R_x \sim \text{LogUniform}(-5, 2)$
SD of colour distribution	R_c	0.1	$R_c \sim \text{LogUniform}(-5, 2)$
ΛCDM cosmology parameters			
Matter energy density	Ω_M	0.3	$\Omega_M \sim \text{Uniform}(0, 2)$
Dark energy density	Ω_Λ	1 - Ω_M	-
Hubble-Lemaître constant [km/s/Mpc]	H_0	72	Fixed
Anisotropic expansion parameters			
Galactic longitude of dipole [rad]	l	$\pi/4$	$l \sim \text{Uniform}(0, \pi)$
Transition parameter	ϵ_0	$\{1, 0.85\}$	$\epsilon_0 \sim \text{Uniform}(0, 1)$

Table 6.2: Here we list the difference between the 1D marginal posterior mean and the true value used to simulated JLA-like data i.e., the bias, averaged over 10 data realizations each. $\Delta x \equiv |x_{\text{true}} - \bar{x}|$ for each parameter x . In parenthesis, the difference is expressed in units of the 1D posterior standard deviation. The top row shows the case where Λ CDM with no anisotropy ($\epsilon = 1$) is used for the simulation; the bottom row assumes Λ CDM with anisotropy ($\epsilon = 0.85$) for the simulation.

Model	$\Delta\Omega_M$	$\Delta b_a(\text{rad})$	$\Delta\epsilon_0$
(i) Λ CDM, $\epsilon_0 = 0.85$	0.01(0.28 σ)	0.04(0.65 σ)	0.02(0.14 σ)
(ii) Λ CDM, $\epsilon_0 = 1$	0.01(0.35 σ)	Unconstrained	< 0.951

and comparison in Section 6.2.3.

When performing the inference on the simulated data we recover the posteriors displayed in Fig. 6.5. The grey solid lines indicate the values we used in simulations. The black solid line is the average over the 10 data realisations of the recovered posterior. When calculating the mean of the recovered posteriors and their bias we find, when averaged over 10 realisations, the recovery has less than 1 sigma bias amongst all parameters as displayed in Table. 6.2. The posterior is sufficiently well-formed and this indicates that BAHAMAS is able to sufficiently infer the degree of anisotropy created by having $\epsilon_0 = 0.85$. However we wish to ascertain what the posterior looks like when no anisotropy is present. This is shown in the bottom of Fig. 6.5. The cosmological parameter Ω_M is still faithfully recovered with less than 1 sigma bias, and we see that b_a is unconstrained, as one would expect for an isotropic universe. We also see that the 1D marginal posterior distribution for ϵ_0 peaks at its true value of $\epsilon_0 = 1$, as expected. From here, we place a 95% credible interval at $\epsilon_0 \geq 0.961$ where 95% of the samples are greater than this value.

6.2.3 Parameters Constraints and Model Selection from the JLA Data

Finally, we are in a position to deploy BAHAMAS on the real JLA data after verifying its parameter recovery on simulated data. The posterior results for this are shown in 6.6. We can see similar results to the simulated results at the bottom of Fig. 6.5, the 1D marginal distribution for ϵ_0 peaks around 1 with 95% credible interval at $\epsilon_0 > 0.970$.

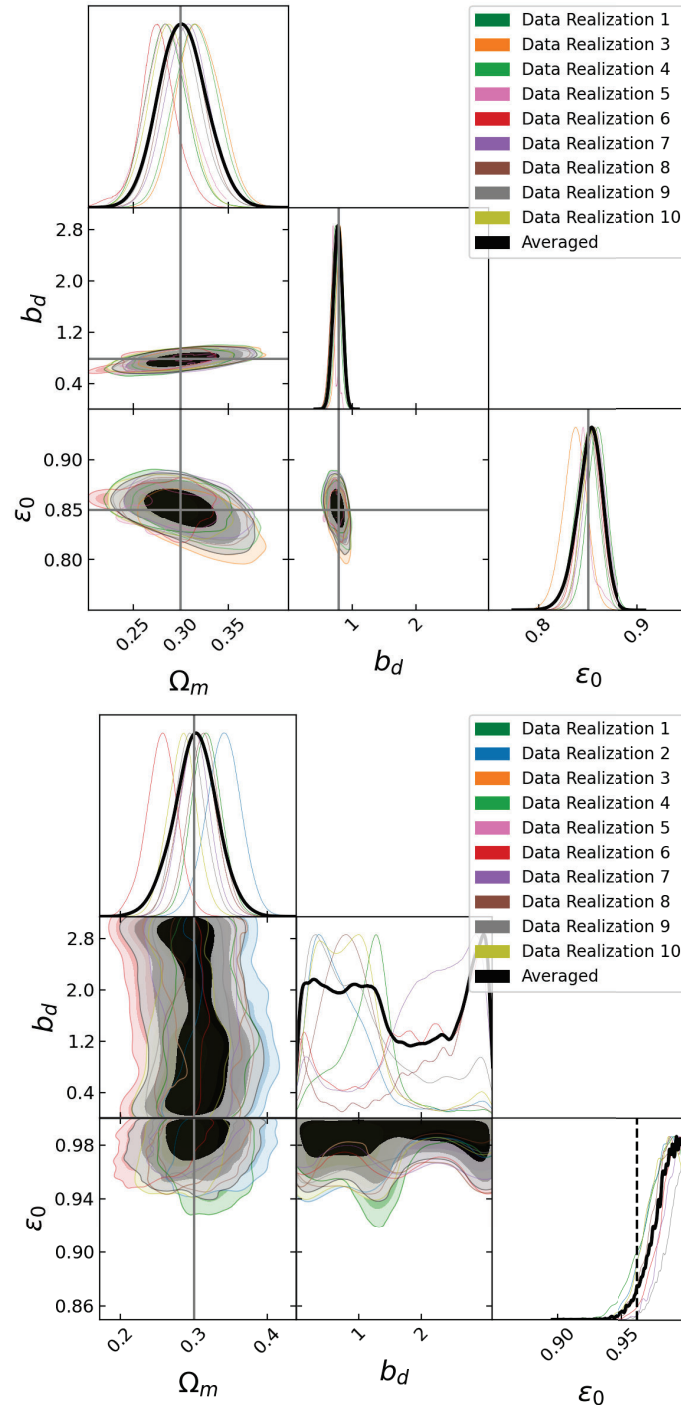


Figure 6.5: Results from simulated data. The top figure is the data with $\epsilon_0 = 0.85$ in the direction of $b_a = \pi/4$. The bottom figure is for data with no anisotropy and $\epsilon_0 = 1$. The grey solid lines indicate the values used in simulations. The black solid line is the average over the 10 data realisations of the recovered posterior. The bottom figure has a black dashed line in the lower right of the corner plot at $\epsilon_0 = 0.961$ which indicates the credible interval where 95% of the samples lay to the right.

The equivalent posterior plot for b_a can be seen to have a peak, but the peak of the posterior does not appear to be well-formed and is asymmetric. This, together with the lack of constraint

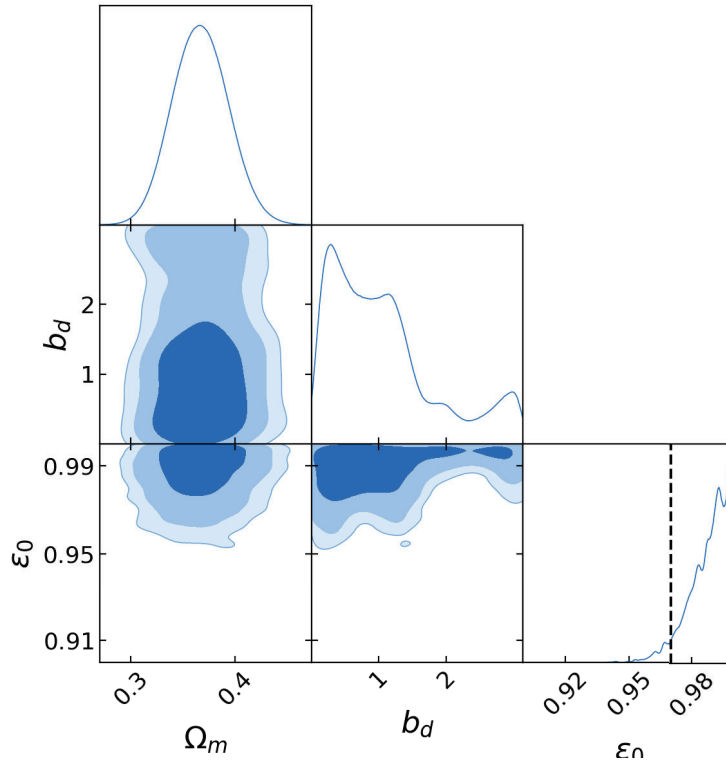


Figure 6.6: Results on JLA data for the anisotropic model with a 95% lower bound for the posterior plot on ϵ_0 in dashed black at $\epsilon_0 = 0.970$.

on ϵ_0 indicates a lack of evidence for significant detection in b_a as well.

We compare the isotropic Λ CDM expansion model to the alternative anisotropic model considered in this chapter via Bayesian model comparison, and calculate the difference in the log of the Bayesian evidence (i.e., the log of the Bayes factor) as

$$\Delta \ln(Z) = \ln B_A - \ln B_I, \quad (6.26)$$

where B_I is the Bayesian evidence for the isotropic model and B_A is the evidence for the anisotropic model, with priors as in Table 6.1. The quantity $\exp(\Delta \ln(Z)) = B_A/B_I$ gives the posterior odds between the isotropic and the anisotropic expansion models. In this particular case, $\Delta \ln(Z) = -4.03$ case which corresponds to odds of 56:1 for moderate to strong evidence in favour of the isotropic model. The evidence is estimated with `PyMultinest`, which was run with 400 live points and an evidence tolerance of 0.5.

6.3 Summary and Conclusions

In this chapter, I tested whether there is a late-time anisotropy in the present universe by using data from Supernovae Type Ia to serve as distance indicators. I have presented a simple model where isotropy is broken along a single direction and where the amount of isotropy-breaking is captured by two free parameters b_a and ϵ_0 . This model was compared with the standard Λ CDM cosmological model. I have shown that by applying the Bayesian Hierarchical Model **BAHAMAS** we are unable to constrain any anisotropy using the JLA dataset. When comparing the Bayesian Evidence, we find also that isotropic Λ CDM model is favoured over our anisotropic model at odds of 56:1, indicating moderate to strong evidence in favour of the isotropic model on the Jaynes scale. A 95% credible interval lower bound on ϵ_0 is placed at 0.970. While the size of the current datasets are not enough to disfavour our isotropic model, upcoming will be aimed at seeing if datasets such as LSST can provide more decisive evidence in favour of or against the cosmological model presented in this chapter. For this we will employ the expected utility formalism of [Trotta et al. \(2011\)](#).

Chapter 7

Selection Effects

In this chapter, I build on the BAHAMAS formalism of [March et al. \(2011\)](#); [Shariff et al. \(2016\)](#), to demonstrate how to include into the method, the treatment of general selection effects, i.e., the non-detection of data, that is present in SNIa datasets. This is done using a Bayesian prescription so that it can be combined with the cosmological inference step of BAHAMAS. The approach is then tested on simulated SNIa-like standard candle data sets. I first present a brief overview of selection effects in Supernovae data and current methods to handle this in [Section 7.1](#) and the Bayesian statistical approach of this chapter in [Section 7.2](#). I discuss the application of this Bayesian approach on SNIa-like simulated standard candles in [7.3](#) where I also compare our approach with the methods of [Rubin et al. \(2015\)](#) (from henceforth referred to as R15). Finally, I discuss my conclusions and future applications of this methodology in [Section 7.4](#).

7.1 Selection Effects Review

Selection effects in data is the effect of missing data which can happen in many scenarios that are related to the data collection procedure. The data can be considered ‘censored’ where the values of the missing data are unknown, for example, due to sample selection, but we do know that they are missing. The data can also be ‘truncated’ where both the values and the number

of missing data points are unknown and can occur, for example, due to a dependence on a dependent variable being measured. In the latter case, the data collection procedure is not random (see [Kelly \(2007\)](#) for more details).

Selection effects, and the associated bias they cause, can occur in many different fields such in Learning-To-Rank (LTR) systems used for recommendation and search engines ([Ovaisi et al., 2020](#)), accounting for survey bias in sociological and longitudinal studies ([Nohr & Liew, 2018](#)) and biological data sets that are often truncated ([Colchero & Clark, 2011](#)). Outside of SNIa cosmology but within astrophysics they are still heavily important in areas such as gravitational waves ([Messenger & Veitch, 2013](#)), trans-neptunian objects ([Loredo, 2004](#)) and more.

Malmquist bias, is the name given to a type of selection bias that occurs when systematically brighter objects are observed as distance increases ([Malmquist, 1922](#)). In SNIa cosmology, as SNIa are dependent on being bright enough to be observed, it is usual that Malmquist bias, is present in magnitude space for a SNIa sample from objects that are missed from not being bright enough. This means selection effects in SNIa are from data truncation and non-random. Malmquist bias appears in many areas of Astronomy and Astrophysics and has been investigated in various forms (see [Butkevich et al. \(2005\)](#) for a full historical account).

The cause of Malmquist bias in SNIa datasets can be multi-faceted. In the case of SNIa, aside from object brightness, they are subject to telescope sensitivity and environmental factors that affect observing conditions. In addition, it can be shown that whether a SNIa is missed or not can be caused by intrinsic variability in SNIa populations, with ‘bluer’ (smaller colour, c) objects being brighter and more easily detected and ‘slower’ (larger stretch, x_1) declining objects staying above the detection threshold for longer increasing the likelihood they are observed ([Phillips, 1993](#); [Hamuy et al., 1996](#); [Phillips et al., 1999](#)). Furthermore, box cuts are often made on SNIa data based on limiting the values of the SALT2 summary statistics to be within a given range and having a minimum number of observations within a set number of days before and after the observed peak of SNe. For example, cuts are often made such that $|c - 0.1| < 0.4$ and $|x_1| < 3$ (see [Kessler & Scolnic \(2017\)](#); [Scolnic & Kessler \(2016\)](#)). Figures 7.2 and 7.1 displays what effect the selection efficiency has on the distribution of the data based

on a simulation of the kind employed in this work (discussed further in Section 7.3).

Applying cosmological fits on data with Malmquist bias can significantly impact cosmological inference (see, for example, [Betoule et al. \(2014\)](#), where the bias correction on JLA data which is dominated by selection effects accounts for $\sim 16\%$ of the uncertainty on Ω_M). We also show in Section 7.2, if it is completely ignored, severe biases appear in the inference of cosmological parameters. One workaround to this is to use complete data in a lower redshift regime which does not suffer from magnitude limited selection effects, but then there will be a loss of high redshift information. It would be more appropriate to come up with a model for the missing data so that all the information in the observed data can be fully utilised. Before we discuss this works' method for doing this, I briefly summarise the procedure used in traditional analysis to correct for selection bias.

7.1.1 Previous Approaches to Handling Selection Effects

Many attempts have been made to attempt to correct for this selection bias with procedures that encompass, both, Bayesian and non-Bayesian methods. The most notable method being the BBC (BEAMS with Bias Corrections) ([Scolnic & Kessler, 2016](#); [Kessler & Scolnic, 2017](#)) a non-Bayesian method. Here we refer to BBC as non-Bayesian as the step we are primarily interested in, correcting for selection effects (as well as cosmological inference), is done in a non Bayesian manner. However, there is a Bayesian component within BBC where they use Bayesian Estimation for Multiple Species (BEAMS) ([Kunz et al., 2007](#)) to estimate probability of an object being of Type Ia which is used in their analysis. For the treatment of selection effects, BBC applies a correction method to the likelihood to 'correct back' data to account for the average of both magnitude and colour-induced selection effects. This method relies on simulating data to correct the observations, however, the correction is not applied to individual SNIa but rather apparent magnitude corrections are made to redshift 'bins' of data to account for the bias from missing data and then propagating the uncertainty to a systematics covariance matrix. In particular, BBC uses a five dimensional model bias model that depends on redshift, color, stretch and the nuisance parameters α and β . The main downsides associated with

simulation based methods however is that the simulations have to assume a cosmology which could bias inferences if our real cosmology does not closely match the simulated data. It can also be difficult propagate the uncertainty associated with these corrections and correlations these corrections have with other parameters unless it is known a-priori. Furthermore, if one now attempts to fit a cosmological model different from Λ CDM or w CDM as is typically assumed in this correction, this is difficult to do in a way that makes the resultant parameter inference valid. In the case of model comparison between exotic cosmological models such as those considered in Chapter 6 it is also not clear using data that applies these adhoc corrections can be sufficiently justified on a principled level. Therefore, having a procedure that infers these selection bias corrections during the cosmological fitting procedures is of high importance.

Some authors have attempted to include the modelling of selection effects within Bayesian methods as we do in this chapter. One of the notable examples of which is UNITY by [Rubin et al. \(2015\)](#). Unity extends the BHM of [March et al. \(2011\)](#) and also attempts to include many sources of systematics into the model, including selection effects. The method of including selection effects in UNITY makes approximations on the missing data distributions that can lead to severe biases in special cases (as demonstrated later in Section 7.2 and 7.3). STEVE by [Hinton et al. \(2018\)](#) is also a similar BHM that builds off the model of [March et al. \(2011\)](#) and similar to UNITY includes many sources of systematics and attempts to improve on the modelling of these systematics and SNIa populations with respect to UNITY, but still makes the same simplifying assumptions as UNITY in its treatment of Malmquist bias that can lead to severe biases in the final parameter inference. [March et al. \(2018\)](#) attempts to do away with the simplifying approximations of UNITY/STEVE, but uses a selection function based on a hard-cut threshold, which is unrealistic compared to the real-life smooth probabilistic transition around the threshold of observation that is expected. In the work presented in this chapter, our formalism improves upon each of the downsides presented in previous works.

7.2 A Bayesian Formalism for Selection Effects

Since this chapter is focused on presenting the conceptual methodology to implementing selection effects in the hierarchical model of BAHAMAS, in the following we simplify our statistical model and formalism by removing the colour and stretch corrections. In practice, this can be achieved by replacing the intrinsic scatter of the absolute magnitudes, M_i , with the residual intrinsic scatter after colour and stretch corrections for standardizable candles, M_i^ϵ which is given by Eq. (3.7). The inclusion of Phillips corrections in our formalism is straightforward, and is presented in Appendix B.4 for completeness. In the following, we perform numerical tests of our method on simulations from the true model, which does not include here the Phillips corrections. We demonstrate that even in this simplified scenario, our proposed method is superior to other approaches of implementing selection effects in a Bayesian hierarchical framework. We also ignore the systematic covariance matrix, and take each SNIa measurement to be independent from all the others. An implementation of our method on realistic SNANA simulations, including the full Phillips corrections and realistic selection effects, will be presented in an upcoming work.

March et al. (2011) have shown that when using spectroscopic redshifts, the latent (true) redshift value can be safely replaced by its observed value, \hat{z}_i , as the difference is negligible when compared with other sources of uncertainty in the analysis. However, for the purposes of accounting for selection effects keeping track of redshift information is crucial for a correct treatment, as we demonstrated below. Hence, in our simplified model we re-introduce explicitly redshift as one of the latent variables, z_i , with a corresponding measured value \hat{z}_i . For our simplified model without stretch and colour correction, but with the explicit addition of redshift, the data is assumed to be drawn from the distribution

$$[\hat{m}_{B,i}, \hat{z}_i] \sim \mathcal{N}([m_{B,i}, z_i], \hat{C}_i), \quad (7.1)$$

where $\hat{C}_i = \text{diag}(\hat{\sigma}_{m_{B,i}}^2, \hat{\sigma}_{z_i}^2)$. We denote the measured data as $\hat{\mathcal{D}}_{\text{obs}, i} := \{\hat{m}_{B,i}, \hat{z}_i, \hat{C}_i\}$.

The statistical model further has population SNIa parameters $\Phi = \{M_0^\epsilon, \sigma_{\text{int}}\}$, where M_0^ϵ is

the mean absolute magnitude of the SNIa population, which is here understood to be after Phillips corrections, and intrinsic residual scatter σ_{int} . The latent (i.e., before measurement noise) apparent magnitude m_i of SNIa i is drawn from

$$m_i|z_i, \mathcal{C}, \Phi \sim \mathcal{N}(\mu(\mathcal{C}, z_i) + M_0^\epsilon, \sigma_{\text{int}}^2). \quad (7.2)$$

In terms of the notation we use for this chapter, the quantity we are interested in for inference is the posterior pdf for the parameters, which, in the absence of selection effects is given by, via Bayes' theorem:

$$p(\mathcal{C}, \Phi | \hat{\mathcal{D}}_{\text{obs}}) = \frac{p(\hat{\mathcal{D}}_{\text{obs}} | \mathcal{C}, \Phi) p(\mathcal{C}, \Phi)}{p(\hat{\mathcal{D}}_{\text{obs}})} \quad (7.3)$$

where $p(\hat{\mathcal{D}}_{\text{obs}} | \mathcal{C}, \Phi)$ is the likelihood, $p(\mathcal{C}, \Phi)$ is the prior and $p(\hat{\mathcal{D}}_{\text{obs}})$ is an irrelevant (for our purposes) normalization constant. Here $\hat{\mathcal{D}}_{\text{obs}}$ is our data, \mathcal{C} , are the cosmological params of interest, and Φ the population parameters that govern how the data is generated. We now turn to the modeling of selection effects.

To account for selection effects, we firstly introduce the notion of *complete data* \mathcal{D} , namely, the data that would be observed if there were no selection effects. In the presence of selection effects, the *observed data* are a subset of the complete data: $\hat{\mathcal{D}}_{\text{obs}} \subset \mathcal{D}$. Selection effects are survey-dependent, so in the following our formalism is understood to apply to a given survey. The total data vector of observed SNIa's is $\hat{\mathcal{D}}_{\text{obs}} = \{\hat{\mathcal{D}}_{\text{obs},1}, \dots, \hat{\mathcal{D}}_{\text{obs},N_{\text{obs}}}\}$ (of dimension N_{obs}), while the complete data, including both observed and unobserved SNIa's, is $\mathcal{D} = \{\mathcal{D}_1, \dots, \mathcal{D}_n\}$, with unknown $n > N_{\text{obs}}$. For each SNIa i , we adopt an indicator variable describing whether the associated data point is observed:

$$I_i := \begin{cases} 1 & \text{if } \mathcal{D}_i \text{ is observed} \\ 0 & \text{if } \mathcal{D}_i \text{ is unobserved.} \end{cases}$$

The probability of it being observed, i.e., the selection function, can be parameterized in two different ways: the first is conditional on the *noisy value of the observables*, $\hat{\mathcal{D}}_i$, which is what

happens in practice when a decision is made whether or not to observe a given object (for example, when the signal-to-noise is too low because of the large apparent magnitude of the object). Notice that we distinguish here explicitly between $\hat{\mathcal{D}}_i$, representing the noisy values of the observables but without the condition that these have actually been observed (i.e., selected), and $\hat{\mathcal{D}}_{\text{obs},i}$, which denote observable, noisy values which have also been selected. Formally, $\hat{\mathcal{D}}_{\text{obs},i} = \hat{\mathcal{D}}_i | I_i = 1$. In the following, we call $\hat{\mathcal{D}}_i$ “the observable data” (i.e., including measurement noise) and $\hat{\mathcal{D}}_{\text{obs},i}$ “the observed” data (i.e., actually selected).

The selection function conditional on the observable data is written as

$$p(I_i = 1 | \hat{\mathcal{D}}_i, \Psi_{\text{noisy}}) \quad (7.4)$$

parameterized by a set of hyperparameters Ψ_{noisy} , which describe the selection function *as a function of the observable data*. For example, a simple selection function is a normal CDF as a function of the observed apparent magnitude, \hat{m}_i , with parameters $\Psi_{\text{noisy}} = \{m_{\text{cut}}^{\text{noisy}}, \sigma_{\text{cut}}^{\text{noisy}}\}$, so that

$$p(I_i = 1 | \hat{m}_i, \Psi_{\text{noisy}}) = \text{CDF}_{\mathcal{N}} \left(m_{\text{cut}}^{\text{noisy}} - \hat{m}_i, \sigma_{\text{cut}}^{\text{noisy}} \right). \quad (7.5)$$

This selection function tends to 1 for $\hat{m}_i \ll m_{\text{cut}}^{\text{noisy}}$ (i.e., SNIa much brighter than $m_{\text{cut}}^{\text{noisy}}$ are observed) and to 0 for $\hat{m}_i \gg m_{\text{cut}}^{\text{noisy}}$ (i.e., SNIa much dimmer than $m_{\text{cut}}^{\text{noisy}}$ are missed). The parameter $\sigma_{\text{cut}}^{\text{noisy}}$ controls the width of the transition. More sophisticated distributions can be obtained numerically from simulations, the subject of a future work.

The second way of parameterizing the selection function is conditional on the *latent values of the data* (i.e., before observation noise). This is convenient because in SNANA¹ simulations one has easy access to latent values for all SNIa’s, but only selected SNIa’s are given measurement noise. Hence it is difficult to derive from such simulations a selection function of the form given by Eq. (7.5). In this approach, the selection function is written as

$$p(I_i = 1 | \mathcal{D}_i, \Psi), \quad (7.6)$$

¹SNANA (SuperNova ANALysis) is a widely used tool for simulating SNe - <https://snana.uchicago.edu/>

which is now conditional on the latent variables, \mathcal{D}_i , and parameterized with a set of parameters Ψ whose numerical value differs from Ψ_{noisy} . For a normal CDF selection function, this leads to

$$p(I_i = 1|m_i, \Psi) = \text{CDF}_{\mathcal{N}}(m_{\text{cut}} - m_i, \sigma_{\text{cut}}). \quad (7.7)$$

The two expressions for the selection function, Eq. (7.5) and Eq. (7.7) are related by:

$$p(I_i = 1|m_i, \Psi) = \int d\hat{m}_i p(I_i = 1, \hat{m}_i|m_i, \Psi) \quad (7.8)$$

$$= \int d\hat{m}_i p(I_i = 1|\hat{m}_i, \Psi_{\text{noisy}}) p(\hat{m}_i|m_i). \quad (7.9)$$

This shows that the selection function over the latent variables, $p(I_i = 1|m_i, \Psi)$, is the convolution of the selection function over observable data, $p(I_i = 1|\hat{m}_i, \Psi_{\text{noisy}})$, and the measurement noise, $p(\hat{m}_i|m_i) = \mathcal{N}_{\hat{m}_i}(m_i, \hat{\sigma}_{m,i}^2)$. As such, the selection function over the latent variables has a less sharp transition between 1 and 0, as the scale of the transition region is of order $\sigma_{\text{cut}}^2 \approx \sigma_{\text{cut}}^{\text{noisy}^2} + \hat{\sigma}_{m,i}^2$ to account for the uncertainty about the value of the latent variables. It should be noted here that the value of $\hat{\sigma}_{m,i}^2$ in general differs from SNIa to SNIa meaning that when parameterized on the latent values, the selection function has a different shape depending on measurement noise and is not a universal function to be applied to each SNe. Therefore one needs to marginalise out the unknown value of $\hat{\sigma}_{m,i}^2$ which is dependent on the latent magnitude to perform the integration over the dummy SNe such that

$$p(I_i = 1|m_i, \Psi) = \int d\hat{m}_i \int d\hat{\sigma}_{m,i} p(I_i = 1|\hat{m}_i, \Psi_{\text{noisy}}) p(\hat{m}_i|m_i, \hat{\sigma}_{m,i}) p(\hat{\sigma}_{m,i}|m_i). \quad (7.10)$$

The functional form of the observational error model $p(\hat{\sigma}_{m,i}|m_i)$ is instrument dependent. As the observing instruments for various surveys are well studied, their observational error model is usually available a-priori.

It should be noted this does not mean one can freely swap m_i and \hat{m}_i in Eq. (7.8) and Eq. (7.9). While the two forms can be related, to invert the relation, Bayes' theorem is needed as the two selection functions on either the observed data or latent variables are different and not equivalent. If the variables are naively swapped rather than properly inverting with Bayes'

theorem, the width of the selection function will be incorrect. Related to this, one needs to be careful during forward simulations in latter sections that the selection function applied, $p(I_i = 1|\hat{m}_i, \Psi_{\text{noisy}})$ or $p(I_i = 1|m_i, \Psi)$, is consistent with your choice of parameterisation and applied on the right variables in the hierarchy. For this work, I choose to parameterise my selection function on the noisy observed data.

In terms of the BAHAMAS DAG in Figure 4.1, Ψ is associated with the top level of our DAG, together with the cosmological and population parameters from its dependence on the observed data and latent variables. The exact conditional dependence is decided by whether one parameterized on the observed data or latent variables.

Similar to equation Eq. (7.3), we now seek to find the posterior distribution for the parameters, conditional on the observed data, $\hat{\mathcal{D}}_{\text{obs}}$.

$$\begin{aligned} p(\mathcal{C}|\hat{\mathcal{D}}_{\text{obs}}) &= \int d\Phi d\Psi p(\mathcal{C}, \Phi, \Psi|\hat{\mathcal{D}}_{\text{obs}}) \\ &= \int d\Phi d\Psi \frac{p(\hat{\mathcal{D}}_{\text{obs}}|\mathcal{C}, \Psi, \Phi)}{p(\hat{\mathcal{D}}_{\text{obs}})} p(\mathcal{C})p(\Psi)p(\Phi) \end{aligned}$$

(assuming separable priors for \mathcal{C}, Φ, Ψ)

We now turn our attention on the observed data likelihood, $\mathcal{L}_{\text{obs}} = p(\hat{\mathcal{D}}_{\text{obs}}|\mathcal{C}, \Psi, \Phi)$. This can be obtained from the complete data likelihood, $p(\mathcal{D}|\mathcal{C}, \Psi, \Phi)$, via explicit marginalization over the missing data. The details of the calculation are presented in Appendix B.2, where our method is compared in detail with that in R15. More simply, the final result can be obtained via a straightforward application of Bayes' theorem as follows. Let $\hat{\mathcal{D}}_{\text{obs},i}$ be the observed data for a randomly selected SNIa from the complete data \mathcal{D} . In the absence of selection effects, the observed data would be a random subsample of size N_{obs} from the complete data, i.e.

$$\hat{\mathcal{D}}_{\text{obs},i} \sim p(\mathcal{D}|\mathcal{C}, \Phi), \quad i = 1, \dots, N_{\text{obs}}. \quad (7.11)$$

However, we only observe a SNIa with probability given by either Eq. (7.5) or Eq. (7.7). Since

$\hat{\mathcal{D}}_{\text{obs},i} = \hat{\mathcal{D}}_i | I_i = 1$, it follows from Bayes' theorem that

$$p(\hat{\mathcal{D}}_i | I_i = 1, \mathcal{C}, \Phi, \Psi) = \frac{p(I_i = 1 | \hat{\mathcal{D}}_i, \Psi_{\text{noisy}}) p(\hat{\mathcal{D}}_{\text{obs},i} | \mathcal{C}, \Phi)}{p(I_i = 1 | \mathcal{C}, \Phi, \Psi)}. \quad (7.12)$$

Further assuming that the SNIa observations are independent², we obtain

$$\begin{aligned} \mathcal{L}_{\text{obs}} &= p(\hat{\mathcal{D}}_{\text{obs}} | \mathcal{C}, \Phi, \Psi) \propto \\ &\prod_{i=1}^{N_{\text{obs}}} \frac{p(I_i = 1 | \hat{\mathcal{D}}_i, \Psi) p(\hat{\mathcal{D}}_{\text{obs},i} | \mathcal{C}, \Phi)}{p(I_i = 1 | \mathcal{C}, \Phi, \Psi)^{N_{\text{obs}}}}, \end{aligned} \quad (7.13)$$

where the numerator is a product of the selection model Eq. (7.5) and the standard likelihood ignoring selection effects, $p(\hat{\mathcal{D}}_{\text{obs},i} | \mathcal{C}, \Phi)$. Notice that, if the selection function parameters, Ψ , are known, then the first term in the numerator is a constant (independent of \mathcal{C}), and can thus be dropped. However, this is no longer true in when there is uncertainty in the selection function parameters – a case we will return to below.

The denominator of Eq. (7.13) is no longer ignorable, as it depends on the parameters being inferred and it represents the probability of making an observation. Its calculation is generally difficult, for it requires marginalization over all latent variables, including redshift (which does not appear in the denominator, an important point we will discuss more fully below). This marginalization can be expressed in two different ways, as either a function of the noisy or latent variables with

$$p(I_1 = 1 | \mathcal{C}, \Phi, \Psi) = \quad (7.14)$$

$$\int d\hat{\mathcal{D}}_1 p(I_1 = 1 | \hat{\mathcal{D}}_1, \Psi_{\text{noisy}}) p(\hat{\mathcal{D}}_1 | \mathcal{C}, \Phi) \quad (7.15)$$

$$\int d\hat{\mathcal{D}}_1 p(I_1 = 1 | \hat{\mathcal{D}}_1, \Psi_{\text{noisy}}) \int d\mathcal{D}_1 p(\hat{\mathcal{D}}_1 | \mathcal{D}_1) p(\mathcal{D}_1 | \mathcal{C}, \Phi) \quad (7.16)$$

$$\int d\mathcal{D}_1 p(I_1 = 1 | \mathcal{D}_1, \Psi) p(\mathcal{D}_1 | \mathcal{C}, \Phi), \quad (7.17)$$

where in the last equality we have used the relationship given by Eq. (7.8). If one chooses to

²This is not necessarily always true and testing the efficacy of our formalism on realistic data with potential correlations between SNIa from simulations by SNANA is the subject of future work

evaluate the integral Eq. (7.17), it is required to have a formulation of the selection function as the probability of observation *as a function of the latent variables*. This is usually accessible via numerical forward simulation as in simulated data (such as those from SNANA), we know what the latent variable values are. It should be noted here, if one chooses to parameterize with the latter expression which marginalises over the latent variables one would also require a likelihood over the latent variables as opposed to the noisy data.

R15 derive an observed data likelihood that in our notation reads

$$\mathcal{L}_{\text{R15}} \propto \prod_{i=1}^{n_{\text{obs}}} \frac{p(I_i = 1 | \hat{\mathcal{D}}_{\text{obs},i}, \Psi) p(\hat{\mathcal{D}}_{\text{obs},i} | \mathcal{C}, \Phi)}{p(I_i = 1 | \hat{z}_{\text{obs},i}, \hat{\sigma}_{z_i}^2, \mathcal{C}, \Phi, \Psi)} \quad (7.18)$$

which is very similar to our Eq. (7.13), except it conditions on the observed redshift values in the denominator, while our likelihood does not. A similar expression to Eq. (7.18) appears in [Hinton et al. \(2018\)](#). We now show below that it is incorrect to include this conditional if one wishes to calculate $p(\hat{\mathcal{D}}_{\text{obs}} | \mathcal{C}, \Phi, \Psi)$.

R15 present an expression for $p(\hat{\mathcal{D}}_{\text{obs}} | \mathcal{C}, \Phi, \Psi, n)$ that is identical to our Eq. (B.9), but without the marginalization over n . In order to then carry out the marginalization over n , R15 make the crucial assumption that each unobserved SNIa has a redshift that can be approximated by the redshift of an observed SNIa. This represents an a-priori assumption that the missing data and observed data follow the same distribution in redshift which is evidently not true (see the outcome of simulations in Fig. 7.2). To do this, they make the assumption that:

$$p(n_i) \propto \frac{1}{n_i},$$

where n_i is the number of unobserved SNIa at redshifts laying in a bin that can be approximated by $z_{\text{obs},i}$. Our own assumption, that $p(n) \propto 1/n$, is instead an uninformative prior on the *total number of unobserved SNe* in each survey, without making unwarranted assumptions about their redshift distribution.

We show in Appendix B.3 that the relation between our observed data likelihood and the

expression presented in R15, \mathcal{L}_{R15} , is given by

$$p(\hat{\mathcal{D}}_{\text{obs}}|\mathcal{C}, \Phi, \Psi) = \mathcal{L}_{\text{R15}}p(\hat{z}_{\text{obs}}|\mathcal{C}, \Phi, \Psi).$$

This shows that the likelihood in R15, \mathcal{L}_{R15} , is missing some information, namely the conditional distribution of observed redshifts, \hat{z}_{obs} , given the cosmological parameters, the SNIa population parameters and the selection function parameters, $p(\hat{z}_{\text{obs}}|\mathcal{C}, \Phi, \Psi)$. Since this is a function of the parameters of interest, it cannot be ignored. The reader is referred to Appendix B.2 and B.3 for further discussion on why this term cannot be ignored.

We demonstrate that our observed data likelihood correctly recovers unbiased parameter estimates (differently from the expression used in Rubin et al. (2015); Hinton et al. (2018)) in the next section.

7.3 Applications to the SNIa-like Standard Candles

Here, we demonstrate our method on a simplified cosmological example, employing SNIa-like standard candles rather than standardizable candles meaning they require no extra corrections for stretch and colour. Extension of this work to the realistic case where the SNIa are standardized via linear correlations with observed quantities is straightforward and will be addressed in a future, dedicated work. though the general methodology for doing so is outlined in Section B.4.

7.3.1 Known Selection Function and No Measurement Error

Let us consider first the case where the parameters of the selection function are known exactly³, and measurement error is negligible (we relax these assumptions later). We wish to compare our likelihood with R15's, as well as with the standard method for dealing with selection effects,

³One caveat of this method however is that it assumes the functional form of the selection function is correct. For example, at lower redshift, saturation issues can be present and some authors find it more appropriate to use a skew normal function instead of a normal CDF as in STEVE for these data points. When applied to realistic datasets, one should always check if the functional form of the selection function is appropriate.

namely, taking the likelihood that ignore selection effects, \mathcal{L}_{ign} , and apply it to data that have been “corrected back” as follows. The observed magnitudes, $m_{\text{obs},i}$ in a given redshift bin are shifted by the difference $\Delta(m_{\text{obs}})$ computed from a large simulated data set, between the means of the simulated, complete data in that redshift bin and the mean of the observed magnitudes in the bin. The naive likelihood, which ignores selection effects, follows the **BAHAMAS** methodology wherein the parameters not of interest are marginalized out.

The likelihood ignoring selection effects is given by

$$\mathcal{L}_{\text{ign}} \propto \prod_i p(m_i = m_{\text{obs},i} | z_i = z_{\text{obs},i}, \mathcal{C}, \Phi) \quad (7.19)$$

where

$$p(m_i = m_{\text{obs},i} | z_i = z_{\text{obs},i}, \mathcal{C}, \Phi) \propto \prod_i \mathcal{N}(m_{\text{obs},i} - [\mu(\mathcal{C}, z_{\text{obs},i}) + M_0], \sigma_{\text{int}}^2) \quad (7.20)$$

and the standard method applies this likelihood to the data after they have been “corrected back”:

$$\mathcal{L}_{\text{ah}} = \mathcal{L}_{\text{ign}}(m_{\text{obs},i} \rightarrow m_{\text{obs},i} + \Delta(m_{\text{obs},i})), \quad (7.21)$$

where the redshift bins used to determine Δ are logarithmically spaced with widths, in $\log(z)$ -space

$$\lambda_{\log(z)} = 0.1. \quad (7.22)$$

R15’s likelihood reads

$$\mathcal{L}_{\text{R15}} \propto \frac{\mathcal{L}_{\text{ign}}}{\prod_i p(I_i = 1 | z_i = z_{\text{obs},i}, \mathcal{C}, \Phi, \Psi)}, \quad (7.23)$$

where

$$\begin{aligned}
 p(I_i = 1 | z_i = z_{\text{obs},i}, \mathcal{C}, \Phi, \Psi) \\
 &= \text{CDF}_{\mathcal{N}}(m_{\text{cut}} - [\mu(\mathcal{C}, z_{\text{obs},i}) + M_0], \sigma) \\
 \sigma &:= \sqrt{(\sigma^{\text{int}})^2 + (\sigma^{\text{cut}})^2}.
 \end{aligned}$$

Our likelihood, derived in Appendix B.2, modifies the likelihood that ignores selection effects by re-weighting it with a \mathcal{C}, Φ (and Ψ)-dependent term, reading:

$$\mathcal{L}_{\text{obs}} \propto \frac{\mathcal{L}_{\text{ign}}}{p(I_1 = 1 | \mathcal{C}, \Phi, \Psi)^{n_{\text{obs}}}}, \quad (7.24)$$

where $p(I_1 = 1 | \mathcal{C}, \Phi, \Psi)$ is the probability of observing a SNIa in the survey *irrespective of its latent redshift or magnitude* given cosmological parameters \mathcal{C} , population parameters Φ and selection function parameters Ψ , and n_{obs} is the number of SNIae observed in the survey. In the above equation, we have dropped ignorables constants.

The denominator in Eq. (7.24) cannot be calculated analytically due to the non-uniform distribution of the independent variable z , which needs to be marginalized over. We take the z -distribution of the SNIae to be proportional to their volumetric rate:

$$p(z_i) \propto (1 + z_i)^\gamma \quad (7.25)$$

with $\gamma = 1.5$ from Dilday et al. (2008). The denominator of Eq. (7.24) can then be calculated by marginalizing out z :

$$\begin{aligned}
 p(I_1 = 1 | \mathcal{C}, \Phi, \Psi) \\
 &= \int dz_1 p(I_1 = 1 | z_1, \mathcal{C}, \Phi, \Psi) p(z_1),
 \end{aligned}$$

which can be done with a simple 1 dimensional trapezoidal integration.

The priors we adopt are as follows:

$$\Omega_M, \Omega_\Lambda \sim \mathcal{U}(0, 1) \quad (7.26)$$

$$h \sim \mathcal{N}(0.72, 0.1^2) \quad (7.27)$$

$$M_0 \sim \mathcal{N}(-19.3, 0.5^2) \quad (7.28)$$

$$\log \sigma^{\text{int}} \sim \mathcal{U}(-5, 0). \quad (7.29)$$

A realization of the data with

$$\begin{aligned} \Omega_M &= 0.3, & \Omega_\Lambda &= 0.7, & h &= 0.72, \\ M_0 &= -19.3, & \sigma^{\text{int}} &= 0.1, \\ m_{\text{cut}} &= 24, & \sigma^{\text{cut}} &= 0.5 \end{aligned} \quad (7.30)$$

is shown in the top panel of Fig. 7.1, where we have chosen the selection function parameters so as to give a number of observed SNIa (~ 1000) similar in size to current compilation of surveys (such as Pantheon and JLA). Typically, approximately 20% of SNIa are selected based on our selection function parameterization so to ensure approximately 1000 objects are detected we simulate 5000 objects for each realisation of data before applying a selection function on them to decide what is observed. For clarity of visualisation however, we have further thinned this by a factor of 5 so that only 1000 objects in total between observed and unobserved are displayed. Fig. 7.2 shows histograms of the complete and observed simulated data sets, as well as the residuals after the standard data correction procedure is carried out.

To perform our inference on these simulated datasets we use Multinest as before with 400 live points and an evidence tolerance of 0.5. In Fig. 7.3 we compare the posterior distributions when using our likelihood, R15's likelihood, as well as a likelihood ignoring selection effects but adopting corrected data using the standard method. The chains are 'averaged' over 100 data realizations by concatenating all 100 chains. The result is such that if each chain has n samples for a parameter w , then the final concatenated chain has $100 * n$ samples for the parameter w and is used to generate the posterior plots. As each realisation of the data set generates

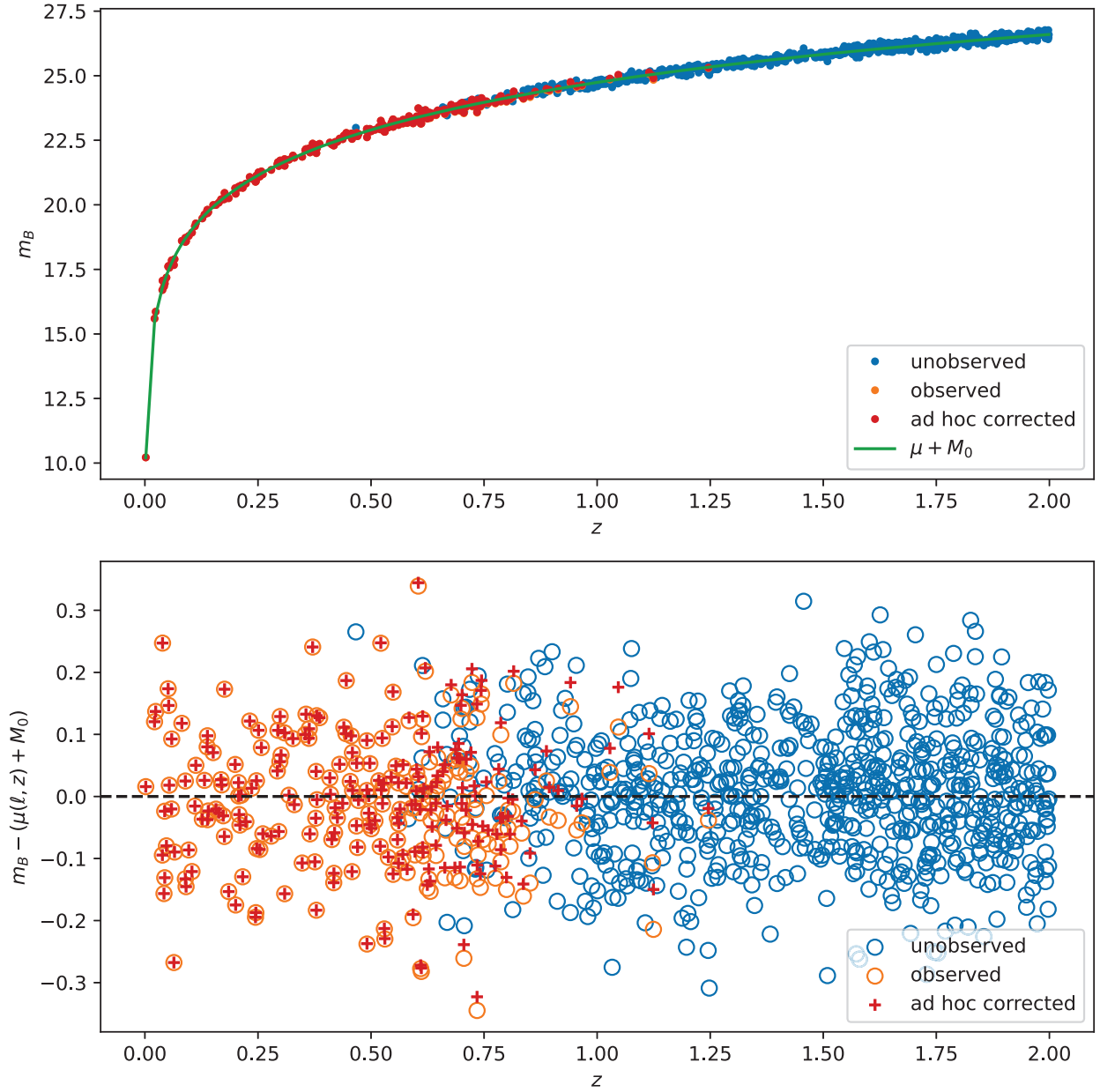


Figure 7.1: A Simulated data set of 5000 SNIa for simplified standard candles, without measurement errors. 219 observed SNIa are shown in orange and 781 unobserved SNIae are shown in blue after the data has been randomly thinned by a factor of 5 for visual clarity. An example visualisation before thinning is presented in Appendix B.1. The standard method corrects for selection effects by shifting the observed apparent magnitudes $m_{\text{obs},i}$ to the values shown in red crosses, then ignores selection effects in the likelihood. In the bottom chart the value of $\mu(\ell, z_i) + M_0$ is subtracted off for clarity. The top chart also has a line in green which represents the fiducial Λ CDM value expected as a function of redshift.

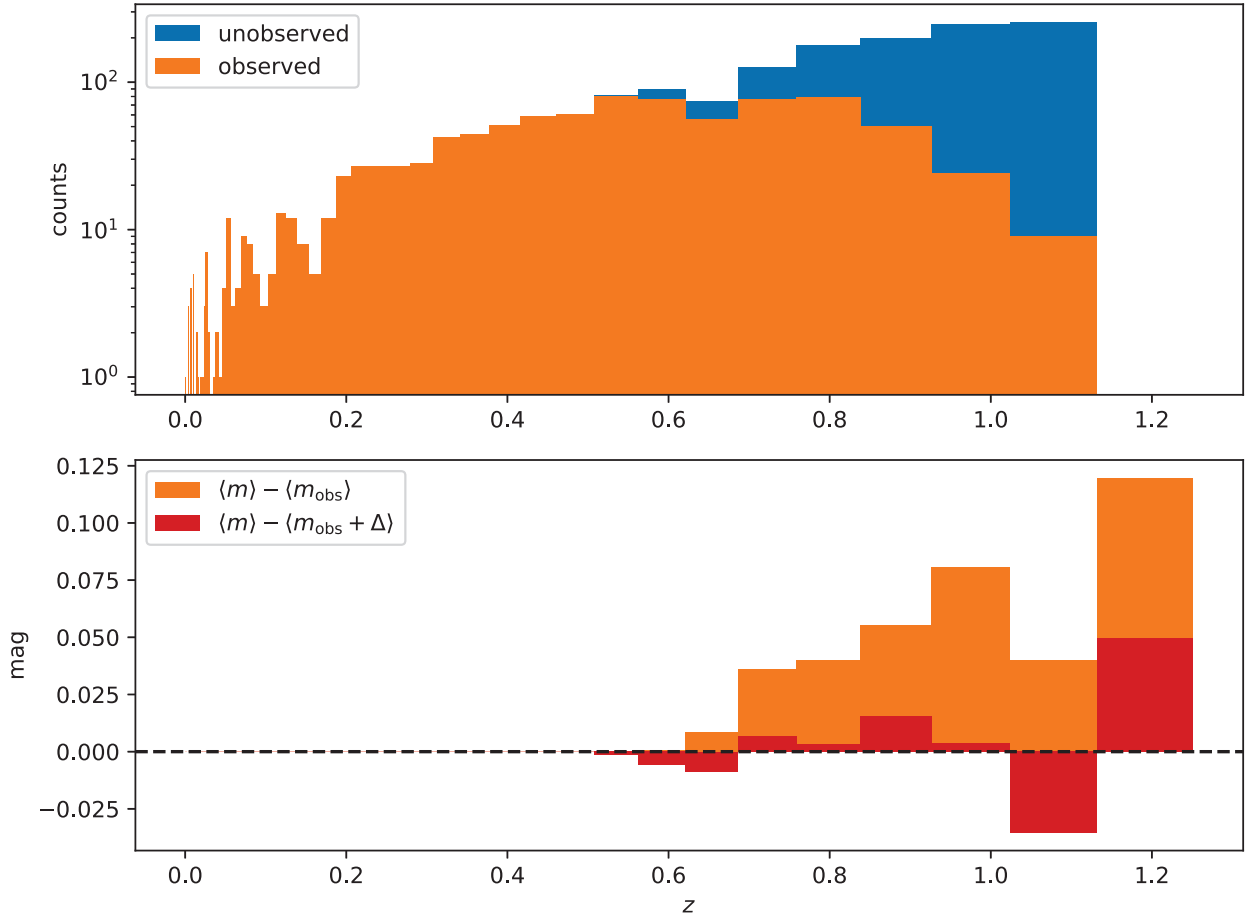


Figure 7.2: Top: Histogram of the complete (red) and observed (green) simulated SNIae. Middle and bottom: difference between binned means of the complete dataset’s apparent magnitudes, m , and the observed apparent magnitudes, m_{obs} , before and after selection effects correction. The bins have width Eq. (7.22) in $\log(z)$ -space with the right-most edge determined by $\max(m_{\text{sim,obs}})$.

a random bias in the parameter inference, plotting this concatenated chain averages this bias such that the posteriors are appropriately centered on their true value. Of course, we only have one data set in real life so this averaging of the realisation noise is not possible for the real universe, and mainly serves to diagnose the efficacy of our model in its ability to recover unbiased posteriors under simulation on average. For the case where the selection function is known and measurement error is negligible, the averaged posteriors for R15 and this works methods are very similar, with the data correction method's posterior being biased low in the cosmological parameters. It is interesting to note, this bias is in the opposite direction to the naive likelihood. This could be a feature of the fact that sometimes adhoc bias correction can cause a magnitude bias in the opposite direction to the pre-correction bias (see Fig. 7.2 bottom panel).

The above analysis represents an ideal scenario where the selection function is exactly known. Knowing the selection model parameters exactly however is not a realistic assumption. In reality, simulations of the observations (e.g. via the SNANA package) can only deliver an approximation of the selection effects model, which may be biased or otherwise inaccurate. Thus, we also wish to understand how inferences on the parameters of interest might be biased when assuming incorrect values for the selection effects model.

Testing this is easily done by substituting biased values of Ψ into equations Eq. (7.19)-(7.24). As a test we assume a value for m_{cut} that is biased high by 0.2 mag, a reasonable value in consideration of the intrinsic dispersion of SNIa magnitudes being of order 0.1 mag and the magnitude bias in some bins being as high as 0.1 mag owing to selection effects (see bottom panel of Fig. 7.7). The results wrongly assuming a known (but incorrect) value for m_{cut} are shown in Fig. 7.4. When using biased values of m_{cut} , our likelihood and the likelihood of R15 recover significantly biased Ω_M, Ω_Λ . It can be seen here also, using our likelihood tends to bias the posterior in the opposite direction compared to using the likelihood of R15. The ad hoc likelihood gives even more strongly biased reconstructions of the cosmological parameters.

With the consideration that potential biases can be present if the parameters of the selection function are fixed a-priori then the reasonable course of action is to include the selection function

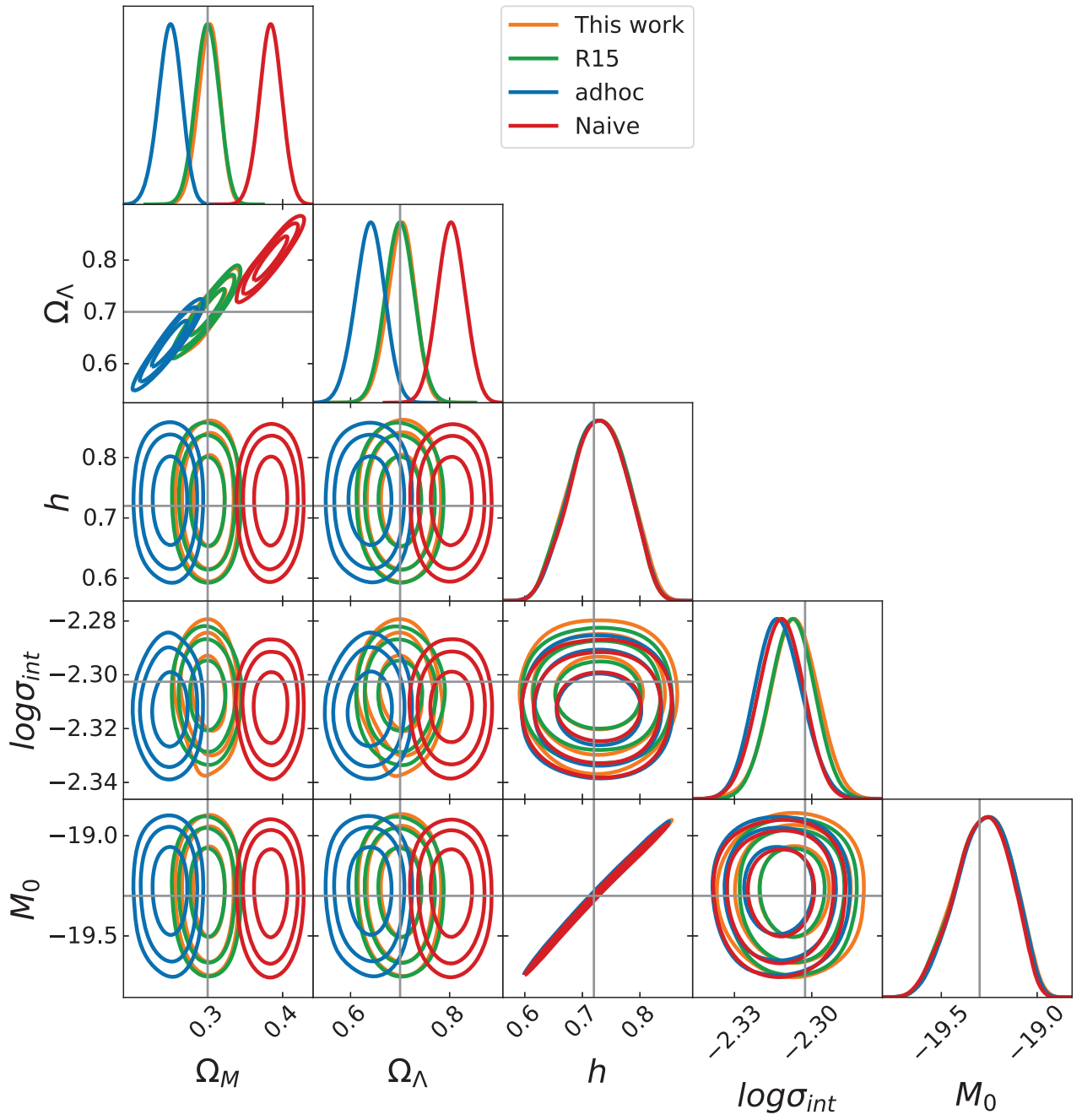


Figure 7.3: Mean over 100 reconstructions for the cosmological example, comparing the naive BAHAMAS likelihood, Rubin’s, and this work’s likelihood assuming no measurement error and a known selection function. The vertical and horizontal line show the true underlying parameter values.

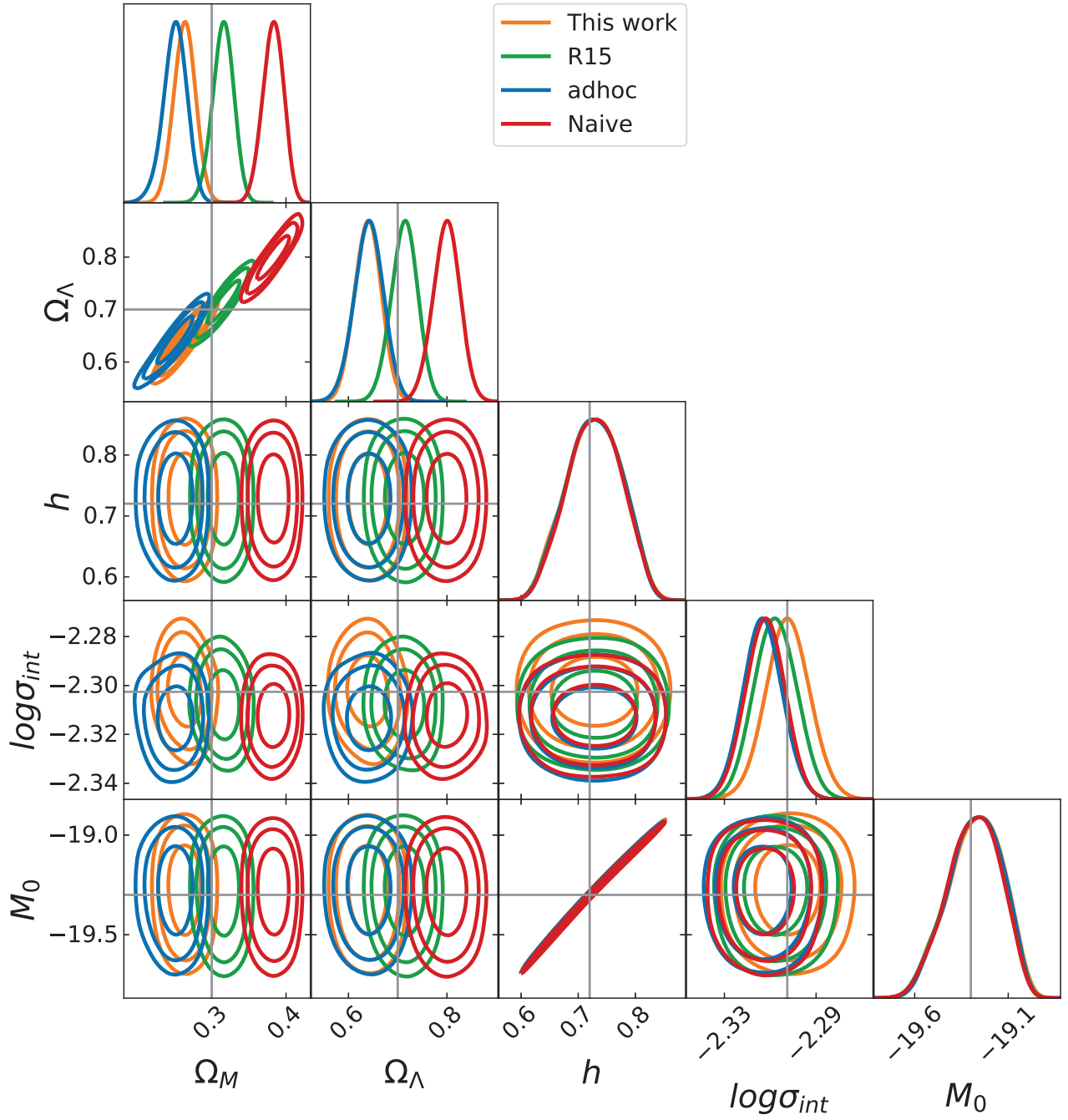


Figure 7.4: Biased parameter reconstruction for the cosmological example, when wrongly assuming a known (but incorrectly too high) value for m_{cut} by 0.2 mag. The vertical and horizontal lines denote the values used to simulate the data

parameters as part of the list of parameters to infer over rather than keep them fixed⁴.

In this case, inferring the value of Ψ requires modifying the likelihoods Eq. (7.23) and Eq. (7.24) by a multiplicative factor

$$\prod_i p(I_i = 1 | m_i = m_{\text{obs},i}, \Psi)$$

which is no longer an ignorable constant. Additionally to the priors of Eqs. (7.26)-(7.29), we now also have the following priors for Ψ ⁵:

$$\begin{aligned} m^{\text{cut}} &\sim \mathcal{U}(21, 27) \\ \log \sigma^{\text{cut}} &\sim \mathcal{U}(-3, 2). \end{aligned}$$

The results of keeping the selection parameters free to infer over is shown in Fig. 7.5. Our likelihood appears to recover approximately unbiased posteriors. In contrast, the R15 likelihood is not able to meaningfully constrain any of the cosmological parameters of interest (Note that the M_0 and h values should already be well constrained by the priors.) and for the selection function parameters, while the peaks can be seen to be close to truth values in R15's case, the posteriors are significantly wider and asymmetric compared to our constraints.

7.3.2 Including Measurement Noise

To be more realistic, we now add an additional layer to the hierarchy, distinguishing the latent, unobserved variables $\{z, m\}$, from the observable (though not necessarily observed) data, which include measurement noise, $\{\hat{m}, \hat{z}\}$, where the measurement uncertainties are given by, $\{\sigma_{z,i}^{\text{meas}}, \sigma_{m,i}^{\text{meas}}\}$. In order to include measurement errors in our analysis, we draw the measure-

⁴I add here also that the proposed methodology of R15 is to approximate the values of Ψ a-priori using simulations from SNANA and then keeping them fixed in the inference

⁵These priors can often be made more informative for realistic datasets by using SNANA simulations to approximate the selection parameters.

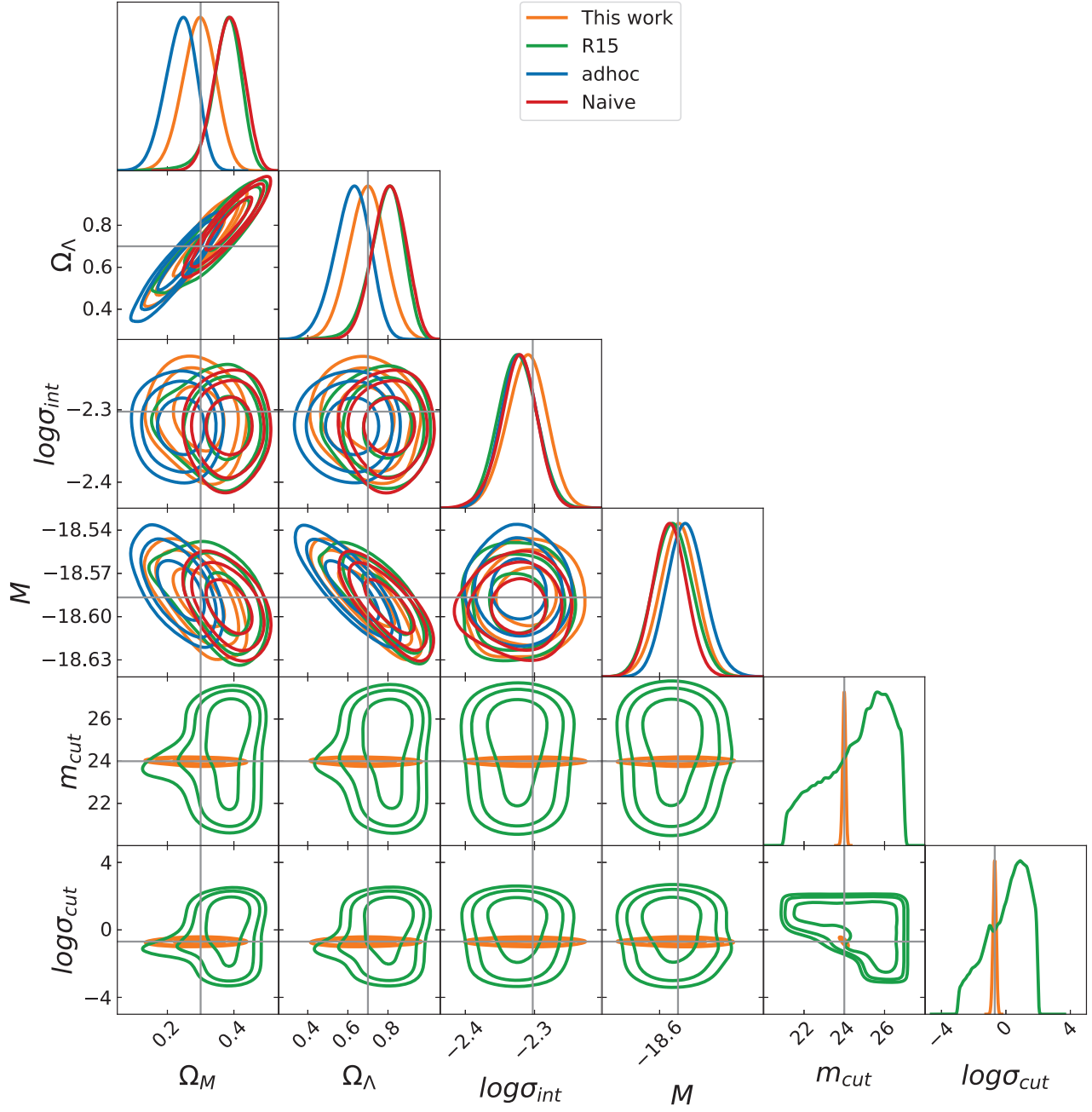


Figure 7.5: Parameter reconstructions averaged over 100 data realizations where we also simultaneously infer the selection parameters m_{cut} and σ_{cut} . Assumes no measurement error.

ment errors $\sigma_{z,i}^{\text{meas}}, \sigma_{m,i}^{\text{meas}}$ from

$$\log \sigma_{z,i}^{\text{meas}} | z_i \sim \mathcal{N}(t_z(z_i - u_z), s_z) \quad (7.31)$$

$$\log \sigma_{m,i}^{\text{meas}} | m_i \sim \mathcal{N}(t_m(m_i - u_m), s_m). \quad (7.32)$$

The measured values \hat{m}_i, \hat{z}_i are then normal-distributed around the latent variables m_i, z_i

$$\hat{m}_i | m_i, \sigma_{m,i}^{\text{meas}} \sim \mathcal{N}(m_i, \sigma_{m,i}^{\text{meas}}) \quad (7.33)$$

$$\hat{z}_i | z_i, \sigma_{z,i}^{\text{meas}} \sim \mathcal{N}(z_i, \sigma_{z,i}^{\text{meas}}) \quad (7.34)$$

where $\sigma_{m,i}^{\text{meas}}$ and $\sigma_{z,i}^{\text{meas}}$ is given by Eq. (7.32) and Eq. (7.31) respectively. The selection model is now a function of the measured values \hat{m}_i

$$p(I_i = 1 | \hat{m}_i, \Psi) = \text{CDF}_{\mathcal{N}}(m^{\text{cut}} - \hat{m}_i, \sigma^{\text{cut}}). \quad (7.35)$$

A realization of the simulated dataset is shown in Figures 7.6 and 7.7 with noise parameters

$$\begin{aligned} t_z &= 0.5, & u_z &= 11.5, & s_z &= 0.2, \\ t_m &= 0.1, & u_m &= 56, & s_m &= 0.6. \end{aligned} \quad (7.36)$$

Notice how Fig. 7.6 exhibits greater dispersion in the bottom panel compared to Fig. 7.1 because of the measurement error. Similarly the bias caused by selection effects in Fig. 7.7 is larger in magnitude than Fig. 7.2, including after applying an ad hoc data correction.

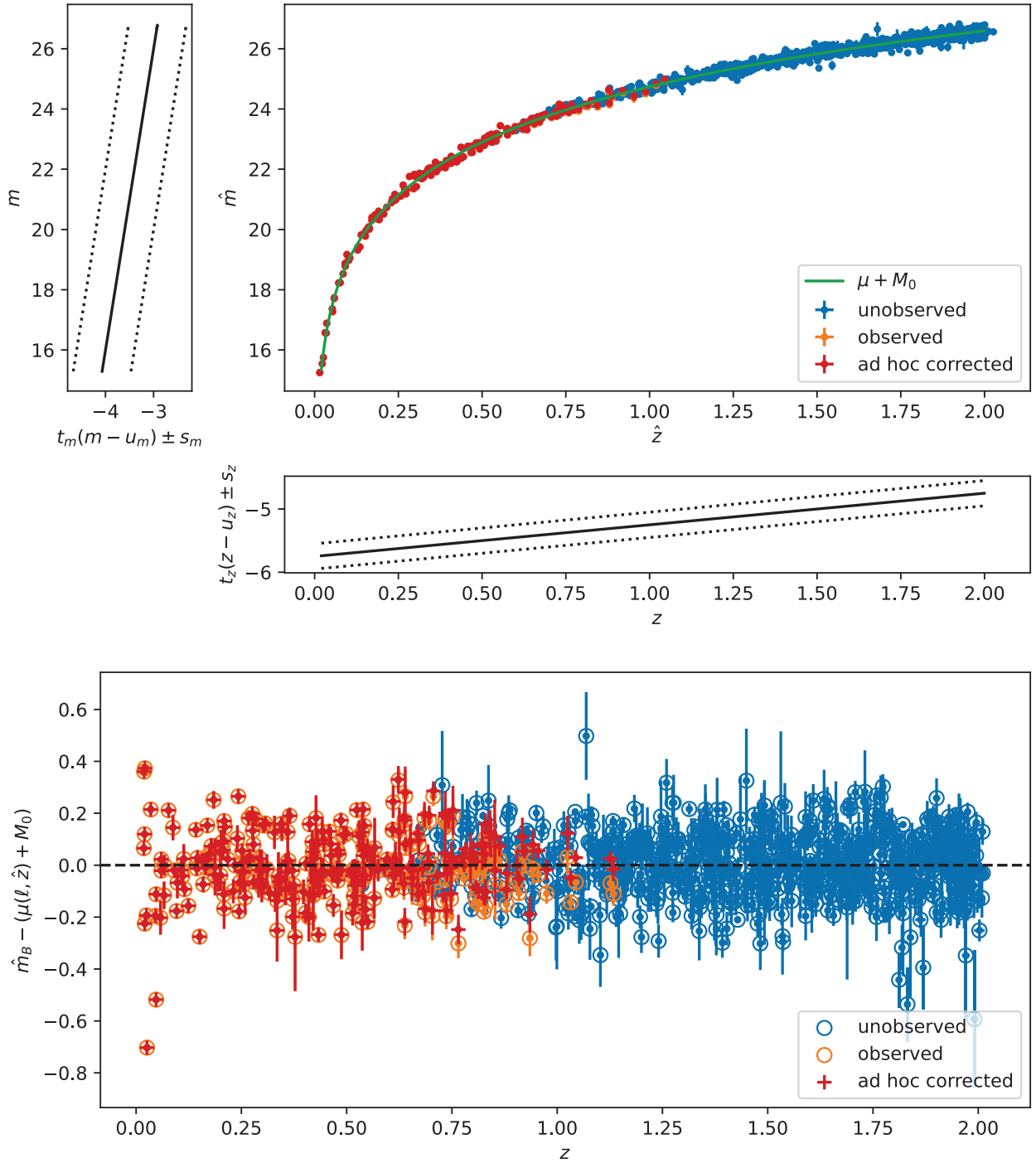


Figure 7.6: A Simulated data set of 5000 SNIa for simplified standard candles, with measurement errors. 262 observed SNIa are shown in orange and 738 unobserved SNIae are shown in blue after the data has been thinned randomly by a factor of 5 for visual clarity. An example visualisation before thinning is presented in Appendix B.1. The vertical bars on the data points represents the measurement error. The standard method corrects for selection effects by shifting the observed apparent magnitudes $m_{\text{obs},i}$ to the values shown in red crosses, then ignores selection effects in the likelihood. In the bottom chart the value of $\mu(\ell, z_i) + M_0$ is subtracted off for clarity. The top chart also has a line in green which represents the fiducial Λ CDM value expected as a function of redshift.

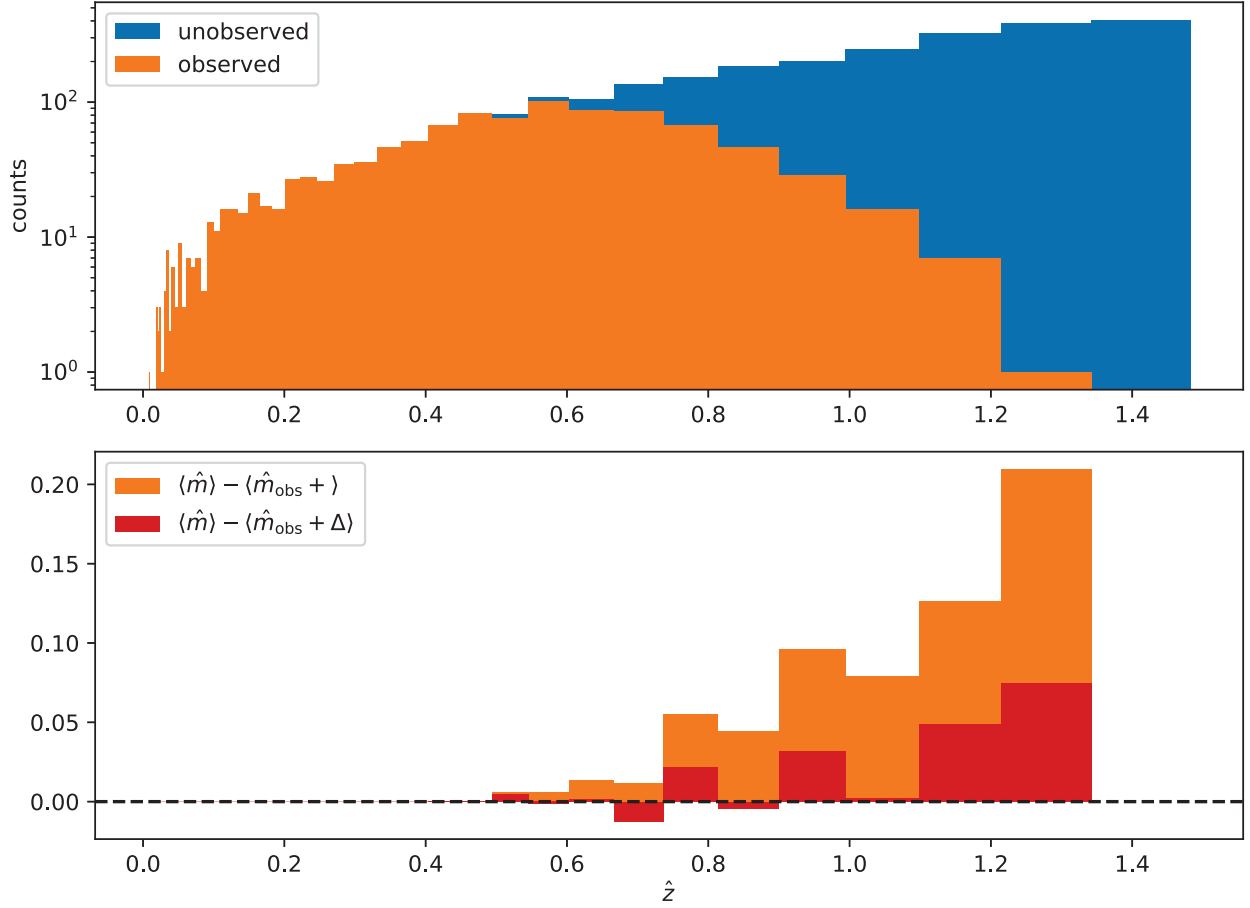


Figure 7.7: Top: Histogram of the complete (red) and observed (green) simulated SNIa with measurement error. Middle and bottom: difference between binned means of the complete dataset's apparent magnitudes, m , and the observed apparent magnitudes, m_{obs} , before and after selection effects correction. The bins have width Eq. (7.22) in $\log(z)$ -space with the right-most edge determined by $\max(m_{\text{sim,obs}})$.

We must modify the likelihoods $\mathcal{L}_{\text{ign}}, \mathcal{L}_{\text{R15}}, \mathcal{L}_{\text{obs}}$ to reflect these changes. Firstly,

$$\begin{aligned}\mathcal{L}_{\text{ign}} &= \prod_i p(D_i = D_{\text{obs},i} | \mathcal{C}, \Phi) \\ &= \prod_i \int dz_i dm_i p(D_i = D_{\text{obs},i} | m_i, z_i) p(m_i | z_i, \mathcal{C}, \Phi) p(z_i)\end{aligned}\quad (7.37)$$

$$\propto \prod_i \int dz_i dm_i \mathcal{N}(\hat{z}_{\text{obs},i} - z_i, \sigma_{z,i}^{\text{meas}}) \quad (7.38)$$

$$\times \mathcal{N}(\hat{m}_{\text{obs},i} - m_i, \sigma_{m,i}^{\text{meas}}) \quad (7.39)$$

$$\times \mathcal{N}(\log \sigma_{z,i}^{\text{meas}} - t_z(z_i - u_z), s_z) \quad (7.40)$$

$$\times \mathcal{N}(\log \sigma_{m,i}^{\text{meas}} - t_m(m_i - u_m), s_m) \quad (7.41)$$

$$\times \mathcal{N}(m_i - [\mu(\mathcal{C}, z_i) + M_0], \sigma^{\text{int}}) (1 + z_i)^\gamma \quad (7.42)$$

$$\propto \prod_i \int dz_i \mathcal{N}(\hat{z}_{\text{obs},i} - z_i, \sigma_{z,i}^{\text{meas}}) \quad (7.43)$$

$$\times \mathcal{N}(\log \sigma_{z,i}^{\text{meas}} - t_z(z_i - u_z), s_z) \quad (7.44)$$

$$\times \mathcal{N}\left(\mathbf{m}_i - [\mu(\mathcal{C}, z_i) + M_0], \sqrt{(\sigma^{\text{int}})^2 + (\boldsymbol{\sigma}_{m,i})^2}\right) \quad (7.45)$$

$$\times (1 + z_i)^\gamma, \quad (7.46)$$

where

$$\begin{aligned}\mathbf{m}_i &= \frac{(\sigma_{m,i}^{\text{meas}})^2 \left(\frac{\log \sigma_{m,i}^{\text{meas}}}{t_m} + u_m \right) + \left(\frac{s_m}{t_m} \right)^2 \hat{m}_{\text{obs},i}}{(\sigma_{m,i}^{\text{meas}})^2 + \left(\frac{s_m}{t_m} \right)^2} \\ \boldsymbol{\sigma}_{m,i}^2 &= \frac{(\sigma_{m,i}^{\text{meas}})^2 \cdot (s_m/t_m)^2}{(\sigma_{m,i}^{\text{meas}})^2 + (s_m/t_m)^2}.\end{aligned}$$

To evaluate the integral over z_i , we will assume that $\sigma_{z,i}^{\text{meas}}$ is small. Then expanding $\log(1 + z_i)$ about $\hat{z}_{\text{obs},i}$ gives

$$\begin{aligned}(1 + z_i)^\gamma &\approx \exp \left\{ \gamma \left[\log(1 + \hat{z}_{\text{obs},i}) + \frac{z_i - \hat{z}_{\text{obs},i}}{1 + \hat{z}_{\text{obs},i}} - \frac{(z_i - \hat{z}_{\text{obs},i})^2}{2(1 + \hat{z}_{\text{obs},i})^2} \right] \right\}\end{aligned}$$

and hence

$$\begin{aligned}
& \mathcal{N}(\hat{z}_{\text{obs},i} - z_i, \sigma_{z,i}^{\text{meas}}) (1 + z_i)^\gamma \\
& \propto \mathcal{N}(\tilde{z}_i - z_i, \tilde{\sigma}_{z,i}) \\
& \mathcal{N}(\hat{z}_{\text{obs},i} - z_i, \sigma_{z,i}^{\text{meas}}) (1 + z_i)^\gamma \\
& \times \mathcal{N}(\log \sigma_{z,i}^{\text{meas}} - t_z(z_i - u_z), s_z) \\
& \propto \mathcal{N}(\mathbf{z}_i - z_i, \boldsymbol{\sigma}_{z,i})
\end{aligned} \tag{7.47}$$

where

$$\begin{aligned}
\tilde{z}_i &= \hat{z}_{\text{obs},i} + \frac{(1 + \hat{z}_{\text{obs},i})\gamma(\sigma_{z,i}^{\text{meas}})^2}{(1 + \hat{z}_{\text{obs},i})^2 + \gamma(\sigma_{z,i}^{\text{meas}})^2} \\
\tilde{\sigma}_{z,i}^2 &= \frac{(1 + \hat{z}_{\text{obs},i})^2(\sigma_{z,i}^{\text{meas}})^2}{(1 + \hat{z}_{\text{obs},i})^2 + \gamma(\sigma_{z,i}^{\text{meas}})^2} \\
\mathbf{z}_i &= \frac{(\tilde{\sigma}_{z,i})^2 \left(\frac{\log \sigma_{z,i}^{\text{meas}}}{t_z} + u_z \right) + \left(\frac{s_z}{t_z} \right)^2 \tilde{z}_i}{(\tilde{\sigma}_{z,i})^2 + \left(\frac{s_z}{t_z} \right)^2} \\
\sigma_{z,i}^2 &= \frac{(\tilde{\sigma}_{z,i})^2 \cdot (s_z/t_z)^2}{(\tilde{\sigma}_{z,i})^2 + (s_z/t_z)^2}.
\end{aligned}$$

By linearizing μ about \mathbf{z}_i similar to what was done in [March et al. \(2011\)](#), we consequently have

$$\begin{aligned}
\mathcal{L}_{\text{ign}} &\propto \prod_i \mathcal{N} \left(\mathbf{m}_i - [\mu(\mathcal{C}, \mathbf{z}_i) + M_0], \right. \\
& \left. \sqrt{(\sigma^{\text{int}})^2 + [\partial_z \mu(\mathcal{C}, \mathbf{z}_i) \sigma_{z,i}]^2 + (\sigma_{m,i})^2} \right),
\end{aligned} \tag{7.48}$$

where the derivative of $\mu(\mathcal{C}, \mathbf{z}_i)$ is given by

$$\begin{aligned}
\partial_z \mu(\mathcal{C}, z) &= \frac{5}{\log 10} \cdot \frac{\partial_z D_L(\mathcal{C}, z)}{D_L(\mathcal{C}, z)} \\
\partial_z D_L(\mathcal{C}, z) &= \frac{D_L(\mathcal{C}, z)}{1 + z} \\
&+ \frac{10^{-5} c (1 + z)}{h} H(\mathcal{C}, z) \cosh \left[\sqrt{\Omega_\kappa} \int_0^z dz' H(\mathcal{C}, z') \right].
\end{aligned}$$

In the standard, ad hoc procedure one attempts to correct for selection effects by shifting the \hat{m}_{obs} in redshift bins

$$\mathcal{L}_{\text{ah}} = \mathcal{L}_{\text{ign}}(\hat{m}_{\text{obs},i} \rightarrow \hat{m}_{\text{obs},i} + \Delta(\hat{m}_{\text{obs},i})). \quad (7.49)$$

R15's likelihood in the presence of measurement errors becomes

$$\mathcal{L}_{\text{R15}} \propto \mathcal{L}_{\text{ign}} \prod_i \frac{p(I_i = 1 | \hat{m}_i = \hat{m}_{\text{obs},i}, \Psi)}{p(I_i = 1 | \hat{z}_i = \hat{z}_{\text{obs},i}, \sigma_{z,i}^{\text{meas}}, \sigma_{m,i}^{\text{meas}}, \Phi, \mathcal{C}, \Psi)}, \quad (7.50)$$

where the denominator can be obtained as

$$\begin{aligned} & p(I_i = 1 | \hat{z}_i = \hat{z}_{\text{obs},i}, \sigma_{z,i}^{\text{meas}}, \sigma_{m,i}^{\text{meas}}, \Phi, \mathcal{C}, \Psi) \\ & \approx \text{CDF}_{\mathcal{N}} \left(\frac{m^{\text{cut}} - [\mu(\mathcal{C}, \mathbf{z}_i) + M_0]}{\sqrt{(\sigma^{\text{cut}})^2 + (\sigma^{\text{int}})^2 + [\partial_z \mu(\mathcal{C}, \mathbf{z}_i) \sigma_{z,i}]^2 + (\sigma_{m,i}^{\text{meas}})^2}} \right). \end{aligned}$$

Our likelihood is given by:

$$\mathcal{L}_{\text{obs}} \propto \mathcal{L}_{\text{ign}} \frac{\prod_i p(I_i = 1 | \hat{m}_i = \hat{m}_{\text{obs},i}, \Psi)}{p(I_1 = 1 | \Phi, \mathcal{C}, \Psi)^{n_{\text{obs}}}}. \quad (7.51)$$

Again, the denominator

$$\begin{aligned} & p(I_1 = 1 | \Phi, \mathcal{C}, \Psi) \\ & = \int d\hat{m}_1 \int d\sigma_{m,1}^{\text{meas}} \int dm_1 \int dz_1 p(I_1 = 1 | \hat{m}_1, \Psi) \\ & \quad \times p(\hat{m}_1 | m_1, \sigma_{m,1}^{\text{meas}}) p(\sigma_{m,1}^{\text{meas}} | m_1) p(m_1 | z_1, \Phi, \mathcal{C}) p(z_1) \\ & = \int d\sigma_{m,1}^{\text{meas}} \int dm_1 \int dz_1 \text{CDF}_{\mathcal{N}} \left(\frac{m^{\text{cut}} - m_1}{\sqrt{(\sigma^{\text{cut}})^2 + (\sigma_{m,1}^{\text{meas}})^2}} \right) \\ & \quad \times p(\sigma_{m,1}^{\text{meas}} | m_1) p(m_1 | z_1, \Phi, \mathcal{C}) p(z_1). \end{aligned} \quad (7.52)$$

requires Monte Carlo integration where we:

1. draw a large sample of z_1 from Eq. (7.25) and draw a corresponding sample of m_1 from

- Eq. (7.2) using the values of Φ and \mathcal{C} conditioned upon in the denominator of \mathcal{L}_{obs} ,
2. draw a sample of $\sigma_{m,1}^{\text{meas}}$ from Eq. (7.32) using the sample of m_1 , and finally
 3. estimate $p(I_1 = 1 | \Phi, \mathcal{C}, \Psi)$ to be the mean of the $\text{CDF}_{\mathcal{N}}$ in Eq. (7.52) computed over the $m_1, \sigma_{m,1}^{\text{meas}}$ samples.

Notice also, in the presence of measurement errors, we also have to marginalize out the unknown measurement error of the dummy SNIa we integrate over which is done numerically. We choose not to do an analytic marginalization as the error model is dependent on the observing instrument so our analytic marginalisation may not be valid for realistic data. Future work will aim to use realistic simulations from SNANA to determine an error model and check if this realistic error model can be analytically marginalised out.

The posteriors resulting from using the likelihoods \mathcal{L}_{ign} , \mathcal{L}_{R15} , \mathcal{L}_{ah} and \mathcal{L}_{obs} are compared in Figure 7.8, where the cut parameters are inferred along with the other model parameters. As before, the naive likelihood ignoring selection effects leads to greatly biased reconstructions. When the cut parameters need to be inferred, using \mathcal{L}_{R15} produces highly unconstrained posteriors. On the other hand, \mathcal{L}_{obs} produces reasonably unbiased and constrained results.

Summary statistics of the mean bias and mean squared error are displayed Table 7.1. It can easily be seen that our method shows the smallest biases compared to the adhoc and unity likelihood with all parameters recovered with less than 1 sigma bias unlike the other two. In addition, our method shows smaller or similar mean squared error in all parameters of interest. Overall therefore, our method demonstrates superior performance. Our formalism is easily extended to include the remaining linear Phillips corrections to be more realistic (see Appendix B). Applying this extension is left to future work where it will be applied to realistic data sets as the focus of this chapter was to only outline the methodology for the inclusion of selection effects.

In all previous plots, one could have also considered sampling over the variable $M = M_0 - 5 \log_{10} h$ or leaving one of the parameters fixed (as in Chapter 5 where h was kept fixed). As the degeneracy cannot be broken between M_0 and h with just supernovae alone, these

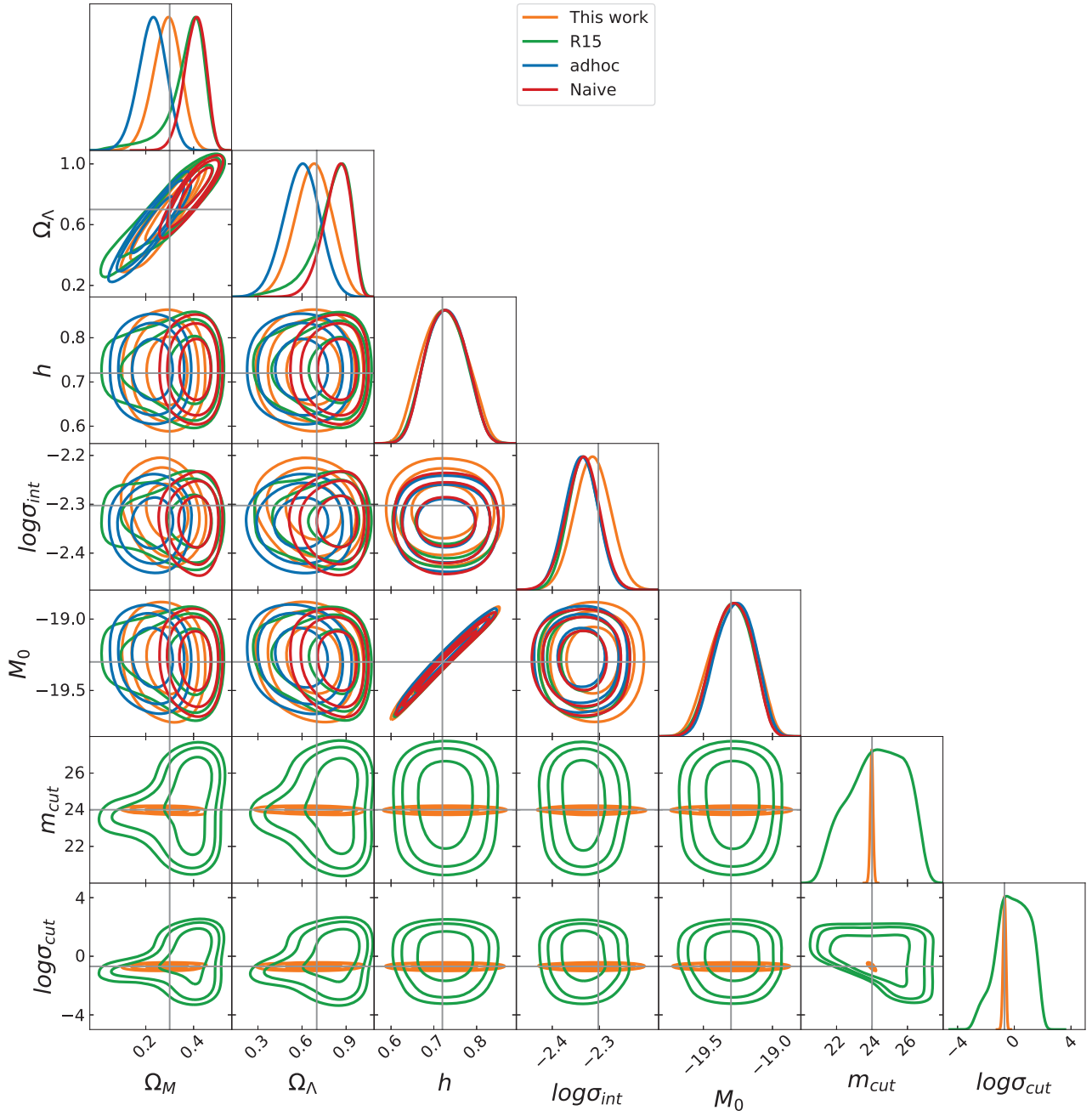


Figure 7.8: Parameter reconstructions averaged over 100 data realizations. Measurement error exists and cut parameters are inferred along with other model parameters, using the likelihoods Eq. (7.49), Eq. (7.50), and Eq. (7.51). The vertical and horizontal lines denote the values used to simulate the data.

parameters were not of primary focus in this paper. Furthermore, as sampling and inference is relatively quick in our toy single parameter model (less than one minute for our sample plots) There is little efficiency in sampling gain from switching to this sampling over M or keeping one parameter, like h , fixed. Therefore, I did not proceed with such a mechanism. This will be revisited during future works on realistic data.

7.4 Summary and Conclusion

In this chapter, we have presented a complete formalism for handling selection effects in a Bayesian way. The efficacy of our method was compared with the methodology of R15 and an adhoc correction procedure similar to what is traditionally employed and we found our method exhibited superior performance in terms of mean bias and mean squared error in both cases. With this methodology in hand, future work will attempt to apply this formalism to realistic datasets such as SNANA simulations, DES ([DES Collaboration et al., 2018](#)) and LSST ([Ivezić et al., 2019a](#)) data. In addition, the formalism presented here includes the uncertainty related to redshift errors, a concern that will be relevant in future surveys when photometric redshifts are only available from larger surveys such as LSST which have comparatively larger errors compared to spectroscopic errors which are typically negligible. Finally, by bypassing the use of an adhoc correction procedure and using a Bayesian formalism, we can more appropriately apply model selection mechanisms to discriminate between various cosmologies.

Mean Bias	Ω_M	Ω_Λ	h	M_0	$\log \sigma_{\text{int}}$	m_{cut}	$\log \sigma_{\text{cut}}$
Corrected data	$-0.066(1.65\sigma)$	$-0.084(1.02\sigma)$	$0.008(0.18\sigma)$	$-0.035(1.38\sigma)$	$0.002(0.19\sigma)$	-	-
Rubin et al. (2015)	$0.071(1.23\sigma)$	$0.102(1.07\sigma)$	$0.007(0.16\sigma)$	$-0.031(1.21\sigma)$	$0.008(0.06\sigma)$	$0.406(0.29\sigma)$	$0.608(0.55\sigma)$
This work	$-0.003(0.06\sigma)$	$-0.006(0.07\sigma)$	$0.005(0.10\sigma)$	$-0.011(0.42\sigma)$	$0.011(0.07\sigma)$	$-0.010(0.21\sigma)$	$-0.015(0.24\sigma)$
Mean Squared Error	Ω_M	Ω_Λ	h	M_0	$\log \sigma_{\text{int}}$	m_{cut}	$\log \sigma_{\text{cut}}$
Corrected data	0.008	0.023	0.002	0.002	0.018	-	-
Rubin et al. (2015)	0.011	0.027	0.002	0.002	0.018	2.360	1.701
This work	0.004	0.016	0.002	0.001	0.022	0.006	0.010

Table 7.1: Mean bias (top) where in parenthesis, the difference is expressed in units of the 1D posterior standard deviation. Mean Squared Error (bottom)

Chapter 8

Conclusion

8.1 Summary of Thesis Achievements

Here I outline the the main achievements of this thesis in the context of the motivations and objectives stated in Chapter 1.

In this thesis, I have provided an extensive literature view surveying the current state of SNIa cosmology. In Chapters 2, 3 and 4, I provided a background that outlines the context in which we analyse SNIa in a Bayesian formalism and our Bayesian Hierarchical Model **BAHAMAS**. This context sets the stage of thesis by surmising the current issues in the field of SNIa cosmology as well as future problems that will arise which are used to justify the motivation behind the work presented in this thesis.

After this review, I have conducted two investigations into the anisotropic expansion of the universe using the standardisable candles, Supernovae Type Ia, as well as how to account for Malmquist bias caused by selection effects in SNIa datasets within a Bayesian formalism.

In the first investigation into the anisotropic expansion, I explored the question of a dipolar anisotropy in the expansion of the universe, and derived new constraints on a possible dipolar modulation from SNIa data by building on **BAHAMAS** and extending it to include a new approximate correction for residual colour-based selection effects. As part of this analysis, the

treatment of peculiar velocities and host galaxy redshifts of SNIa from the original JLA paper by Betoule et al. (2014) were updated by adopting state-of-the-art flow models constrained using the 2M++ galaxy catalogue. Finally, I improved the treatment of both statistical and systematic uncertainties pertaining to peculiar velocity corrections – the dominant source of error for $z \lesssim 0.1$ SNIa, which are crucial for a robust, accurate and precise measurement of anisotropy in the local expansion.

No evidence for a deviation from isotropy, either in the framework of Λ CDM or in the phenomenological Cosmographic expansion could be found. Tight constraints were placed on the amplitude of a possible dipole both in the distance modulus and on the deceleration parameter. The upper bounds on any possible anisotropy are more stringent by a factor of ~ 2 than the results previously obtained from the same data sets with a comparable approach.

Bayesian model comparison was used to evaluate the Bayes factor between models featuring a dipole and an isotropically expanding universe (both in Λ CDM and in the Cosmographic expansion). We found moderate to strong Bayesian evidence against an anisotropic expansion. We have also evaluated the evidence in favour of acceleration, finding that a non-zero cosmological constant is preferred, using JLA SNe alone, by odds of 120:1, a result corroborated by a more traditional p -value approach based on a frequentist hypothesis test, which rejects $\Omega_\Lambda = 0$ with a p -value of 5.9×10^{-3} . In the Cosmographic expansion, a decelerating universe is disfavoured with odds of almost 1100 : 1 w.r.t. an accelerating one.

For the second investigation into anisotropic expansion, in Chapter 6, I further expanded on the works of Chapter 5, but this time I breakaway from analysing within a Λ CDM context and consider instead a phenomenological model that naturally allows for anisotropic expansion that occurs based on a transition period that occurs around dark matter–dark energy equality in redshift. From constructing the luminosity distance within this model, we find again this phenomenological model is not preferred over Λ CDM but neither is it decisively discarded based on the Bayes factor between the models. Current work is underway to apply a Bayesian forecasting method based on the expected utility formalism of Trotta et al. (2011) to determine the efficacy with which future surveys can provide more discriminating evidence to decide

between the models.

Finally, in Chapter 7, I outline the Bayesian formalism to fully account for selection effects caused by Malmquist bias in SNIa datasets. I also indicate the flaw in reasoning when [Rubin et al. \(2015\)](#) used a similar formalism for their selection effects model in their BHM, UNITY. Notably, I show in the realistic case where the selection function applied on to SNIa data is unknown, the formalism presented in this thesis recovered the selection function parameters with superior performance compared to UNITY/STEVE in terms of the mean bias and mean squared error of the recovered parameters. In addition, the model includes how to handle redshift uncertainties to make it suitable for the future of SNIa cosmology where large photometric samples will be available where the uncertainty in the measurements are no longer negligible in comparison to their spectroscopic samples.

8.2 Future Work

To handle the future of SNIa cosmology, there are still many areas to be improved upon within the BAHAMAS model. Future work will focus on accounting for the contamination of non-SNIa in data that can occur due to the lack of spectroscopic confirmation of their typing. One methodology to this, is using the Bayesian Estimation For Multiple Species (BEAMS) formalism ([Kunz et al., 2007](#)). BEAMS is designed to reduce biased parameter estimates made from contaminant supernovae in data that are not of Type Ia. This is especially important for future survey data from the Large Synoptic Survey Telescope (LSST) where multiple types of objects are observed and can not reliably have their object type determined from the photometry alone. The lack of spectroscopic follow up to confirm the typing of the object is due to the large volumes of potential SNIa candidates LSST is expected to observe ($\sim 10^4 - 10^5$ per year) such that it becomes infeasible for every object to have a spectroscopic confirmation.

As input, BEAMS requires classification probabilities for an object being an SNIa or not. The Type Ia probability is expected to be sourced from either extensions to the STACCATO algorithm by [Revsbech et al. \(2018\)](#), which currently does binary classification or another

algorithm. In particular, we are interested in algorithms that provide classification probabilities for objects expected to be observed by LSST. Many potential candidates have appeared after the end of the Kaggle PLaSTiCC competition ([The PLaSTiCC team et al., 2018](#)) such as AVOCADO by [Boone \(2019\)](#) which shares many similarities with STACCATO and RAPID by [Muthukrishna et al. \(2019\)](#) where both are designed to work with LSST-like data.

Our analysis of future data will also be affected by systematics which will require more precise modelling to resolve. For future work, improving our modelling will involve accounting for and investigating the effect of non-gaussianities such as in the colour population of SNIa. Another aspect of investigating the population distributions which the SALT2 summary statistics are drawn from is the potential redshift dependency they may have as claimed by [Rubin & Hayden \(2016\)](#). Based on this analysis, appropriate modifications to BAHAMAS may need to be made.

In addition, there are currently available datasets such as the low redshift Foundation sample [Foley et al. \(2018\)](#); [Jones et al. \(2019\)](#) which we will seek to analyse within our Bayesian formalism. At the low redshifts of the Foundation sample, peculiar velocity uncertainty is one of the dominant systematics. [Dam \(2020\)](#) illustrates a way to infer peculiar velocity corrections in the context of a BHM to allow propagation of correlations with cosmological parameters, a novelty that can also be pursued within BAHAMAS. For testing our selection effects formalism, other higher redshift samples such as those from DES ([DES Collaboration et al., 2018](#)) will also be used as well as realistic simulations from SNANA.

In closing, this is certainly an exciting transition point in cosmology, where the statistical power from future surveys is expected to provide a significant leap in our understanding of cosmology where we can delve deep into unravelling the mysteries behind our Universe and discriminate between the many proposed explanations for it.

Appendix A

Derivation and Test of Method of Moments

We present here the derivation of the the first and second moment of the moments generating function. We also demonstrate that our method of moments correctly recovers the selection function from simulations and that inference from replica of the data under the model is unbiased.

Consider the distribution of the random variable C , denoting the observed colour within a single survey and redshift bin sj . From Eqs. (5.26) and (5.27), we wish to compute the moment generating function,

$$M_C(t) = \int_{-\infty}^{\infty} e^{t\hat{c}} f_C(\hat{c}) d\hat{c} \quad (\text{A.1})$$

$$\begin{aligned} &= \frac{1}{p(I=1|\Psi, \Theta)} \times \\ &\int_{-\infty}^{\infty} e^{t\hat{c}} \cdot \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(\hat{c}-c_\star)^2} \cdot \Phi\left(\frac{c_{sj}^{\text{obs}} - \hat{c}}{\sigma_{sj}^{\text{obs}}}\right) d\hat{c}. \end{aligned} \quad (\text{A.2})$$

where $\sigma^2 \equiv R_c^2 + \sigma_{\hat{c}}^2$, and $\sigma_{\hat{c}}$ is the average measurement noise for colour observations (which we approximate as being the same for all data points in a given survey and redshift bin).

The above can be recast as:

$$M_C(t) = \frac{e^{c_\star t + \frac{1}{2}\sigma^2 t^2}}{p(I=1|\Psi, \Theta)} \times \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(\hat{c} - (c_\star + \sigma^2 t))^2} \Phi\left(\frac{c_{sj}^{\text{obs}} - \hat{c}}{\sigma_{sj}^{\text{obs}}}\right) d\hat{c} \quad (\text{A.3})$$

$$= \frac{e^{c_\star t + \frac{1}{2}\sigma^2 t^2}}{p(I=1|\Psi, \Theta)} \int_{-\infty}^{\infty} \mathcal{N}_{\hat{c}}(c_\star + \sigma^2 t, \sigma^2) \cdot \Phi\left(\frac{c_{sj}^{\text{obs}} - \hat{c}}{\sigma_{sj}^{\text{obs}}}\right) d\hat{c} \quad (\text{A.4})$$

$$= \frac{e^{c_\star t + \frac{1}{2}\sigma^2 t^2}}{p(I=1|\Psi, \Theta)} \Phi\left(\frac{c_{sj}^{\text{obs}} - (c_\star + \sigma^2 t)}{\sqrt{\sigma^2 + \sigma_{sj}^{\text{obs}2}}}\right). \quad (\text{A.5})$$

We now compute the first and second moments, set $h(t) \equiv e^{c_\star t + \frac{1}{2}\sigma^2 t^2}$ and $g(t) = \frac{c_{sj}^{\text{obs}} - (c_\star + \sigma^2 t)}{\sqrt{\sigma^2 + \sigma_{sj}^{\text{obs}2}}}$.

A dash (') symbol indicates a derivative with respect to t . Hence:

$$\begin{aligned} \left. \frac{dM_C(t)}{dt} \right|_{t=0} &= \frac{1}{p(I=1|\Psi, \Theta)} \frac{d}{dt} [h(t)\Phi(g(t))]_{t=0} \\ &= \frac{1}{p(I=1|\Psi, \Theta)} [h'(t)\Phi(g(t)) + h(t)\Phi'(g(t))]_{t=0} \end{aligned} \quad (\text{A.6})$$

and

$$\begin{aligned} \left. \frac{d^2 M_C(t)}{dt^2} \right|_{t=0} &= \frac{1}{p(I_i=1|\Psi, \Theta)} \times \\ &\quad [h''(t)\Phi(g(t)) + 2h'(t)\Phi'(g(t)) + h(t)\Phi''(g(t))]_{t=0} \end{aligned} \quad (\text{A.7})$$

We derive each of the terms $h(t)$, $h'(t)$, $h''(t)$, $\Phi(g(t))$, $\Phi'(g(t))$ and $\Phi''(g(t))$ evaluated at $t=0$:

$$h(0) = 1. \quad (\text{A.8})$$

$$h'(0) = h(t)(c_\star + \sigma^2 t)|_{t=0} = c_\star. \quad (\text{A.9})$$

$$h''(0) = h(t)\sigma^2 + h'(t)(c_\star + \sigma^2 t)|_{t=0} = \sigma^2 + c_\star^2. \quad (\text{A.10})$$

To determine $\Phi'(g(t))$ and $\Phi''(g(t))$ we use the Leibniz rule for differentiating under an integral (the CDF). As a reminder, our CDF is from the integral of $\mathcal{N}_x(0, 1)$ from $-\infty$ up to $g(t)$ with

c_{sj}^{obs} and σ_{sj}^{obs} used to control the width as opposed to the normal distribution hyperparameters.

This gives:

$$\Phi(g(t=0)) = \int_{-\infty}^{g(0)} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx \quad (\text{A.11})$$

$$\Phi'(g(t=0)) = \frac{d}{dt} \left[\int_{-\infty}^{g(t)} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx \right]_{t=0} \quad (\text{A.12})$$

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}g(t)^2} g'(t)|_{t=0} \quad (\text{A.13})$$

$$= -\frac{\sigma^2}{\sqrt{2\pi}\sqrt{\sigma^2 + \sigma_{sj}^{\text{obs}2}}} e^{-\frac{1}{2}\frac{(c_{sj}^{\text{obs}} - c_{\star})^2}{\sigma^2 + \sigma_{sj}^{\text{obs}2}}} \quad (\text{A.14})$$

where $g(0) = \frac{c_{sj}^{\text{obs}} - c_{\star}}{\sqrt{\sigma^2 + \sigma_{sj}^{\text{obs}2}}}$ and $g'(t)|_{t=0} = -\frac{\sigma^2}{\sqrt{\sigma^2 + \sigma_{sj}^{\text{obs}2}}}|_{t=0} = -\frac{\sigma^2}{\sqrt{\sigma^2 + \sigma_{sj}^{\text{obs}2}}}$. Finally,

$$\Phi''(t=0) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}g(t)^2} \cdot -g(t) \cdot g'^2(t) + \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}g(t)^2} g''(t)|_{t=0} \quad (\text{A.15})$$

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}g(t)^2} \left(g''(t) - g(t)g'^2(t) \right) \Big|_{t=0}. \quad (\text{A.16})$$

Given that, $g''(t)|_{t=0} = 0$, this reduces to

$$\Phi''(g(t))|_{t=0} = -\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}g(0)^2} \left(g(0)g'^2(0) \right). \quad (\text{A.17})$$

This leads to the first and second moments:

$$\frac{dM_C(t)}{dt} \Big|_{t=0} = \frac{1}{p(I=1|\Psi, \Theta)} \left(c_{\star} \Phi(g(0)) - \frac{\sigma^2}{\sqrt{2\pi}\sqrt{\sigma^2 + \sigma_{sj}^{\text{obs}2}}} e^{-\frac{1}{2}\frac{(c_{sj}^{\text{obs}} - c_{\star})^2}{\sigma^2 + \sigma_{sj}^{\text{obs}2}}} \right) \quad (\text{A.18})$$

$$\frac{d^2 M_C(t)}{dt^2} \Big|_{t=0} = \frac{1}{p(I=1|\Psi, \Theta)} \times \quad (\text{A.19})$$

$$\left((\sigma^2 + c_{\star}^2) \Phi(g(0)) - \frac{2c_{\star}\sigma^2}{\sqrt{2\pi}\sqrt{\sigma^2 + \sigma_{sj}^{\text{obs}2}}} e^{-\frac{1}{2}\frac{(c_{sj}^{\text{obs}} - c_{\star})^2}{\sigma^2 + \sigma_{sj}^{\text{obs}2}}} - \frac{\sigma^4}{\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(c_{sj}^{\text{obs}} - c_{\star})^2}{\sigma^2 + \sigma_{sj}^{\text{obs}2}}} \left(\frac{c_{sj}^{\text{obs}} - c_{\star}}{(\sigma^2 + \sigma_{sj}^{\text{obs}2})^{3/2}} \right) \right) \quad (\text{A.20})$$

where the normalization constant is given by Eq. (5.25).

We tested our method of moments to reconstruct the selection function on a suite of simulations, with $c_\star = 0.0$, $R_c = 0.1$, and three different choices of selection function parameters: $\{c_{sj}^{\text{obs}}, \sigma_{sj}^{\text{obs}}\} = \{[-0, 1, 0.02], [-0.1, 0.10], [0.0, 0.06]\}$, chosen to span the parameter space of interest in our application. We show the results of the reconstructed selection function for $N_{sj} = 30, 50, 200$ (from top to bottom) in Fig. A.1. The results show that the reconstruction, when averaged over realizations, is extremely close to the underlying true selection function, thus validating the method.

We tested parameter inference in the presence of residual colour-based selection effects data simulated according to the method presented in section 5.3.1, with colour-based selection effects as described in section 5.2.4, with selection function parameters for each survey being:

$$\text{SDSS} = \{(-0.5, 3.4), (-0.5, 0.57), (-0.35, 0.29), (0.20, 0.20), (0.20, 0.20)\},$$

$$\text{SNLS} = \{(0.20, 0.20), (-0.50, 1.17), (0.14, 0.17), (-0.06, 0.13), (-0.18, 0.14)\},$$

$$\text{Low-z} = \{(-0.50, 2.95), (-0.50, 4.47), (0.09, 0.01), (0.017, 0.01)\},$$

$\text{HST} = \{(-0.01, 0.12)\}$, where each tuple gives the values of $(c_{sj}^{\text{obs}}, \sigma_{sj}^{\text{obs}})$ in order from lowest redshift bin to highest within each survey. In the reconstruction, we estimate the selection function parameters as described above, and present 1- and 2-D marginal posteriors on all parameters in Fig. A.2. The posterior distributions have been averaged over $N = 100$ replicas. We observe that the posterior for all of the parameters has a mode very close to the true value, thus validating our methodology.

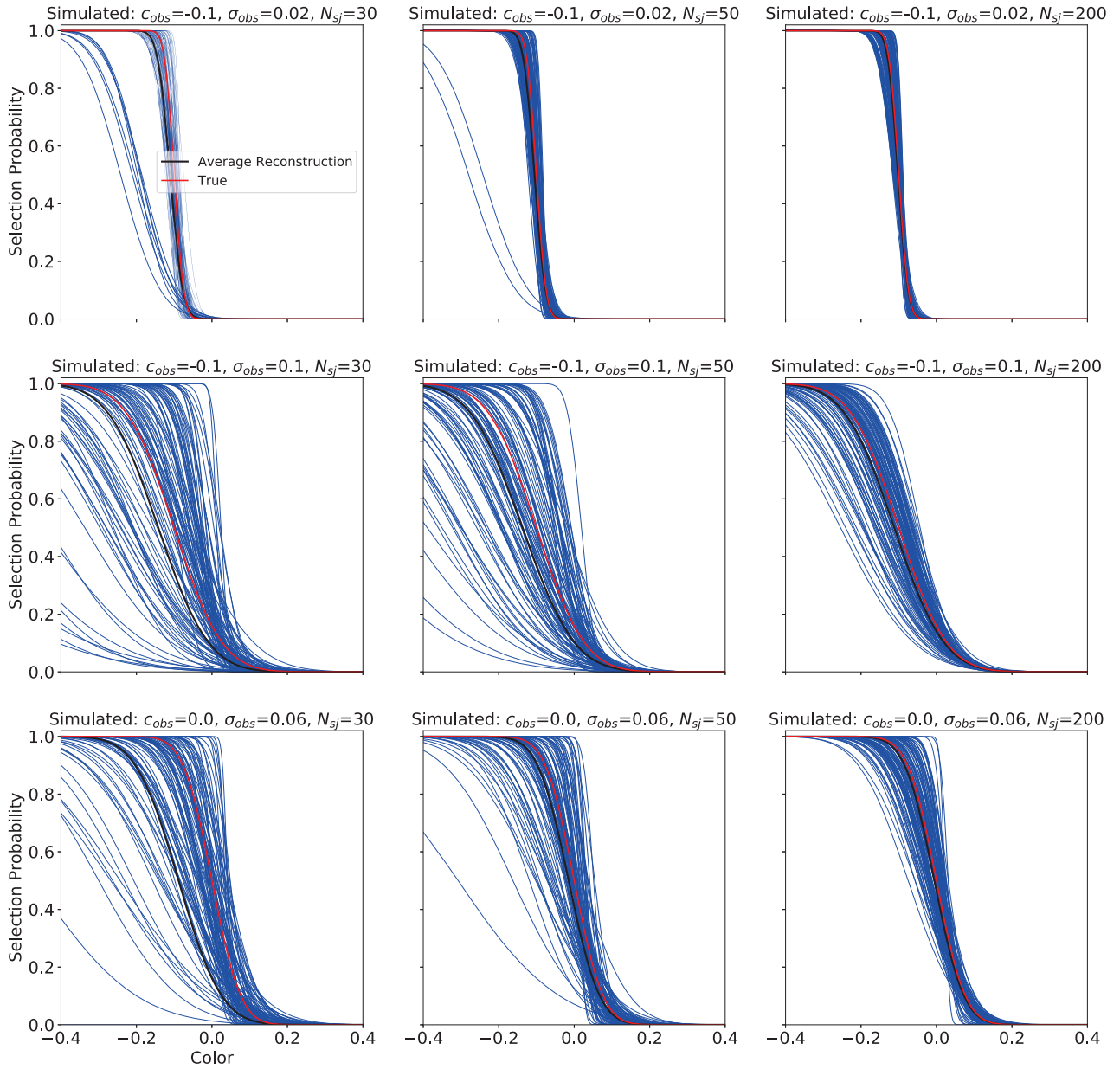
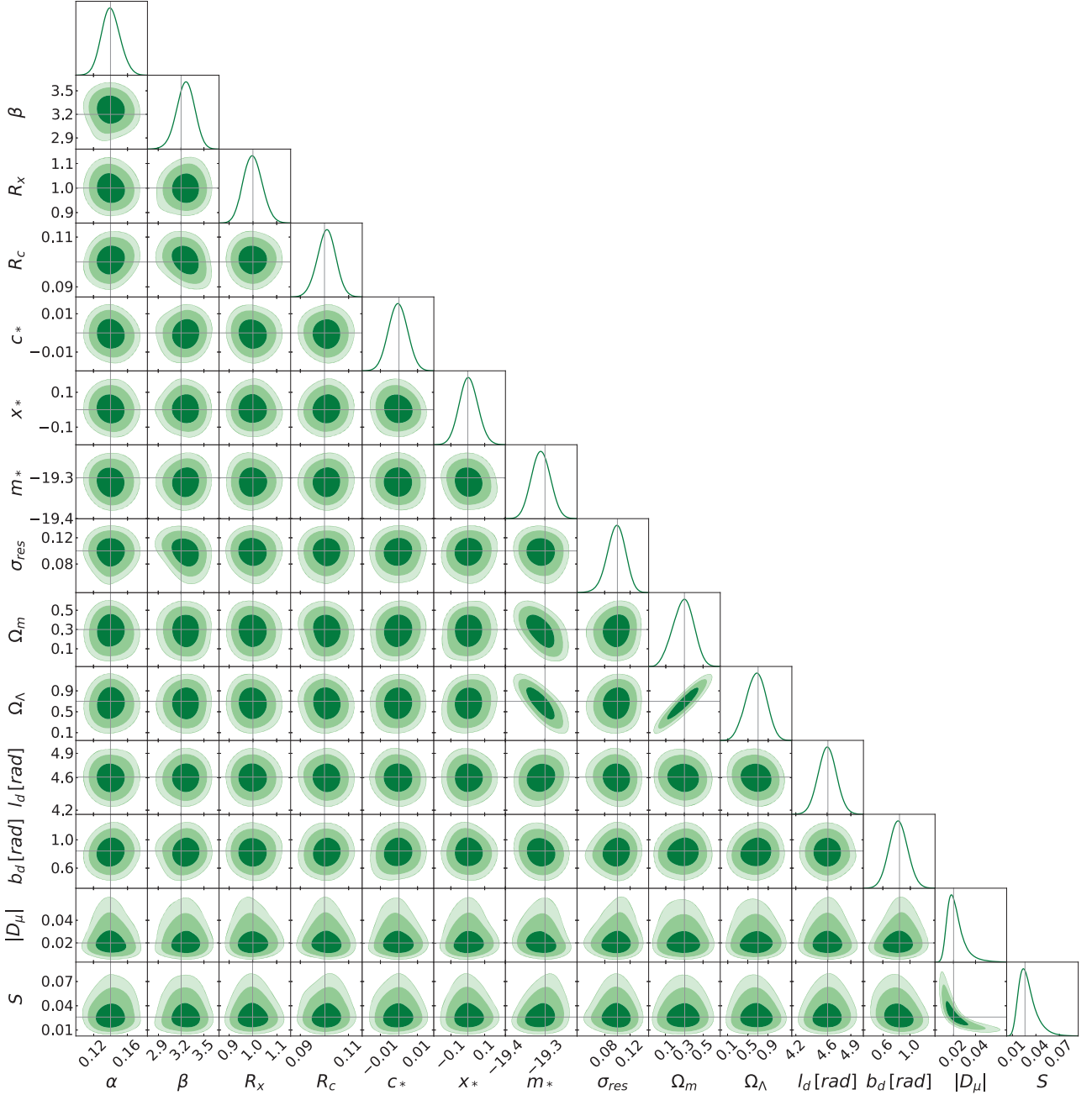


Figure A.1: Reconstructions of the selection function from the two first moments of the moment generating function from simulated data: we show three representative choices for the selection function parameters, $c_{sj}^{obs}, \sigma_{sj}^{obs}$ (columns) and three different sample sizes (SNe per bin, $N_{sj} = 30, 50, 100$, top to bottom in rows). Blue lines are the individual reconstruction from each of $N_{sim} = 100$ simulations and solid red is the true selection function.

Figure A.2: Posterior marginal distributions on simulated data with colour-based selection effects, and a simulated dipole, averaged over $N = 100$ data realizations. The posterior includes a correction for colour-based selection effects according to our method. Vertical lines give the true value of the parameters.



Appendix B

Selection Effects for SNIa

B.1 Simulations of Standard Candles

Here, I provide supplementary plots that indicate what a sample standard candle simulation looks like before they were thinned by a factor of 5 for clarity in Figs. 7.1 and 7.6. Figs. B.1 and B.2 show the charts for simulations without and with measurement errors respectively

B.2 Derivation of the Observed Data Likelihood

We write the observed data likelihood, $p(D_{\text{obs}}|\ell, \Phi, \Psi)$, as a marginalization over the unknown total number of SNIa, n :

$$p(D_{\text{obs}}|\ell, \Phi, \Psi) = \sum_{n=0}^{\infty} p(D_{\text{obs}}, D_{\text{obs}}^{\text{---}}|\ell, \Phi, \Psi, n)p(n). \quad (\text{B.1})$$

Here, $D_{\text{obs}}^{\text{---}}$ is the objects that are *not* observed. To compute $p(D_{\text{obs}}|\ell, \Phi, \Psi, n)$, we introduce the indicator variable I (linking D_{obs} to D) and marginalize it out:

$$p(D_{\text{obs}}|\ell, \Phi, \Psi, n) = \sum_I p(D_{\text{obs}}, I|\ell, \Phi, \Psi, n).$$

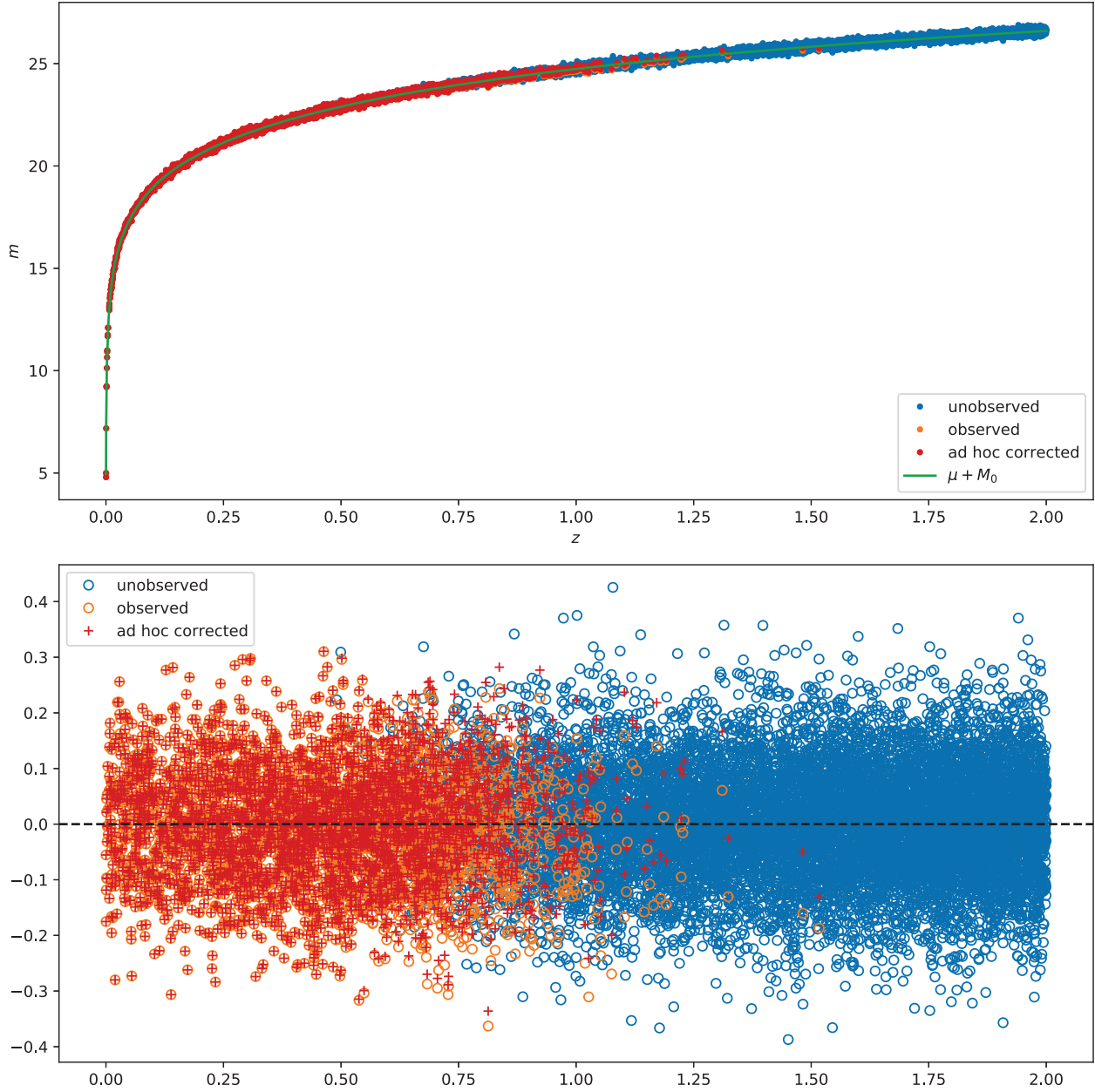


Figure B.1: A Simulated data set of 5000 SNIa for simplified standard candles, without measurement errors. 1038 observed SNIa are shown in orange and 3962 unobserved SNIae are shown in blue. The standard method corrects for selection effects by shifting the observed apparent magnitudes $m_{\text{obs},i}$ to the values shown in red crosses, then ignores selection effects in the likelihood. In the bottom chart the value of $\mu(\ell, z_i) + M_0$ is subtracted off for clarity. The top chart also has a line in green which represents the fiducial Λ CDM value expected as a function of redshift.

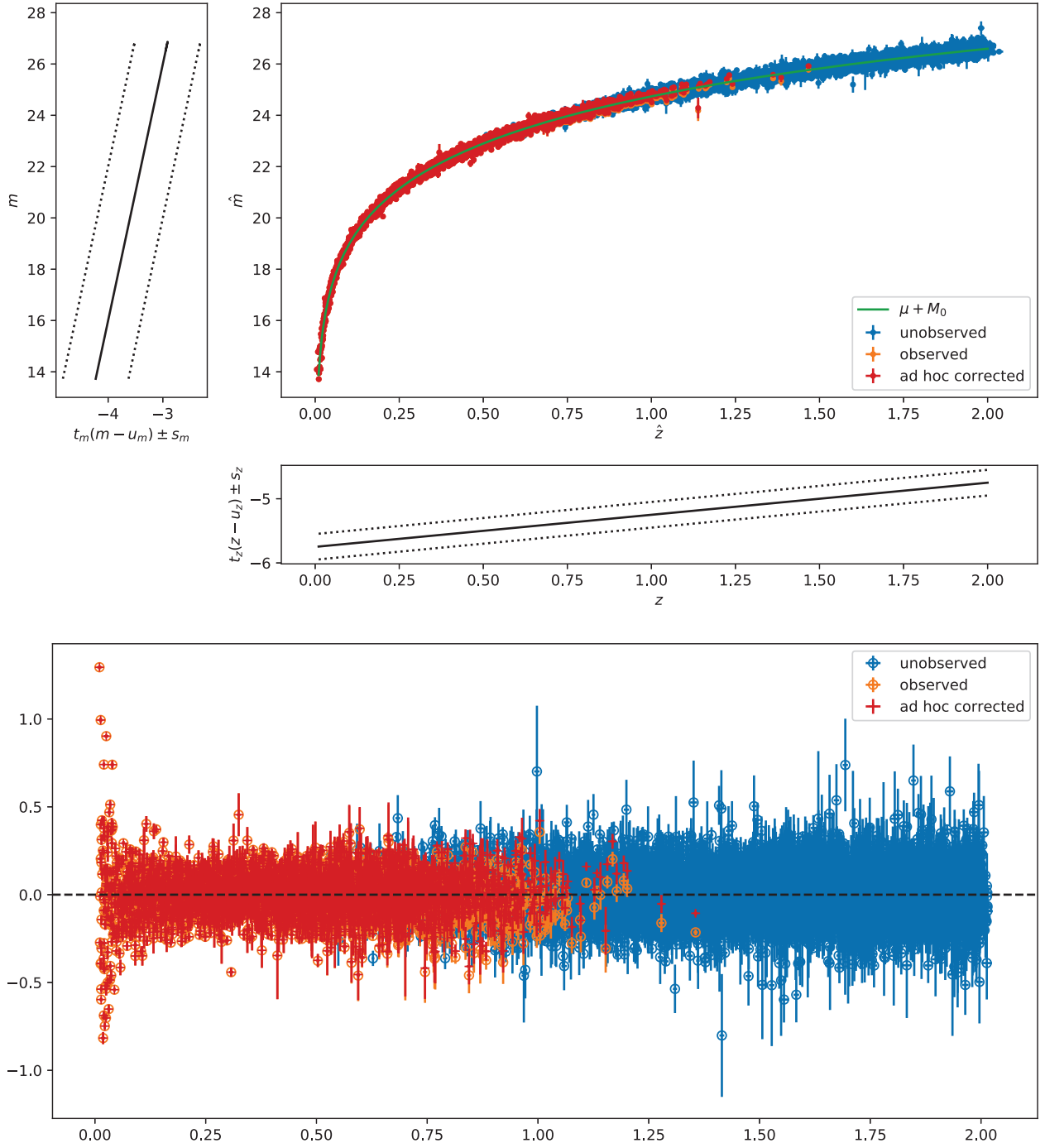


Figure B.2: A Simulated data set of 5000 SNIa for simplified standard candles, with measurement errors. 1153 observed SNIa are shown in orange and 3847 unobserved SNIae are shown in blue. The standard method corrects for selection effects by shifting the observed apparent magnitudes $m_{\text{obs},i}$ to the values shown in red crosses, then ignores selection effects in the likelihood. In the bottom chart the value of $\mu(\ell, z_i) + M_0$ is subtracted off for clarity. The top chart also has a line in green which represents the fiducial Λ CDM value expected as a function of redshift.

But, for a given number of observed SNIa, n_{obs} (which is determined by the length of D_{obs} and the parameters $\ell, \Phi, \text{and } \Psi$) and a given n , the indicator variable I only chooses which of the slots in D are filled by D_{obs} . Thus, the summand is 0 if I is not compatible with D_{obs} and n ; otherwise, the summand is independent of I . Hence,

$$\begin{aligned} p(D_{\text{obs}}|\ell, \Phi, \Psi, n) \\ = \left[\prod_k \binom{n_k}{n_{\text{obs},k}} \right] p(D_{\text{obs}}, I = \mathbb{1}_{n_{\text{obs}}}|\ell, \Phi, \Psi, n) \end{aligned}$$

where $\mathbb{1}_{n_{\text{obs}}}$ is the choice of I that is 1 in the first n_{obs} entries of each survey (indexed by k) and 0 elsewhere. We can rewrite this as:

$$p(D_{\text{obs}}, I = \mathbb{1}_{n_{\text{obs}}}|D, \Psi) \tag{B.2}$$

$$= \left[\prod_k \prod_{i=1}^{n_{\text{obs},k}} \delta(D_{ki} - D_{\text{obs},ki}) \right] p(I = \mathbb{1}_{n_{\text{obs}}}|D, \Psi). \tag{B.3}$$

We now introduce and marginalize out the complete data, D :

$$p(D_{\text{obs}}|\ell, \Phi, \Psi, n) = \int dD p(D_{\text{obs}}, D|\ell, \Phi, \Psi, n) p(D|\ell, \Phi, \Psi, n), \tag{B.4}$$

and the delta-functions appearing in Eq. (B.3) fill in the $n_{\text{obs},k}$ observed values, D_{ki} , for each survey k , thus reducing the integration to be only over the remaining unobserved data:

$$p(D_{\text{obs}}|\ell, \Phi, \Psi, n) \quad (\text{B.5})$$

$$\begin{aligned} & \left[\prod_k \binom{n_k}{n_{\text{obs},k}} \right] \int dD \, p(D_{\text{obs}}, I = \mathbb{1}_{n_{\text{obs}}}|D, \Psi) p(D|\ell, \Phi, \Psi, n) \\ &= \left[\prod_k \binom{n_k}{n_{\text{obs},k}} \right] \int dD_{\text{obs}} \left[p(I = \mathbb{1}_{n_{\text{obs}}}|D = (D_{\text{obs}}, D_{\text{obs}}), \Psi) p(D = (D_{\text{obs}}, D_{\text{obs}})|\ell, \Phi, \Psi, n) \right] \\ &= \prod_k \left\{ \binom{n_k}{n_{\text{obs},k}} \left[\prod_{i=1}^{n_{\text{obs},k}} p(I_{ki} = 1|D_{ki} = D_{\text{obs},ki}, \Psi_k) p(D_{ki} = D_{\text{obs},ki}|\ell, \Phi) \right] \right. \\ & \quad \times \underbrace{\int dD_{\text{obs}} \left[\prod_{j=n_{\text{obs},k}}^{n_k} p(I_{kj} = 0|D_{kj} = D_{\text{obs},kj}, \Psi_k) p(D_{kj} = D_{\text{obs},kj}|\ell, \Phi) \right]}_{I_{k,\text{obs}}} \left. \right\} \quad (\text{B.6}) \end{aligned}$$

where in the last line we have used the fact that observed and unobserved data are independent.

The integral over unobserved data can be computed as follows:

$$I_{k,\text{obs}} = \prod_{j=n_{\text{obs},k}}^{n_k} \left[\int dD_{\text{obs},kj} p(I_{kj} = 0, dD_{\text{obs},kj}|\ell, \Phi, \Psi_k) \right] \quad (\text{B.7})$$

$$= p(I_{k\mathcal{J}_k} = 0|\ell, \Phi, \Psi_k)^{n_k - n_{\text{obs},k}} \quad (\text{B.8})$$

where the last equality follows from the fact that the integral does not depend on the SNIa index j (as it should, since these SNIae are unobserved) and hence we can replace the SNIa index j by a common value \mathcal{J}_k (say, 1) for survey k . Notice that here the unobserved data contain *all* latent variables describing the unobserved SNIae, including in particular redshift (an important point we will return to later). Using this result in Eq. (B.6) and plugging the resulting expression into (B.1), we obtain for $p(D_{\text{obs}}|\ell, \Phi, \Psi)$:

$$\begin{aligned} p(D_{\text{obs}}|\ell, \Phi, \Psi) &= \prod_k \left\{ \left[\prod_{i=1}^{n_{\text{obs},k}} p(I_{ki} = 1|D_{ki} = D_{\text{obs},ki}, \Psi_k) p(D_{ki} = D_{\text{obs},ki}|\ell, \Phi) \right] \right. \\ & \quad \times \left[\sum_{n_k=n_{\text{obs},k}}^{\infty} \binom{n_k}{n_{\text{obs},k}} p(I_{k\mathcal{J}_k} = 0|\ell, \Phi, \Psi_k)^{n_k - n_{\text{obs},k}} p(n_k) \right] \left. \right\}, \quad (\text{B.9}) \end{aligned}$$

The convenient choice of a uniform-log prior $p(n_k) \propto 1/n_k$ for the total (unobserved) number of SNIae in each survey k renders the summand into an unnormalized negative binomial distribution. This can be seen by writing, with $p \equiv p(I_{k\mathcal{J}_k} = 1|\ell, \Phi, \Psi_k)$:

$$\sum_{n_k=n_{\text{obs},k}}^{\infty} \binom{n_k}{n_{\text{obs},k}} p^{n_k-n_{\text{obs},k}} \frac{1}{n_k} \quad (\text{B.10})$$

$$= \sum_{n_k=n_{\text{obs},k}}^{\infty} \binom{n_k-1}{n_{\text{obs},k}} p^{n_k-n_{\text{obs},k}} \frac{(1-p)^{n_{\text{obs},k}}}{(1-p)^{n_{\text{obs},k}}} \quad (\text{B.11})$$

$$= (1-p)^{-n_{\text{obs},k}} \sum_{l=0}^{\infty} \binom{l+r-1}{r} p^l (1-p)^r \quad (\text{B.12})$$

$$= (1-p)^{-n_{\text{obs},k}} \quad (\text{B.13})$$

where we have defined $l = n_k - n_{\text{obs},k}$, $r = n_{\text{obs},k}$ and the last equality follows from the normalization property of the negative binomial distribution for the number of successes, l , in a sequence with r failures (where p is the probability of success). Since $1-p = p(I_{k\mathcal{J}_k} = 1|\ell, \Phi, \Psi_k)$ is the probability that a SNIa is observed in survey k , we obtain for the likelihood:

$$p(D_{\text{obs}}|\ell, \Phi, \Psi) \propto \prod_k \frac{\prod_{i=1}^{n_{\text{obs},k}} p(I_{ki} = 1|D_{ki} = D_{\text{obs},ki}, \Psi_k) \times p(D_{ki} = D_{\text{obs},ki}|\ell, \Phi)}{p(I_{k1} = 1|\ell, \Phi, \Psi_k)^{n_{\text{obs},k}}}, \quad (\text{B.14})$$

where the numerator is a product of the selection model and the standard likelihood ignoring selection effects.

B.3 Connection Between our Result and R15

Here we show that the likelihood in R15 approximates the quantity

$$\mathcal{L}_{\text{R15}} \propto p(D_{\text{obs}}|\hat{z}_{\text{obs}}, \ell, \Phi, \Psi) \not\propto p(D_{\text{obs}}|\ell, \Phi, \Psi) \quad (\text{B.15})$$

(where z_{obs} , as n_{obs} -length array, should be understood as containing the information determining n_{obs}).

To see this, consider the expression for $p(\hat{z}_{\text{obs}}|\ell, \Phi, \Psi)$, analogous to (B.9):

$$p(\hat{z}_{\text{obs}}|\ell, \Phi, \Psi) \propto \prod_k \prod_{i=1}^{n_{\text{obs},k}} [p(I_{ki} = 1|\hat{z}_{ki} = \hat{z}_{\text{obs},ki}, \ell, \Phi, \Psi_k)p(\hat{z}_{ki} = \hat{z}_{\text{obs},ki}|\ell, \Phi, \Psi_k)] \\ \times p(I_{k1} = 1|\ell, \Phi, \Psi_k)^{-n_{\text{obs},k}}. \quad (\text{B.16})$$

We now additionally assume that $p(\hat{z}_{ki} = \hat{z}_{\text{obs},ki}|\ell, \Phi, \Psi)$ is independent of ℓ, Φ, Ψ . This is saying that, if one only knows that the (k, i) th SN exists and the values of the parameters ℓ, Φ, Ψ , then values of \hat{z}_{ki} are unaffected. However, $p(\hat{z}_{\text{obs}}|\ell, \Phi, \Psi)$ is not independent of ℓ, Φ, Ψ , since knowing ℓ, Φ, Ψ should significantly constrain the length n_{obs} of the arrays \hat{z}_{obs} (see derivation in Appendix B.2).

With this (incorrect) assumption, we have

$$p(I_{ki} = 1|\hat{z}_{ki} = \hat{z}_{\text{obs},ki}, \ell, \Phi, \Psi) \propto p(I_{ki} = 1, \hat{z}_{ki} = \hat{z}_{\text{obs},ki}|\ell, \Phi, \Psi).$$

Combining this with (B.16) and Rubin's likelihood (7.18) gives

$$p(\hat{z}_{\text{obs}}|\ell, \Phi, \Psi)\mathcal{L}_{\text{R15}} \\ \propto \prod_k \frac{\prod_{i=1}^{n_{\text{obs},k}} p(I_{ki} = 1|D_{ki} = D_{\text{obs},ki}, \Psi_k) \times p(D_{ki} = D_{\text{obs},ki}|\ell, \Phi)}{p(I_{k1} = 1|\ell, \Phi, \Psi_k)^{n_{\text{obs},k}}} \\ = p(D_{\text{obs}}|\ell, \Phi, \Psi),$$

giving (B.15). If $p(\hat{z}_{\text{obs}}|\ell, \Phi, \Psi)$ is conditionally independent on ℓ, Φ, Ψ (which we argued is generally not the case), then R15's likelihood and our likelihood are proportional to each other.

B.4 Including Phillips' Corrections

To generalize our selection effects model to include the remaining Phillips' corrections, let us suppose that the latent colour, stretch, and apparent magnitude, $d_i = (c_i, x_i, m_i)$, are drawn from

$$(c_i, x_i, M_i) | \Phi \sim \mathcal{N}((c_*, x_*, M_0), \text{COV}(c, x, M)), \quad (\text{B.17})$$

$$p(m_i | c_i, x_i, M_i, z_i, \ell) = \delta[\mu(\ell, z) + M - \alpha x_{1i} + \beta c_i - m_i], \quad (\text{B.18})$$

where $\ell = (\mathcal{C}, \alpha, \beta)$ now includes the Tripp coefficients α, β . The measured values $\hat{d}_i = (\hat{c}_i, \hat{x}_i, \hat{m}_{B,i})$ are then drawn from a normal distribution centered about d_i with some measurement covariance Σ_i^{meas} . Instead of (7.32), we require that Σ_i^{meas} be drawn from some d_i -dependent distribution for positive definite matrices $p(\Sigma_i^{\text{meas}} | d_i)$, e.g. a Wishart distribution with suitable parameters.

We also generalize the selection function appearing in our likelihood (7.13), making the probability for selection dependent on \hat{c}_i, \hat{x}_i :

$$\begin{aligned} p(I_i = 1 | D_i = D_{\text{obs},i}, \Psi) \\ = \text{CDF}_{\mathcal{N}} \left(\frac{m^{\text{cut}} - (\hat{m}_i + a_{\text{cut}} \hat{x}_{1i} - b_{\text{cut}} \hat{c}_i)}{\sigma_{\text{cut}}} \right), \end{aligned}$$

where $a_{\text{cut}}, b_{\text{cut}}$ are constant cut parameters included in $\Psi = (a_{\text{cut}}, b_{\text{cut}}, m_{\text{cut}}, \sigma_{\text{cut}})$.

The naive part is the regular likelihood from March et al. (2011); Shariff et al. (2016) that ignores selection effects and is given by

$$\begin{aligned} \mathcal{L}_{\text{ign}} &= \prod_i p(D_i = D_{\text{obs},i} | \ell, \Phi) \\ &= \prod_i \int (dz_i) (d^3 d_i) p(D_i = D_{\text{obs},i} | d_i, z_i) p(d_i | z_i, \ell, \Phi) p(z_i) \end{aligned} \quad (\text{B.19})$$

where in BAHAMAS $p(D_i = D_{\text{obs},i} | d_i, z_i)$ is approximated as

$$p(D_i = D_{\text{obs},i} | d_i, z_i) \propto p(\hat{d}_i = \hat{d}_{\text{obs},i}, \hat{z}_i = \hat{z}_{\text{obs},i} | d_i, z_i, \Sigma_i^{\text{meas}}, \sigma_i^{\text{meas}}). \quad (\text{B.20})$$

Finally, the denominator of our likelihood (7.13) reads

$$\begin{aligned} & p(I_1 = 1 | \Phi, \ell, \Psi) \\ &= \int (d^3 \hat{d}_1) (d \Sigma_1^{\text{meas}}) (d^3 d_1) (dz_1) p(I_1 = 1 | \hat{d}_1, \Psi) \\ & \quad \times p(\hat{d}_1 | d_1, \Sigma_1^{\text{meas}}) p(\Sigma_1^{\text{meas}} | d_1) p(d_1 | z_1, \Phi, \ell) p(z_1) \\ &= \int (d \Sigma_1^{\text{meas}}) (d^3 d_1) (dz_1) \\ & \quad \times \text{CDF}_{\mathcal{N}} \left(\frac{m^{\text{cut}} - \psi \cdot d_1}{\sqrt{(\sigma^{\text{cut}})^2 + \psi^T \Sigma_1^{\text{meas}} \psi}} \right) \\ & \quad \times p(\Sigma_1^{\text{meas}} | d_1) p(d_1 | z_1, \Phi, \ell) p(z_1), \end{aligned} \quad (\text{B.21})$$

where

$$\psi = \begin{bmatrix} -b^{\text{cut}} \\ a^{\text{cut}} \\ 1 \end{bmatrix}.$$

Computation of (B.21) requires Monte Carlo integration, as in the simplified standard candle case (7.52):

1. draw a large sample of z_1 from (7.25) and draw a corresponding sample of d_1 from (B.17),(B.18) using the values of Φ and ℓ conditioned upon in the denominator of \mathcal{L}_{obs} ,
2. draw a sample of Σ_1^{meas} from $p(\Sigma_1^{\text{meas}} | d_1)$, and finally
3. estimate $p(I_1 = 1 | \Phi, \ell, \Psi)$ to be the mean of the $\text{CDF}_{\mathcal{N}}$ in (B.21) computed over the $d_1, \Sigma_1^{\text{meas}}$ samples.

Bibliography

- Aghanim N., et al., 2020, *Astronomy & Astrophysics*, 641, A6
- Akrami Y., Fantaye Y., Shafieloo A., Eriksen H. K., Hansen F. K., Banday A. J., Górski K. M., 2014, *The Astrophysical Journal Letters*, 784, L42
- Albrecht A., et al., 2006, arXiv e-prints, pp astro-ph/0609591
- Alexeyev E. N., Alexeyeva L. N., Krivosheina I. V., Volchenko V. I., 1988, *Physics Letters B*, 205, 209
- Amanullah R., et al., 2010, *The Astrophysical Journal*, 716, 712
- Amanullah R., et al., 2015, *Monthly Notices of the Royal Astronomical Society*, 453, 3301
- Anderson L., et al., 2012, *The Monthly Notices of the Royal Astronomical Society*, 427, 3435
- Andrade U., Bengaly C. A. P., Alcaniz J. S., Santos B., 2018a, *Physical Review D*, 97, 083518
- Andrade U., Bengaly C. A. P., Santos B., Alcaniz J. S., 2018b, *Astrophys. J.*, 865, 119
- Annis J., Evans N. J., Miller B. J., Palmeri T. J., 2019, *Journal of Mathematical Psychology*, 89, 67
- Antoniou I., Perivolaropoulos L., 2010, *Journal of Cosmology and Astroparticle Physics*, 2010, 012
- Arcavi I., 2017, *Hydrogen-Rich Core-Collapse Supernovae*. Springer International Publishing, Cham, pp 239–276, doi:10.1007/978-3-319-21846-5_39, https://doi.org/10.1007/978-3-319-21846-5_39

- Ariew R., 1976, PhD thesis, University of Illinois at Urbana-Champaign
- Arnett W. D., 1982, *The Astrophysical Journal*, 253, 785
- Astier P., et al., 2006, *Astronomy and Astrophysics*, 447, 31
- Azzalini A., Arellano-Valle R. B., 2012, arXiv e-prints, p. arXiv:1203.2376
- Baade W., 1948, *Publications of the Astronomical Society of the Pacific*, 60, 230
- Baade W., 1952, *Proceedings of the International Astronomical Union*, 72, 169
- Baade W., 1956, *Scientific American*, 195, 92
- Baade W., Zwicky F., 1934, *Physical Review*, 46, 76
- Balbi A., et al., 2000, *The Astrophysical Journal Letters*, 545, L1
- Bautista J. E., et al., 2020, *Monthly Notices of the Royal Astronomical Society*, 500, 736
- Bayes T., Price R., 1763, *Royal Society letters*, Reproduced in *Biometrika*(1958),, 45, 293
- Begelman M. C., 2003, *Science*, 300, 1898
- Bennett C. L., et al., 2013, *The Astrophysical Journal Supplement Series*, 208, 20
- Bergström L., Goliath M., Goobar A., Mörtzell E., 2000, *Astronomy and Astrophysics*, 358, 13
- Bernal C., Cárdenas V. H., Motta V., 2017, *Physics Letters B*, 765, 163
- Bernardo J. M., 1979, *Journal of the Royal Statistical Society: Series B (Methodological)*, 41, 113
- Betoule M., et al., 2014, *Astronomy and Astrophysics*, 568, A22
- Birrer S., et al., 2020, *Astronomy & Astrophysics*, 643, A165
- Blandford R., Bühler R., 2017, in Alsabti A. W., Murdin P., eds, , *Handbook of Supernovae*. Springer, p. 83, doi:10.1007/978-3-319-21846-5_46
- Bondi H., Gold T., 1948, *Monthly Notices of the Royal Astronomical Society*, 108, 252

- Boone K., 2019, *The Astronomical Journal*, 158, 257
- Boruah S. S., Hudson M. J., Lavaux G., 2020a, arXiv e-prints, p. arXiv:2010.01119
- Boruah S. S., Hudson M. J., Lavaux G., 2020b, *Monthly Notices of the Royal Astronomical Society*, 498, 2703
- Branch D., 2006, *Nature*, 443, 283
- Branchini E., et al., 1999, *The Monthly Notices of the Royal Astronomical Society*, 308, 1
- Brout D., Scolnic D., 2021, *The Astrophysical Journal*, 909, 26
- Brout D., et al., 2022, arXiv e-prints, p. arXiv:2202.04077
- Buchner J., 2021, *Journal of Open Source Software*, 6, 3001
- Buchner J., et al., 2014, *Astronomy and Astrophysics*, 564, A125
- Burns C. R., et al., 2010, *The Astronomical Journal*, 141, 19
- Butkevich A. G., Berdyugin A. V., Teerikorpi P., 2005, *Monthly Notices of the Royal Astronomical Society*, 362, 321
- Cai R.-G., Tuo Z.-L., 2012, *Journal of Cosmology and Astroparticle Physics*, 2012, 004
- Calcino J., Davis T., 2017, *Journal of Cosmology and Astroparticle Physics*, 2017, 038
- Campanelli L., Cea P., Fogli G. L., Marrone A., 2011, *Phys. Rev. D*, 83, 103503
- Campbell H., Fraser M., Gilmore G., 2016, *Monthly Notices of the Royal Astronomical Society*, 457, 3470
- Carrick J., Turnbull S. J., Lavaux G., Hudson M. J., 2015, *Monthly Notices of the Royal Astronomical Society*, 450, 317
- Casella G., Berger R., 2001, *Statistical Inference*. Duxbury Resource Center
- Chambers K. C., et al., 2016, arXiv e-prints, p. arXiv:1612.05560

- Chevallier M., Polarski D., 2001, *International Journal of Modern Physics D*, 10, 213
- Childress M., et al., 2013, *The Astrophysical Journal*, 770, 108
- Childress M. J., Wolf C., Zahid H. J., 2014, *Monthly Notices of the Royal Astronomical Society*, 445, 1898
- Chugai N. N., 1991, *Monthly Notices of the Royal Astronomical Society*, 250, 513
- Colchero F., Clark J. S., 2011, *Journal of Animal Ecology*, 81, 139
- Colin J., Mohayaee R., Rameez M., Sarkar S., 2019a, arXiv e-prints, p. arXiv:1912.04257
- Colin J., Mohayaee R., Rameez M., Sarkar S., 2019b, *Astronomy & Astrophysics*, 631, L13
- Conley A., et al., 2011, *The Astrophysical Journal Supplement Series*, 192, 1
- Cooke R., Lynden-Bell D., 2010, *The Monthly Notices of the Royal Astronomical Society*, 401, 1409
- Cooke J., et al., 2012, *Nature*, 491, 228
- Curtin C., et al., 2019, *The Astrophysical Journal Supplement Series*, 241, 17
- D'Andrea C. B., et al., 2011, *The Astrophysical Journal*, 743, 172
- DES Collaboration et al., 2018, arXiv e-prints
- Dahlen T., Strolger L.-G., Riess A. G., 2008, *The Astrophysical Journal*, 681, 462
- Dam L., 2020, *Monthly Notices of the Royal Astronomical Society*, 497, 1301
- Dam L. H., Heinesen A., Wiltshire D. L., 2017, *The Monthly Notices of the Royal Astronomical Society*, 472, 835
- Davis M., Nusser A., Willick J. A., 1996, *The Astrophysical Journal*, 473, 22
- Davis T. M., et al., 2011, *The Astrophysical Journal*, 741
- Decourchelle A., 2017, in Alsabti A. W., Murdin P., eds, , *Handbook of Supernovae*. Springer, p. 117, doi:10.1007/978-3-319-21846-5_48

- Dey D., 2005, *Bayesian thinking : modeling and computation*. Elsevier, Amsterdam Boston
- Dhawan S., et al., 2019, *Monthly Notices of the Royal Astronomical Society*
- Diehl R., Timmes F. X., 1998, *Publications of the Astronomical Society of the Pacific*, 110, 637
- Dilday B., et al., 2008, *Astrophys. J.*, 682, 262
- Dodelson S., 2003, *Modern cosmology*. Elsevier Academic Press
- Dodelson S., 2021, *Modern cosmology*, 2 edn. Academic Press, London, United Kingdom
- Draine B. T., 2003, *Annual Review of Astronomy and Astrophysics*, 41, 241
- Duane S., Kennedy A., Pendleton B. J., Roweth D., 1987, *Physics Letters B*, 195, 216
- Einstein A., 1915, *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften* (Berlin, pp 844–847
- Einstein A., 1917, *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften* (Berlin, pp 142–152
- Eisenstein D. J., Hu W., 1998, *The Astrophysical Journal*, 496, 605
- Eisenstein D. J., et al., 2005, *The Astrophysical Journal*, 633, 560–574
- Etherington I. M. H., 1933, *Philosophical Magazine*, 15, 761
- Feroz F., Hobson M. P., 2008, *Monthly Notices of the Royal Astronomical Society*, 384, 449
- Feroz F., Hobson M. P., Bridges M., 2009, *Monthly Notices of the Royal Astronomical Society*, 398, 1601
- Feroz F., Hobson M. P., Cameron E., Pettitt A. N., 2013, *arXiv e-prints*,
- Filippenko A. V., 1997, *Annual Review of Astronomy and Astrophysics*, 35, 309
- Foglizzo T., 2016, *Explosion Physics of Core-Collapse Supernovae*. Springer International Publishing, Cham, pp 1–21, doi:10.1007/978-3-319-20794-0_52-1, https://doi.org/10.1007/978-3-319-20794-0_52-1

- Foley R. J., Kirshner R. P., 2013, *The Astrophysical Journal*, 769, L1
- Foley R. J., et al., 2013, *The Astrophysical Journal*, 767, 57
- Foley R. J., et al., 2018, *The Monthly Notices of the Royal Astronomical Society*, 475, 193
- Fragoso T. M., Bertoli W., Louzada F., 2017, *International Statistical Review*, 86, 1
- Friday T., Clowes R. G., Williger G. M., 2022, *Monthly Notices of the Royal Astronomical Society*, 511, 4159
- Friedmann A., 1922, *Zeitschrift fur Physik*, 10, 377
- Friedmann A., 1924, *Zeitschrift fur Physik*, 21, 326
- Frieman J. A., et al., 2008, *The Astronomical Journal*, 135, 338
- Gal-Yam A., 2012, *Science*, 337, 927
- Gallagher J. S., 2008, PhD thesis, University of Notre Dame
- Gamezo V. N., Khokhlov A. M., Oran E. S., 2004, *Physical Review Letters*, 92
- Gardner J. P., et al., 2006, *Space Science Reviews*, 123, 485
- Gaskell C. M., Cappellaro E., Dinerstein H. L., Garnett D. R., Harkness R. P., Wheeler J. C., 1986, *The Astrophysical Journal Letters*, 306, L77
- Gelman A., Carlin J. B., Stern H. S., Dunson D. B., Vehtari A., Rubin D. B., 2013, *Bayesian Data Analysis*. Chapman and Hall/CRC, doi:10.1201/b16018, <https://doi.org/10.1201/b16018>
- Geman S., Geman D., 1984, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6, 721
- Goliath M., Mörtzell E., 2000, *Physics Letters B*, 486, 249
- Gonçalves R. S., Carvalho G. C., Bengaly C. A. P., Carvalho J. C., Alcaniz J. S., 2018, *The Monthly Notices of the Royal Astronomical Society*, 481, 5270

- Gonçalves R. S., Carvalho G. C., Andrade U., Bengaly C. A. P., Carvalho J. C., Alcaniz J., 2020, arXiv e-prints, p. arXiv:2010.06635
- Goobar A., et al., 2017, *Science*, 356, 291
- Gupta S., Saini T. D., Laskar T., 2008, *Monthly Notices of the Royal Astronomical Society*, 388, 242
- Gutierrez J., Garcia-Berro E., Icko J. I., Isern J., Labay J., Canal R., 1996, *The Astrophysical Journal*, 459, 701
- Guy J., Astier P., Nobili S., Regnault N., Pain R., 2005, *Astronomy & Astrophysics*, 443, 781
- Guy J., et al., 2007, *Astronomy and Astrophysics*, 466, 11
- Hamuy M., Phillips M. M., Maza J., Suntzeff N. B., Schommer R. A., Aviles R., 1995, *The Astronomical Journal*, 109, 1
- Hamuy M., Phillips M. M., Suntzeff N. B., Schommer R. A., Maza J., Aviles R., 1996, *The Astronomical Journal*, 112, 2391
- Handley W. J., Hobson M. P., Lasenby A. N., 2015, *Monthly Notices of the Royal Astronomical Society*, 453, 4385
- Hastings W. K., 1970, *Biometrika*, 57, 97
- Heck D. W., 2018, *British Journal of Mathematical and Statistical Psychology*, 72, 316
- Heneka C., Marra V., Amendola L., 2014, *The Monthly Notices of the Royal Astronomical Society*, 439, 1855
- Hernández J. I. G., Ruiz-Lapuente P., Tabernero H. M., Montes D., Canal R., Méndez J., Bedin L. R., 2012, *Nature*, 489, 533
- Hicken M., Wood-Vasey W. M., Blondin S., Challis P., Jha S., Kelly P. L., Rest A., Kirshner R. P., 2009a, *The Astrophysical Journal*, 700, 1097

- Hicken M., Wood-Vasey W. M., Blondin S., Challis P., Jha S., Kelly P. L., Rest A., Kirshner R. P., 2009b, *The Astrophysical Journal*, 700, 1097
- Hill R., et al., 2018, *Monthly Notices of the Royal Astronomical Society*, 481, 2766
- Hillebrandt W., Niemeyer J. C., 2000, *Annual Review of Astronomy and Astrophysics*, 38, 191
- Hinshaw G., et al., 2009, *The Astrophysical Journal Supplement Series*, 180, 225
- Hinton S. R., et al., 2018, arXiv e-prints, p. arXiv:1811.02381
- Hoang T., 2017, *The Astrophysical Journal*, 836, 13
- Hobson M. P., Jaffe A. H., Liddle A. R., Mukherjee P., Parkinson D., eds, 2009, *Bayesian Methods in Cosmology*. Cambridge University Press, doi:10.1017/cbo9780511802461, <https://doi.org/10.1017/cbo9780511802461>
- Hollinger A. M., Hudson M. J., 2021, *The Monthly Notices of the Royal Astronomical Society*, 502, 3723
- Holz D. E., Linder E. V., 2005, *The Astrophysical Journal*, 631, 678
- Howell D. A., 2017, *Superluminous Supernovae*. Springer International Publishing, Cham, pp 431–458, doi:10.1007/978-3-319-21846-5_41, https://doi.org/10.1007/978-3-319-21846-5_41
- Howell D. A., et al., 2006, *Nature*, 443, 308
- Howlett C., Robotham A. S. G., Lagos C. D. P., Kim A. G., 2017, *The Astrophysical Journal*, 847, 128
- Hoyle F., 1948, *Monthly Notices of the Royal Astronomical Society*, 108, 372
- Hu W., White M., 2004, *Scientific American*, 290, 44
- Hubble E., 1929, *Proceedings of the National Academy of Sciences of the United States of America*, 15, 168

- Hubble E., 19251, American Astronomical Society, pp 252–255
- Hudson M. J., 1993, The Monthly Notices of the Royal Astronomical Society, 265, 43
- Huterer D., 2020, arXiv e-prints, p. arXiv:2010.05765
- Huterer D., Shafer D. L., Scolnic D. M., Schmidt F., 2017, Journal of Cosmology and Astroparticle Physics, 2017, 015
- Hutsemékers, D. Braibant, L. Pelgrims, V. Sluse, D. 2014, A&A, 572, A18
- Hutsemékers D., Cabanac R., Lamy H., Sluse D., 2005, Astronomy and Astrophysics, 441, 915
- Ivanov D., Liberati S., Viel M., Visser M., 2018, JCAP, 06, 040
- Ivezić Ž., et al., 2019a, The Astrophysical Journal, 873, 111
- Ivezić v. Z., et al., 2019b, Astrophys. J., 873, 111
- Janka H.-T., 2012, Annual Review of Nuclear and Particle Science, 62, 407
- Javanmardi B., Porciani C., Kroupa P., Pflamm-Altenburg J., 2015, The Astrophysical Journal, 810, 47
- Jeffreys H., 1939, The Theory of Probability. Oxford Classic Texts in the Physical Sciences, Oxford University Press
- Jha S. W., 2017, Handbook of Supernovae, p. 375–401
- Jha S., Riess A. G., Kirshner R. P., 2007, The Astrophysical Journal, 659, 122
- Jiménez J. B., Salzano V., Lazkoz R., 2015, Physics Letters B, 741, 168
- Joglekar H., Gangal K., Vahia M. N., Sule A., 2011, Oldest sky-chart with Supernova record 1, doi:10.1.1.599.8185
- Johansson J., 2015, Doctoral Thesis
- Johansson J., et al., 2021, The Astrophysical Journal, 923, 237

- Jones D. O., Riess A. G., Scolnic D. M., 2015, *The Astrophysical Journal*, 812, 31
- Jones D. O., et al., 2018, *The Astrophysical Journal*, 867, 108
- Jones D. O., et al., 2019, *The Astrophysical Journal*, 881, 19
- Jönsson J., et al., 2010, *The Monthly Notices of the Royal Astronomical Society*, 405, 535
- Jönsson J., Kronborg T., Mörtzell E., Sollerman J., 2008, *Astronomy & Astrophysics*, 487, 467
- Kamiya Y., Tanaka M., Nomoto K., Blinnikov S. I., Sorokina E. I., Suzuki T., 2012, *Astrophys. J.*, 756, 191
- Kato M., Hachisu I., 2003, *The Astrophysical Journal*, 598, L107–L110
- Katsuda S., 2017, in Alsabti A. W., Murdin P., eds, , *Handbook of Supernovae*. Springer, p. 63, doi:10.1007/978-3-319-21846-5_45
- Kelly B. C., 2007, *The Astrophysical Journal*, 665, 1489
- Kelly P. L., et al., 2015, *Science*, 347, 1123
- Ken'ichi N., Shigeyama T., 1988, in Kafatos M., Michalitsianos A. G., eds, *Supernova 1987A in the Large Magellanic Cloud*. pp 273–288
- Kenworthy W. D., et al., 2021, *The Astrophysical Journal*, 923, 265
- Kessler R., Scolnic D., 2017, *The Astrophysical Journal*, 836, 56
- Kessler R., et al., 2009, *The Astrophysical Journal Supplement Series*, 185, 32
- Kim Y.-L., Smith M., Sullivan M., Lee Y.-W., 2018, *The Astrophysical Journal*, 854, 24
- Kolatt T. S., Lahav O., 2001, *The Monthly Notices of the Royal Astronomical Society*, 323, 859
- Kothes R., 2017, in Alsabti A. W., Murdin P., eds, , *Handbook of Supernovae*. Springer, p. 97, doi:10.1007/978-3-319-21846-5_47
- Kowalski M., et al., 2008, *The Astrophysical Journal*, 686, 749

- Kragh H., 2015, *Masters of the universe : conversations with cosmologists of the past*. Oxford University Press, Oxford
- Krause E., et al., 2017, arXiv e-prints, p. arXiv:1706.09359
- Krisciunas K., Phillips M. M., Suntzeff N. B., 2004, *The Astrophysical Journal*, 602, L81
- Krisciunas K., Prieto J. L., Garnavich P. M., Riley J.-L. G., Rest A., Stubbs C., McMillan R., 2006, *The Astronomical Journal*, 131, 1639
- Kullback S., Leibler R. A., 1951, *The Annals of Mathematical Statistics*, 22, 79
- Kunz M., Bassett B. A., Hlozek R. A., 2007, *Physical Review D*, 75, 103508
- Kunz M., Hlozek R., Bassett B. A., Smith M., Newling J., Varughese M., 2013, in Hilbe J. M., ed., *Astrostatistical Challenges for the New Astronomy*, Edited by Joseph M. Hilbe. Springer, 2013, p. 1013, p. 63-86. p. 1013 (arXiv:1210.7762), doi:10.1007/978-1-4614-3508-2_4
- LSST Dark Energy Science Collaboration 2012, preprint (arXiv:1211.0310)
- LSST Science Collaboration et al., 2009, *LSST Science Book, Version 2.0* (arXiv:0912.0201)
- Lampeitl H., et al., 2010, *The Astrophysical Journal*, 722, 566
- Lavaux G., Hudson M. J., 2011, *Monthly Notices of the Royal Astronomical Society*, 416, 2840
- Leavitt H. S., 1908, *Annals of Harvard College Observatory*, 60, 87
- Leavitt H. S., Pickering E. C., 1912, *Harvard Obs. Circ.*, 173, 1
- Léget P.-F., et al., 2020, *Astronomy and Astrophysics*, 636, A46
- Lemaître G., 1931, *The Monthly Notices of the Royal Astronomical Society*, 91, 490
- Lemaître G., 1933, *Annales de la Société Scientifique de Bruxelles*, 53, 51
- Li X. D., van den Heuvel E. P. J., 1997, *Astronomy and Astrophysics*, 322, L9
- Li W., et al., 2011, *Monthly Notices of the Royal Astronomical Society*, 412, 1441

- Liddle A. R., 1999, *An introduction to modern cosmology*. Wiley
- Liddle A. R., Mukherjee P., Parkinson D., 2009, in Hobson M. P., Jaffe A. H., Liddle A. R., Mukherjee P., Parkinson D., eds, , *Bayesian Methods in Cosmology*. Cambridge University Press, pp 79–98, doi:10.1017/cbo9780511802461.005, <https://doi.org/10.1017/cbo9780511802461.005>
- Lilow R., Nusser A., 2021, arXiv e-prints, p. arXiv:2102.07291
- Lin H.-N., Wang S., Chang Z., Li X., 2016a, *Monthly Notices of the Royal Astronomical Society*, 456, 1881
- Lin H.-N., Li X., Chang Z., 2016b, *The Monthly Notices of the Royal Astronomical Society*, 460, 617
- Lindley D. V., 2000, *Journal of the Royal Statistical Society: Series D (The Statistician)*, 49, 293
- Loredo T. J., 1992, in Feigelson E. D., Babu G. J., eds, , *Statistical Challenges in Modern Astronomy*. Springer New York, New York, NY, pp 275–297, doi:10.1007/978-1-4613-9290-3_31, https://doi.org/10.1007/978-1-4613-9290-3_31
- Loredo T. J., 2004, in *AIP Conference Proceedings*. AIP, doi:10.1063/1.1835214, <https://doi.org/10.1063/1.1835214>
- Maartens R., 2011, *Philosophical Transactions of the Royal Society of London Series A*, 369, 5115
- Macauley E., et al., 2020, *Monthly Notices of the Royal Astronomical Society*, 496, 4051
- Maguire K., 2017, *Handbook of Supernovae*, pp 293–316
- Malmquist K. G., 1922, *Meddelanden fran Lunds Astronomiska Observatorium*, 100, 1
- Malz A. I., et al., 2019, *The Astronomical Journal*, 158, 171
- Mandel K. S., Wood-Vasey W. M., Friedman A. S., Kirshner R. P., 2009, *The Astrophysical Journal*, 704, 629

- Mandel K. S., Narayan G., Kirshner R. P., 2011, *The Astrophysical Journal*, 731, 120
- Mandel K. S., Scolnic D. M., Shariff H., Foley R. J., Kirshner R. P., 2017, *The Astrophysical Journal*, 842, 93
- Mandel K. S., Thorp S., Narayan G., Friedman A. S., Avelino A., 2021, *Monthly Notices of the Royal Astronomical Society*, 510, 3939
- Mannucci F., Valle M. D., Panagia N., Cappellaro E., Cresci G., Maiolino R., Petrosian A., Turatto M., 2005, *Astronomy & Astrophysics*, 433, 807
- March M. C., Trotta R., Berkes P., Starkman G. D., Vaudrevange P. M., 2011, *The Monthly Notices of the Royal Astronomical Society*, 418, 2308
- March M., Wolf R., Sako M., D'Andrea C., Brout D., 2018, *arXiv: Cosmology and Nongalactic Astrophysics*
- Mariano A., Perivolaropoulos L., 2012, *Phys. Rev. D*, 86, 083517
- Mather J. C., et al., 1994, *The Astrophysical Journal*, 420, 439
- Mazzali P. A., Röpke F. K., Benetti S., Hillebrandt W., 2007, *Science*, 315, 825
- McGrayne S., 2011, *The theory that would not die : how Bayes' rule cracked the enigma code, hunted down Russian submarines, and emerged triumphant from two centuries of controversy*, 1st edn. Yale University Press, New Haven Conn
- Messenger C., Veitch J., 2013, *New Journal of Physics*, 15, 053027
- Metcalf R. B., 1999, *Monthly Notices of the Royal Astronomical Society*, 305, 746
- Metropolis N., Rosenbluth A. W., Rosenbluth M. N., Teller A. H., Teller E., 1953, *The Journal of Chemical Physics*, 21, 1087
- Minkowski R., 1941, *Publications of the Astronomical Society of the Pacific*, 53, 224
- Mitton S. A., 2020, *arXiv*

- Moriya T., Tominaga N., Tanaka M., Maeda K., Nomoto K., 2010, *The Astrophysical Journal Letters*, 717, L83
- Moriya T. J., Sorokina E. I., Chevalier R. A., 2018, *Space Science Reviews*, 214
- Mörtsell E., Gunnarsson C., Goobar A., 2001, *The Astrophysical Journal*, 561, 106
- Mukhanov V., 2005, *Physical Foundations of Cosmology*. Cambridge University Press, doi:10.2277/0521563984
- Müller-Bravo T. E., Sullivan M., Smith M., Frohmaier C., Gutiérrez C. P., Wiseman P., Zontou Z., 2021, *Monthly Notices of the Royal Astronomical Society*
- Muthukrishna D., Narayan G., Mandel K. S., Biswas R., Hložek R., 2019, *Publications of the Astronomical Society of the Pacific*, 131, 118002
- Neill J. D., Hudson M. J., Conley A., 2007, *The Astrophysical Journal*, 661, L123
- Nielsen J. T., Guffanti A., Sarkar S., 2016, *Scientific Reports*, 6, 35596
- Nohr E. A., Liew Z., 2018, *Acta Obstetricia et Gynecologica Scandinavica*, 97, 407
- Nomoto K., Kondo Y., 1991, *The Astrophysical Journal*, 367, L19
- Nomoto K., Sugimoto D., 1977, *Publications of the Astronomical Society of Japan*, 29, 765
- Nomoto K., Sugimoto D., Neo S., 1976, *Astrophysics and Space Science*, 39, L37
- Nomoto K., Iwamoto K., Suzuki T., 1995, *Physics Reports*, 256, 173
- Nomoto K., Yamaoka H., Shigeyama T., Iwamoto K., 1996, *International Astronomical Union Colloquium*, 145, 49–68
- Nussbaumer H., Bieri L., 2011, Who discovered the expanding universe?, doi:10.48550/ARXIV.1107.2281, <https://arxiv.org/abs/1107.2281>
- Oguri M., 2019, *Reports on Progress in Physics*, 82, 126901
- Oke J. B., Gunn J. E., 1983, *The Astrophysical Journal*, 266, 713

- Opik E., 1922, *The Astrophysical Journal*, 55, 406
- Ovaisi Z., Ahsan R., Zhang Y., Vasilaky K., Zheleva E., 2020, in *Proceedings of The Web Conference 2020*. ACM
- Pakmor R., Kromer M., Taubenberger S., Sim S. A., Röpke F. K., Hillebrandt W., 2012, *The Astrophysical Journal*, 747, L10
- Peacock J. A., 1999, *Cosmological Physics*. Cambridge Academic Press
- Penzias A. A., Wilson R. W., 1965, *The Astrophysical Journal*, 142, 419
- Perlmutter S., et al., 1997, *ApJ*, 483, 565
- Perlmutter S., et al., 1999, *The Astrophysical Journal*, 517, 565
- Phillips M. M., 1993, *Astrophysical Journal*, Part 2 - Letters, 413, L105
- Phillips M. M., Lira P., Suntzeff N. B., Schommer R. A., Hamuy M., Maza J., 1999, *The Astrophysical Journal*, 118, 1766
- Piersanti L., Bravo E., Straniero O., Cristallo S., Dominguez I., 2022, *The Astrophysical Journal*
- Pike R. W., Hudson M. J., 2005, *The Astrophysical Journal*, 635, 11
- Planck Collaboration et al., 2016, *Astronomy and Astrophysics*, 594, A16
- Planck Collaboration et al., 2018, *Astronomy and Astrophysics*, 641, A1
- Poludnenko A. Y., Chambers J., Ahmed K., Gamezo V. N., Taylor B. D., 2019, *Science*, 366
- Prajs S., et al., 2016, *Monthly Notices of the Royal Astronomical Society*, 464, 3568
- Pskovskii Y. P., 1977, *SvA*, 21, 675
- Pskovskii Y. P., 1984, *Soviet Astronomy*, 28, 658
- Quimby R. M., et al., 2011, *Nature*, 474, 487
- Radburn-Smith D. J., Lucey J. R., Hudson M. J., 2004, *The Monthly Notices of the Royal Astronomical Society*, 355, 1378

- Rahman W., Trotta R., Boruah S. S., Hudson M. J., van Dyk D. A., 2021, Arxiv eprint (submitted to MNRAS)
- Rakavy G., Shaviv G., 1967, *The Astrophysical Journal*, 148, 803
- Refsdal S., 1964, *Monthly Notices of the Royal Astronomical Society*, 128, 307
- Revsbech E. A., Trotta R., van Dyk D. A., 2018, *The Monthly Notices of the Royal Astronomical Society*, 473, 3969
- Riess A. G., Press W. H., Kirshner R. P., 1996, *The Astrophysical Journal*, 473, 88
- Riess A. G., Davis M., Baker J., Kirshner R. P., 1997, *The Astrophysical Journal Letters*, 488, L1
- Riess A. G., et al., 1998, *The Astronomical Journal*, 116, 1009
- Riess A. G., et al., 2007, *The Astrophysical Journal*, 659, 98
- Riess A. G., et al., 2016, *The Astrophysical Journal*, 826, 56
- Riess A. G., Casertano S., Yuan W., Macri L. M., Scolnic D., 2019, *The Astrophysical Journal*, 876, 85
- Rigault M., et al., 2015, *The Astrophysical Journal*, 802, 20
- Rigault M., et al., 2020, *Astronomy & Astrophysics*, 644, A176
- Roberts E., Lochner M., Fonseca J., Bassett B. A., Lablanche P.-Y., Agarwal S., 2017, *Journal of Cosmology and Astroparticle Physics*, 2017, 036–036
- Robertson H. P., 1935, *The Astrophysical Journal*, 82, 284
- Roman M., et al., 2018, *Astronomy & Astrophysics*, 615, A68
- Rowan-Robinson M., 1985, *The cosmological distance ladder : distance and time in the universe.* W.H.Freeman and Co. Ltd
- Rubin D., Hayden B., 2016, *The Astrophysical Journal*, 833, L30

- Rubin D., Heitlauf J., 2020, *The Astrophysical Journal*, 894, 68
- Rubin D., et al., 2015, *The Astrophysical Journal*, 813, 137
- Rust B. W., 1974, PhD thesis, Oak Ridge National Lab., TN.
- Ryden B., 2016, *Introduction to Cosmology*, 2 edn. Cambridge University Press, doi:10.1017/9781316651087
- Ryle M., Clarke R. W., 1961, *The Monthly Notices of the Royal Astronomical Society*, 122, 349
- Said K., Colless M., Magoulas C., Lucey J. R., Hudson M. J., 2020, *Mon. Not. Roy. Astron. Soc.*, 497, 1275
- Saio H., Nomoto K., 1985, *Astronomy and Astrophysics*, 150, L21
- Saio H., Nomoto K., 2004, *The Astrophysical Journal*, 615, 444
- Sako M., et al., 2018, *Publications of the Astronomical Society of the Pacific*, 130, 064002
- Sarkar S., Pandey B., Khatri R., 2019, *The Monthly Notices of the Royal Astronomical Society*, 483, 2453
- Scalzo R. A., et al., 2010, *The Astrophysical Journal*, 713, 1073
- Schlafly E. F., et al., 2016, *The Astrophysical Journal*, 821, 78
- Schwarz D. J., Weinhorst B., 2007, *Astronomy and Astrophysics*, 474, 717
- Schwarz D. J., Copi C. J., Huterer D., Starkman G. D., 2016, *Classical and Quantum Gravity*, 33, 184001
- Scolnic D., Kessler R., 2016, *The Astrophysical Journal Letters*, 822, L35
- Scolnic D., et al., 2017, preprint ([arXiv:1710.00845](https://arxiv.org/abs/1710.00845))
- Scolnic D. M., et al., 2018a, *The Astrophysical Journal*, 859, 101
- Scolnic D., et al., 2018b, *Astrophys. J.*, 859, 101

- Scolnic D., et al., 2019, The Next Generation of Cosmological Measurements with Type Ia Supernovae (arXiv:1903.05128)
- Secrest N., von Hausegger S., Rameez M., Mohayaee R., Sarkar S., Colin J., 2020, arXiv e-prints, p. arXiv:2009.14826
- Shah P., Lemos P., Lahav O., 2022, Monthly Notices of the Royal Astronomical Society, 515, 2305
- Shapley H., 1918, Publications of the Astronomical Society of the Pacific, 30, 42
- Shariff H., Jiao X., Trotta R., van Dyk D. A., 2016, The Astrophysical Journal, 827, 1
- Sivia D. S., Skilling J., 2006, Data Analysis - A Bayesian Tutorial, 2nd edn. Oxford Science Publications, Oxford University Press
- Skilling J., 2004, in AIP Conference Proceedings. AIP, doi:10.1063/1.1835238, <https://doi.org/10.1063/1.1835238>
- Skilling J., 2006, Bayesian Analysis, 1
- Slipher V. M., 1913, Lowell Observatory Bulletin, 1, 56
- Slipher V. M., 1915, Popular Astronomy, 23, 21
- Slipher V. M., 1917, The Observatory, 40, 304
- Smith M., et al., 2013, The Astrophysical Journal, 780, 24
- Smith M., et al., 2020, Monthly Notices of the Royal Astronomical Society, 494, 4426
- Soltis J., Farahi A., Huterer D., Liberato C. M., 2019, Physical Review Letters, 122, 091301
- South L. F., Riabiz M., Teymur O., Oates C. J., 2021, Annual Review of Statistics and Its Application, 9
- Speagle J. S., 2020, Monthly Notices of the Royal Astronomical Society, 493, 3132
- Spergel D., et al., 2015, arXiv e-prints, p. arXiv:1503.03757

- Stahl B. E., de Jaeger T., Boruah S. S., Zheng W., Filippenko A. V., Hudson M. J., 2021, The Monthly Notices of the Royal Astronomical Society
- Strauss M. A., Willick J. A., 1995, Physics Reports, 261, 271
- Stritzinger M. D., et al., 2011, The Astronomical Journal, 142, 156
- Sullivan M., et al., 2006, The Astrophysical Journal, 648, 868
- Sullivan M., et al., 2010, Monthly Notices of the Royal Astronomical Society, 406, 782
- Sullivan M., et al., 2011, The Astrophysical Journal, 737, 102
- Sun Z. Q., Wang F. Y., 2018a, preprint ([arXiv:1804.05191](https://arxiv.org/abs/1804.05191))
- Sun Z. Q., Wang F. Y., 2018b, The Monthly Notices of the Royal Astronomical Society, 478, 5153
- Sun Z. Q., Wang F. Y., 2018c, Monthly Notices of the Royal Astronomical Society, 478, 5153–5158
- Sun Z. Q., Wang F. Y., 2019, European Physical Journal C, 79, 783
- Suyu S. H., Chang T.-C., Courbin F., Okumura T., 2018, Space Science Reviews, 214
- Taubenberger S., et al., 2011, Monthly Notices of the Royal Astronomical Society, 412, 2735
- The PLAsTiCC team et al., 2018, arXiv e-prints, p. [arXiv:1810.00001](https://arxiv.org/abs/1810.00001)
- Thorp S., Mandel K. S., 2022, Monthly Notices of the Royal Astronomical Society, 517, 2360
- Thorp S., Mandel K. S., Jones D. O., Ward S. M., Narayan G., 2021, arXiv e-prints, p. [arXiv:2102.05678](https://arxiv.org/abs/2102.05678)
- Tripp R., 1998, Astronomy and Astrophysics, 331, 815
- Trotta R., 2008, Contemporary Physics, 49, 71
- Trotta R., Kunz M., Liddle A. R., 2011, Monthly Notices of the Royal Astronomical Society, 414, 2337–2344

- Turnbull S. J., Hudson M. J., Feldman H. A., Hicken M., Kirshner R. P., Watkins R., 2012, The Monthly Notices of the Royal Astronomical Society, 420, 447
- Uddin S. A., Mould J., Lidman C., Ruhlmann-Kleider V., Zhang B. R., 2017, The Astrophysical Journal, 848, 56
- Uddin S. A., et al., 2020, The Astrophysical Journal, 901, 143
- Vaart A. W. v. d., 1998, Asymptotic statistics. Cambridge University Press, Cambridge, UK
New York, NY, USA
- VandenBerg D. A., Brogaard K., Leaman R., Casagrande L., 2013, The Astrophysical Journal, 775, 134
- Verde L., Treu T., Riess A. G., 2019, Nature Astronomy, 3, 891
- Vink J., 2017, in Alsabti A. W., Murdin P., eds, , Handbook of Supernovae. Springer, p. 139, doi:10.1007/978-3-319-21846-5_49
- Vishwakarma R. G., Narlikar J. V., 2010, Research in Astronomy and Astrophysics, 10, 1195–1198
- Visser M., 2004, Classical and Quantum Gravity, 21, 2603–2615
- Walker A. G., 1937, Proceedings of the London Mathematical Society, 42, 90
- Wang Y., 1999, The Astrophysical Journal, 525, 651
- Wang Y., 2005, Journal of Cosmology and Astroparticle Physics, 2005, 005
- Wang J. S., Wang F. Y., 2014, Monthly Notices of the Royal Astronomical Society, 443, 1680
- Wang X., Wang L., Pain R., Zhou X., Li Z., 2006, The Astrophysical Journal, 645, 488
- Webbink R. F., 1984, The Astrophysical Journal, 277, 355
- Webbink R. F., 1994, in Shafter A. W., ed., Astronomical Society of the Pacific Conference Series Vol. 56, Interacting Binary Stars. p. 440

- Wilks S. S., 1938, *The Annals of Mathematical Statistics*, 9, 60
- Winkler P. F., Gupta G., Long K. S., 2003, *The Astrophysical Journal*, 585, 324
- Wood-Vasey W. M., et al., 2007, *The Astrophysical Journal*, 666, 694
- Wood-Vasey W. M., et al., 2008, *The Astrophysical Journal*, 689, 377
- Woosley S., Janka T., 2005, *Nature Physics*, 1, 147–154
- Woosley S. E., Taam R. E., Weaver T. A., 1986, *The Astrophysical Journal*, 301, 601
- Yamaoka H., 2016, *Discovery, Confirmation, and Designation of Supernovae*. Springer International Publishing, Cham, pp 1–4, doi:10.1007/978-3-319-20794-0_128-1, https://doi.org/10.1007/978-3-319-20794-0_128-1
- Yang X., Wang F. Y., Chu Z., 2014, *Monthly Notices of the Royal Astronomical Society*, 437, 1840
- Zhao F.-Y., Strom R. G., Jiang S.-Y., 2006, *Chinese Journal of Astronomy and Astrophysics*, 6, 635
- Zhao D., Zhou Y., Chang Z., 2019, *The Monthly Notices of the Royal Astronomical Society*, 486, 5679
- Zheng C., et al., 2008, *The Astronomical Journal*, 135, 1766
- de Bernardis P., et al., 2000, *Nature*, 404, 955
- de Sitter W., 1917, *The Monthly Notices of the Royal Astronomical Society*, 78, 3