Measurement of supernova host galaxy properties in next generation surveys

Jamie Dumayne

Physics

Department of Physics

Lancaster University

November 2023

A thesis submitted to Lancaster University for the degree of Doctor of Philosophy in the Faculty of Science and Technology

Supervised by Professor Isobel M. Hook

1. Reviewer: Mathew Smith

2. Reviewer: Kathy Romer

Abstract

Upcoming supernova (SN) surveys will enable an exciting time in cosmology. The Rubin Observatory's 10-year Legacy Survey of Space and Time (LSST) will observe near to 20 billion galaxies. Approximately 10^5 galaxies observed per year will contain Type Ia supernovae (SNe Ia). One of the sub-surveys using the 4-metre Multi-Object Spectroscopic Telescope (4MOST), the Time Domain Extragalactic Survey, will obtain spectra of 35,000 live transients and 50,000 host-galaxies during the first 5 years of 4MOST.

In this thesis we quantify the ability for these two surveys to measure host-galaxy properties, when used together. For each galaxy obsereved by 4MOST and LSST the properties can be inferred, allowing SN host-galaxy properties to be calculated on a large scale. Measuring the properties of SN host-galaxies serves two main purposes. The first is that Type Ia SNe exhibit correlations between host-galaxy properties and the peak luminosities of the SNe, which has implications for their use as standardisable candles in cosmology. Secondly, there are known correlations between host-galaxy type and supernova type, which can be used to aid in the classification of SNe. We have used simulations to quantify the improvement in host-galaxy stellar mass (M_*) measurements when supplementing photometry from Rubin with spectroscopy from the 4MOST instrument. We provide results in the form of expected uncertainties in M_* for galaxies with $0.1 < z < 0.9$ and $18 < r_{AB} < 25$. We show that for galaxies mag 22 and brighter, combining Rubin and 4MOST data reduces the uncertainty measurements of galaxy M_* by more than a factor of 2 compared with Rubin data alone. This applies for elliptical and Sc type hosts. We see a reduction to other galaxy properties, including: galaxy age, metallicity, star

formation rate, star formation timescale and v-band extinction. We demonstrate that the reduced uncertainties in M_* lead to a reduction of 7% in the ambiguity of the application of the "mass step" correction. This leads to a 2% reduction in the uncertainty on w. This is a negligible improvement. However, we would expect a larger effect to w_0 and w_a. This means more precise mass measurements could be crucial in future surveys.

We then demonstrate that improved measurements of redshift, stellar-mass and star formation rate aids in photometric classification of supernovae. To do this, we used a simple lookup table containing the host-galaxy properties and a machine learning algorithm. The machine learning algorithm was trained on 6 different datasets with different information included. We find the machine learning model which only has access to the SNe light curves performs the worst at producing a pure SNe Ia sample. The lookup table performs the best. Whilst a ML model which has access to SNe light curves, spectroscopic redshifts and host-galaxy properties as observed by 4MOST and LSST is able to produced the largest sample of comparable purity. We expect these results will help to improve the constraints on the dark energy of state parameters, w_0 and w_a.

The results presented in this thesis are proof of concept of what will be possible with 4MOST and LSST, and demonstrate the implications to SN cosmology in the near future.

Acknowledgements

I couldn't have completed this PhD without all of the support I have been given over the years. Firstly, I'd like to thank my supervisor Isobel Hook. I appreciate all the guidance and wisdom you've imparted which has been irreplaceable in my research.

I'd like to say a big heartfelt thank you to the entire observational astrophysics group. All of the PhDs, postdocs and lectrurers have all given me so much help. I really wouldn't have been able to get through these last 4 years without it. I'm also thankful for all the oppurtunities which came with the PhD. There were some incredible adventures that I will remember forever.

I would also like to give a big thank you to Andy Blake, Agnieszka Nowak and the whole public engagement team. I've learnt so much from you all. I really appreciate the support all of you gave me, whether it was just giving advice or driving for 8 hours to support my crazy ideas. We did some amazing things, and I'm so grateful I got to be a part of them.

Finally, I'd like to thank my family for all the support they've given me. On top of supporting me through the PhD, I also couldn't have gotten through the Covid pandemic without them. An especially big thank you to Brenda Zambrano for being my biggest supporter.

Declaration

This thesis is my own work and no portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification at this or any other institute of learning.

The research presented in chapter 3 of this thesis has been published in paper format by a peer-reviewed scientific journal as Dumayne et al. (2023, RASTI, 2, 1). As per the policy of the RASTI journal, we have made our major results available. The data used to create Figures 3.5, 3.6 3.9, 3.10, 3.13, 3.14, 3.15 and 3.16 are available, as well as the results obtained for the Sc galaxy, at Lancaster University's data archive `https://doi.org/10.17635/lancaster/researchdata/621`.

Our work has made use of tools and data provided by other researchers. We have detailed all of these below.

- The software we used for SED fitting (Chapter 3), FAST, was created by Mariska Kriek (Kriek et al., 2009), documented by James Aird and adapted to Python by Corentin Schreiber
- The deep learning software we used to classify supernovae using host galaxy properties (Chapter 4) is SuperNnova, created by Anais Möller (Möller & de Boissiè re, 2019a).
- The catalogue of simulated supernovae and host-galaxies we used to train the deep learning model (Chapter 4) was created by Maria Vincenzi and Chris Frohmaier (Frohmaier et al. in prep.). The catalogue was provided internally by the TiDES collaboration.

"Keep moving forward" - Wilbur Robinson

Contents

List of Figures

List of Tables

Relevant Publications by the Author

- Dumayne J., Hook I. M., Williams S. C., Lowes G. A., Head D., Fritz A., Graur O., Holwerda B., Humphrey A., Milligan A., Nicholl M., Roukema B. F., Wiseman P. Using the next generation of telescopes to refine the measurement of supernova host galaxy masses. RAS Techniques and Instruments. Volume 2, Issue 1, Pages 453-369. 2023.

Chapter 1

Introduction

One of the large scale goals of science is to understand the Universe and the physical processes which govern it. This understanding comes from studying the observable Universe. New instruments and facilities are able to push the edge of the observable Universe further, whilst new methods are able to develop a deeper understanding of that which can already be observed. In this thesis we aim to quantify the improvements to the measurement of supernova (SN) host-galaxy properties achievable by the the 4-metre Multi-Object Spectroscopic Telescope (4MOST). We then apply these measurements to quantify the improvements 4MOST will enable to SN classification. These areas are two aspects that impact the use of Type Ia supernovae as a tool to measure cosmological parameters. The longer-term driver behind this work is ultimately to reduce uncertainty on cosmological parameters and hence obtain a better understanding of dark energy.

The following sections introduce our current understanding of the Universe and expected developments in the near future, both of which are key motivations to the work in this thesis. We then finish this section with an overview of machine learning, which we implement in our research.

1.1 Background cosmology

Distant galaxies were seen to be moving away from Earth in the early 20th century (Hubble, 1929). This was observed by seeing the spectral lines of galaxies shifting to redder wavelengths, as the galaxies moved away, observed as a Doppler shift. To measure the rate at which the galaxies are moving away we use redshift. Redshift, z, is the measured Doppler shift of an object's wavelength. It is calculated using the following formula:

$$1 + z = \frac{\lambda_{obs}}{\lambda_{em}} \tag{1.1}$$

where λ_{obs} is the observed wavelength and λ_{em} is the emitted wavelength. When we observe objects we are seeing the light which left that object. Due to the finite speed of light, when we observe objects at higher redshifts we're looking at younger objects. It is through the study of high redshift objects that we can understand the mechanics of how they form.

Currently, the best model which fits the observations of the Universe is the Λ Cold Dark Matter (ΛCDM) cosmological model (e.g. Peebles 1968; Carroll 2001; Peebles & Ratra 2003). The ΛCDM model has a series of parameters associated with it, with the simplest model using only 6. Some commonly used properties are: baryon density, dark matter density, dark energy density, age of the Universe, scalar spectral index, curvature fluctuation amplitude and reionization optical depth. We will cover the most important parameters and parameters which are most relevant to the work of this thesis below.

As the Universe has been observed in more detail we have found that it is expanding (Friedman, 1922). The rate of expansion is called Hubble's constant, H_0. To understand the expansion of the Universe, accurate distance measurements need to be taken. In order to measure distances across vast distances standard candles can be used. A standard candle is an object with a known luminosity, L. The luminosity distance, d_L can be defined as follows

$$d_L = \left(\frac{L}{4\pi f}\right)^{\frac{1}{2}} \tag{1.2}$$

where f is the measured flux. There are a variety of objects which can be used as standard candles. Cepheid variable stars are a reliable standard candle used as distance indicators (see Feast 1999 for a review). The luminosity of a Cepheid can be found through a relationship between period and luminosity (Leavitt & Pickering, 1912). However, Cepheid variable stars can only be used to measure distances within the local Universe. At distances beyond 20 Mpc it is difficult to separate Cepheids from nearby stars with current telescopes (Freedman et al., 2019). Another useful cosmological probe are Type Ia Supernovae (SNe Ia) as they act as standard candles (e.g., Branch & Tammann 1992; Wood-Vasey et al. 2008; Dhawan et al. 2022a). SNe Ia are useful as they allow the luminosity distance to be calculated to an accuracy of $\sim$7% (Phillips et al., 1999) at greater distances than Cepheid variable stars. We can use SNe Ia to construct a Hubble diagram, similar to Hubble's galaxy recession diagrams (Hubble, 1929). The Hubble diagram is a plot of a distance modulus $\mu = 5log_{10}\frac{d_L}{10pc}$, against redshift.

Previously, we introduced the luminosity distance. As we calculate the luminosity distance to objects further and further away, the cosmological parameters will have a larger effect on the calculated value. Since we can observe SNe Ia at large distances, we can use the luminosity distances and redshifts of a sample of SNe Ia to constrain cosmological parameters (Riess et al., 2004a; Riess et al., 2007; Kowalski et al., 2008; Brout et al., 2022). The intercept of the Hubble diagram gives the Hubble constant (H_0), which is the rate of expansion of the Universe. A Hubble diagram created using SNe Ia can be seen in Figure 1.1. When SNe Ia are used as standard candles a value for H_0 is found of $73.6 \pm 1.1 kms^{-1}Mpc^{-1}$ (Brout et al., 2022), assuming a flat ΛCDM model. We explain the concept of a flat ΛCDM model in the next paragraphs. To effectively use SNe Ia as standard candles the light curves need to first be standardised. We discuss the standardisation of SNe Ia in Section 1.1.3.

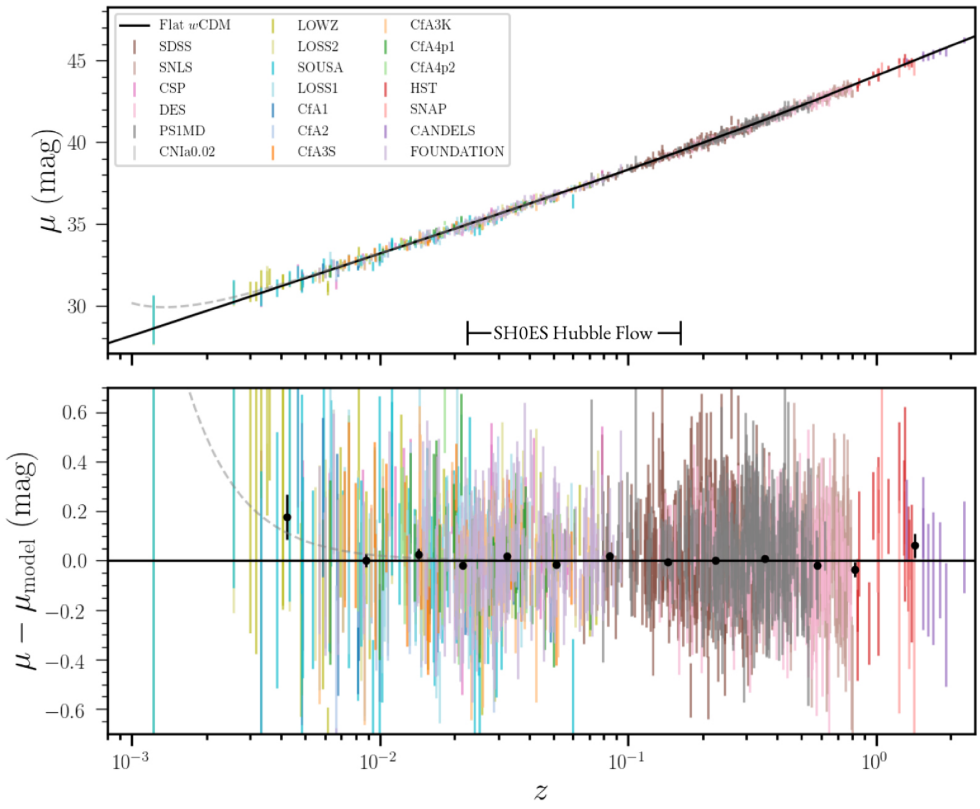

Figure 1.1: Top panel: A Hubble diagram taken from Brout et al. 2022. The Hubble diagram is produced by plotting a standardised distance modulus vs redshift. Bottom panel: distance modulus residuals relative to a best-fit cosmological model.

The Universe's current density parameter, Ω_0, is a very important parameter. It can be defined using the following equation:

$$\Omega_0 = \frac{\rho}{\rho_c} \tag{1.3}$$

where ρ is the total density of the Universe and ρ_c is the critical density (the density required for the Universe to halt its expansion). As ρ is the sum of everything in the Universe, Ω_0 can be split into multiple parameters each representing a different component. The subset parameters include: Baryon density parameter (Ω_b), Dark matter density parameter (Ω_c), Matter density parameter (Ω_m) and Dark energy density parameter (Ω_{DE}). Some of these parameters can be used in tandem to calculate others, for example $\Omega_c \approx \Omega_m - \Omega_b$. We first explore the implications of different value of Ω_0 below, then we cover Ω_m and Ω_{DE} in more detail.

The precise value of Ω_0 is important as its value can lead to 3 different models of the Universe. If $\Omega_0 > 1$ then the Universe is closed. The Universe's density would be great enough to overcome the force of expansion and the Universe will eventually stop expanding, collapsing in on itself. Space-time would be folded back on itself to be a finite size with no center or edge, and would appear to be shaped like a sphere. If $\Omega_0 < 1$ then the Universe is referred to as open. The density of the Universe is low and would be unable to stop its expansion. An open Universe would appear saddle shaped. Finally, if $\Omega_0 = 1$ then the Universe is flat, with the shape being similar to a piece of paper. The Universe would have enough matter to stop the expansion, but not so much that it would then collapse in on itself. Most widely accepted measurements give Ω_0 to be close to 1 (within uncertainties), implying that the Universe is flat.

The energy density of all forms of matter is represented by Ω_m. There are several ways to estimate matter density, all of which come from surveys (see Freedman 1998 for a review of Ω_m). One method is to compare the apparent magnitude of low and high-redshift SNe Ia (see Riess et al. 1998 for an overview of this method). Luminosity distances derived from a sample of SNe Ia can be used to constrain Ω_m (e.g. Riess et al. 2004b). The most recent measurement of Ω_m found a value of 0.298 ± 0.022, this value was measured using only SNe and assumed a flat ΛCDM

model of the Universe (Scolnic et al., 2018).

Ω_{DE} represents the fraction of the Universe attributed to dark energy. The current value is 0.679 ± 0.013 (Planck Collaboration et al., 2020), meaning 68 % of the Universe is dark energy.

The equation of state of dark energy is defined as:

$$p = w\rho c^2 \tag{1.4}$$

where ρ is the uniform mass density, p is the uniform pressure, c is the speed of light, and w is a constant referred to as the dark energy equation of state parameter. In some models, a time component is added as:

$$w = w_0 + (1 - a)w_a \tag{1.5}$$

where $a = (1 + z)^{-1}$ which represents the scale factor of the universe, w_0 is the current value of the equation of state parameter, w_a is a parameter for the evolution over time (Chevallier & Polarski, 2001; Linder, 2003).

w can be any value, however there are 3 values that produce interesting situations. If $w = 0$ then the pressure becomes 0, which implies that the Universe is dominated by matter. This would also mean that the energy of the Universe is conserved as it expands. If w is $\frac{1}{3}$ then the Universe is radiation dominated. Energy would be lost as the Universe expands. From observations we have seen that the Universe is expanding at an accelerated rate (Riess et al., 1998; Perlmutter et al., 1999). Therefore we expect w to be negative. The simplest explanation is a cosmological constant, and if dark energy is in the form of a cosmological constant then w is exactly -1 everywhere and at all times. Recent work in cosmology has focused on measuring w to understand if w is a constant value (e.g., Garnavich et al. 1998; Scolnic et al. 2018; Chen et al. 2022).

The ΛCDM model seems to be a very good fit to the data. However, assuming that $w = -1$ raises a concern as it appears as a vacuum energy density with density $\approx 10^{-47} GeV^4$. This value disagrees with the standard model of particle

physics which predicts a density $\approx 10^{74} GeV^4$. The disagreement between these two values is a difference of 120 orders of magnitude. One solution to solving this issue is to consider alternative models of dark energy, such as the quintessence model which assumes that dark energy is dynamic and changes over time (Ratra & Peebles, 1988; Caldwell et al., 1998).

The ΛCDM model is not perfect. There still remain many questions it cannot answer, for example the exact nature of dark matter or if dark energy is a constant value. It is only through further observations of the Universe, and developing the standard model that our understanding of the Universe will progress.

1.1.1 Current state of cosmology

We have discussed each of the important and most relevant cosmological parameters in the previous Section. In this section we shall discuss recent developments in cosmology, as well as any disagreements or points of contention.

The biggest disagreement in cosmology is over the specific value of the Hubble constant. This is referred to as the Hubble tension. The original measurements of the Hubble constant had very large uncertainties associated with it. However, with advances in techniques and facilities this uncertainty became smaller and smaller. This has lead to an issue in the early 21st century where different methods for measuring H_0 found different values. The Hubble tension arises as the measurements of H_0 from the local universe are higher than the measurements made using the CMB. The most recent calculated value using the local universe is $73.6 \pm 1.1 kms^{-1} Mpc^{-1}$ (Brout et al., 2022), whilst the most recent CMB calculated value is $67.4 \pm 0.5 kms^{-1} Mpc^{-1}$ (Planck Collaboration et al., 2020).

Researchers wish to understand if w has a value of -1 everywhere in space, for all of time. If w changes for different times or at different locations in the Universe then dark energy cannot be a cosmological constant. There have been several large surveys which have calculated the cosmological parameters using SNe Ia. The Supernova Legacy Survey (SNLS) collected spectra of approximately 500 SNe Ia in the early 2000s (e.g. Balland et al. 2009; Sullivan & Balland 2008). Using the 3-year data a value of w was found to be $-0.91^{+0.16}_{-0.20}$ (Conley et al., 2011). The SNLS value for w is derived using only SNe Ia and makes the assumptions that the

Universe is homogeneous on large scales and general relativity is correct. Another survey which ran in the early 2000s was the Sloan Digital Sky Survey (SDSS), it has collected 1.5 million galaxy spectra (e.g. York et al. 2000; Alam et al. 2015). It also collected 484 spectra of SNe (Frieman et al., 2008; Sako et al., 2008). A recent publication using 20-years of observations found a w value of -0.909 ± 0.081, assuming a flat ΛCDM (Alam et al., 2021). SDSS will continue to observe, after recently being upgraded to a robotic fiber positioner array (Grossen et al., 2020; Sayres et al., 2021). There are expected to be future publications using SDSS data, which will calculate the cosmological parameters with more observations.

A final survey we wish to cover is the Dark Energy Survey (DES), an optical-near infrared survey (The Dark Energy Survey Collaboration, 2005). Several different sub-surveys were carried out, but the survey of interest to our work is the Supernova survey. This survey utilised 10% of the total DES time to observe SNe Ia, which as of the 3-year data release had observed 207 SNe Ia [1]. The 3-year results of DES and the Planck Cosmic Microwave Background (CMB) observations were combined to find a value of w of -0.978 ± 0.059, for a flat wCDM model (Abbott et al., 2019). The DES survey finished collecting observations in 2019. However, the final analysis containing all of the observations has yet to be published.

The Supernovae and H_0 for the Equation of State of dark energy (SH0ES) team has been attempting to break the degeneracy among cosmological parameters, by using SNe Ia and Cepheid variable stars (Macri & Riess, 2009). Whilst the research aims to investigate the impact w has on H_0, the most recent publication by Riess et al. focused on q_0, the deceleration parameter. Future publications from the SH0ES team could have an impact on cosmological parameter measurements.

The most recent calculation of w using SNe Ia, uses the Pantheon+ data set, found a value of $-0.978^{+0.024}_{-0.031}$ (Brout et al., 2022). The Pantheon+ dataset represents the largest dataset of SNe Ia to date, but only contains SNe at $z \leq 2.3$. As new surveys are carried out, and we observe more SNe Ia it is expected we will be able to understand if w is the same in all parts of the Universe. It is hoped that more observations will explain the disagreement with the standard particle

[1] https://des.ncsa.illinois.edu/releases/sn

physics model.

A collection of optical light curves, redshifts and classifications for 1058 SNe Ia makeup the Pantheon dataset. Using Pantheon with the Planck CMB constraints, a w value of -1.026 ± 0.041 was found (Scolnic et al., 2018). Pantheon was later expanded upon by combining multiple SNe Ia surveys together to create Pantheon+. Pantheon+ contains 1550 SNe Ia (Scolnic et al., 2022). The value of w found using Pantheon+ and Planck is $-0.982^{0.022}_{0.038}$ (Brout et al., 2022). The Pantheon and Pantheon+ results assume a flat wCDM model of the universe. Pantheon+ is expected to be replaced by other surveys which will discover significantly more SNe Ia, such as: the Zwicky Transient Facility (Dhawan et al., 2022b), the Legacy Survey of Space and Time (LSST) (Ivezić et al., 2019) and the Nancy Grace Roman Space Telescope (Hounsell et al., 2018).

There are many upcoming facilities which aim to measure cosmological values to a higher precision. We discuss in Chapter 2 those which will have the biggest impact on SN cosmology, such as: the Vera Rubin Observatory, the 4-metre Multi-Object Spectroscopic Telescope (4MOST), the Euclid satellite and the Nancy Grace Roman space telescope. There will also be new facilities for CMB cosmology. These next-generation CMB facilities will be able to measure H_0 and w more precisely. The measurement of H_0 will hopefully clear up the disagreement with the SNe Ia derived values. An improved measurement of w using the CMB can be combined with SNe measurements of w to produce a more precise value. The Planck satellite produced its final data release in 2018. There are other facilities which are close to completion of construction. The LiteBird satellite will launch in the late 2020s, and aims to improve upon Planck by studying the polarisations of the CMB with a higher degree of precision (Hazumi et al., 2020). The Simons Observatory is a ground based observatory due to finish in 2024 [1]. The Simons Observatory will have ten times the sensitivity and five times the angular resolution of Planck [2]. Simulations of objects observed by the Simons Observatory predict that a value of H_0 will be measured with an uncertainty of $0.3 kms^{-1}Mpc^{-1}$ (Ade et al., 2019). A secondary science goal of the Simons Observatory aims to measure the dark energy of state, although simulations have not yet been carried

[1] https://simonsobservatory.org/doing-business-with-so/
[2] https://simonsobservatory.org/science-goals/

out to find the expected effect on w to date.

There are several limits on the cosmology achievable with SNe Ia. The biggest limiting factor is the number of observed SNe Ia to date. This limit should be drastically reduced with upcoming surveys such as LSST and 4MOST. Another limit is that accurate redshift for each SNe Ia needs to be collected (Barris & Tonry, 2004). The most accurate way to find SN redshift is to use spectroscopy. However, spectroscopy is expensive meaning only limited observations can be carried out. Our work in Section 4 tests the possibility of using spectroscopic observations of the host-galaxy, which is easier to carry out, rather than observations of the SN event.

We have covered here the most recent developments in cosmology which are relevant to the work in this thesis. There are other methodologies which can be used to measure the cosmological parameters. These include using: gravitational lensing (Grillo et al., 2008), baryon acoustic oscillations (Addison et al., 2013; Aubourg et al., 2015) and strong gravitational waves as standard sirens (Holz & Hughes, 2005). Future results from each of the mentioned research areas need to be carefully monitored for any improvement which could impact or improve SN cosmology.

1.1.2 Supernova types

We have previously discussed the use of SNe Ia as standard candles. However there are many types of SNe. It is important to understand the differences between SNe to be able to classify them correctly. Misclassified SNe can affect cosmological calculations being carried out with a dataset. The work of this thesis is concerned with trying to reduce the contamination of a SNe Ia sample, as even a small contamination from other SN types can affect the measurement of cosmological parameters (e.g. Riess et al. 2004a; Barris & Tonry 2004; Homeier 2005). We will briefly discuss the differences between different SNe in this section.

There are several distinct categories of supernovae, categorized primarily by the triggering mechanism behind the supernova event. SNe Ia occur when a white dwarf acquires sufficient mass from a companion star to exceed the Chandrasekhar mass limit, 1.4 $M_{\odot}$ (e.g. Chandrasekhar 1931; Hoyle & Fowler 1960; Branch et al.

1995; Yungelson & Livio 1998). This additional mass can be acquired by merging with another white dwarf or collecting accreting gas from a companion star. However, no progenitor of a SN Ia has been clearly observed (see Maoz et al. 2014 for an overview). Type Ib, Ic, and II SNe form when a star's core collapses due to depletion of nuclear fuel. Core collapse SNe leave either a neutron star or a black hole as a remnant (e.g., Bethe et al. 1979; Arnett et al. 1989). SNe can be distinguished from one another based on their spectral lines (Filippenko et al., 1993). Type Ia SNe can be identified due to a silicon absorption line in their spectra (Filippenko, 1997). Type II SNe exhibit hydrogen lines in their spectra (Minkowski, 1941). Type Ib SNe display helium lines in their spectra (Harkness et al., 1987), whilst Type Ic SNe have neither hydrogen nor helium lines in their spectra (Wheeler & Harkness, 1986). These spectral traits reflect the composition of the star's outer layers during the supernova explosion. The outer hydrogen-rich envelope remains intact in Type II SNe, whilst Type Ib SNe lose their hydrogen envelope. Type Ic SNe shed both hydrogen and helium layers (Modjaz et al., 2014a).

The vast majority of SN observations are carried out using photometric facilities, due to spectroscopy being expensive. These photometric facilities will take a series of observations of a SN over different dates, often taken through different filters. The sequential data can then be plotted, to show the evolution of the output light of the object over time. The differences between all SN types are reflected in their light curves. A comparison of the different SN light curves can be seen in Figure 1.2.

Type II SNe can be categorised further into IIb IIP, IIL and IIn. SNe IIP and IIL are classified based on the shape of their light curves, after the initial peak. SNe IIL show a consistent linear decline after the peak, whilst SNe IIP light curves plateau during the decline (see Barbon et al. 1979 for the separation of these two sub-classes). SNe IIn are distinguished from the other type II classes by the presence of narrow hydrogen emission lines in the spectra (Schlegel, 1990). SNe IIb have a weak hydrogen line in the initial spectrum, which becomes undetectable later in the SN's life (Filippenko et al., 1993).

The types of SNe mentioned are the most common, but they are not the only types of SNe. Other categories include: Calcium-rich transient, peculiar SNIax, peculiar type Ia-91bg, Super Luminous supernovae, as well as the transient event

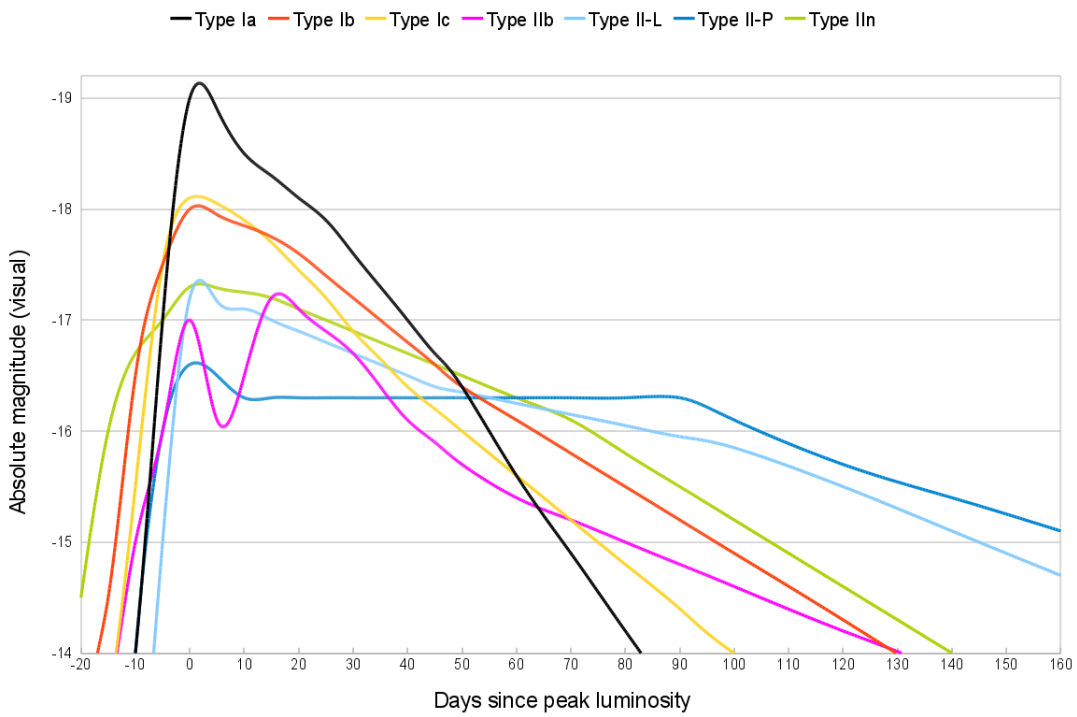

Figure 1.2: A comparison of the light curves of the major SN types, showing how the emitted light evolves over time. The x-axis has a zero value for each SN at the peak magnitude. This plot was created for Wikipedia by user 'Lithopsian' from data provided by Karttunen (2007); Modjaz et al. (2014b); Nyholm et al. (2020).

Tidal Disruption Events. To date these additional classes have been observed in small quantities. More of these objects will be observed with future surveys. All non-Ia SNe are potential contaminants in SN Ia samples.

1.1.3 Standardising Type Ia SNe

SNe Ia as standard candles have a dispersion of approximately 0.8 magnitudes at peak luminosity (Phillips, 1993; Riess et al., 1996). To reduce the dispersion, which would make SNe Ia more effective standard candles, and allow the accurate calculation of cosmological parameters SNe Ia need to be standardised. The standardisation then allows a standardised distance modulus to be calculated. A popular standardisation uses the SALT2 SED model (Guy et al., 2007, 2010). This model uses several parameters: a stretch parameter x_1, a colour parameter c and host-galaxy stellar mass M_*. To calculate the standardised distance modulus (μ_B), from the SN's apparent magnitude, m_B (at peak, in the B filter band):

$$\mu_B = m_B - M + \alpha x_1 + \beta c + \gamma M_* \tag{1.6}$$

α, β and γ are correction terms for stretch, colour and stellar mass respectively and M is the absolute magnitude. There is a correlation between SNe Ia luminosity and light-curve width (Phillips, 1993). Generally, more luminous SNe Ia have broader light curves, whilst fainter SNe Ia have narrower light curves. Stretch is accounted for using the α correction term. The next standardisation relationship is between the luminosity of a SN object and the SN B-V colour at the peak luminosity. It has been found that bluer SNe are brighter than red SNe (Riess et al., 1996). The correction for colour is accounted for using the β term. These correlations were first observed from a small sample of SNe Ia. Larger SN Ia samples have further proven these standardisation relationships, and allowed the inference of SN light-curve parameters and measure distances μ.

The correction for light-curve width and colour are well established (e.g. Tripp 1998). However the correction based on stellar mass is a recent development. It has been found that host-galaxy properties correlate with the peak magnitude of

SNe Ia (e.g., Kelly et al. 2010; Lampeitl et al. 2010; Sullivan et al. 2010), this is often called the "mass step", and corrections need to be made in order to avoid biases in measurements of cosmological parameters. The mass step is observed when the Hubble residual (the difference between the distance modulus to the SN and the predicted value by a cosmological model at the SN's redshift (e.g., Jha et al. 2007; Gallagher et al. 2008; Kelly et al. 2010) is plotted as a function of the host-galaxy stellar mass (M_*). Observing the Hubble residual as a function of host-galaxy M_* shows that SNe Ia in high mass galaxies ($> 10^{10} M_\odot$) are brighter than SNe Ia in low mass galaxies ($< 10^{10} M_\odot$), after correction for stretch and colour (Childress et al., 2013). Empirical evidence has shown the mass step appears at approximately $10^{10} M_\odot$ (Sullivan et al., 2010; Uddin et al., 2017). Therefore a correction can be applied as a step function, depending on which side of $10^{10} M_\odot$ the host-galaxy's M_* falls.

To date no significant correlation has been found between the width of a SN light curve and the SN colour at peak. However, a correlation between SN stretch and the stellar mass of the host galaxy has been observed (Riess et al., 1996; Hamuy et al., 2000; Sullivan et al., 2010; Smith et al., 2020; Wiseman et al., 2021). Additionally, a dependency has been observed between the colour correction and host-galaxy stellar mass (Sullivan et al., 2010; González-Gaitán et al., 2021).

To improve the ability of SNe Ia to act as standard candles new SNe Ia corrections are being tested, as stretch, colour and host-galaxy M_* are not the only property that can lead to an improvement. Gallagher et al. (2008) found a correlation between the Hubble residual and metal abundance. Additionally, Wolf et al. (2016) and Rigault et al. (2020) found a correlation between Hubble residual and specific star formation rate. Briday et al. (2022) and Wiseman et al. (2023) demonstrate stellar-population age is the galaxy parameter that drives the observed step. Metallicity, star-formation rate and host-galaxy mass are linked, so these results are to be expected (Speagle et al., 2014; Li et al., 2022). Another method to improve the standardisation of SNe Ia is to understand the cause of the mass-step. Brout & Scolnic (2021) report that the mass step can be explained by introducing a new SN colour model, by modelling different dust distributions. Chung et al. (2023) suggest that the mass step attributed to a non-linear relationship between mass and age, reflecting the difference between old and young

galaxies. Many of these suggested causes of the mass step can be measured by fitting a combination of photometric and spectroscopic data of the host galaxies (Jones et al., 2022; Lower et al., 2020).

A new technique, which is still in its infancy, was recently presented. Brout & Scolnic (2021); Boone et al. (2021a,b) presented a technique which made use of identifying SNe twins, SNe whose optical spectra are similar at peak luminosity. This method appears to reduce SN Ia dispersion to less than 0.06-0.07 magnitudes. This novel method has the potential to significantly reduce the systematic uncertainty on distances calculated from SNe Ia.

1.1.4 Systematic uncertainties

Systematic uncertainties make up a large portion of the uncertainty associated with w. Below the key sources of the systematic uncertainty are discussed.

Photometry and calibration. Accurate photometric calibration is challenging for SN cosmology. A survey will calibrate the data to the specific instrument first, and then correct to a generalized standard system Bohlin et al. (2014); Bohlin et al. (2020).

Light-curve modelling. In order to standardise SNe Ia, light-curve fitting techniques are used. No technique is perfect and each presents its own issues. The most widely used technique is SALT 2 (Guy et al., 2007, 2010). Additionally, light-curve fitting techniques are trained on SN data samples which could be biased, or affected by calibration and instrument uncertainties.

Bias corrections. To account for biases, bias corrections can be estimated from large simulations. However, simulations can introduce new uncertainties. The simulations themselves may carry out incomplete modelling of SN Ia intrinsic scatter and stretch/colour populations, as well as be unable to simulate unknown dependencies of SN Ia property. These issues increase the systematic uncertainty on w. However upcoming large surveys, such as 4MOST, will aim to reduce the bias correction issue by observing greater number of SNe.

Our galaxy. As we observe SNe by looking out through the Milky Way galaxy, we need to correct for extinction present within the Milky Way. This can be corrected for using extinction maps Schlegel et al. (1998); Schlafly et al. (2010);

Schlafly & Finkbeiner (2011), but this correction will also add additional uncertainty.

Peculiar velocities. Spectroscopic redshifts are measure with a high level of precision. However, an uncertainty on the value can be caused by the peculiar velocity of the host galaxy. The peculiar velocity can be estimated and corrected using matter density maps and estimates of the growth rate of structures. This correction will also introduce an uncertainty.

Contamination. Samples of SNe Ia are not always completely pure. Most SN Ia cosmology uses spectroscopically classified SNe Ia, where contamination from other SN types is relatively small. However, when using photometrically classified SNe the rate of contamination is likely to be higher. This is issue is the focus of our work in Chapter 4.

1.1.5 SN rates of appearance within galaxies

The work carried out in this thesis uses host-galaxy properties to improve the classification of SNe, to produce purer SNe Ia data sets. It has previously been found that there are correlations between SN classes and host galaxy properties (e.g. Hamuy et al. 2001; Galbany et al. 2014; Gagliano et al. 2021) which can be used to aid the classification of SNe. SNe Ia appear more frequently in star-forming galaxies than passive galaxies (e.g. Oemler & Tinsley 1979; Botticella et al. 2017), with 10 times as many SNe Ia appearing in strongly star-forming galaxies compared with passive galaxies (Sullivan et al., 2006). The work of Graur et al. (2017a) found that SNe Ia rates anti-correlate with the host-galaxy mass. Their follow up work confirmed that SNe Ia are more common in low-mass galaxies (Graur et al., 2017b). Type Ia, Ib/c and II SNe are more common in late-stage spiral galaxies than early-stage spiral galaxies (Mannucci et al., 2005). Type Ic SNe have host-galaxies with high specific star formation rates and low metallicities (Modjaz et al., 2020). Therefore by measuring the properties of host-galaxies such as star formation rate, we hope to be able to improve the classification of transients from LSST even when there is no transient spectrum available.

1.1.6 Measuring galaxy properties

In order to measure the properties of galaxies in this thesis we make use of Spectral Energy Distribution (SED) fitting. It is important to understand SED fitting as we use it extensively in Chapter 3. SED fitting allows the calculation of properties from a galaxy's photometric light curve or spectrum, due to the physical properties shaping the output light (e.g. Spinrad 1972). To begin with, a simple stellar population model is used. The stellar population shows the evolution in time of the SED. A simple stellar population requires: a stellar evolution theory, stellar spectral libraries and an Initial Mass Function (IMF). The stellar evolution theory is provided by a star formation history model. Although the specific model needs to be chosen carefully as it can introduce biases (de Barros et al., 2014). Stellar spectral libraries convert the output of stellar evolution (which make use of effective temperatures and gravities as a function of metallicity) into observable SEDs. There is not a single stellar spectral library which covers all of the parameter space, so multiple libraries need to be utilised and combined depending on the work being carried out. Finally, the IMF is the initial distribution of stellar masses along the main sequence. A review of IMFs by Bastian et al. (2010) showed that the IMF does not vary and a standard model can be used.

There are many properties that can be measured using SED fitting. We focus on: age, dust, metallicity, stellar mass and star formation rate. Beginning with stellar mass, this can be estimated using the stellar mass-to-light ratio. There are three ways to do this, using relations between colour and stellar mass-to-light ratio (Bell & de Jong, 2001; Bruzual & Charlot, 2003), modeling broadband photometry (Giallongo et al., 1998; Sawicki et al., 1999) and modeling moderate resolution spectra (e.g. Kauffmann et al. 2003). The stellar mass-to-light ratio is then multiplied by the observed luminosity. One problem with the estimation of galaxy masses is that young stars can outshine older stars (Papovich et al., 2001).

Dust is an important factor when observing a galaxy. It both obscures light in the UV through to the near-infrared and emits light in the infrared. Dust grains obscure light by absorbing and scattering light. The obscuring of light is referred to as attenuation or extinction interchangeably in the literature. However there is a subtle difference. Attenuation includes effects from the distribution of stars and

dust, where as extinction only measures total loss of light (Salim & Narayanan, 2020). Attenuation and extinction are measured by obtaining a spectrum of a galaxy and compared with the expected spectrum of the galaxy without dust (Conroy, 2013). The emitted light from dust can be calculated by utilising one of a variety of models, which have been created that combine grain size distribution, grain optical properties and models of starlight to calculate the infrared emission (examples include Desert et al. 1990; Silva et al. 1998; Groves et al. 2008).

The stellar ages of a galaxy can be measured using spectral features. A common procedure to measure age is to fit a single-age simple stellar population to the spectra. In this method, the hydrogen Balmer lines provide the main constraint on the age (Conroy, 2013).

The metallicity has an affect on the SED of a galaxy. If the effective temperature of the galaxy is fixed, a higher metallicity causes generally redder colours. However, the age of a galaxy can also cause the SED to appear redder. Therefore there is a degeneracy between age and metallicity (e.g. Worthey 1994). Worthey found that the degeneracy is less pronounced at low metallicities. They also found hydrogen Balmer lines were more sensitive to age than metallicity. Other work has found that Fe4668 and Fe5270 spectral lines are more sensitive to metallicity, and thus can break the degeneracy for spectroscopic SED fitting (see O'Connell 1980; Rabin 1982 for early attempts to break the degeneracy). For photometric observations, it has been shown that near-infrared observations can break the degeneracy (e.g. Bell & de Jong 2000; MacArthur et al. 2003; chul Lee et al. 2007).

The Star Formation Rate (SFR) of a galaxy can be difficult to estimate from SED fitting. The choice of dust model and Star Formation History library imposes biases on the calculated age and SFR (Papovich et al., 2001). There is also a degeneracy between age-dust-metallicity, which makes it hard to determine age and thus SFR. A common practice to measure the SFR is to make use of a relationship between SFR and Hydrogen-α emission line. Hα is emitted by nebulae, so the larger the line the more star formation taking place (see Kennicutt 1998 for a review of SFR measurement techniques). In galaxies with low levels of star formation indicators can be contaminated by other factors, such as AGN activity. An accurate solution to measure SFR has not yet been found. A simple solution often used is to separate star forming and quiescent galaxies from one another using a

UV-optical colour boundary (e.g. Jeong et al. 2009). The specific star formation rate can also be measured, which is a measurement of the star formation per unit mass. Specific star formation rate is calculated by dividing the SFR by the stellar mass.

Chapter 2

Next Generation Facilities

In this chapter we first discuss data science, machine learning and specific models used for classifying SNe. Then we discuss upcoming instruments and facilities that will further the research of SNe Ia cosmology. In this thesis we use simulated observations from the Vera Rubin observatory and the 4-metre Multi-Object Spectroscopic Telescope. We discuss surveys carried out by these two facilities, upcoming satellites relevant to our research, and for each show their expected impact on cosmology.

2.1 The world of big data

Our society is in the information age. There is more data available than ever before. This is particularly true in the field of astronomy. Large sky surveys such as LSST carried out by the Vera Rubin Observatory will acquire 500 petabytes of data over the lifetime of the survey (Ivezić et al., 2019). It will observe 20 billion galaxies over the 10-year survey[1]. Similarly, 4MOST will collect 13 million spectra of galaxies [2]. With the sheer volume of data being taken, machine learning data processing techniques need to be implemented. In this thesis we use a deep learning recurrent neural network to classify a large number of SNe. In this Section we give a general

[1]https://www.lsst.org/scientists/keynumbers
[2]https://www.4most.eu/cms/science/exgalconsurv/

overview of machine learning, and cover key concepts. Then we will delve into the inner workings of neural networks, specifically focusing on convolutional neural networks and recurrent neural networks. These two types of neural networks are widely employed for the classification of SN events.

2.1.1 Introduction to machine learning

The broad definition of machine learning is any computer program that can make decisions on its own. The benefit of using machine learning (ML) is that a machine learning model is able to see connections within the data that a human cannot (see Huppenkothen et al. 2023 for best practices on the implementation of ML models in astronomy). A machine learning model will be given data to train on, with the columns of a data table often being referred to as 'features'. A model will be trained to recognise patterns within the features. Most machine learning training consists of minimising a loss function on a training dataset of examples. The model will then be able to apply what it has learnt to a new data set, called the 'test data'. A third data set can be created called the 'validation data'. The validation data is used to test the model, after it has been trained, to check for overfitting. If there is overfitting the model is altered before being presented with the test data.

There are many different types of machine learning models. Most machine learning models are referred to as 'supervised learning models'. This is due to the user providing labelled training data with features for the model. A subset of machine learning is deep learning. Deep learning consists of models which use neural networks. Neural networks consist of models which use unsupervised learning techniques. Unsupervised learning differs from supervised learning as the user does not provide data with features, instead the model will decide from the raw data which portions of the data set are relevant. We will explain briefly the most widely used models below. We cover neural networks in Section 2.1.2 due to their relevance to the work in this thesis.

The simplest form of machine learning is regression analysis (also known as linear regression) (Plackett, 1972). Regression analysis minimises a cost function to fit a line to a dataset. Gradient descent is a widely used technique for minimising

the cost function. It starts with random coefficients for the fitted line and adjusts them iteratively by a fixed amount, altering the gradient to reach the minimum cost value. The learning rate determines how fast the minimum is found; higher rates speed up the process but can lead to overshooting.

Another simple machine learning model is k-nearest neighbours (KNN) (Altman, 1992). A KNN model works by assigning k points within the data's parameter space. The distance from each data point to each k-point is then calculated, and all data close to a k-point is grouped together. The value of k is a hyperparameter which needs to be optimised for every data set. KNN models are most often used for classification, however they can be used for regression too. During classification a machine learning model will assign a probability that each object belongs to a class.

An additional model in use is Support-vector machines (SVMs) (Cortes & Vapnik, 1995). These work in a similar way to KNN models. Data is mapped to a high-dimensional feature space, allowing different categories to be separated. New data can then be assigned a class by using characteristics to assign each data point to a group. SVMs are useful for classification in high dimensional spaces, such as handwriting recognition. However, they can have issues of over fitting if there are too many dimensions.

A final machine learning model we will discuss is the decision tree model (Friedman, 2002). Its name comes from the way the model works through data in the same way a regular decision tree works, in a flowchart-like structure. The difference between a decision tree ML model and a regular decision tree is the ML model will decide what questions should be used to most effectively classify data. Decision tree ML models can be combined to improve performance, creating a new type of model called a random forest. Decision trees and random forests are most commonly used for classification.

Machine learning models have hyperparameters. These are values which dictate how the model learns and are different in every machine learning scenario. Hyperparameters need to be carefully tuned to prevent overfitting or underfitting of the model to the training data. If a model overfits to the data, this means that the model is matched too closely to the provided data and will struggle to adapt when the model is presented with new data. A model underfits when the model

does not understand the data well enough to carry out accurate analysis. An example of a hyperparameter is the learning rate of a model. The optimum learning rate can be found by observing the training and validation data. The number of epochs to train a model needs to be found, where an epoch is one pass through the entire training dataset. The optimum number can also be found from observing the training and validation data. The optimum number of epochs is reached when the validation curve reaches its minimum. Any additional epochs after this will result in the model overfitting to the training data, too few epochs and the model will underfit. Both of these situations will reduce the accuracy of a model. Another hyperparameter to consider is the batch size, which determines the sample size of training data utilised for updating a model's weights. A machine learning model uses weights to set the influence of input data on the classification value. An epoch may consist of multiple batches, depending on the dataset size and chosen batch size. The final hyperparameter we will cover is the split of data into training, testing and validation datasets. The size of the dataset will change the percentage which needs to be selected for each data set. A large dataset ($> 1,000,000$ of each class) will use a smaller percentage for training, to prevent overfitting. On the other hand, a small dataset (~ 100 of each class) will need a higher percentage of the data to train the model, to prevent underfitting. A final hyperparameter that is crucial for machine learning classification is the threshold at which an object is assigned to a class. The optimal threshold value is unique to every situation.

In order to evaluate the performance of a machine learning model a variety of success metrics have been created. Machine learning models can perform well at one task, and poorly at another. Success metrics need to be specific enough to highlight a model's strengths and weaknesses, to evaluate if a model is performing optimally for the desired task. Metrics are usually calculated using the true positive rate, false positive rate, true negative rate and false negative rate. In a 2 class classification, the true positive rate is the number of objects which have been correctly classified in the desired class. False positive rate is the number of objects which have been incorrectly classified as the desired class. True negative rate is the number of objects correctly classified in the second classification class. False negative rate is the number of objects which have been misclassified into the second classification class. A good all-round model will have high true positive

and true negative values, with low false positive and false negative values. All 4 values can be seen easily in a confusion matrix, which displays them in a grid. A common success metric is the receiver operating characteristic (ROC) curve. This is a plot of false positive rate against true positive rate. For a model which classifies between two classes, a diagonal line of $y = x$ represents the performance of the model if it classified objects at random. The further a model's ROC curve is above the diagonal line, the better the model's performance. As well as graphically showing the ROC, we can quantify the ROC curve of machine learning models by calculating the area under the curve (AUC). An AUC value of 1 represents a perfect model, whilst a value of 0 represents a model which misclassifies every object. This is only a brief summary of success metrics for evaluating machine learning models. In Section 4.1.3 we cover success metrics relevant to our work.

A hot topic in machine learning is ethics. The ethics of machine learning aims to understand when a machine learning model should and shouldn't be used, and what biases needed to be prevented. There have been examples of models showing racial or gender biases, which need to be prevented (e.g. Angwin et al. 2016; Buolamwini & Gebru 2018; Vilarino & Vicente 2021). The ethics of machine learning is applicable when a model uses data about humans or makes decisions for a human. As we are using astronomical data these ethical biases are not of concern to us. However, the machine learning model we used took steps to ensure our classes had equal sample sizes. The best way to handle the imbalanced classes in real-life situations is still being actively discussed (Johnson & Khoshgoftaar, 2019).

2.1.2 Introduction to deep learning with neural networks

Deep learning is the branch of machine learning which uses neural networks. Neural networks have recently been embraced by the machine learning community, due to increased computer processing power. There are many different types of neural network models, each optimised to a particular situation. An overview of several neural networks relevant to the work in this thesis is presented below. It should be noted that the research of neural networks is still in its infancy, with new models being developed every year.

First we give an overview of how a neural network operates. A neural network consists of nodes, and connections between each node. They are designed to work similar to how a brain works (Rosenblatt, 1960). An example of a neural network can be seen in Figure 2.1. Neural networks have input nodes, where data is passed into the neural network. The input data is passed to the hidden nodes, which then passes to the output nodes. With each connection between a node the data is multiplied by a weight and a bias is added. A weight is multiplied by the data value to convey the importance of the value. A bias is used to shift the activation function. All weight and bias values are determined by the neural network. After the data value being passed between the nodes has a weight and bias assigned, it is then passed into an activation function. The activation function is a way to control the calculated values within a network to be specified values. For example, the sigmoid activation function will transform all negative values into 0 and all positive values into 1. There are many different activation functions, each with a unique purpose. The correct activation function is a hyperparameter, as are the number of hidden nodes and layers. Some nodes will have multiple connections passing data into them. All connections will be summed, with each connection being multiplied by a weight. All weight and bias values are decided upon by the neural network. The values which are calculated at the output nodes become the prediction for each of the input data points. As more data points are input into the neural network, the network will start to map the input data. As more data points are presented to the neural network more accurate predictions can be made.

The initial weight and bias values are chosen at random, resulting in poorly optimised values. In order to optimise the weights and biases, backpropagation can be carried out (Werbos, 1974). For a data point weights and biases are calculated for each stage, as described previously. Then, the data received in the output node is compared with the data in the input node. If the two values are different then the weights and biases are adjusted by minimising the sum of squared residuals using gradient descent. The backpropagation process begins with the final weight and bias values. Backpropagation then works from right to left, working backwards through the neural network using gradient descent to adjust each weight and bias. A similar optimisation is forward propagation which is carried out after backpropagation, but works from left to right.

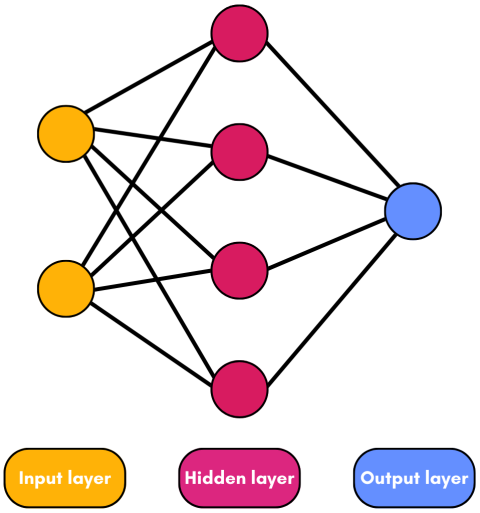

Figure 2.1: The connections within a neural network. Each of the circles represent a node, and the lines represent the connections between the nodes. The data is imported using the input layer. The weights and biases in the connection are applied as the data is passed between the hidden nodes. Finally, the data is passed to the output layer where a classification can be made. Neural networks are highly customisable depending on the situation they are to be used in. The neural network shown is a simple model.

A convolutional neural network (CNN) is able to recognise visual patterns (Fukushima, 1975), and is almost exclusively used to analyse images. To carry out image analysis, convolution layers are placed between the input layer and the hidden layer. The convolution layers break down images into components, allowing the neural network to build up an understanding of different types of images. The neural network can then recognise patterns, enabling it to recognise similar images and classify them.

A recurrent neural network (RNN) is a type of neural network that is able to analyse sequential data (see Brush 1967 for the history of RNNs). RNNs have a feedback loop which enables RNNs to use sequential data to make predictions. The first piece of data is passed through the network and the weights are saved. Then the model loops back and takes in the next piece of data. A second set of weights is calculated for the second piece of data. Then the two sets of weights are combined, and can be used to predict what the third data input will be. As more data is input into a RNN, the weights are recalculated and combined with what's been previously calculated and the model improves at predicting what will come next. An RNN can be trained to look for patterns, such as what a supernova light curve looks like. Then when a lightcurve is passed in, as the network passes through the observations across different nights it will be able to more accurately classify an object. RNNs are predominantly used to analyse videos, but can be used to analyse any sequential data with a time component.

There is an issue with RNNs that as they process large quantities of data the gradient within the gradient descent can become very large or very small, preventing gradient descent from working effectively. These situations are known as the exploding gradient and vanishing gradient problems (Hochreiter & Bengio, 2001). A basic RNN can be modified to fix these issues by using Long Short-Term Memory (LSTM) (Hochreiter & Schmidhuber, 1997). Instead of a simple loop that a basic RNN uses, a LSTM uses a long-term value and a short-term value. These values are interwoven to update each other with each iteration of the model, and prevent the gradient from becoming too small or too large. We shall explain this process below. A flowchart representation of how this process works can be seen in Figure 2.2.

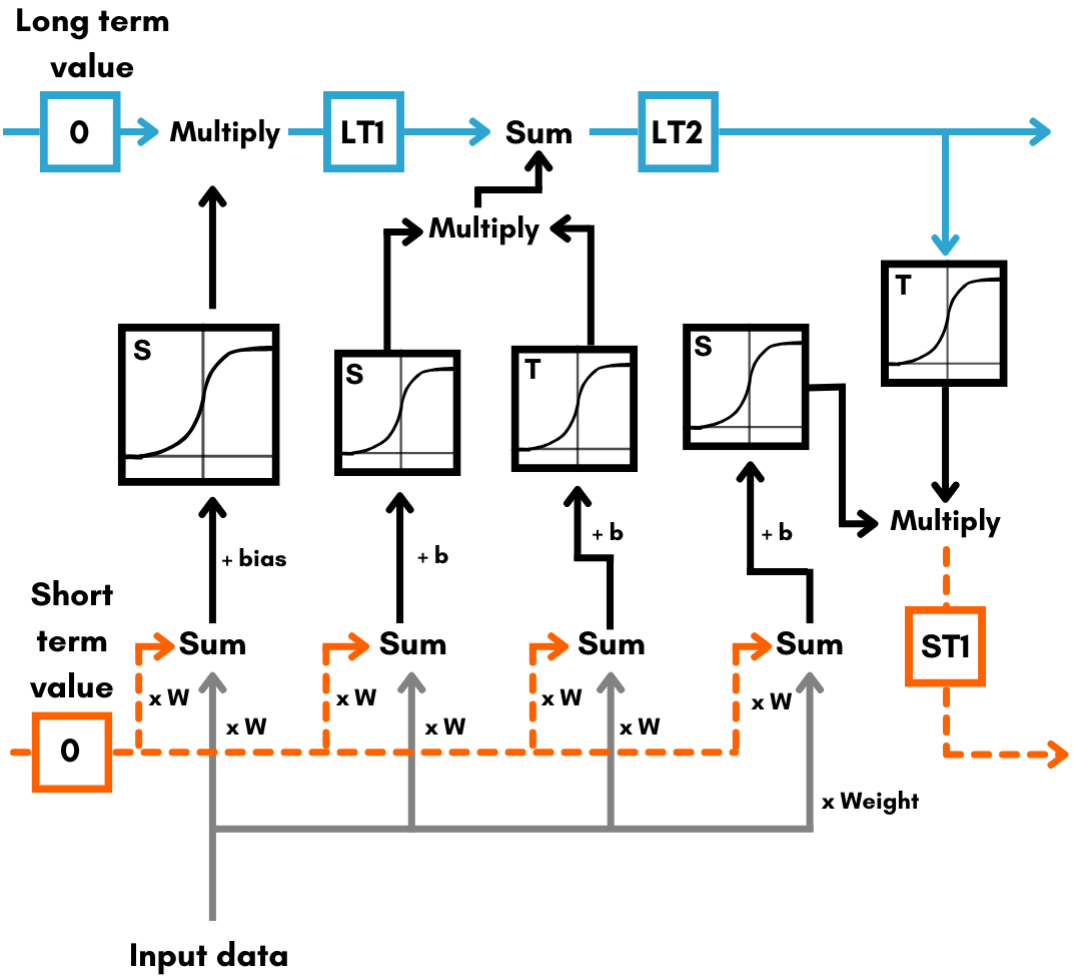

Figure 2.2: A flowchart of how the long-term (blue) and short-term (dashed orange) values are calculated. If this is the first data point then the initial short-term and long-term values will both be 0. The short-term value is multiplied by a weight (W), before being combined with the input data point. Every weight and bias (b) value is unique. The combined value is then passed through several activation functions before being used to calculate the long-term (LT2) and updated short-term value (ST1). Two different activation functions are used, a sigmoid function and tanh function, which are denoted with s and t respectively. The process shown is carried out for a single data point. For the next data point LT2 and ST1 will become the initial long-term and short-term values. The entire process will be repeated for each data point in the dataset. A detailed description of the process can be found in Section 2.1.2.

A short-term and long-term value is stored from any previous iterations of the RNN. If this is the first iteration of the model then the short-term and long-term values will both be set to 0. The current data input value is put through the network, to calculate the weights. The weights are then combined with weights calculated using the current short-term value. This combined output is input into a sigmoid activation function. The value produced by the sigmoid activation function is then multiplied by the current long-term value to create a new value (LT1). Then in parallel the input data is put through two different versions of the network. One version uses a sigmoid activation function, and the other uses a *tanh* activation function. The outputs from the activation functions are multiplied together, and then summed with LT1. This creates an updated long-term value (LT2). LT2 is used to calculate a new short-term value as follows. The current data point is input into the network to calculate weights, these weights are combined with weights calculated from the short-term value. The combined weights are then input into a sigmoid activation. In parallel, LT2 is input into a *tanh* activation function. The outputs from the activation functions are multiplied together, to create a new short-term value (ST1). ST1 and LT2 are both stored, and become the new starting long-term and short-term values when the next data point is input into the RNN. When the next next data point is input, the entire process to calculate the long-term and short-term values is repeated. After passing all of the data into the RNN, the long-term value will become the classification. The process of recalculating the short term and long term values ensures the full breadth of the data is captured, whilst preventing gradient descent from becoming too large or too small.

2.1.3 Machine learning classification of SNe

The volume of transient observations from upcoming surveys will make it impossible for each transient to be classified by people. Machine learning models will allow objects to be classified significantly faster. As mentioned previously, spectroscopy is expensive which means spectroscopic observations are limited to a small sample. As most observations are carried out photometrically, a lot of work has gone into developing classification tools which can classify SNe using only the photometric

light curves. Classification tools may be able to make use of the redshift of the object or its host-galaxy. The redshift can be from photometry or spectroscopy, depending on the observations carried out.

A number of machine learning classification models have been developed, each with their own strengths and weaknesses. The Real-time Automated Photometric Identification (RAPID) model was developed by Muthukrishna et al. (2019a). RAPID uses a recurrent neural network (RNN) to classify SN type based on early light curves. They achieve an AUC value of 0.95 for early light curves, and a value of 0.98 for full light curves. Similar to RAPID, SCONE uses a CNN to classify SNe from early photometric light curves, developed by Qu & Sako (2022). SCONE improves upon RAPID's ability to classify type Ibc and type II SNe, whilst RAPID is able to outperform SCONE at classifying early type Ia light curves. SCONE achieves an average AUC value 0.88 for all SNe types at the date of SN detection. The average AUC value rises to 0.97 50 days after the initial explosion. However, when redshift is included the AUC values increases. The average AUC value at the date of SN detection is > 0.9. The average AUC value increases to 0.975 50 days after the initial explosion. The Deep Automated Supernova and Host classifier (DASH) was developed by Muthukrishna et al. (2019b). It uses a convolutional neural network (CNN) to classify a SN type and calculate the age and redshift of the host galaxy from a spectrum. The authors of DASH found DASH to be significantly faster than SNID and Superfit. The latter two compare a spectrum to a catalogue of template spectra which decreases performance speed. DASH is able to classify most SN classes to a high degree of accuracy, with 10 of the 16 classes used for classification having $<= 10\%$ contamination from other SNe types.

One issue facing SN machine learning classifiers is the small sample sizes of certain SN types. SNID and Superfit have been found to classify SNe of types with small data sets, which also have a low signal-to-noise, as SNe Ia (Blondin & Tonry, 2007). To improve the performance of machine learning models, some studies have used data augmentation to improve performance (e.g. Revsbech et al. 2017; Boone 2019; Carrick et al. 2021). Data augmentation makes datasets artificially more representative.

To prepare for LSST, a data challenge was run for 3 months to push the development of transient classification models, the Photometric LSST Astronomical

Time Series Classification Challenge (PLAsTiCC) (The PLAsTiCC team et al., 2018; Hložek et al., 2023). PLAsTiCC contained 3.5 million simulated LSST objects. It was won by Avocado (Boone, 2019), which outperformed all other models by using a boosted decision tree trained on light curves augmented to cover a range of redshifts and observing conditions. The challenge is being recreated with the Extended LSST Astronomical Time-series Classification Challenge (ELAsTiCC) (Narayan & ELAsTiCC Team, 2023). ELAsTiCC furthered the original PLAsTiCC challenge, by creating and testing an end-to-end real-time pipeline for time-domain science. ELAsTiCC features 4.3 million objects streamed over 3 months, which finished in January 2023.

We have covered a small selection of machine learning models used for classifying SNe. The number of models is increasing rapidly, and will only further increase as more surveys come online.

2.2 The Vera Rubin Observatory

The Vera C. Rubin observatory is expected to begin collecting observations for LSST in 2025. The observatory will survey a large portion of the southern hemisphere. The main telescope of the Rubin observatory is the Simonyi Survey Telescope. The telescope has an 8.4 metre primary mirror, with the world's largest CCD camera [1]. The Rubin Observatory will observe a 9.62 square-degree area of sky at a single pointing,[2] with the entire survey covering ∼30,000 deg^2 (Ivezić et al., 2019). The Wide-Fast-Deep (WFD) survey, which is the primary survey, will have a declination range of -65 to +5 degrees (Marshall et al., 2017), although it should be noted that the final survey design has not been decided. In addition to the WFD survey, LSST will spend 10-20% of observing time dedicated to other programs, including intensive observations of a set of Deep Drilling Fields (DDFs) [3]. Although the number of DDFs to be visited is yet to be determined. The Rubin Observatory is expected to observe about 10^5 SNe Ia per year for 10 years (Ivezić

[1] The specifications for the Simonyi Survey Telescope can be found on the Rubin observatory website (https://www.lsst.org/about).

[2] https://www.lsst.org/about/tel-site/optical_design

[3] https://www.lsst.org/scientists/survey-design/ddf

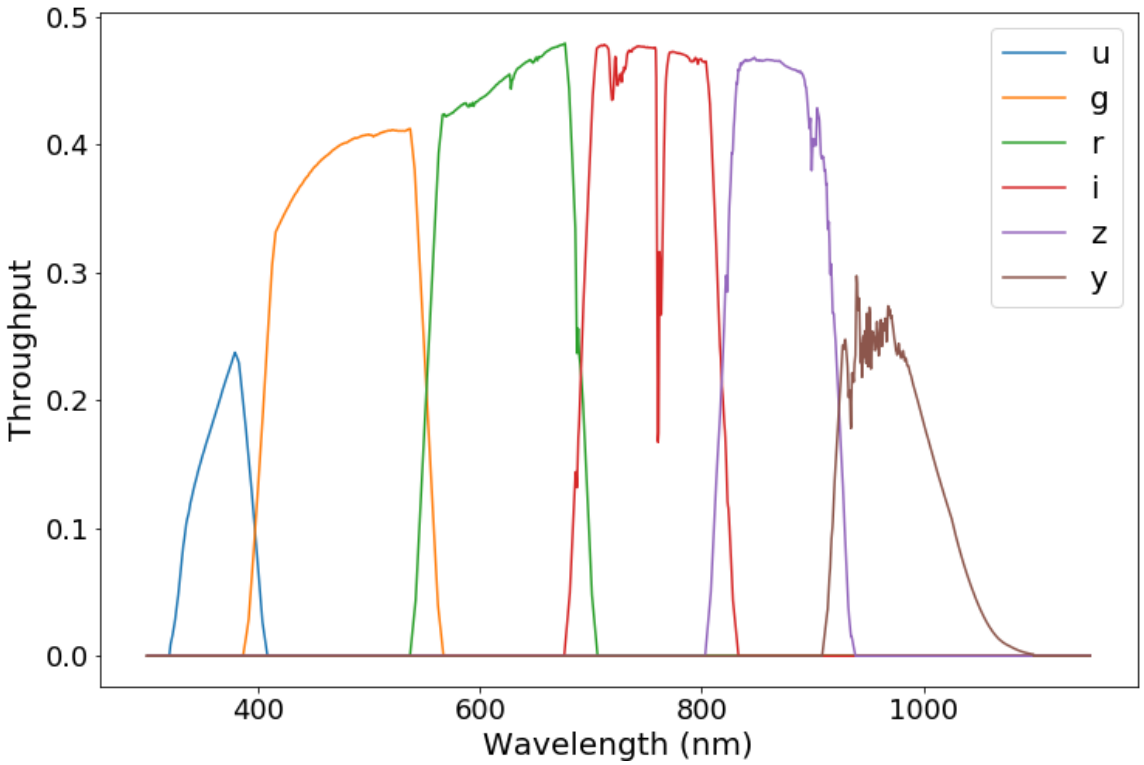

Figure 2.3: The throughput of the filters used by the Vera Rubin observatory (u, g, r, i, z, y).

& the LSST Science Collaboration, 2013), of which approximately 112,000 will be suitable for cosmology (The LSST Dark Energy Science Collaboration et al., 2018a). Upon the completion of the 10-year survey, the observed SNe Ia are expected to improve the precision of w_0 and w_a by 100-1000 times more than LSST's precursors (Ivezić et al., 2019). LSST will achieve this by observing large quantities of SNe Ia, with good imaging quality, to minimise the statistical uncertainty on w. In this thesis we are concerned with measurements of the host galaxies. In order to estimate the quality of photometry from LSST, we assume the 10-year, 5-σ survey depths of the Wide Fast Deep survey, namely u=26.1, g=27.4, r=27.5, i=26.8, z=26.1 and y=24.9 mags.[1] We use AB magnitudes throughout this thesis.

[1]https://www.lsst.org/scientists/keynumbers

2.3 The 4-metre Multi-Object Spectroscopic Telescope (4MOST)

4MOST is a new high-multiplex, wide-field spectroscopic survey facility under development for the 4m VISTA Telescope (Guiglion et al., 2019a; de Jong et al., 2019). 4MOST is due to begin operations in 2024. It will be a fibre-fed spectrograph, with 2,436 fibres in an approximately 4 square degree field-of-view.[1]. Each fibre will have a diameter of 1.45 arcseconds [2]. A third of the fibres will be connected to a High-Resolution Spectrograph (HRS), with the remaining two-thirds of fibres being connected to two Low-Resolution Spectrographs (LRS). The LRS will observe with a resolution of 5,000 at a wavelength range of 370-950 nm. The HRS will have a resolution of approximately 20,000 and will observe at 392.6-435.5, 516-573 and 610-679 nanometres (de Jong et al., 2016). 4MOST will be able to reach magnitudes as faint as $r_{AB} \approxeq 22.5$. In a 5-year survey 4MOST will be able to cover approximately 21,000 square-degrees of sky, covering a declination range of -70 $<$ dec $<$ 5 degrees (Guiglion et al., 2019b). This declination range has a significant overlap with LSST. 4MOST will visit both the WFD survey locations and the DDFs observed by LSST. The LSST DDFs will be visited by 4MOST at least every 14 days, which will allow more opportunities for spectroscopic classification of transients with improved light curves, compared to the WFD fields. With repeat observations in the DDFs, spectra of SN host-galaxies can be stacked to observe more distant targets [3].

During the 5-year 4MOST survey, approximately 20 million low-resolution and 3 million high-resolution spectra will be observed.[4] This assumes an exposure time of 2 hrs. The observing strategy of the 4MOST surveys are still to be decided.

[1]The specifications for 4MOST can be found on the ESO website (www.eso.org/sci/facilities/develop/instruments/4MOST.html#BasSpec).

[2]Details about 4MOST's fibres can be found in the 4MOST user manual https://www.4most.eu/cms/facility/overview/.

[3]https://docushare.lsst.org/docushare/dsweb/Get/Document-37640/Frohmeier_TiDES.pdf

[4]See footnote 1.

2.3.1 The Time Domain Extragalactic Survey (TiDES)

The Time Domain Extragalactic Survey (TiDES) is one of 25 surveys within the 4MOST consortium [1]. The TiDES science program contains three sub-surveys: (i) TiDES-Live, (ii) TiDES-Hosts and (iii) TiDES-RM. TiDES-Live will observe live transients by carrying out spectroscopic follow up of LSST, and other photometric survey discoveries. The TiDES-Hosts sub-survey will measure the spectroscopic redshift for SNe host galaxies, for which live spectra were not captured. In addition to measuring redshift, the observations of TiDES-Hosts will lead to the calculation of galaxy properties such as star-formation rate and age. The approach of TiDES-Hosts has been used successfully by the Dark Energy Survey (DES). DES combined deep host-galaxy photometry (Wiseman et al., 2020) with fibre-fed spectroscopy from the Anglo-Australian Telescope (Lidman et al., 2020). TiDES-Hosts will play a key role in the LSST SN Ia cosmology programme (The LSST Dark Energy Science Collaboration et al., 2018b). TiDES-Live aims to collect 35,000 live transients, whilst TiDES-Hosts aims to observe 50,000 host galaxies during the first 5 years of 4MOST (Swann et al., 2019; Frohmaier et al. (in prep)). The final survey, TiDES-RM will collect spectra of AGN, to use the AGN broad-line lags to build a Hubble diagram to $z = 2.5$. AGN are powerful standard candles similar to SNe Ia (e.g. Kaspi 2001; Bentz et al. 2013; Peterson et al. 2014).

Simulations have been carried out to quantify the effect TiDES will have on SN observations. One simulation has explored the ability of TiDES combined with LSST observations to constrain w. The simulation has shown that TiDES with LSST will be able to measure w with an uncertainty of 0.02, w_0 with an uncertainty of 0.054 and w_a with an uncertainty of 0.342 (Vincenzi, 2022). These values assume a flat $w_0 w_a$ CDM model, and combine SNeIa measurements with CMB and Baryonic Acoustic Oscillation measurements.

TiDES is different from other 4MOST surveys because it will not drive the pointing of 4MOST. Instead, TiDES relies on the fact that wherever 4MOST points LSST will have recently observed transients. TiDES-Live will not be able to observe every object observed by LSST, due to different observing strategies and 4MOST's longer exposure time. A selection function will be applied to SNe

[1] https://www.eso.org/sci/observing/PublicSurveys/4most-surveys-projects.html

identified from the LSST data-stream. The selection criteria a SN must reach are: (i) only data points from *griz* are considered, (ii) the transient must be detected to $> 5\sigma$ in at least 3 bands, (iii) $> 5\sigma$ detections must occur on at least 2 distinct nights and (iv) the object must reach 22.5 mag in any *griz* filter (Frohmaier et al. in prep). TiDES utilises a selection criteria instead of a machine learning model. This is due to TiDES having enough fibre hours to target all SNe which meet the selection criteria within a 4MOST field.

2.4 Other upcoming instruments

As well as the facilities we have used in our work, there are several upcoming satellites which will have a large impact on SN cosmology. We give an overview of the capabilities of the recently launched Euclid satellite and the soon to be launched Nancy Grace Roman Space Telescope.

2.4.1 Euclid

Euclid is a recently launched European Space Agency space telescope (Laureijs et al., 2011). It is designed to explore the evolution of the dark Universe. The primary mirror is 1.2 meters in diameter [1]. Euclid has a photometry instrument (530-920nm) and both a near infrared spectrometer (926nm - 1892nm) and a photometer (949.6-2021.4nm) [2]. Euclid will obtain median 5-σ survey depths of the Wide survey of i=26.2, y=24.3, j=24.5 and h=24.4 mags (Euclid Collaboration et al., 2022). The near infrared observations will be complementary to LSST, as ground based observatories struggle to carry out deep infrared observations due to the thermal background on Earth.

At present Euclid has two surveys planned. The wide survey will observe 15,000 deg^2 of sky. The Deep survey will observe a 0.57 square degree area of the sky approximately every hour [3]. At present a transient survey is not planned but one was proposed by Astier et al. (2014). The issue with using Euclid to observe

[1] https://www.esa.int/Science_Exploration/Space_Science/Euclid_overview
[2] https://www.euclid-ec.org/science/overview/
[3] https://www.euclid-ec.org/public/data/surveys/

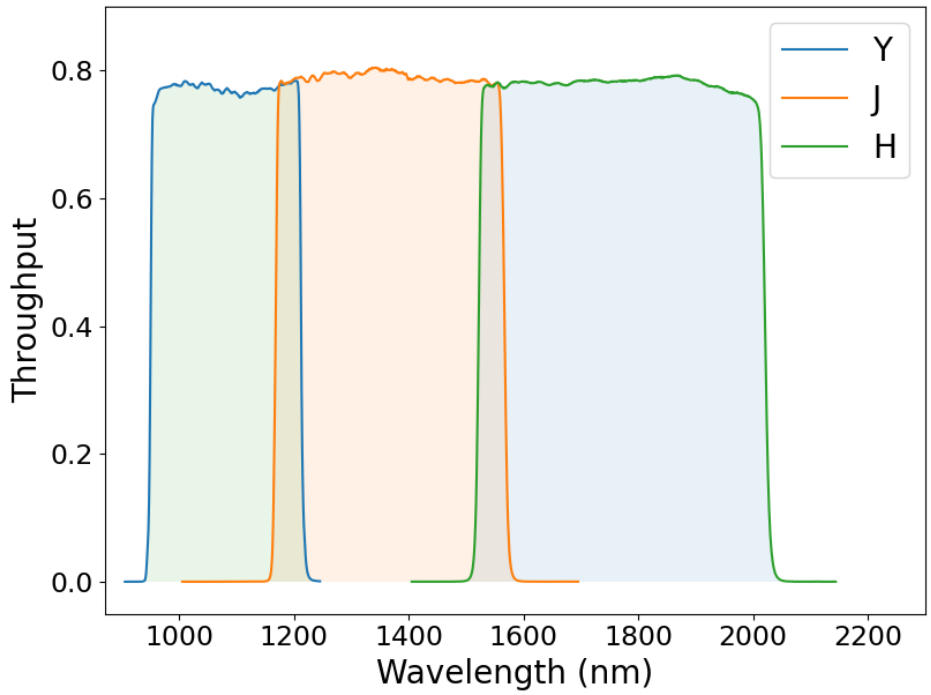

Figure 2.4: The throughput of the near-infrared filters used by the Euclid space telescope (Y, J, H).

SNe Ia in the deep fields is that the preliminary observing strategy is to perform a series of observations of the same area of sky for a week, and then not return for 6 months. These observations would struggle to capture the full details of a SN light curve. One way to fix this issue is to combine Euclid observations with observations from another survey. Bailey et al. (2023) simulated 5-years of Euclid deep observations. They predict that LSST will observe approximately 3,700 SNe that will have at least five Euclid detections with SNR > 3, and 1,900 SNe which will have at least ten detections with SNR > 3 from Euclid.

Euclid will be able to measure some of the cosmological parameters to a high degree of precision. In particular, it is expected that Euclid will be able to measure w_0 with an uncertainty between 0.05 - 0.07 and w_a with an uncertainty between 0.16 - 0.3 (Amendola et al., 2013, 2018). However, these values are measured using weak lensing and Baryonic Acoustic Oscillations.

2.4.2 The Nancy Grace Roman Space Telescope

The Nancy Grace Roman Space Telescope (Roman) is NASA's next premier space-based instrument, expected to launch in 2027. Roman will carry out near infrared imaging and spectroscopic observations. The Roman telescope will have a 2.4 meter mirror [1]. Roman will have a field of view of 0.28 square degrees with each pointing [2]. There is synergy with Roman and other surveys. Roman's survey will be narrower and deeper than Euclid's survey, making the two instruments complementary (Spergel et al., 2015). A wide survey using Roman will be able to provide high-resolution imaging for $> 95\%$ of the LSST 10-year survey, within 5-months (Eifler et al., 2021).

Roman is expected to be useful for the study of SNe. Roman will observe between 3600 and 8800 SNe Ia during it's mission (Hounsell et al., 2018; Akeson et al., 2019). The observed SNe will be at redshifts ≤ 2. A potential SN observing strategy is presented by Spergel et al. (2015). One of the aims of Roman is to carry out a survey of SNe Ia which will reduce the uncertainty of the dark energy equation of state by 70% [3]. Using SNe Ia the uncertainty of w_0 is expected to be measured to 0.068, whilst the measured uncertainty for w_a is expected to be 0.2 (Hounsell et al., 2018) using a flat ΛCDM model. These results only use Roman data, the best improvements to dark energy will come from a combined analysis using Euclid, LSST and Roman data (Spergel et al., 2015).

In addition to the cosmological survey, Roman will observe $10^8 - 10^9$ galaxies over the mission lifetime, and measure the spectroscopic redshift of each galaxy [4]. This large survey of spectroscopic galaxy observations will prove useful to the classification of SNe observed by Rubin, in the area of overlap.

[1] https://roman.gsfc.nasa.gov/science/observatory_technical.html

[2] https://www.stsci.edu/files/live/sites/www/files/home/roman/_documents/roman-science-sheet.pdf

[3] https://www.stsci.edu/files/live/sites/www/files/home/roman/_documents/cosmology-with-roman.pdf

[4] https://www.stsci.edu/files/live/sites/www/files/home/roman/_documents/roman-capabilities-galaxies.pdf

Chapter 3

Measuring host galaxy properties

This Chapter is dedicated to the work we carried out quantifying the use of 4MOST spectra with Rubin photometry to measure galaxy properties more precisely than if photometry alone is used to measure galaxy properties. We are interested in measuring physical properties, such as: stellar mass, star formation rate and galaxy age.

3.1 Methodology

3.1.1 Simulating observations

As 4MOST and the Vera Rubin Observatory are not operational yet, we use simulated data. To assess both instruments' ability to measure galaxy properties, we need to start with a spectrum of a galaxy with values of physical properties that we adopt as the ground truth for this experiment. We use an example output spectrum of the Fitting Assessment of Synthetic Templates (FAST) SED fitting code (Kriek et al., 2009), which has associated M_*, star formation rate and other galaxy property values. In order to generate the target galaxy properties with which we can compare our results, we ran FAST for an initial pass (we later use FAST in a second pass to analyse the simulated 4MOST output). For this we gave FAST an input spectrum to fit. We used the Kinney et al. (1996) elliptical tem-

plate which is built into the 4MOST Exposure Time Calculator (4MOST ETC)[1]. In later stages of the work we carry out the process with an Sc galaxy, however the initial results are obtained for the elliptical galaxy. We only used an elliptical and an Sc galaxy. As elliptical galaxies have the least amount of star formation (Kokusho et al., 2017), comparing this to a star forming spiral galaxy will allow us to test the most extreme situations. In the following subsections we describe the workflow from the Kinney et al. spectrum to the template spectrum with known properties. This also provides an overview of the method which will be elaborated in the remainder of Section 3.1.

3.1.2 Producing a template spectrum

We begin with the Kinney et al. (1996) elliptical spectrum redshifted to 0.3, approximately the middle of the expected TiDES redshift range. The Kinney et al. spectrum was used as it had a good signal-to-noise ratio and was readily available. The elliptical galaxy spectrum was normalised to r = 21 mag and an observation was simulated using the 4MOST ETC code (see Section 3.1.5 for the parameters assumed). As a reminder, we use AB magnitude scale throughout this work. The raw output from the 4MOST ETC and the corresponding noise spectrum were processed to generate realistic-flux-calibrated spectra (see Section 3.1.5 for details). Synthetic photometry was also created, and the process to do this is explained in Section 3.1.4.

The spectrum and photometry were then fit with FAST. FAST takes a parameter file which defines the settings to be applied, as shown in table 3.1. Once defined these are not changed. This initial run through FAST produced a best-fit spectrum that acts as the template which later fitted-galaxy-property values are compared with. From this calculated value we could extrapolate additional target values of stellar mass for redshift = 0.3, by using the linear relationship of $\log(M_*)$ and magnitude.

[1]We used the internal PYTHON based 4MOST ETC; nonetheless, the public web-based ETC provides the same results (https://etc.eso.org/observing/etc/fourmost).

Parameter	Option or Range chosen
Number of simulations	10,000
Confidence interval	68%
Stellar population library	Bruzual & Charlot (2003)
Stellar initial mass function	Chabrier (2003)
Star formation history	Delayed exponential SFH
SFR Average	0 (Instantaneous SFR)
Method to find best-fit	Median of Monte Carlo
Dust law	Kriek & Conroy (2013)
$\log(\tau)$ [$\log(\tau/\mathrm{yr})$]	6.5 - 11
log(age) [log(age/yr)]	8.0 - 9.8
V-band extinction (A_V)	0.0 - 3.0
Metallicity	0.004, 0.008, 0.02, 0.05
Hubble Constant [km/s/Mpc]	70.0
Ω_M	0.3
Ω_Λ	0.7

Table 3.1: The parameters that we chose to use when running FAST to find the best-fitting galaxy parameters for input photometry and/or spectra.

3.1.3 SED fitting code - FAST

FAST is a Spectral Energy Distribution (SED) fitting code. FAST works by fitting stellar population synthesis templates to a spectrum and/or broadband photometry. FAST then returns values for the galaxy properties of the best-fit galaxy. The galaxy properties FAST can calculate are: metallicity, stellar age, V-band extinction (A_V), M_*, star formation rate, specific star formation rate, star formation timescale (τ) and the ratio of age to star formation timescale. FAST can also calculate the redshift of the galaxy. However, since we define the redshift this was not necessary. As mentioned previously, FAST takes a parameter file that defines the allowed ranges of each parameter over which it searches (relating to FAST's library of galaxy spectra). The chosen parameters can be seen in Table 3.1. The best-fitting galaxy properties are found by taking the median of the distribution of 10,000 runs of FAST. In each run FAST alters the photometry and/or spectral flux values within the corresponding error values. FAST then fits these new photometry values in each run. After 10,000 runs, there is a range of calculated galaxy

property values. FAST then finds the upper and lower limits that contain 68% of the data, to find the equivalent (Gaussian interpretation) 1-σ uncertainty range for each galaxy property. FAST does not use a minimum-searching algorithm, instead it fits every point of the parameter space.[1] FAST was used for this project due to its ability to fit both a spectrum and photometry data.

3.1.4 Synthetic photometry

The next stage of the simulations takes the noiseless-template spectrum (from Section 3.1.2) to produce synthetic photometry and a spectrum as observed by 4MOST. During the work described in this Section and 3.1.5, we shifted the template to different redshifts ($0.1 < z < 0.7$) and normalised it to different desired magnitudes in the range ($18 < \mathrm{r} < 25$). We call these shifted template spectra 'target spectra'. For each case the Python package Pyphot (version 1.0) was used to calculate the flux within the filter bands u, g, r, i, z and y. The throughput received after passing through the atmosphere, the filters and the detectors can be seen in Figure 3.1. Pyphot calculates the flux by integrating each template through the given filter bands. The Rubin filter response functions provided on the Vera Rubin Observatory's website [2] are used for this process, which we have added into both Pyphot and FAST. For each filter band we calculate the 1-σ sky noise corresponding to the 10-year LSST depth, by scaling from the 5-σ depths given in Section 2.2, assuming that sky noise dominates the photometric error. For each photometry point, we use the sky noise in the corresponding filter as the photometric uncertainty, unless the sky noise is less than 1% of the flux, in which case we set the photometric error to be 1% of the flux. This ensures that we do not use unrealistically small photometry errors.

3.1.5 Simulated 4MOST spectra

The template spectrum is redshifted and magnitude normalised to each of the required values, then input into the 4MOST Facilities Simulator Exposure Time Cal-

[1] https://github.com/jamesaird/FAST
[2] https://www.lsst.org/scientists/keynumbers

culator (4FS ETC). The 4FS ETC is a software tool that estimates what 4MOST would see with specific observing conditions given an input spectrum with a specific magnitude. During this research we used many different versions of the 4FS ETC, as it was updated. Our results shown here made use of version 2.04 of the 4FS ETC. There is a newer version of the 4FS ETC which was released during the process of this work, version 2.4. The newer version produces additional outputs compared to version 2.04 (which we do not use). The outputs used from version 2.04 were compared to the equivalent output from version 2.4. and found to be identical. The observing conditions were kept constant during this study. The 4FS ETC was set to have an airmass of 1.2, seeing of 0.8 arcseconds and a dark moon. We use an exposure time of 2 hrs, assuming 2 1-hour field visits based on Tempel et al. (2020). Later in the process we allow for the fact that brighter objects will be removed from the observing queue after reaching the spectral success criterion (see the final paragraph in this Section).

The 4FS ETC v2.04 produces: spectrograph gain [electrons/adu], target signal count [electrons], sky background count [electrons], CCD dark current [electrons], CCD readout noise [electrons], noise count [electrons], efficiency [electrons/photon] and spectral bin width [nm]. All of these are produced as a spectrum with wavelength units of nanometres, which we convert to the units of angstrom. We will call the target-signal-count spectrum 'object spectrum', to prevent confusion with our shifted target spectrum. The object spectra are produced separately for blue, green and red wavelength ranges corresponding to the three arms of the 4MOST spectrograph. The top and middle panel of Figure 3.2 show examples of input and output of the 4FS ETC.

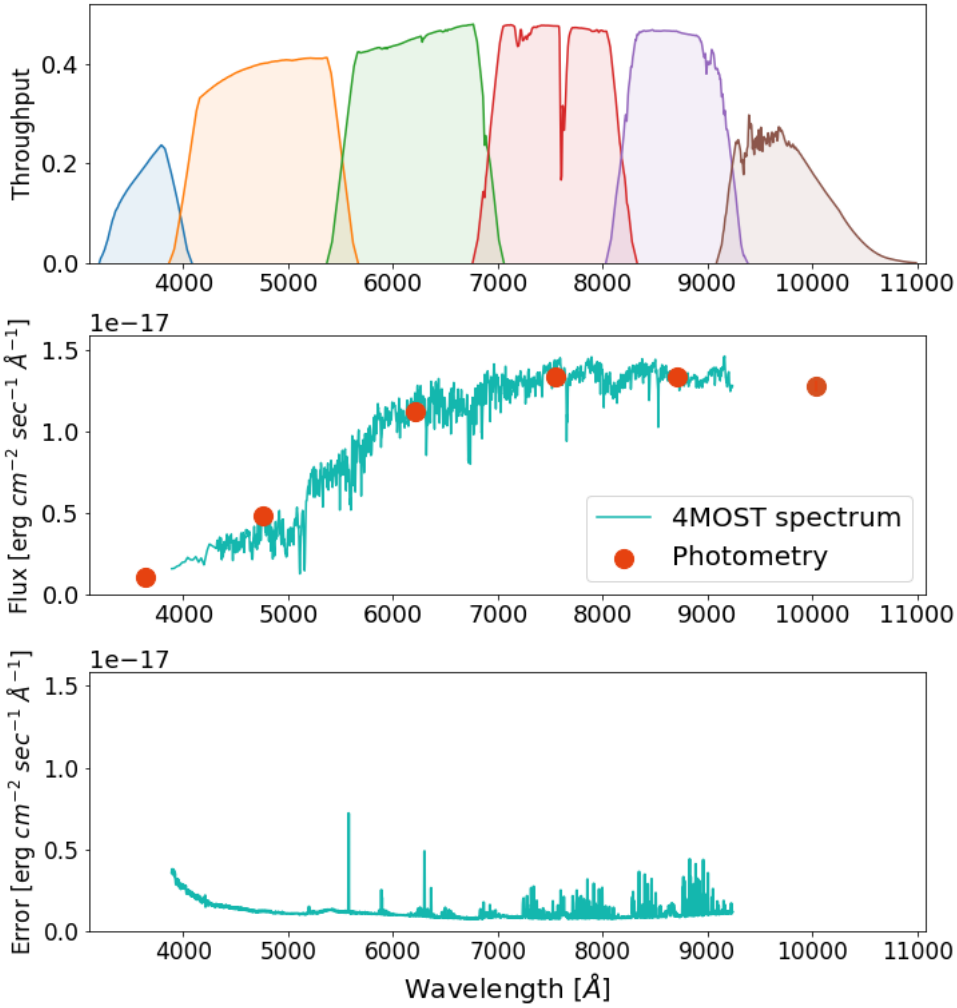

Figure 3.1: The top panel shows the throughput received, after passing through the atmosphere, filters and detectors. The middle panel shows an example of the 4MOST spectral flux produced by the 4MOST ETC and combined into a single spectrum for magnitude 21, $z = 0.3$. The photometry for magnitude 21 can also be seen as the circles plotted on top of the 4MOST spectrum. The photometry uncertainties are also plotted, however they are smaller than the photometry symbol and so cannot be seen. Finally, the lower plot is the error values that correspond to the spectrum.

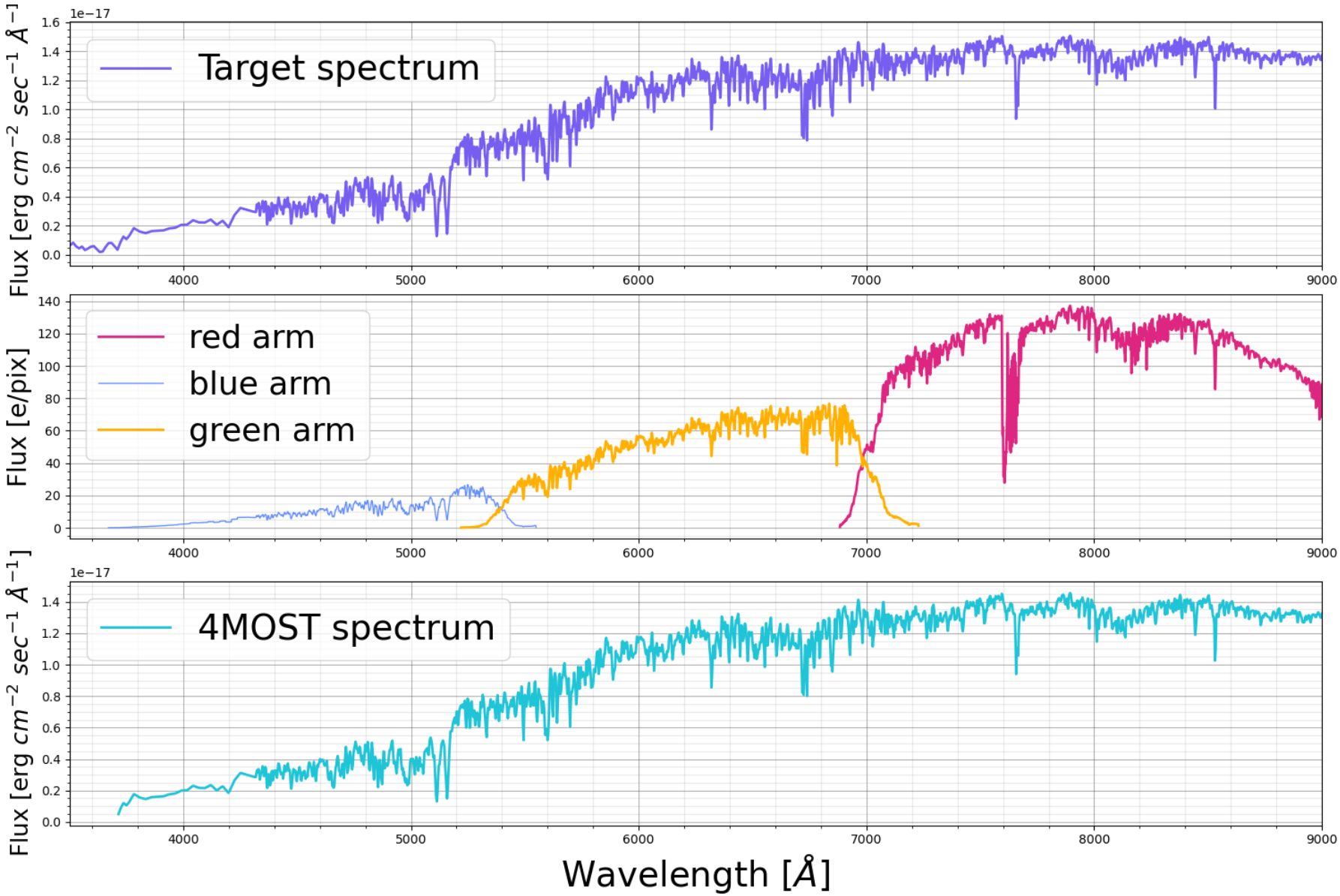

Figure 3.2: The spectrum at different stages throughout the process of simulating a spectrum from 4MOST. The top plot is the best-fit spectrum produced during the target stage. The middle plot is the target best-fit spectrum after it has been put through the 4FS ETC to produce a noiseless spectrum. Finally, the bottom plot is the flux-calibrated spectrum. This is the flux-calibrated object spectrum after it has been binned and combined into a single spectrum.

To mimic flux calibration we multiplied each object spectrum by the corresponding gain, and then divided by the instrument response function. In the absence of simulated-spectral-standard stars we determined the response function using the shifted target spectrum itself, as shown as follows

$$\mathrm{res}(\lambda_i) = \frac{\mathrm{obj}(\lambda_i).g(\lambda_i)}{\mathrm{targ}(\lambda_i).t_{\mathrm{exp}}} \tag{3.1}$$

where $\mathrm{res}(\lambda_i)$ is the response at the wavelength of the ith pixel, $\mathrm{obj}(\lambda_i)$ is the object spectrum, t_{exp} is exposure time, g is gain at each wavelength value, $\mathrm{targ}(\lambda_i)$ is the shifted target spectrum before it was input into the ETC. This effectively assumes perfect flux calibration. In practice we expect 4MOST to produce relative flux calibration (which is what is important to this application), and absolute calibration will be provided by comparison to LSST photometry. The object spectra produced by the ETC are noiseless (however the ETC also produces noise values for each wavelength value). A calibrated noise spectrum is calculated for each of the three wavelength ranges by dividing the noise values for each wavelength by its corresponding response value.

To calculate the galaxy's properties using FAST, the three sections of each object spectrum and calibrated noise spectrum must be combined into a continuous spectrum. Since the three arms of 4MOST have different wavelength binning, a binning function is used to place all three sections on a regular wavelength step of 3 Å. This binning process is also applied to the weight spectra. The weight spectrum, $\mathrm{w}(\lambda_j)$, is defined as follows

$$w(\lambda_j) = \left(\frac{n(\lambda_j)}{\mathrm{res}(\lambda_j)}\right)^{-2} \tag{3.2}$$

where $n(\lambda_j)$ is the noise spectrum and all spectra are defined on the new wavelength spacing, λ_j. The weighting function ensures the extreme start and end values of a spectrum are given less weighting, where the throughput of the instrument is small. The weighting function was then applied to combine the spectra in the regions where the spectra overlap through a weighted average, i.e.

$$f(\lambda_j) = \frac{w_1(\lambda_j).\mathrm{obj}_1(\lambda_j).g_1(\lambda_j) + w_2(\lambda_j).\mathrm{obj}_2(\lambda_j).g_2(\lambda_j)}{w_1(\lambda_j) + w_2(\lambda_j)} \tag{3.3}$$

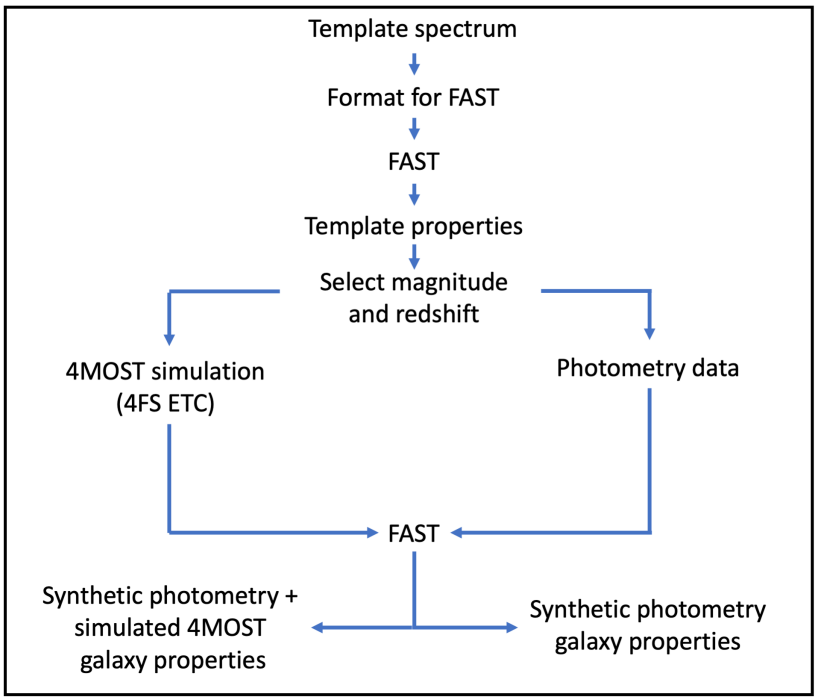

Figure 3.3: The steps carried out to calculate galaxy properties for magnitudes 18 $<$ r $<$ 25 and redshifts 0.1 $< z <$ 0.7.

where $f(\lambda_j)$ is the final combined flux, obj_1 and obj_2 are the two spectra to be joined, w_1 and w_2 are their corresponding weight spectra and g_1 and g_2 are their corresponding gain spectra. The error values in the overlap section are summed in quadrature to produce a single continuous error spectrum. An example of the final combined spectrum is shown in the bottom panel of Figure 3.2.

Once the photometry and 4MOST spectrum have been produced, along with their corresponding error values, they are input into FAST for the main comparison between photometry only (referred to from here on as "phot") and photometry with 4MOST spectra (referred to from here on as "phot + 4MOST"). An example of the input is shown in Figure 3.1. One version of results was produced from inputting the photometry in to FAST alone and another version of results was produced by inputting the photometry and 4MOST spectrum together. A flowchart of the process can be seen in Figure 3.3. An example of the photometry and spectrum input into FAST, compared with the best-fit produced by FAST can be seen in Figure 3.17.

To account for bright objects being removed from the queue we assumed an uncertainty floor. TiDES uses a selection criteria that does not allow an observed galaxy to have a signal-to-noise ratio smaller than 3. When an object reaches

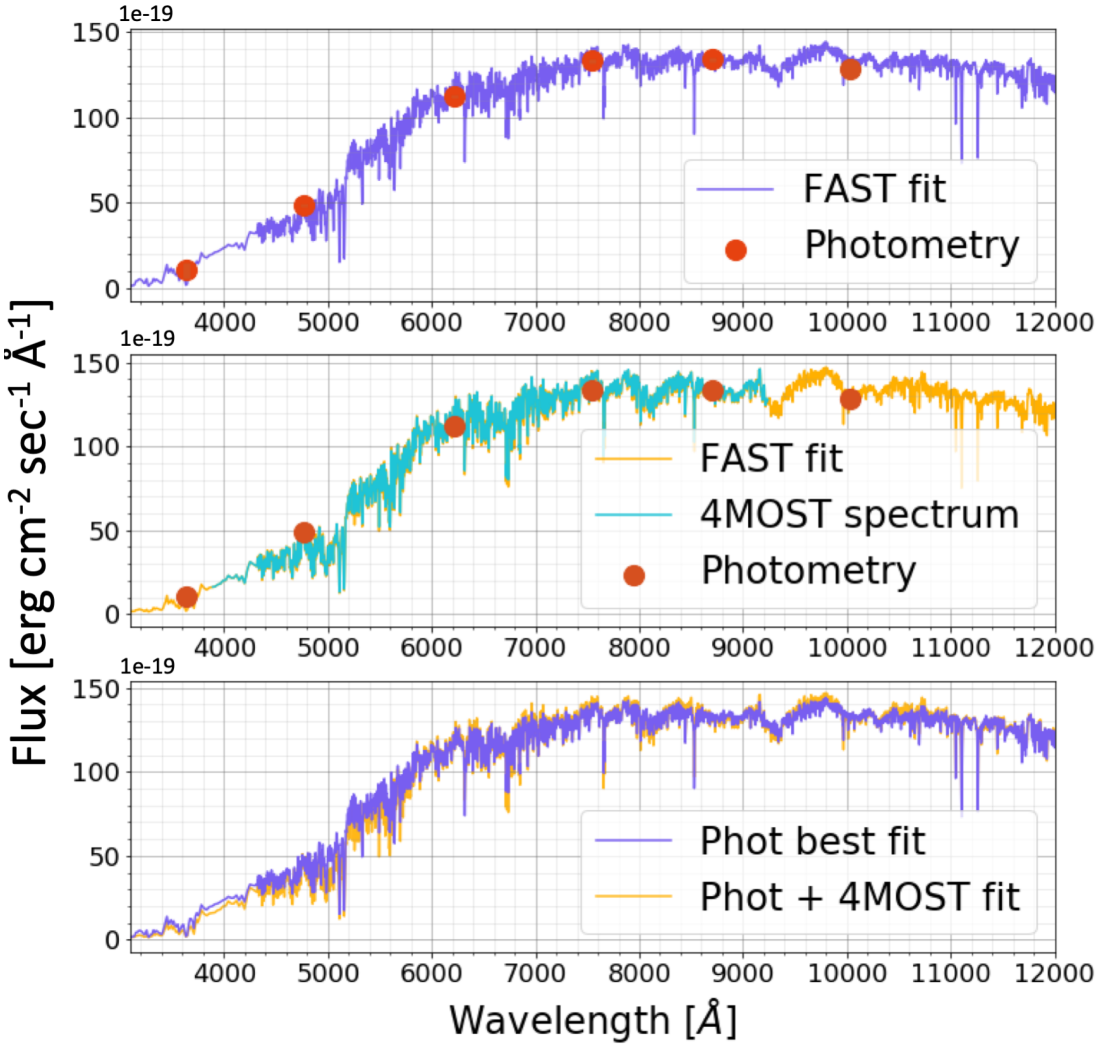

Figure 3.4: The top panel shows the best-fit found by FAST for photometry alone (purple line), with the photometry points (orange circles) plotted at the central wavelengths of the filters. Full profiles of the Rubin filters are shown in Figure 3.1. The middle panel shows the best-fit found for photometry with a 4MOST spectrum (gold line). The photometry is plotted on top (orange circle) with the 4MOST spectrum (blue line). The bottom panel shows a comparison of the two best-fits. As can be seen the two best-fits diverge below 6000 angstroms. All of the plots are for an elliptical galaxy at magnitude 21 and redshift 0.3.

this value it is removed from 4MOST's observing queue. The magnitude for each galaxy type to reach the signal-to-noise criteria, for each redshift, was measured. For each galaxy property calculated from spectra, the recorded uncertainty at this magnitude was applied to all brighter objects of the same redshift.

3.2 Results

We are interested in the values FAST calculated for galaxy M_*, as this property has been found to correlate with SN properties. The results for galaxy M_* can be seen in Figure 3.5 and Table 3.3 in Section 3.5. The results for the template spectrum, simulated photometry and simulated spectroscopy at $r = 21$ and redshift 0.3, agree within the 68 percentile confidence ranges reported by FAST. This gives us confidence in the fitting process. The results show that for galaxies with $r = 22$ mag and brighter, combining Rubin photometry and 4MOST spectroscopy reduces the uncertainty around the best-fit measurement of a galaxy's M_* by 49 – 95% compared to using photometry alone. There is also an improvement at magnitudes 23 to 25, however the improvement is not as significant.

At fixed redshift the best-fit log (M_*) drops linearly as the apparent magnitude of the fitted galaxy increases, as shown in Figure 3.5. This is expected as FAST models the brightness of a galaxy as being proportional to the M_* of the galaxy for a given stellar population. The best-fit galaxy parameters are calculated based on a multi-dimensional search with each of FAST's galaxy parameters. We find that at brighter magnitudes ($r \leq 22$) some of the mass uncertainties reported by FAST are exactly zero, which is unrealistic. We believe that this is because the space between models in galaxy parameter space contained within FAST is not fine enough for FAST to find another model, within the 68% confidence of the best-fit mass, with the small uncertainties provided by 4MOST.

The results for the Sc galaxy can be seen in Figure 3.6 and Table 3.4 (in the appendix). The results are comparable with the elliptical galaxy. Phot + 4MOST reduces the uncertainty for magnitudes ≤ 22 by 79 – 95%, whilst at magnitudes fainter than 22 the uncertainty is reduced by 8 – 68%. The slope produced when M_* is plotted against magnitude for each redshift is different for a Sc galaxy than the slope for the elliptical galaxy. This is due to FAST finding different values for the other parameter for each galaxy, thus changing the slope of the plot by a small amount.

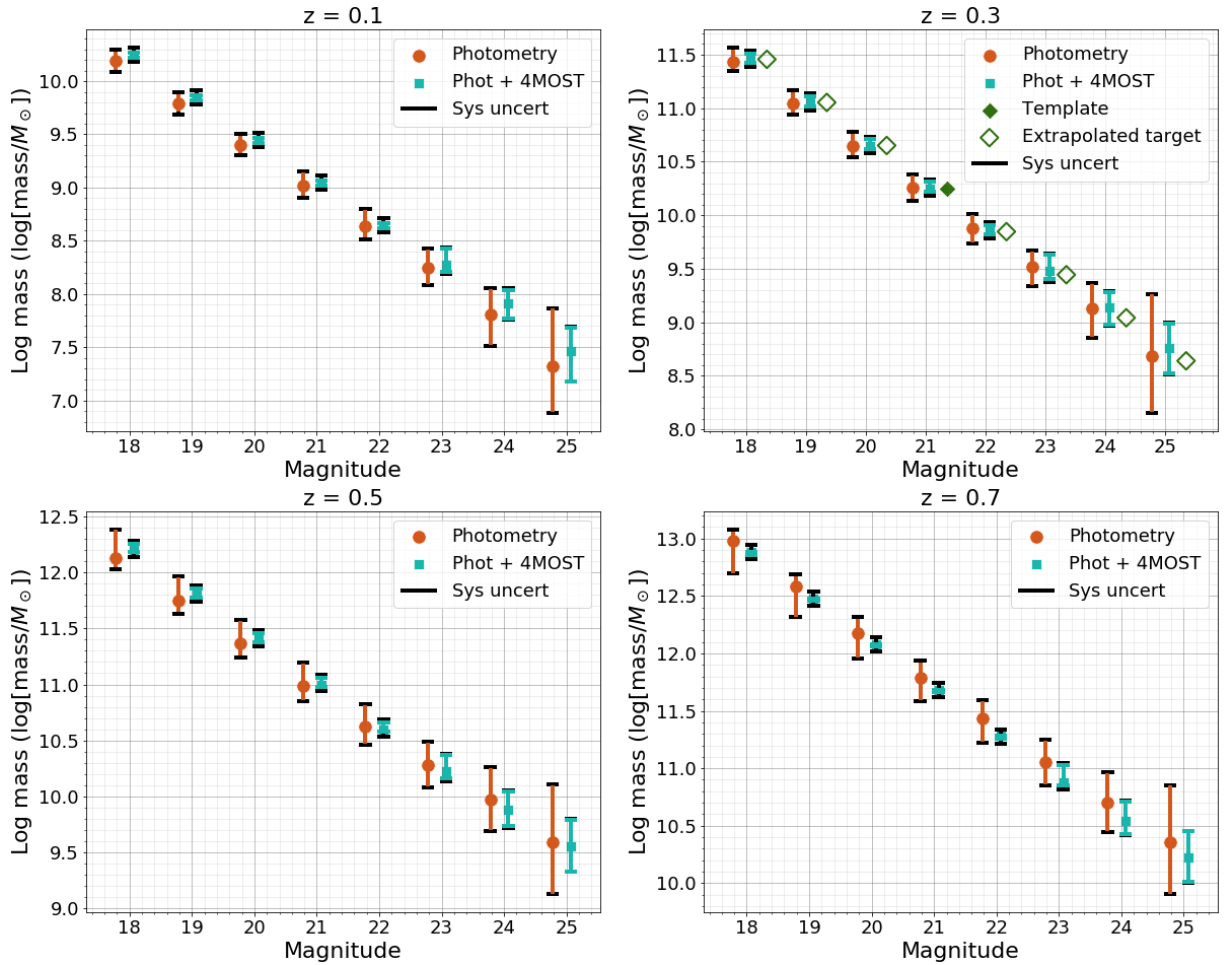

Figure 3.5: Derived log(M_*) as a function of magnitude for LSST photometry only (orange circles) and for LSST phot + 4MOST spectroscopy combined (blue squares) for an elliptical galaxy at z=0.1, 0.3, 0.5 and 0.7. The template spectrum's M_* is also shown (filled green diamond). The template mass at other magnitudes were extrapolated (empty green diamonds) from the original template spectrum (filled green diamond). There is an uncertainty associated with the template spectrum, which is the uncertainty that FAST calculates from the input real galaxy spectrum, although the uncertainty is too small to see in this Figure. The extended uncertainties (black line) show the total error bar when an estimate of systematic uncertainty is added in quadrature. The log (M_*) of the galaxies are shown as a function of magnitude and redshift. It can be seen that at all simulated redshifts and magnitudes the precision of the calculated galaxy mass is significantly improved for phot + 4MOST, compared with using photometry alone.

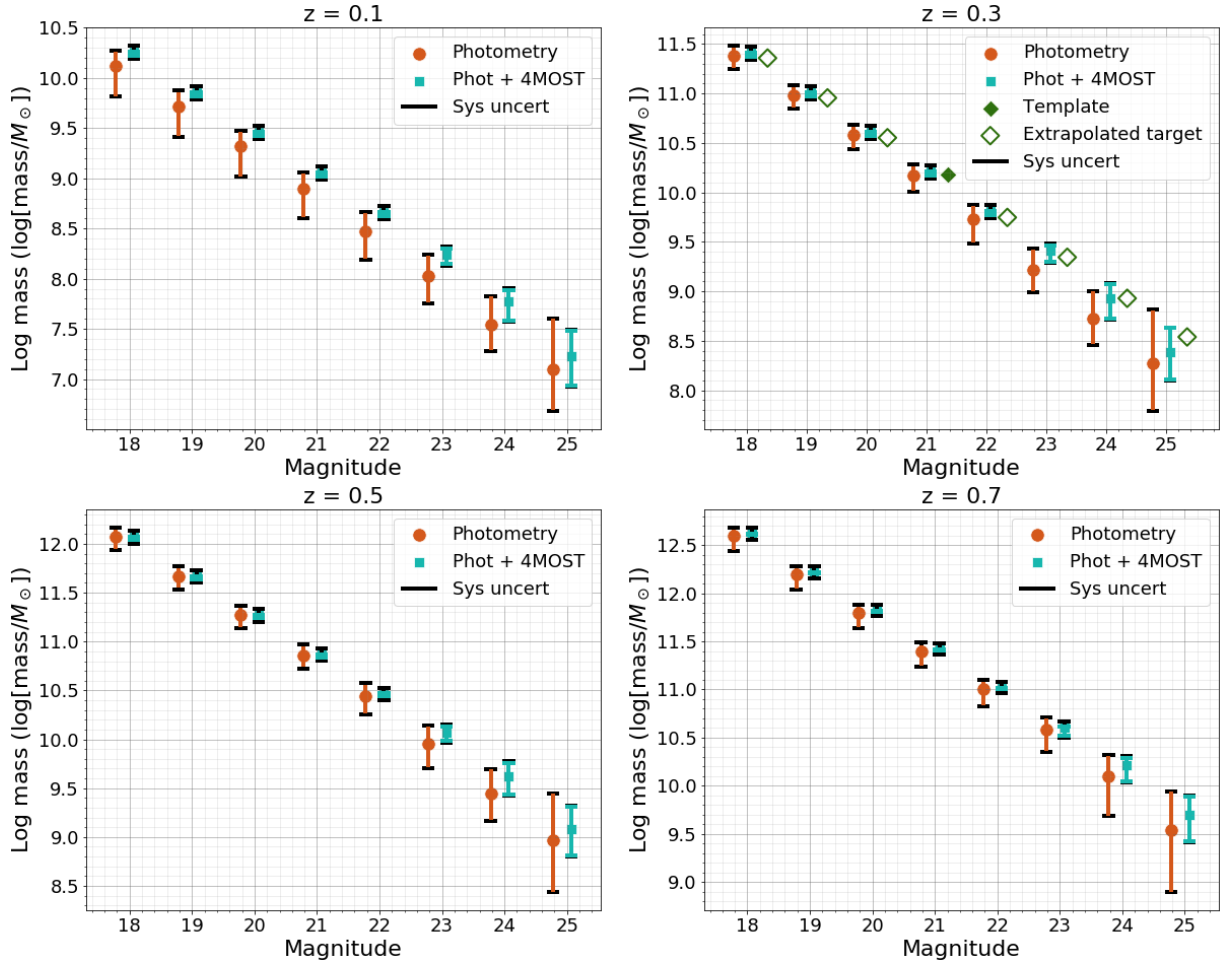

Figure 3.6: The same as figure 3.5 but for a Sc galaxy. It can be seen that at all simulated redshifts and magnitudes the precision of the calculated galaxy mass is significantly improved for phot + 4MOST, compared with using photometry alone.

A recent publication by Pacifici et al. (2023) reported on a systematic uncertainty when using SED fitting to measure galaxy properties. They reported on a systematic uncertainty which accounts for differences in SED fitting, as well as differences in model parameter assumptions. They explored a wide range of SED fitting codes, but did not analyse FAST. We used their median systematic values over all SED codes for mass (0.12 dex), star formation rate (0.27 dex) and V-band extinction (0.27 dex). These systematic values were added in quadrature to our uncertainties calculated from FAST. We have presented these results alongside our own. If a systematic uncertainty specifically for FAST is found in future studies then we can implement it.

Figures 3.7 and 3.8 show the uncertainty measurements on log (M_*) for each of the simulated magnitudes and redshift for the elliptical and Sc respectively. We compare the uncertainties with and without the inclusion of spectroscopy. We see an improvement to the precision of the measured host-galaxy M_* when spectroscopy is included, at all simulated magnitudes and redshifts. A dashed black line across the plot represents the locus of points where $\log(\text{mass}/\text{M}_\odot) = 10$. The 10^{10} stellar mass is the typical location of the mass step, as discussed in Section 1.1.3. This was found using Figure 3.5 and reading off the magnitude and redshift of a 10^{10} $M_\odot$ galaxy on each subplot. Note that the photometry values are calculated using the 10-year LSST depth. 4MOST spectroscopy will produce considerably better results than the intermediate photometry, until LSST reaches the 10th year of the survey.

In addition to the improvement to the precision of host-galaxy mass, all other galaxy parameters saw an improvement for phot + 4MOST compared with photometry alone. The results for an elliptical host-galaxy star formation rate can be seen in Figure 3.9. There is a clear improvement for all magnitudes brighter than 23, and for redshifts 0.3, 0.5 and 0.7. However, at fainter magnitudes the uncertainty for phot + 4MOST is a similar size to the phot results. This is also true for redshift 0.1. A value for the improvement to star formation rate uncertainties is difficult to calculate because for photometry alone the lower value often hits FAST's lower limit of -99. At redshift 0.1 the uncertainty for phot + 4MOST is unconstrained due to a lack of models matching our galaxy in this parameter space. For the Sc galaxy we found that FAST was not programmed to measure the star formation rate from emission lines.

Uncertainties in age see improvements at all simulated redshifts and magnitudes. The results for this can be seen in Figure 3.10. For phot + 4MOST, the uncertainty on the measurement of age was reduced by 56 – 86% for magnitudes $\leq$ 22, while at magnitudes fainter than 22 the uncertainty was reduced by 6 – 77%. The improvements to age could prove to be a crucial detail that 4MOST spectra will be able to provide when calculating galaxy properties. Stellar absorption fea-

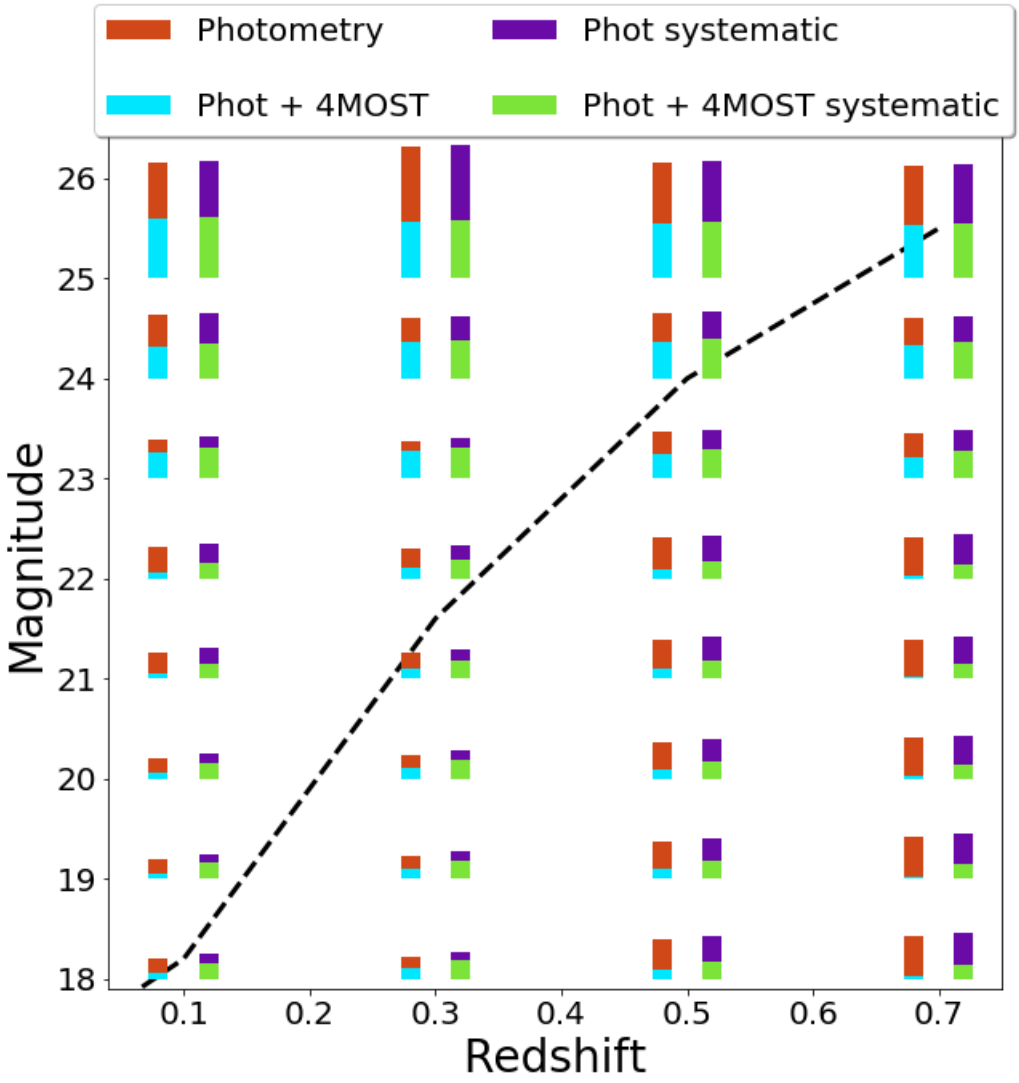

Figure 3.7: A comparison of the uncertainty on log(mass/$M_{\odot}$) for a an elliptical galaxy for a collection of redshifts and magnitudes when phot + 4MOST are used together can be seen in light blue. The additional uncertainty measured when using photometry only can be seen in dark orange. The size of the line is proportional to the size of the uncertainty. The dashed black line shows where the M_* of the galaxy is 10^{10} $M_{\odot}$. Finally, the systematic uncertainty from using SED fitting is added in quadrature to be compared with the uncertainty values. The phot + 4MOST with systematic uncertainties is shown in green. The additional uncertainty for phot with systematic uncertainties is shown in purple.

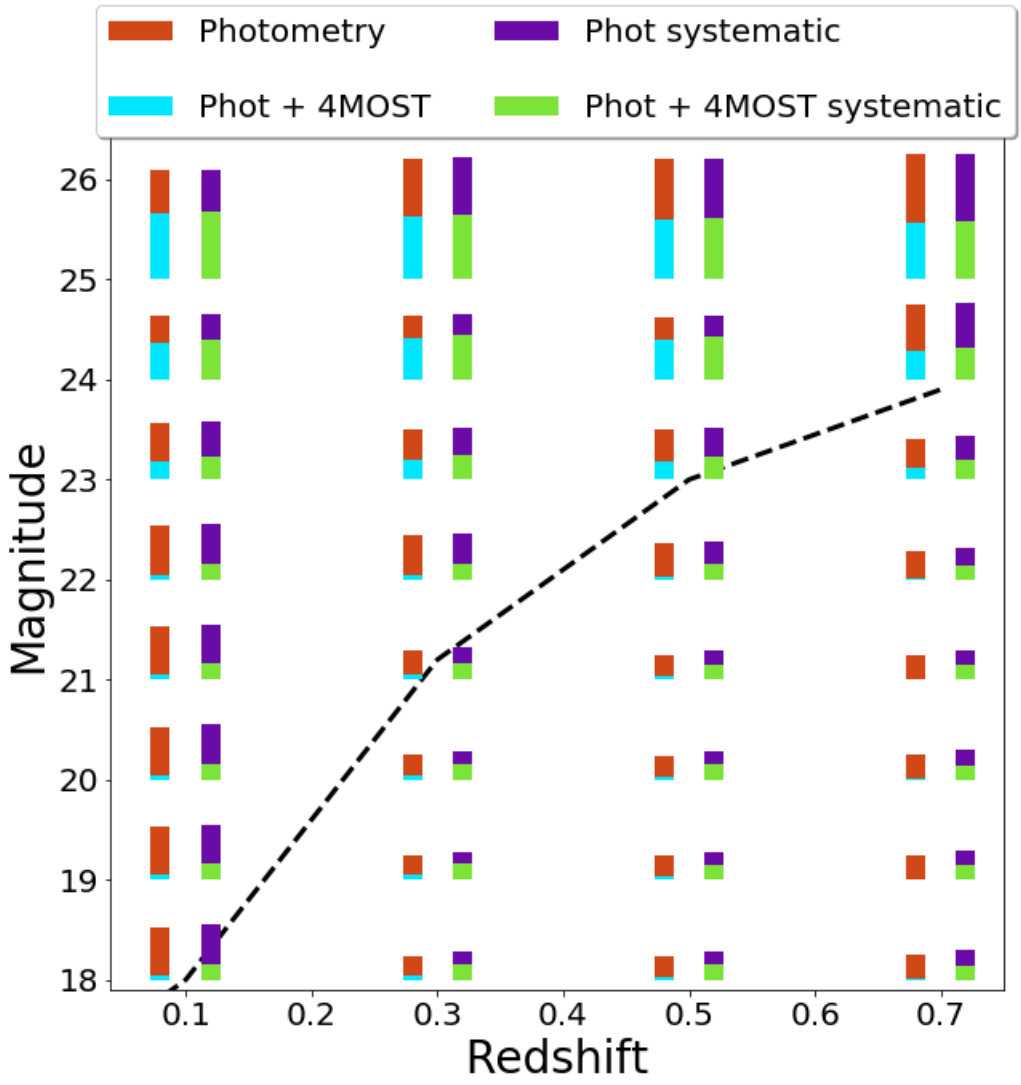

Figure 3.8: The same as Figure 3.7 but for a Sc galaxy.

tures are an indication of age, most notably the hydrogen Balmer lines (Serra & Trager, 2007), which photometry is not able to capture (Salim et al., 2007).

Star formation rate and host-galaxy mass are the most commonly used parameters when making corrections in SNe cosmology (as discussed in Section 1.1.3). However all of the galaxy properties studied here may be of interest in the future because it is still unknown which galaxy parameters drive the correlations with SN properties. The work of Gupta et al. (2011) found that after light curve corrections over-luminous SNe Ia tend to occur in older stellar populations. The work of Rose et al. (2021) found including age as a parameter improved the ability to standardise Type Ia SNe. The results of the study of the remaining galaxy properties calculated by FAST can be seen in the Additional results section. The improved precision of galaxy property measurements could prove to be useful in the future at breaking degeneracies for SNe host-galaxy properties, such as age and metallicity (Worthey, 1994; Walcher et al., 2011). The key addition that 4MOST will provide is a large quantity of spectra with this level of quality. One limitation of our results is that FAST does not capture all of the information that a spectrum contains, such as emission lines. TiDES will be able to make direct measurements of emission lines, without fitting templates. For example, the H-alpha line will provide a further indication of star formation rate.

3.3 Discussion

Our work has shown that SNe host-galaxy M_* can be measured more precisely when 4MOST and the Vera Rubin Observatory are used together, compared with the Rubin observatory alone. We expect to be able to improve the mass step correction by reducing the uncertainty as to which side of the $10^{10} M_\odot$ line a SN host galaxy belongs.

To quantify the extent of the improvement we have used the Supernova Legacy Survey third year data set (Balland et al., 2009), used by Sullivan et al. (2010), and the DES 3 year data (Smith et al., 2020) to create a mass redshift distribution. While newer data sets could also have been used, these two data sets contain host-galaxy magnitudes, are readily available and are large enough to demonstrate our improvements in estimates of the host-galaxy mass. Balland et al. report magnitudes of each of their SNe host-galaxies in i-band. Since our uncertainties correspond to r-band magnitudes, a conversion was required. To calculate the conversion, we used the integrated fluxes of our shifted target spectra through the r and i pass band (see Section 3.1.4) to determine the colour as a function of redshift. We did this for the elliptical galaxy template, used in the analysis presented up to this point, and also a Sc galaxy template. For each SNe in Balland

et al. we assigned a host-galaxy type at random, weighted by the rate at which SNe Ia appear in elliptical and Sc galaxies, as provided by Hakobyan et al. (2012). The i-band magnitudes for the Balland et al. SNe host galaxies are converted to r-band using the colours calculated for the assigned galaxy type. The r-band magnitude of each galaxy was then rounded to the nearest integer value. The DES data provided r-band magnitudes so did not need any additional work. Each host galaxy was assigned mass uncertainties corresponding to its redshift and r-band magnitude bins, as previously shown in Figures 3.5 and 3.6. Mass uncertainties were assigned for phot and phot + 4MOST, to enable a comparison. Host galaxies with redshift of 0.8 or greater were removed and any galaxies with r band magnitudes greater than 25.5 were also removed. Host galaxies at higher redshifts and fainter magnitudes could not be used as the uncertainties start to increase rapidly for both LSST and 4MOST. A separate study of deep fields would be needed to determine the performance of our techniques for the faintest host galaxies.

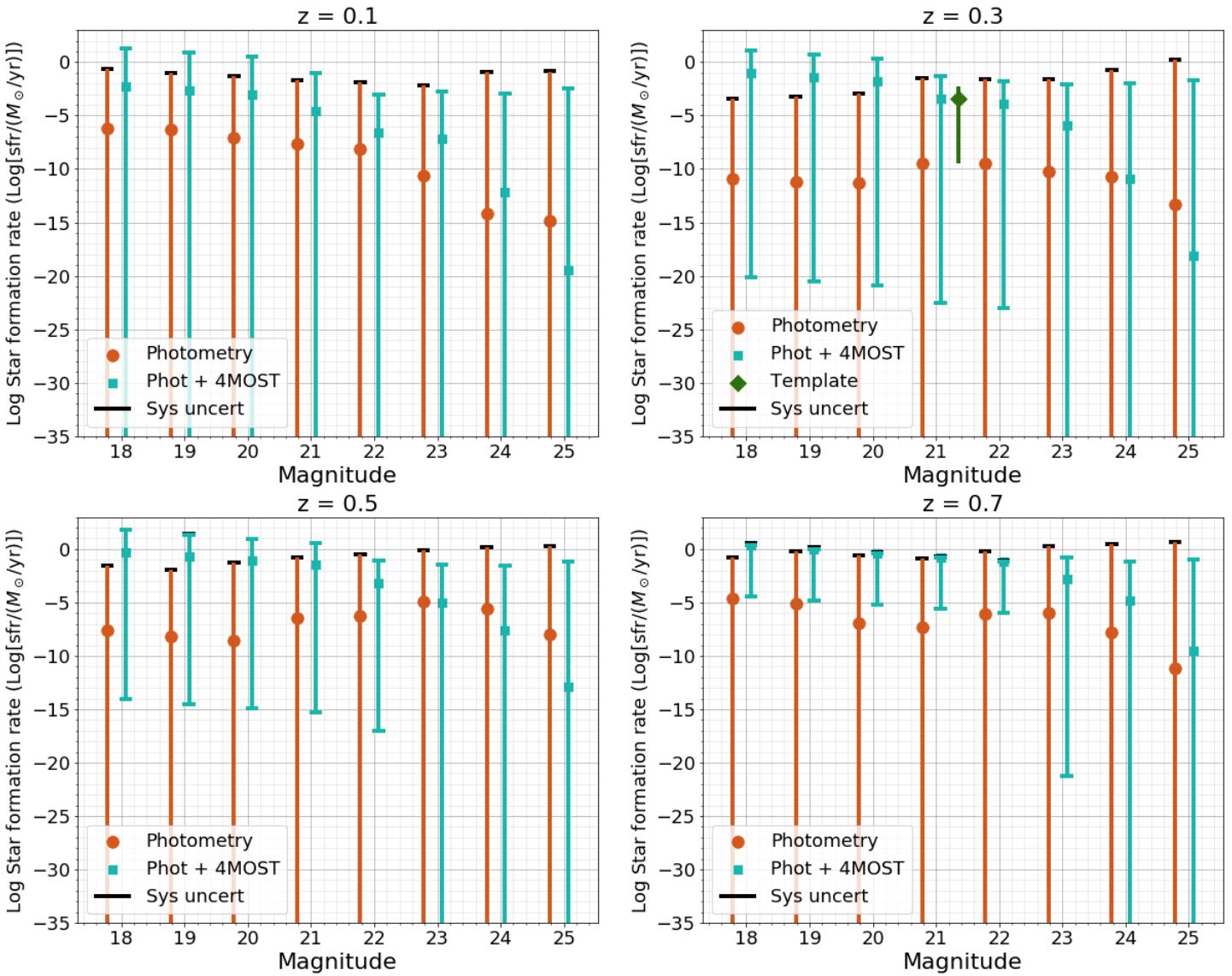

Figure 3.9: The simulated log Star Formation Rate (SFR) of an elliptical galaxy as a function of magnitude and redshift. The symbols and colours are the same as used in Figure 3.5. There is a clear improvement when 4MOST is used, compared with when photometry is used alone. This is true for all magnitudes brighter than 23, and redshifts 0.3, 0.5 and 0.7. However, at fainter magnitudes the uncertainty for 4MOST with photometry is a similar size to the photometry alone results. This is also true for redshift 0.1. A value for the improvement to star formation rate uncertainties is difficult to calculate as with photometry alone the lower value often hits FAST's lower limit of -99.

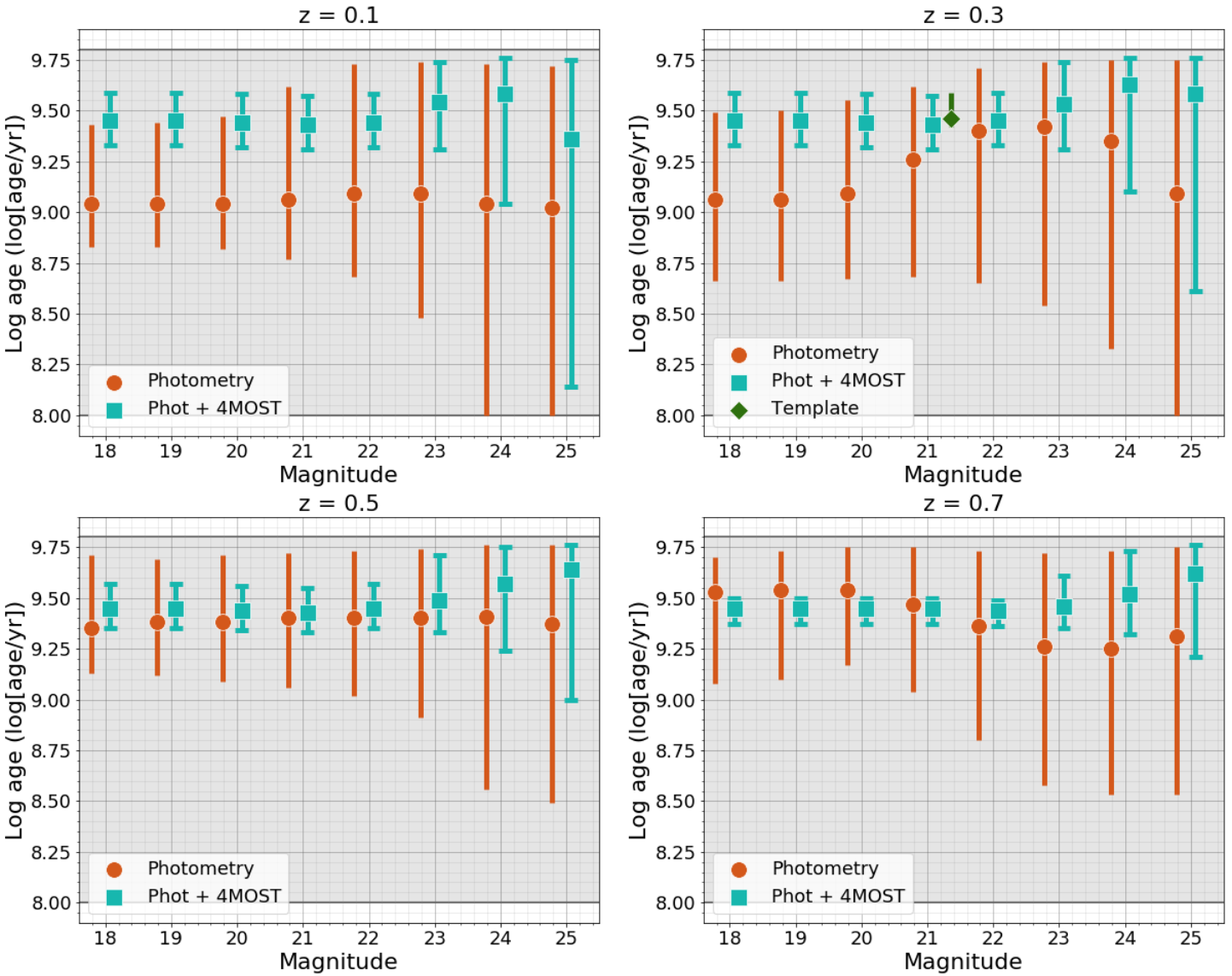

Figure 3.10: Simulated log age of of an elliptical galaxy as a function of magnitude and redshift. The precision increases drastically when 4MOST spectrum is used with photometry, compared with photometry alone. The limits between which the age can be found by FAST are shown by the grey shaded region. The limits are set at 8 and 9.8 (log[age/yr]).

After the cuts, we were left with a sample of 310 host galaxies. The results can be seen in Figure 3.11, where the top panel shows the host galaxies with the uncertainty produced from the photometry. The bottom panel shows the host-galaxies with the uncertainty produced by phot + 4MOST. The black dashed line on both plots is 10^{10} (mass/$M_{\odot}$), the divide for the mass step. With the photometry uncertainty values, 35 of the host-galaxies cross the mass step line, making it unclear which correction term would need to be applied. With the uncertainty values from phot + 4MOST, 13 host-galaxies cross the mass step line. This results in a 7% improvement of SNe from the sample of 310 having the true correction applied. This shows that the uncertainty produced by phot + 4MOST would lead to a more accurate correction being applied than if only the photometry is used. It should be noted that we have assumed all host galaxies have spectra observed by 4MOST. The TiDES-Hosts sub-survey aims to capture spectra of SNe host galaxies, for which live observations were not captured (Swann et al., 2019). This will result in a significant number of SN host galaxies having spectra, which we will be able to select for when using results from 4MOST.

We find on average our photometry mass uncertainty ranges are larger than those observed by Sullivan et al. (2010). This is to be expected as Sullivan et al. incorporates near-infrared filters (J, H, K_s) with photometry to measure a galaxy's mass. At magnitude 20 the SNLS data has an average mass uncertainty size of 0.12 dex, whilst our photometry uncertainty has a size of 0.19 dex. The phot + 4MOST uncertainty for magnitude 20 has a size of 0.08 dex. This further shows the improvement that 4MOST will enable. A similar trend occurs for magnitudes 21 and 22, with phot + 4MOST having a smaller uncertainty range than SNLS and phot only having a larger range. However, at magnitude 23 and beyond the 4MOST uncertainties become larger than the SNLS values (SNLS: 0.1 dex, phot: 0.68 dex, phot + 4MOST: 0.34 dex). Once again this is expected, as magnitude 23 is when the signal-to-noise for 4MOST becomes larger.

A recent study by Galbany et al. (2022) discussed aperture corrections to counteract effects caused by fixed-aperture fibre spectroscopy on host-galaxy correlations, such as the mass step. In this work we have effectively assumed that the magnitude of our host galaxies corresponds to the light which entered a 4MOST fibre. Using the galaxy catalogue published by Karachentsev et al. (2004) we randomly selected several galaxies to calculate their size. We calculated a range of sizes between 2.48 arcseconds and 53.10 arcseconds. We previously stated that each 4MOST fibre will have a diameter of 1.45 arcseconds. As we selected objects randomly, there may be smaller and larger galaxies within the catalogue. We used the catalogue created by Karachentsev et al. (2004) as it was readily accessible. However the catalogue only contained objects at redshift < 0.02. When we have access to real images from LSST we will be able to calculate the amount of light

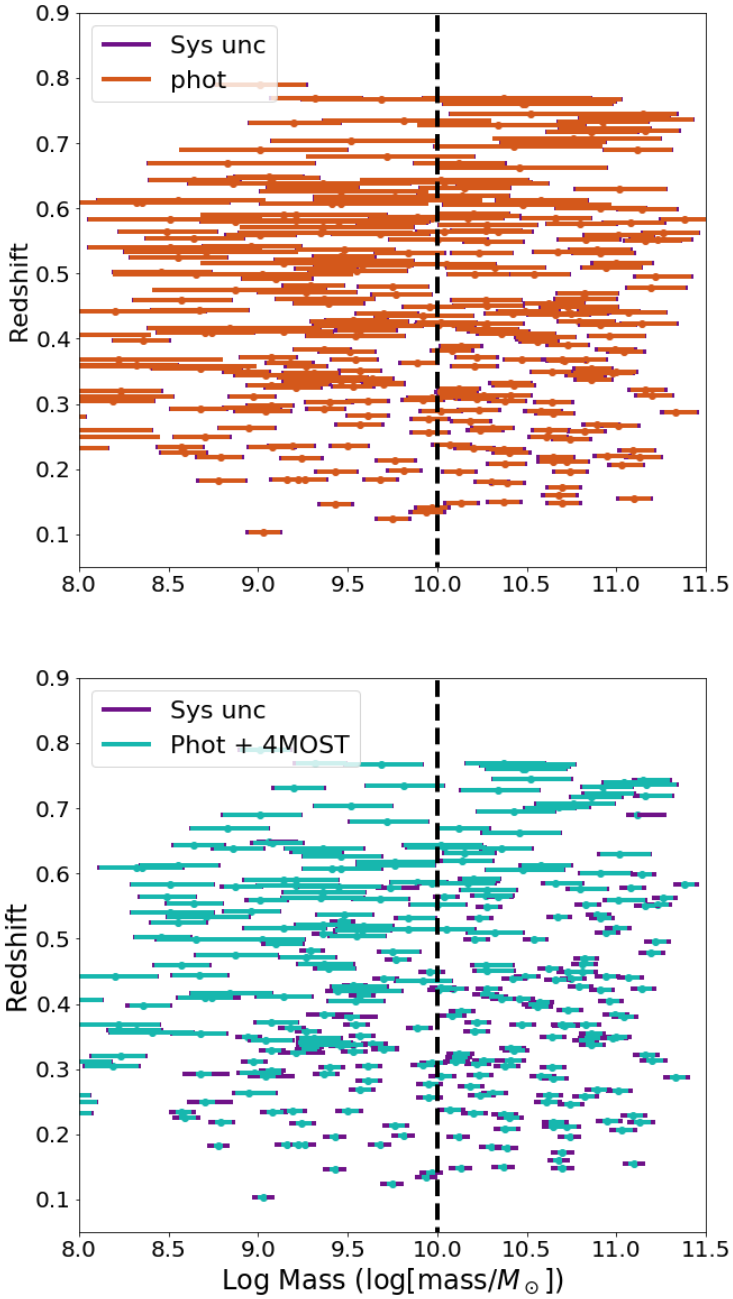

Figure 3.11: The mass-redshift distribution graph is created with our uncertainties. The SNe were taken from the Supernova Legacy Survey third year data set and the DES 3 year data. As our previous work had calculated uncertainties in redshift increments, each SN was assigned a bin to match these increments. Each SN was then assigned an uncertainty calculated from phot (in the top panel) and an uncertainty calculated from phot + 4MOST (in the bottom panel). Both plots have a black dashed line representing 10^{10} ($M_\odot$). There are 310 SNe plotted. The uncertainty values from the photometry lead to 35 SNe crossing the dashed line. Whilst, the uncertainty values from phot + 4MOST lead to 13 SNe crossing the dashed line.

going down each fibre, and therefore apply aperture corrections.

3.3.1 Analysing the effect of our work to cosmology

We now estimate the impact of our results on measurements of cosmological parameters. We investigated this by exploring the impact our results would have on the measurement of w. Measurements of w are sensitive to small changes in distance modulus; a 2% change in w corresponds to a change in distance modulus of only 2.6 millimagnitudes at $z = 0.3$ (assuming a flat ΛCDM cosmology and all other parameters remain constant). We aim to determine how much our changes of mass uncertainties can improve scatter around the best-fit of the distance-redshift relation. We parameterise this using the root mean squared (RMS) scatter of the data points around the Hubble residual = 0 line.

We calculated the Hubble residual of the corresponding SN for each of our host galaxies in Figure 3.11 which appear in the Pantheon compilation (Scolnic et al., 2022), leaving us with 286 objects. The Hubble residual was calculated using a modified version of the Tripp formula (Tripp, 1998), presented by Brout et al. (2019), as shown as follows:

$$\text{Hubble Residual} = m_{\text{B,fix}} - \text{distance modulus}(z) \tag{3.4}$$

$$\text{Where } m_{\text{B,fix}} = (m_B + \alpha x_1 - \beta c) - M_{B,SNIa} \tag{3.5}$$

where m_B is the peak apparent magnitude in the rest frame B-band. α and β are correlation coefficients of x_1 and c. $M_{B,SNIa}$ is the mean absolute magnitude. This value was not provided in the Pantheon+ paper so we calculated it from the weighted average of the m_b_corr column (from Pantheon) minus the distance modulus, weighted with the MU_SHOES_ERR_DIAG column (from Pantheon). The values for m_B, x_1 and c were taken from the Pantheon dataset. The values for α and β were set as 0.1533 and 3.44, respectively (taken from Brout et al. (2022)). The distance modulus was calculated using the distmod function in the Python module Astropy. We then calculated the probability of an object appearing either side of the 10^{10} (mass/$M_\odot$) line using Gaussian uncertainties, assuming our 1σ values from FAST, centered on our best-fit M_* values. This was repeated using our photometry uncertainties, and again for the phot + 4MOST uncertainties. We used the distance of each galaxy from the 10^{10} (mass/$M_\odot$) line and a probability distribution to assign a probability of each host galaxy crossing the line. Each galaxy was then randomly assigned a right/wrong classification term, weighted by the probabilities for a galaxy to be given the right or wrong correction term.

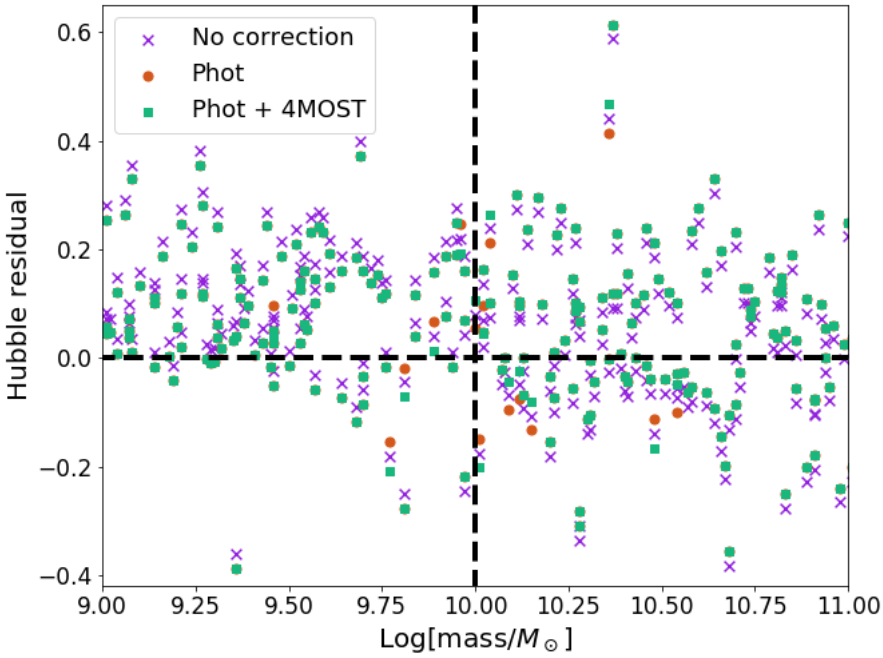

Figure 3.12: A diagram to illustrate the effect of the improved precision of galaxy stellar mass on Hubble residual. For each object in our selected sample three corrections are plotted, based on the stellar mass of the SN host galaxy. The purple crosses indicate no correction, the orange circles represent corrections assigned from photometric mass estimates, and the green squares show corrections assigned from photometric + 4MOST mass estimates. In most cases, the corrections based on phot and phot + 4MOST are the same. The orange circles are visible when different corrections have been applied, otherwise the green squares obscure them. As expected, near the $10^{10}M_\odot$ line it is more common to have disagreements between the correction terms applied. See Section 3.3.1 for more detail.

The host galaxies with $M_* < 10^{10}M_\odot$ were assigned a correction of -0.0265 mag (taken from Brout et al. 2022) whilst galaxies with $M_* >= 10^{10}M_\odot$ were assigned a correction of $+0.0265$ mag. This occurred for galaxies which were assigned the "right" correction term, else they were given the opposite correction value. Note that for simplicity we assumed a step function, whereas Brout et al. assumed a sigmoid function, as described in Brout & Scolnic (2021), which is almost identical.

Our results are shown in Figure 3.12. As expected only those host galaxies closest to the $10^{10}M_\odot$ line are affected. When the phot uncertainty is used to assign the mass correction, more points are assigned the wrong correction than when the phot + 4MOST uncertainties are used. In the figure, when the two corrections disagree the orange circles can be seen. When the corrections agree, the orange circles are obscured by the green squares.

To quantify the effect of the different corrections we calculated the RMS around the Hubble residual = 0 line. With no mass correction applied an RMS of 0.1752 mag was measured. When the mass correction calculated from the phot uncertainties were applied an RMS of 0.1720 mag was calculated. Finally, when the phot + 4MOST uncertainties were used to apply the mass correction, an RMS of 0.1715 mag was calculated. This is only a small improvement but represents the minimum improvement that 4MOST will be able to achieve to the mass correction. The SNe in our sample were taken from DES and SNLS, which observed each SN host galaxy with photometry. Hence, there could be ingrained uncertainty as to which side of the $10^{10} M_{\odot}$ mass line the host galaxy should be. To account for this we altered the way the mass correction is applied. As before, we assigned a right/wrong binary classification to decide if a galaxy should be given the right or wrong correction. We again used the Gaussian errors from phot and phot + 4MOST to assign the right/wrong classification. We then applied a correction based on the sign of the Hubble residual. All galaxies with a Hubble residual $>$ 0 were given a correction of -0.0265 mag and all galaxies with a Hubble residual < 0 were assigned a correction of $+0.0265$ mag. The same as for the mass-based correction, galaxies assigned with the "right" correction term were given these corrections, else they were given the opposite correction value. The RMS was calculated again and found to be 0.1577 mag when the correction was applied using photometry uncertainties. The RMS was found to be 0.1566 mag when phot + 4MOST uncertainties were used. This represents the best improvement 4MOST will be able to accomplish. All of the RMS results are summarised in Table 3.6. We estimate the improvement to the RMS is between 0.0005 - 0.001 mag, which in the best case scenario would result in a 2% improvement to the measurement uncertainty of w. We recognise that this is not significant to zeroth order, when calculating the mass step correction in this way. However, it is possible that such effects might be important when considering changes in w with redshift (for example in $w_0 - w_a$ models), due to evolution of the host-galaxy population with redshift. We have not studied this effect.

The results from our research will also be used to understand whether improved galaxy properties can aid classification of SNe Ia. This work is carried out in Chapter 4 In addition, the results from our research can be improved with the addition of data from future instruments. ESA's *Euclid* satellite (Laureijs et al., 2011) will provide near-infrared photometry for billions of galaxies over a wide area of the sky, and will begin collecting data in the coming months. The *ULTRASAT* satellite will provide the first wide-field ultraviolet time-domain sky survey and is due to launch in 2026 (Ben-Ami et al., 2022). We expect using Euclid NIR photometry and ULTRASAT UV images, together with 4MOST and LSST data, will significantly improve the estimation of stellar mass and other galaxy properties.

Correction	RMS Hubble residual (mag)	RMS Hubble residual from M_* (mag)
None	0.1752	0.1752
Phot	0.1720	0.1577
Phot + 4MOST	0.1715	0.1566

Table 3.2: The calculated RMS scatter values from 3 corrections: no correction, using the phot uncertainty and using phot + 4MOST uncertainty. The first column was calculated by measuring the scatter around the Hubble residual = 0 line, when a correction is applied based on which side of the $10^{10} M_{\odot}$ mass line a SN's host galaxy appears. These values are the minimum improvement which can be achieved. The second column is similar, but has a correction applied based on the sign of the Hubble residual, which is an attempt to account for ingrained uncertainty in the DES and SNLS host measurements. The scatter around the Hubble residual = 0 line is again measured. The second column also represents the maximum possible improvement. From these values, we estimate the improvement to the RMS is between 0.0005 - 0.001 mag.

3.4 Conclusion

We have shown that 4MOST can work in conjunction with imaging telescopes, such as the Vera Rubin observatory, to calculate galaxy properties more precisely than those derived from photometry alone. This was an expected result which we have now quantified. Our results are summarised as follows:

- For elliptical galaxies with brighter magnitudes ($r \leq 22$) the uncertainty for a galaxy's log (M_*) is 49 – 95% smaller when spectroscopy is used with photometry, compared with that derived from photometry alone. The range of improvements depends on the magnitude and redshift.

- At fainter magnitudes ($r \geq 23$) the M_* uncertainty is 24 – 71% smaller for phot + 4MOST, compared to photometry alone. The range depends on magnitude and redshift, with more improvements seen for brighter galaxies.

- We see similar improvements to the precision of Sc host-galaxy masses, when adding 4MOST spectroscopy.

- We applied our uncertainties derived from FAST to real SNe host-galaxy masses. The smaller uncertainties produced when adding 4MOST spectroscopy make it easier to distinguish which side of the 10^{10} ($M_{\odot}$) line a host galaxy falls. This has implications for applying the correct mass step corrections in cosmological analysis.

- Other galaxy properties see a significant improvement in uncertainties, including: star formation rate, age, V-band extinction, metallicity, specific star formation rate and star formation timescale.
- Whilst there is not as significant an improvement to the precision of a galaxy's metallicity, there is still an improvement at most simulated magnitudes and redshifts.

The correlations between SNe Ia peak brightness and host-galaxy properties are one of the main systematic effects in SNe Ia cosmology. The improved galaxy property measurements from 4MOST and Rubin have the potential to improve the corrections used in SN cosmology. We have chosen to focus on the host-galaxy mass and the implications to the mass-step, but a deeper investigation could be carried out with all of the galaxy properties. The galaxy-property precision improvements come from the fact that spectra contain significantly more information than photometry, which leads to the breaking of degeneracies.

3.5 Additional results

r-band magnitude	redshift	Template log(mass) (log[mass/$M_\odot$])	Phot log(mass)	Phot and spectroscopy log(mass)	Phot with Systematic error	Phot and spectroscopy With systematic error
18	0.1	-	$10.19^{10.28}_{10.11}$	$10.25^{10.27}_{10.22}$	$10.19^{10.30}_{10.09}$	$10.25^{10.31}_{10.18}$
19	0.1	-	$9.79^{9.88}_{9.71}$	$9.85^{9.87}_{9.82}$	$9.79^{9.90}_{9.69}$	$9.85^{9.91}_{9.78}$
20	0.1	-	$9.40^{9.49}_{9.32}$	$9.45^{9.47}_{9.42}$	$9.40^{9.51}_{9.32}$	$9.45^{9.51}_{9.38}$
21	0.1	-	$9.02^{9.14}_{8.92}$	$9.05^{9.07}_{9.02}$	$9.02^{9.15}_{8.90}$	$9.05^{9.11}_{8.98}$
22	0.1	-	$8.64^{8.79}_{8.53}$	$8.65^{8.67}_{8.62}$	$8.64^{8.80}_{8.51}$	$8.65^{8.71}_{8.58}$
23	0.1	-	$8.25^{8.42}_{8.09}$	$8.27^{8.43}_{8.21}$	$8.25^{8.43}_{8.08}$	$8.27^{8.44}_{8.19}$
24	0.1	-	$7.81^{8.05}_{7.52}$	$7.91^{8.04}_{7.77}$	$7.81^{8.06}_{7.51}$	$7.91^{8.05}_{7.76}$
25	0.1	-	$7.32^{7.86}_{6.89}$	$7.46^{7.68}_{7.18}$	$7.32^{7.86}_{6.89}$	$7.46^{7.69}_{7.17}$
18	0.3	11.46	$11.43^{11.55}_{11.37}$	$11.45^{11.51}_{11.42}$	$11.43^{11.56}_{11.35}$	$11.45^{11.53}_{11.38}$
19	0.3	11.05	$11.04^{11.15}_{10.96}$	$11.05^{11.11}_{11.02}$	$11.04^{11.16}_{10.94}$	$11.05^{11.13}_{10.98}$
20	0.3	10.65	$10.65^{10.76}_{10.56}$	$10.65^{10.71}_{10.62}$	$10.65^{10.78}_{10.54}$	$10.65^{10.72}_{10.58}$
21	0.3	$\star\ 10.25^{10.26}_{10.24}\ \star$	$10.26^{10.37}_{10.15}$	$10.25^{10.31}_{10.22}$	$10.26^{10.39}_{10.13}$	$10.25^{10.33}_{10.18}$
22	0.3	9.85	$9.88^{10.00}_{9.75}$	$9.85^{9.91}_{9.82}$	$9.88^{10.01}_{9.74}$	$9.85^{9.93}_{9.78}$
23	0.3	9.44	$9.52^{9.66}_{9.35}$	$9.48^{9.63}_{9.40}$	$9.52^{9.67}_{9.34}$	$9.48^{9.64}_{9.38}$
24	0.3	9.04	$9.13^{9.36}_{8.86}$	$9.14^{9.28}_{8.98}$	$9.13^{9.37}_{8.85}$	$9.14^{9.29}_{8.97}$
25	0.3	8.64	$8.68^{9.26}_{8.16}$	$8.76^{8.99}_{8.52}$	$8.68^{9.26}_{8.16}$	$8.76^{9.00}_{8.51}$
18	0.5	-	$12.13^{12.38}_{12.05}$	$12.20^{12.26}_{12.18}$	$12.13^{12.39}_{12.03}$	$12.20^{12.28}_{12.14}$
19	0.5	-	$11.75^{11.96}_{11.65}$	$11.80^{11.86}_{11.78}$	$11.75^{11.97}_{11.63}$	$11.80^{11.88}_{11.74}$
20	0.5	-	$11.37^{11.57}_{11.26}$	$11.40^{11.46}_{11.38}$	$11.37^{11.58}_{11.24}$	$11.40^{11.48}_{11.34}$
21	0.5	-	$10.99^{11.19}_{10.86}$	$11.00^{11.06}_{10.98}$	$10.99^{11.20}_{10.85}$	$11.00^{11.08}_{10.94}$
22	0.5	-	$10.62^{10.81}_{10.47}$	$10.60^{10.66}_{10.58}$	$10.62^{10.82}_{10.46}$	$10.60^{10.68}_{10.54}$
23	0.5	-	$10.28^{10.48}_{10.09}$	$10.22^{10.37}_{10.16}$	$10.28^{10.49}_{10.08}$	$10.22^{10.38}_{10.14}$
24	0.5	-	$9.97^{10.25}_{9.70}$	$9.88^{10.04}_{9.73}$	$9.97^{10.26}_{9.69}$	$9.88^{10.05}_{9.72}$
25	0.5	-	$9.59^{10.10}_{9.13}$	$9.55^{9.79}_{9.33}$	$9.59^{10.10}_{9.13}$	$9.55^{9.80}_{9.32}$

Table 3.3: The best-fit log(M_*) determined by FAST for an elliptical galaxy using photometry and 4MOST with photometry at various magnitudes and redshifts. The log M_* maximum is the 1-σ upper limit on M_* determined by FAST. Likewise, the log (M_*) minimum is the 1-σ lower limit. The ranges reported are the 68th percentile ranges from FAST. A maximum uncertainty was introduced for 4MOST objects with a signal-to-noise that does not reach 3, to prevent the flux values from being unrealistically small. The first column shows our template values. The value at magnitude 21, z=0.3 (highlighted with stars) was calculated as described in section 3.1.2. The remaining values do not have an uncertainty as the values were extrapolated from the calculated value at magnitude 21. The last two columns of the table show our results with a systematic error added in quadrature, based on the estimates reported by Pacifici et al. (2023).

r-band magnitude	redshift	Template log(mass) (log[mass/$M_\odot$])	Phot log(mass)	Phot and spectroscopy log(mass)	Phot with Systematic error	Phot and spectroscopy With systematic error
18	0.7	-	$12.98^{13.06}_{12.70}$	$12.88^{12.88}_{12.86}$	$12.98^{13.08}_{12.69}$	$12.88^{12.94}_{12.82}$
19	0.7	-	$12.58^{12.67}_{12.32}$	$12.48^{12.48}_{12.46}$	$12.58^{12.69}_{12.31}$	$12.48^{12.54}_{12.42}$
20	0.7	-	$12.18^{12.30}_{11.96}$	$12.08^{12.08}_{12.06}$	$12.18^{12.31}_{11.95}$	$12.08^{12.14}_{12.02}$
21	0.7	-	$11.79^{11.92}_{11.59}$	$11.68^{11.68}_{11.66}$	$11.79^{11.93}_{11.58}$	$11.68^{11.74}_{11.62}$
22	0.7	-	$11.43^{11.58}_{11.23}$	$11.28^{11.28}_{11.26}$	$11.43^{11.59}_{11.22}$	$11.28^{11.34}_{11.22}$
23	0.7	-	$11.05^{11.24}_{10.86}$	$10.88^{11.03}_{10.85}$	$11.05^{11.25}_{10.85}$	$10.88^{11.04}_{10.81}$
24	0.7	-	$10.70^{10.96}_{10.45}$	$10.54^{10.71}_{10.43}$	$10.70^{10.97}_{10.44}$	$10.54^{10.72}_{10.41}$
25	0.7	-	$10.36^{10.85}_{9.91}$	$10.22^{10.45}_{10.01}$	$10.36^{10.85}_{9.91}$	$10.22^{10.46}_{10.00}$

Table 3.3: Table 3.3 continued to show redshift 0.7 values.

r-band magnitude	redshift	Template log(mass) (log[mass/$M_\odot$])	Phot log(mass)	Phot and spectroscopy log(mass)	Phot with Systematic error	Phot and spectroscopy With systematic error
18	0.1	-	$10.12^{10.26}_{9.82}$	$10.26^{10.26}_{10.22}$	$10.12^{10.27}_{9.81}$	$10.26^{10.32}_{10.19}$
19	0.1	-	$9.72^{9.86}_{9.42}$	$9.86^{9.86}_{9.82}$	$9.72^{9.87}_{9.41}$	$9.86^{9.92}_{9.79}$
20	0.1	-	$9.32^{9.46}_{9.02}$	$9.46^{9.46}_{9.42}$	$9.32^{9.47}_{9.01}$	$9.46^{9.52}_{9.39}$
21	0.1	-	$8.90^{9.05}_{8.61}$	$9.06^{9.06}_{9.02}$	$8.90^{9.06}_{8.60}$	$9.06^{9.12}_{8.99}$
22	0.1	-	$8.47^{8.65}_{8.20}$	$8.66^{8.66}_{8.62}$	$8.47^{8.66}_{8.19}$	$8.66^{8.72}_{8.59}$
23	0.1	-	$8.03^{8.23}_{7.76}$	$8.24^{8.30}_{8.15}$	$8.03^{8.24}_{7.75}$	$8.24^{8.32}_{8.13}$
24	0.1	-	$7.54^{7.82}_{7.29}$	$7.77^{7.89}_{7.58}$	$7.54^{7.83}_{7.28}$	$7.77^{7.90}_{7.57}$
25	0.1	-	$7.10^{7.60}_{6.69}$	$7.23^{7.48}_{6.93}$	$7.10^{7.60}_{6.67}$	$7.23^{7.49}_{6.93}$
18	0.3	11.46	$11.38^{11.46}_{11.26}$	$11.41^{11.41}_{11.37}$	$11.38^{11.48}_{11.25}$	$11.41^{11.47}_{11.34}$
19	0.3	10.95	$10.98^{11.06}_{10.86}$	$11.01^{11.01}_{10.97}$	$10.98^{11.08}_{10.85}$	$11.01^{11.07}_{10.94}$
20	0.3	10.55	$10.58^{10.66}_{10.45}$	$10.61^{10.61}_{10.57}$	$10.58^{10.68}_{10.44}$	$10.61^{10.67}_{10.54}$
21	0.3	$\star\, 10.18^{10.19}_{10.13}\, \star$	$10.17^{10.26}_{10.02}$	$10.21^{10.21}_{10.17}$	$10.17^{10.28}_{10.01}$	$10.21^{10.27}_{10.14}$
22	0.3	9.75	$9.73^{9.86}_{9.49}$	$9.81^{9.81}_{9.77}$	$9.73^{9.87}_{9.48}$	$9.81^{9.87}_{9.74}$
23	0.3	9.34	$9.22^{9.42}_{9.00}$	$9.41^{9.46}_{9.30}$	$9.22^{9.43}_{8.99}$	$9.41^{9.49}_{9.28}$
24	0.3	8.93	$8.73^{9.00}_{8.47}$	$8.93^{9.07}_{8.72}$	$8.73^{9.01}_{8.46}$	$8.93^{9.08}_{8.71}$
25	0.3	8.54	$8.27^{8.81}_{7.80}$	$8.39^{8.63}_{8.11}$	$8.27^{8.81}_{7.80}$	$8.39^{8.64}_{8.10}$
18	0.5	-	$12.07^{12.15}_{11.95}$	$12.07^{12.07}_{12.04}$	$12.07^{12.17}_{11.94}$	$12.07^{12.13}_{12.00}$
19	0.5	-	$11.67^{11.75}_{11.55}$	$11.67^{11.67}_{11.64}$	$11.67^{11.77}_{11.54}$	$11.67^{11.73}_{11.60}$
20	0.5	-	$11.27^{11.35}_{11.15}$	$11.27^{11.27}_{11.24}$	$11.27^{11.37}_{11.14}$	$11.27^{11.33}_{11.20}$
21	0.5	-	$10.86^{10.95}_{10.74}$	$10.87^{10.87}_{10.84}$	$10.86^{10.97}_{10.73}$	$10.87^{10.93}_{10.80}$
22	0.5	-	$10.44^{10.56}_{10.26}$	$10.47^{10.47}_{10.44}$	$10.44^{10.57}_{10.25}$	$10.47^{10.53}_{10.40}$
23	0.5	-	$9.95^{10.13}_{9.71}$	$10.07^{10.13}_{9.98}$	$9.95^{10.14}_{9.70}$	$10.07^{10.15}_{9.96}$
24	0.5	-	$9.44^{9.69}_{9.17}$	$9.62^{9.76}_{9.43}$	$9.44^{9.70}_{9.16}$	$9.62^{9.77}_{9.43}$
25	0.5	-	$8.97^{9.44}_{8.44}$	$9.08^{9.31}_{8.81}$	$8.97^{9.44}_{8.44}$	$9.08^{9.32}_{8.80}$
18	0.7	-	$12.60^{12.66}_{12.45}$	$12.62^{12.62}_{12.61}$	$12.60^{12.68}_{12.44}$	$12.62^{12.68}_{12.56}$
19	0.7	-	$12.20^{12.26}_{12.05}$	$12.22^{12.22}_{12.21}$	$12.20^{12.28}_{12.03}$	$12.22^{12.28}_{12.16}$
20	0.7	-	$11.80^{11.86}_{11.65}$	$11.82^{11.82}_{11.81}$	$11.80^{11.88}_{11.64}$	$11.82^{11.88}_{11.76}$
21	0.7	-	$11.40^{11.46}_{11.25}$	$11.42^{11.42}_{11.41}$	$11.40^{11.48}_{11.24}$	$11.42^{11.48}_{11.36}$
22	0.7	-	$11.00^{11.08}_{10.84}$	$11.02^{11.02}_{11.01}$	$11.00^{11.10}_{10.83}$	$11.02^{11.08}_{10.96}$
23	0.7	-	$10.58^{11.70}_{10.36}$	$10.61^{10.62}_{10.52}$	$10.58^{10.71}_{10.35}$	$10.61^{10.67}_{10.50}$
24	0.7	-	$10.10^{10.31}_{9.69}$	$10.21^{10.29}_{10.05}$	$10.10^{10.32}_{9.69}$	$10.21^{10.31}_{10.03}$
25	0.7	-	$9.54^{9.94}_{8.90}$	$9.70^{9.89}_{9.42}$	$9.54^{9.94}_{8.90}$	$9.70^{9.90}_{9.41}$

Table 3.4: The same as Table 3.3 but for an Sc galaxy.

Additional galaxy parameters were investigated to observe the effect our method could have on the measured uncertainty. Uncertainties in V-band extinction see improvements at all simulated redshifts and magnitudes. The results for this can be seen in Figure 3.13. The uncertainties measured on the V-band extinction value were reduced by 51 – 95% for magnitudes ≤ 22 when using phot + 4MOST, whilst at fainter magnitudes there is a reduction to the uncertainties of 18 – 87%. The star formation timescale (τ) sees an improvement to its uncertainty at all simulated magnitudes and redshifts. For phot + 4MOST, the uncertainty on the measurement of star formation timescale was reduced by 11 – 77% for magnitudes ≤ 22. Whilst at magnitudes fainter than 22 the uncertainty was reduced by 3 – 59%. We found that at magnitudes 24 and 25 the uncertainty range is constricted by the lower limit possible within FAST. The results for star formation rate timescale can be seen in Figure 3.14. We found that the τ parameter uncertainty range can hit the lower limit of τ values available within FAST's libraries. We found the τ uncertainty would hit the lower limit for magnitude 25 for all simulated redshifts. It also occurred with magnitude 24 for z=0.1, 0.3 and 0.5. The lower limit restriction could prevent the fit from reaching the true value of other galaxy parameters. However, as this only happens at the faintest magnitudes it is not a major concern to this study and could be investigated further in a study of deep fields.

The measurement of specific star formation rate sees an improvement at brighter magnitudes, aside from redshift 0.1, but this is expected as it is a ratio of host-galaxy mass and star formation rate. The specific star formation rate results can be seen in Figure 3.15. Metallicity sees an improvement at most simulated magnitudes and redshifts when 4MOST spectroscopy is used. At magnitudes ≤ 22 there is a reduction in uncertainty of 9 – 37%. However, at magnitudes 24 and 25 the uncertainties are the same size or even larger than when only photometry is used. Our template is a bright galaxy which has been magnitude normalised to fainter magnitudes. As we are calculating many parameters at the same time, it is possible that FAST does not cover the parameter space for this artificial combination of properties. The results for metallicity can be seen in Figure 3.16.

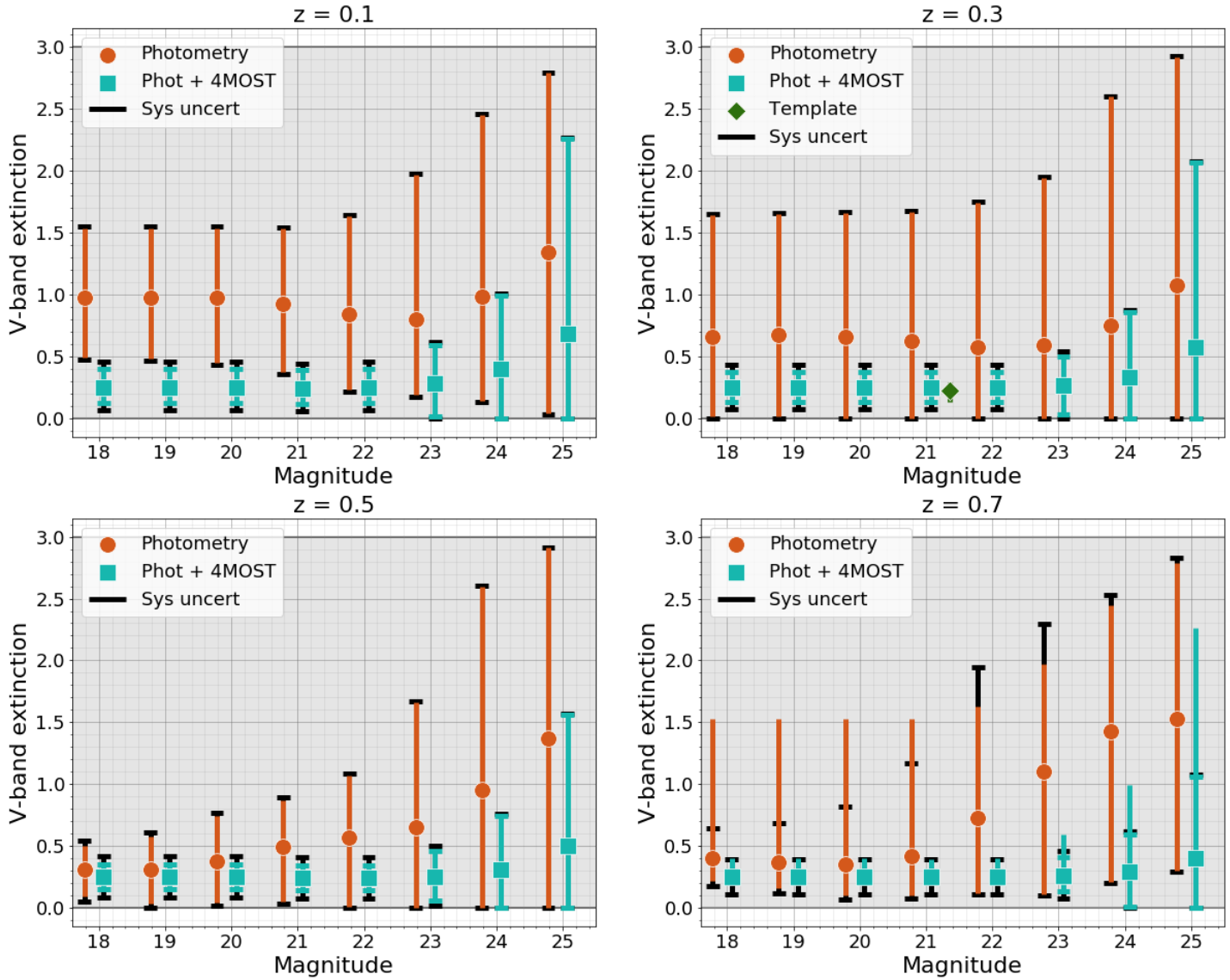

Figure 3.13: Simulated V-band extinction of an elliptical galaxy as a function of magnitude and redshift. The precision for this galaxy parameter also increases when 4MOST is used with photometry. The limits between which the V-band extinction values can be found by FAST are shown by the grey shaded region. The limits are set at 0 magnitude and 3 magnitude.

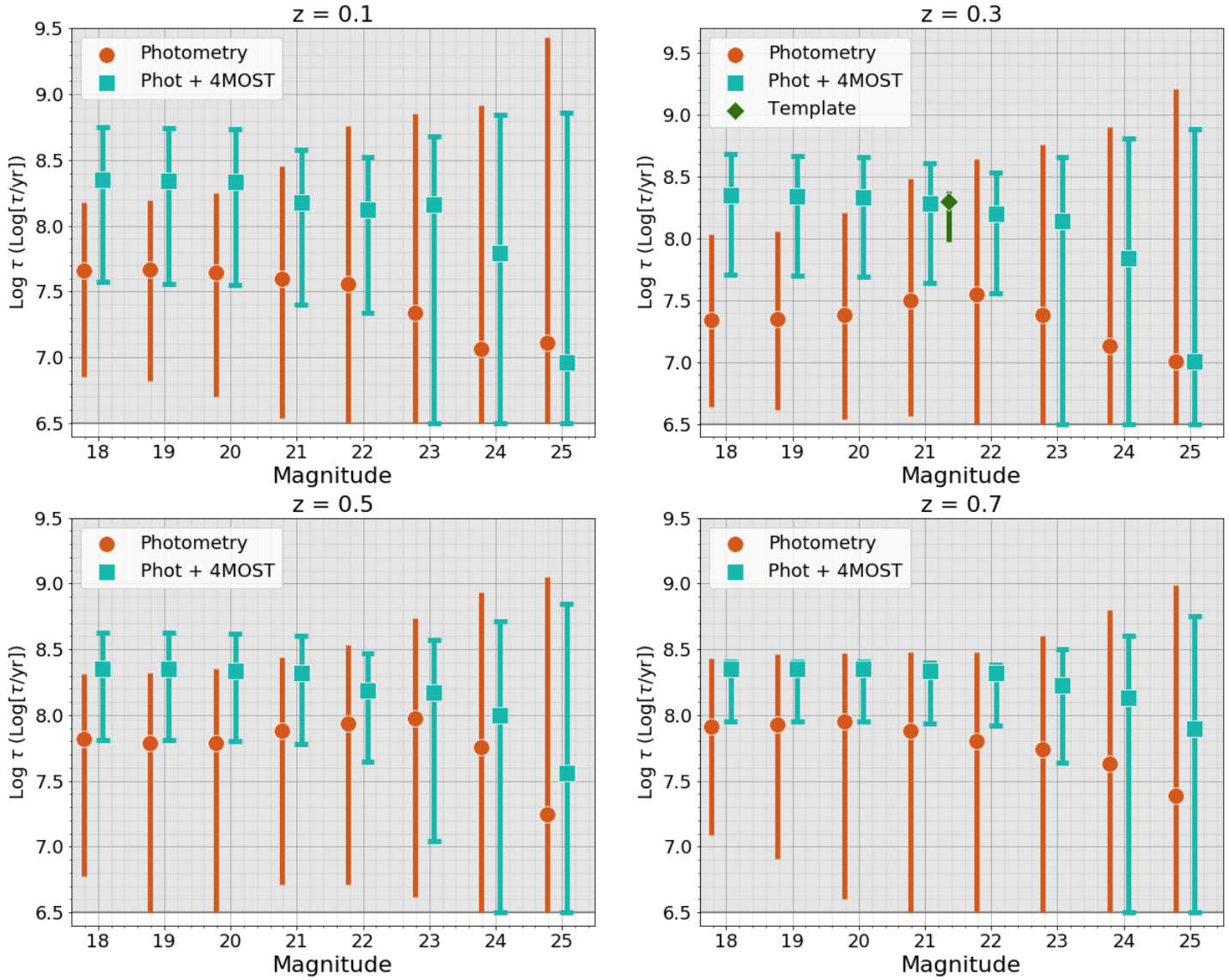

Figure 3.14: The simulated log (star formation timescale) results of an elliptical galaxy as a function of magnitude and redshift. The precision of the value is drastically improved at all simulated magnitudes and redshifts when 4MOST is used with photometry. However, at fainter magnitudes the τ value appears to be restricted by the lower limit. The limits of log τ covered in FAST's libraries is shown by the grey shaded region. The limits are set at 6.5 and 11 (log[τ/yr]).

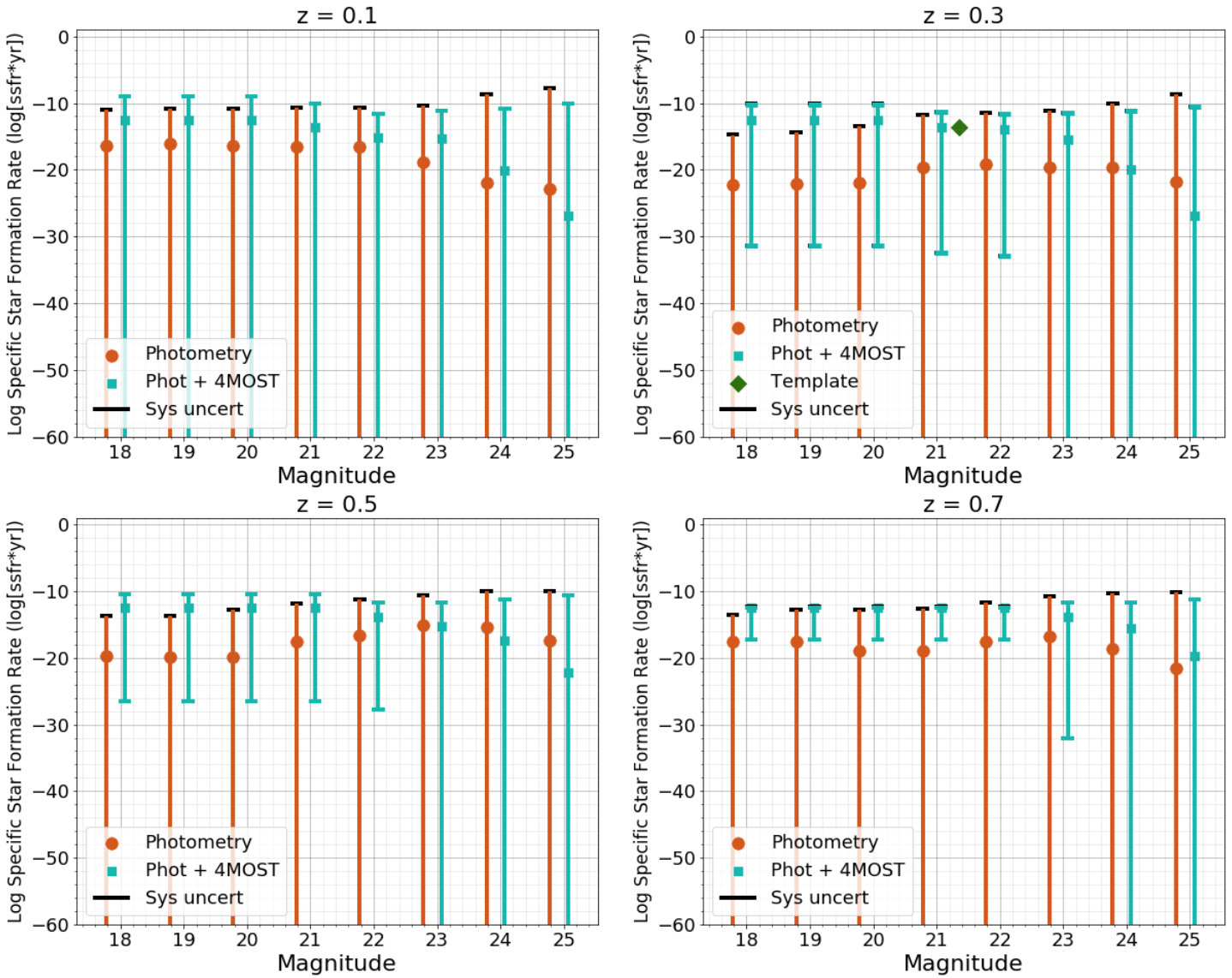

Figure 3.15: The simulated log Specific Star Formation Rate of an elliptical galaxy as a function of magnitude and redshift. The specific star formation rate sees an improvement at brighter magnitudes, for redshifts 0.3, 0.5 and 0.7. This is expected as it matches the host-galaxy mass and star formation rate results, which SSFR is a ratio of.

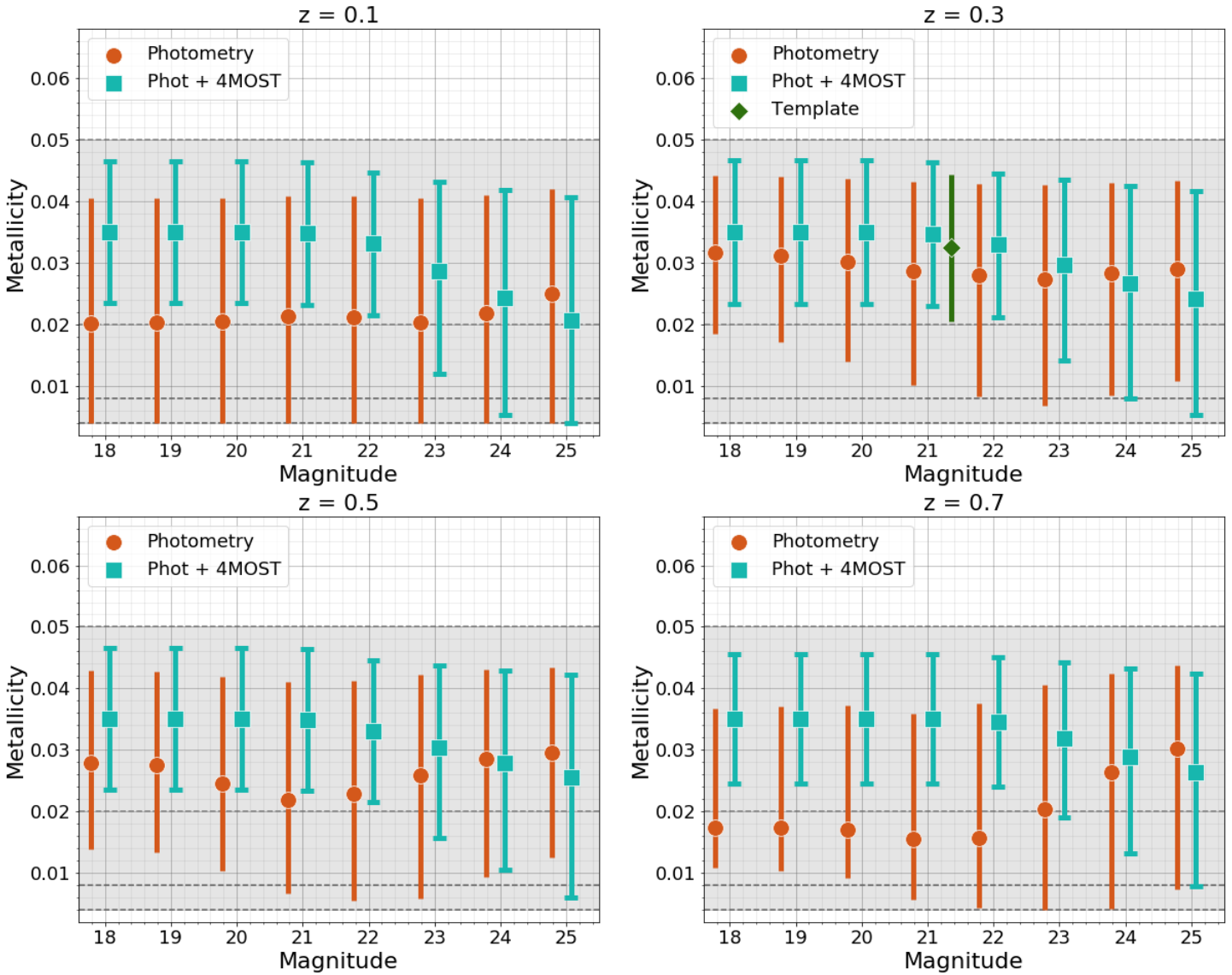

Figure 3.16: Simulated metallicity of of an elliptical galaxy as a function of magnitude and redshift. In the fit, metallicity was constrained to be one of four values, represented by the dashed lines. The possible values were 0.004, 0.008, 0.02, 0.05. The precision slightly increases for most redshifts and magnitudes, but some of the fainter magnitudes have larger uncertainty ranges.

3.5.1 Appendix

The following figures have been added to clarify questions which emerged during the viva.

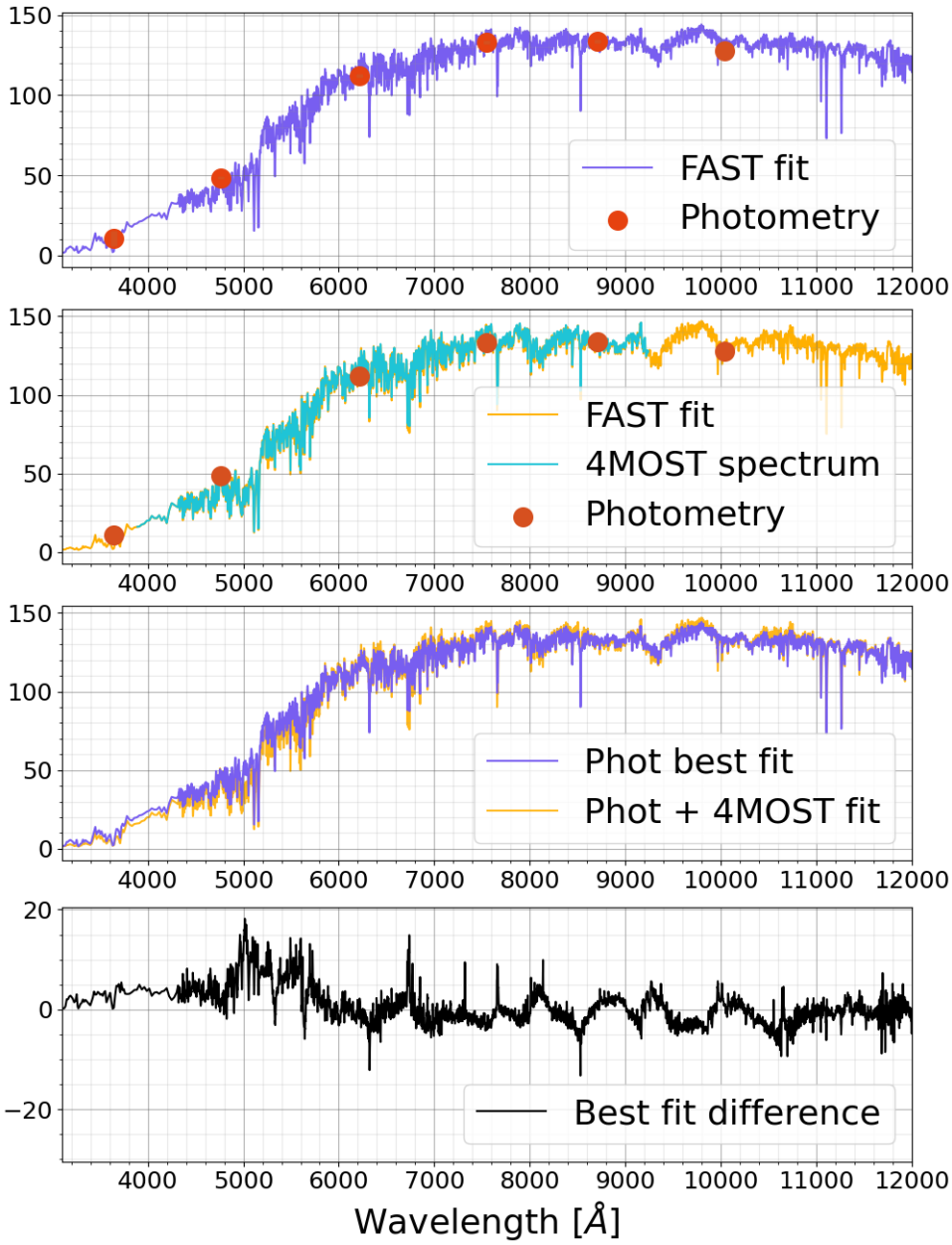

Figure 3.17: The top panel shows the best-fit found by FAST for photometry alone (purple line), with the photometry points (orange circles) plotted at the central wavelengths of the filters. Full profiles of the Rubin filters are shown in Figure 3.1. The second panel shows the best-fit found for photometry with a 4MOST spectrum (gold line). The photometry is plotted on top (orange circle) with the 4MOST spectrum (blue line). The third panel shows a comparison of the two best-fits, whilst the fourth panel shows the residual between the two. As can be seen the two best-fits diverge below 6000 angstroms. All of the plots are for an elliptical galaxy at magnitude 21 and redshift 0.3.

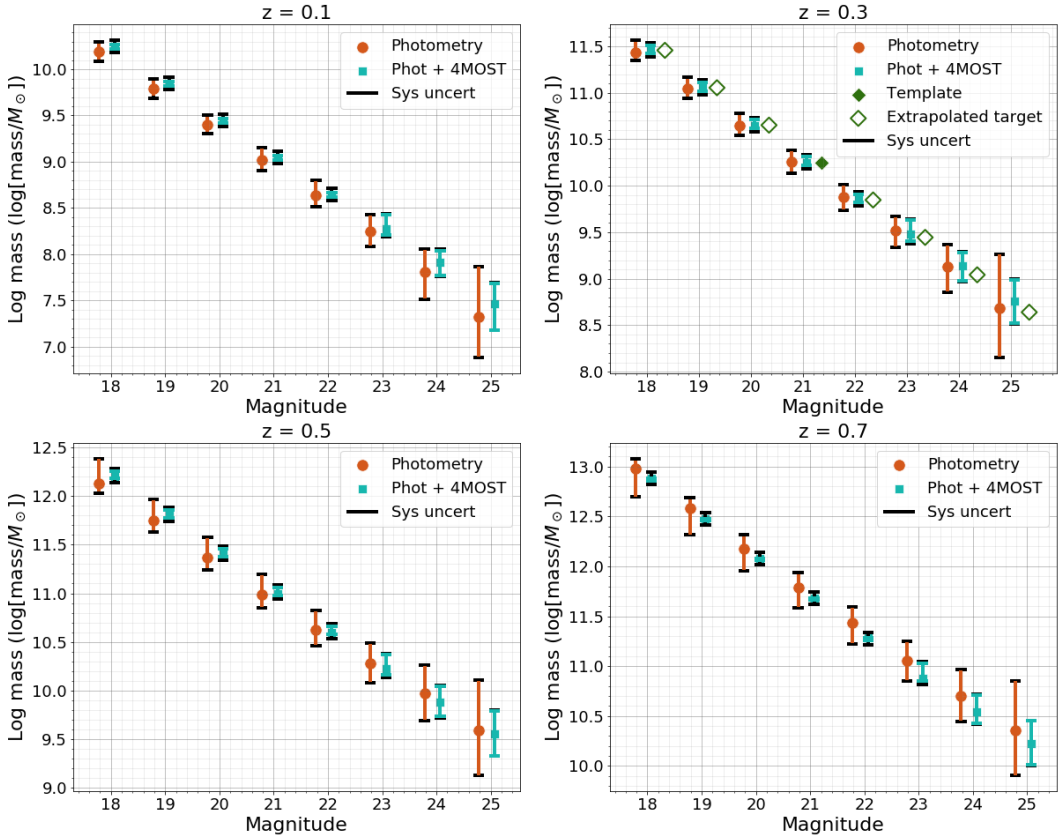

Figure 3.18: Derived log(M_*) as a function of magnitude for LSST photometry only (orange circles) and for LSST phot + 4MOST spectroscopy combined (blue squares) for an elliptical galaxy at z=0.1, 0.3, 0.5 and 0.7. The template spectrum's M_* is also shown (filled green diamond). The template mass at other magnitudes were extrapolated (empty green diamonds) from the original template spectrum (filled green diamond). The points have been artificially separated along the x-axis to make comparison easier. There is an uncertainty associated with the template spectrum, which is the uncertainty that FAST calculates from the input real galaxy spectrum, although the uncertainty is too small to see in this Figure. The extended uncertainties (black line) show the total error bar when an estimate of systematic uncertainty is added in quadrature. The log (M_*) of the galaxies are shown as a function of magnitude and redshift. It can be seen that at all simulated redshifts and magnitudes the precision of the calculated galaxy mass is significantly improved for phot + 4MOST, compared with using photometry alone.

Correction	RMS Hubble residual (mag)	RMS Hubble residual from M_* (mag)
None	0.1752	0.1752
Phot	0.1720	0.1577
Phot + 4MOST	0.1715	0.1566
Perfect	0.1712	0.1560

Table 3.6: Expanding on Table 3.6 a fourth row has been added for the RMS calculated if the perfect correction was applied to each SNe. In order to calculate this we assumed the mass value has no uncertainty. This represents the maximum improvement that would be possible. We find that the perfect results are only a small improvement on Phot + 4MOST.

Figure3.6 sees a systematic offset with the stellar mass uncertainties for a Sc galaxy being lower for Phot only than Phot + 4MOST. There is a difference of approximately 0.2 dex between the two uncertainties, although this varies with magnitude and redshift. The cause of this difference is caused by FAST not being able to measure the emission line of an Sc galaxy correctly, hence why this issue doesn't occur with the elliptical galaxy. It should be noted that in most cases the Phot only uncertainty bars fully contain the Phot + 4MOST uncertainty ranges. There are a few cases where this doesn't occur, most notably at z=0.1. In these situations there could be ramifications for the cosmology results we derive. The lower mass values of the phot only uncertainty range has the potential for SNe host-galaxies to cross the mass step line, and be assigned the wrong mass correction. The wrong corrections would be shown in Figure 3.12 and would also be reflected in the RMS shown in Table 3.6. Future work could study the systematic difference between phot only and phot + 4MOST in depth, with additional SED fitting software to determine if the issue can be remedied.

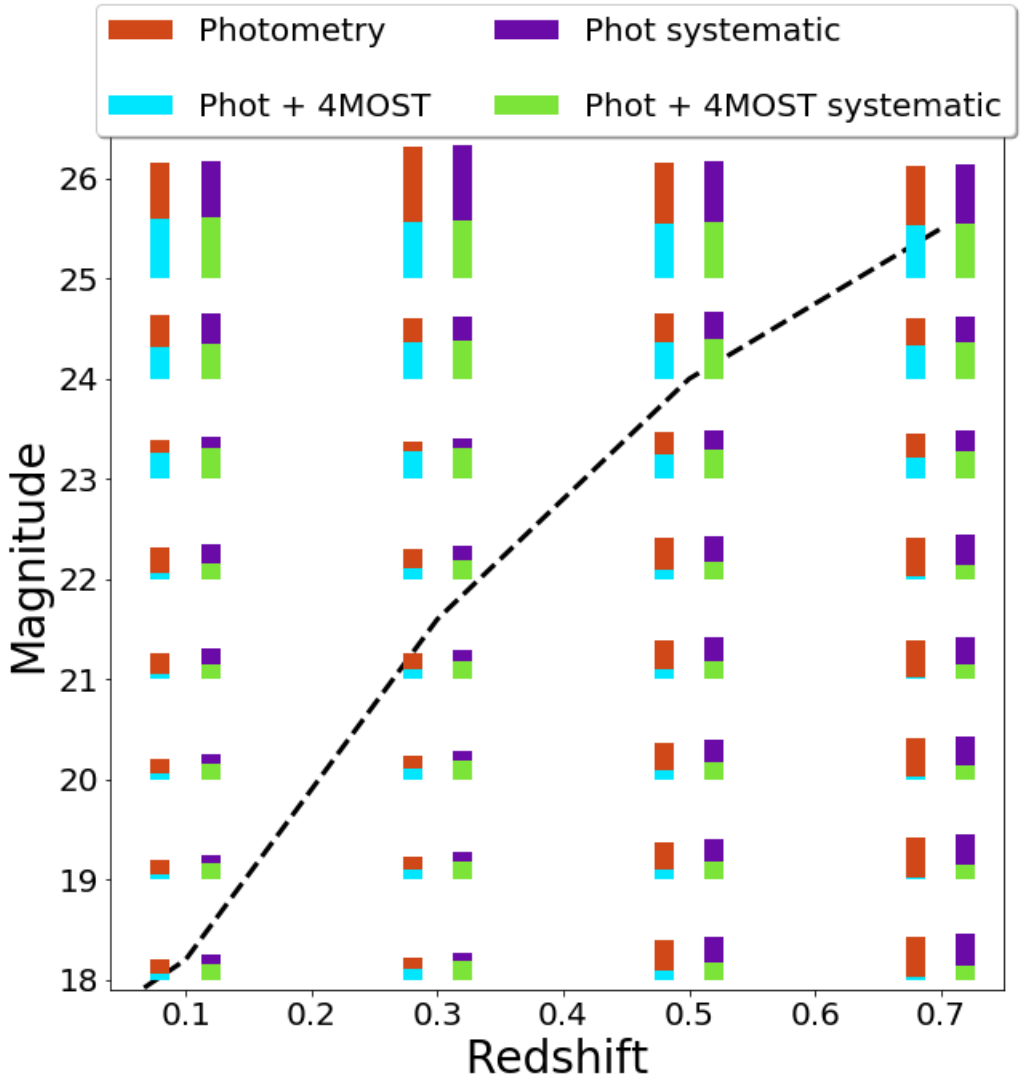

Figure 3.19: A comparison of the uncertainty on log(mass/$M_\odot$) for a an elliptical galaxy for a collection of redshifts and magnitudes when phot + 4MOST are used together can be seen in light blue. The additional uncertainty measured when using photometry only can be seen in dark orange. The size of the line is proportional to the size of the uncertainty. The dashed black line shows where the M_* of the galaxy is 10^{10} $M_\odot$. Finally, the systematic uncertainty from using SED fitting is added in quadrature to be compared with the uncertainty values. The phot + 4MOST with systematic uncertainties is shown in green. The additional uncertainty for phot with systematic uncertainties is shown in purple. The bars for a given magnitude are plotted at the same y-axis value, allowing a comparison of the change of uncertainty size as magnitude and redshift increase

.

Chapter 4

Classifying SNe with deep learning

4MOST's surveys will collect a large number of spectroscopic observations of galaxies, which can be used to calculate each galaxy's stellar mass and SFR, amongst other properties. Previous studies have successfully used apparent galaxy properties to classify SNe (Gagliano et al., 2021). Work carried out by Baldeschi et al. (2020) showed that SFR can be used to improve the accuracy of SNe classification. Therefore, we expect the high precision measurements, carried out using 4MOST, of physical host-galaxy properties will be able to improve the classification of photometrically observed supernovae. This would enable the creation of larger samples of accurately classified SNe, with the spectroscopic resources available. We begin by using a probability calculated using a lookup table. This method has previously been used by Foley & Mandel (2013). We replicate this method with a simplified probability calculation, within a 3-dimensional host parameter space. We then use a Convolutional Neural Network to classify a sample of SNe, using host parameters as additional training information. Both of these methods take the improved host-galaxy properties into a pipeline to classify SNe. We then compare the results of our two methods.

4.1 Method

4.1.1 Data

For this work we used a simulated dataset provided internally by the TiDES collaboration. The dataset was created by Frohmaier et al. (in prep.). This dataset contains realistic simulations of SN light curves carried out using SNANA with host-galaxy properties. SNANA is a general analysis package of supernova lightcurves, which is also capable of simulating SNe lightcurves. (Kessler et al., 2009). We show an example SNe Ia light curve from the data in Figure 4.1. We chose to use this catalogue as it contains a large sample of SNe with simulated host-galaxy properties. We show the distribution of the galaxies across M_*, magnitude, redshift and SFR in Figure 4.2. All SFR values used throughout this Chapter are instantaneous SFR values. The host-galaxy properties were all calculated by Frohmaier et al. as follows. Galaxies were taken from the DES science verification galaxy catalogues. Each galaxy has DES optical photometry and VISTA near-infrared photometry (where available). Each galaxy then had SEDs generated using PEGASE.2, which were fit to the observations using χ^2 minimisation. The host-galaxy properties could then be calculated using SED fitting. Host galaxies are linked to simulated SNe Ia in accordance with the observed rates, which are determined based on host-galaxy properties as measured by Sullivan et al. (2006). For core collapse SNe, the rates and host-galaxy association measurements measured by Graur et al. (2017a) were used. The catalogue provided by Frohmaier notably contains host-galaxy redshift, stellar mass and SFR derived from photometry. The dataset also contains spectroscopic redshift values for the host galaxy, which are derived spectra synthesised using PEGASE.2 (as described in Sullivan et al. 2010). We considered using other SN datasets such as GHOST (Gagliano et al., 2021) and SCOTCH (Lokken et al., 2023) which whilst extensive, focused on using apparent galaxy properties whilst we wish to use physical galaxy properties.

The TiDES dataset contained realistic simulations of the LSST observing strategy with the 4MOST observing strategy simulated on top. For LSST, v2.0 of the publicly available strategy is used which makes use of a rolling coverage of the sky [1]. In the case of 4MOST, objects observed from LSST's wide-fast-deep survey and Deep Drilling Fields are simulated as separate datasets. We chose to use objects from the Deep Drilling Fields. The selected dataset contains 58,356 SNe.

As a reminder, in Chapter 3 we estimated the properties of a set of elliptical and Sc galaxies at different magnitude and redshifts, using SED fitting to photometry and spectra. These properties included: M_*, galaxy age, metallicity, star-formation rate, specific Star Formation Rate, star formation time scale and

[1] http://astro-lsst-01.astro.washington.edu:8080

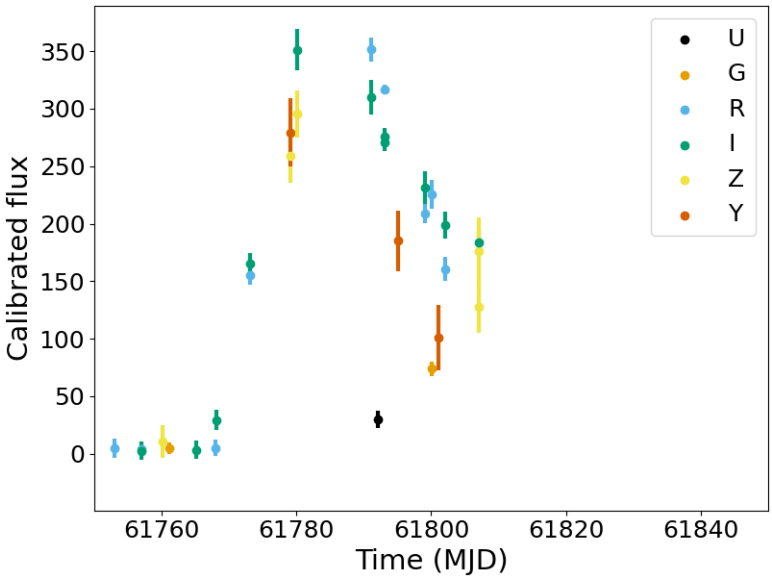

Figure 4.1: An example SNe Ia from the provided dataset. For each object in the data set there are lightcurves in U, G, R, I, Z and Y filters.

V-band extinction. For each property, we calculated the uncertainty as if the property had been derived from photometry alone ("phot"). We then also calculated the uncertainty for each property using the 4MOST spectrum with photometry ("phot + 4MOST"). We wished to apply the M_* and SFR uncertainties to the host-galaxies in the TiDES dataset, to explore how the M_* and SFR uncertainties from phot and phot + 4MOST affected classification of SNe Ia. The M_* and SFR uncertainties were assigned using the magnitude and redshift of the host-galaxy. To begin with, we limited the sample to be in the same magnitude and redshift range as the work in Chapter 3. This limited the magnitude to be between $18 < r < 25$ and the redshift to be < 0.9. Then we placed each host-galaxy in a magnitude and redshift bin, to assign the correct uncertainty for both phot and phot + 4MOST. The TiDES dataset contained many different types of SNe. We selected SNe in the classes: Ia, Ib, Ic, II, IIb and IIn. We combined these classes to simplify the classification to Ia and non-Ia. After these cuts we were left with 24,973 in the Ia class, and 29,530 in the non-Ia class. The sample sizes of 24,973 Ia and 29,530 non-Ia is what is then split into train, test and validation datasets for our machine learning classification. The SNe samples after the cuts were explored, which can be seen in Figure 4.2. The relative rates of the two classes present used is the same the rates presented by (Shivvers et al., 2017), which are SNe rates observed by the Lick Observatory Supernova Search. It can be seen in the top right panel of the figure that a dip in the number of SNe Ia can be seen at approximately 10^{10}. This is due to the equations used to simulate SNe Ia making use

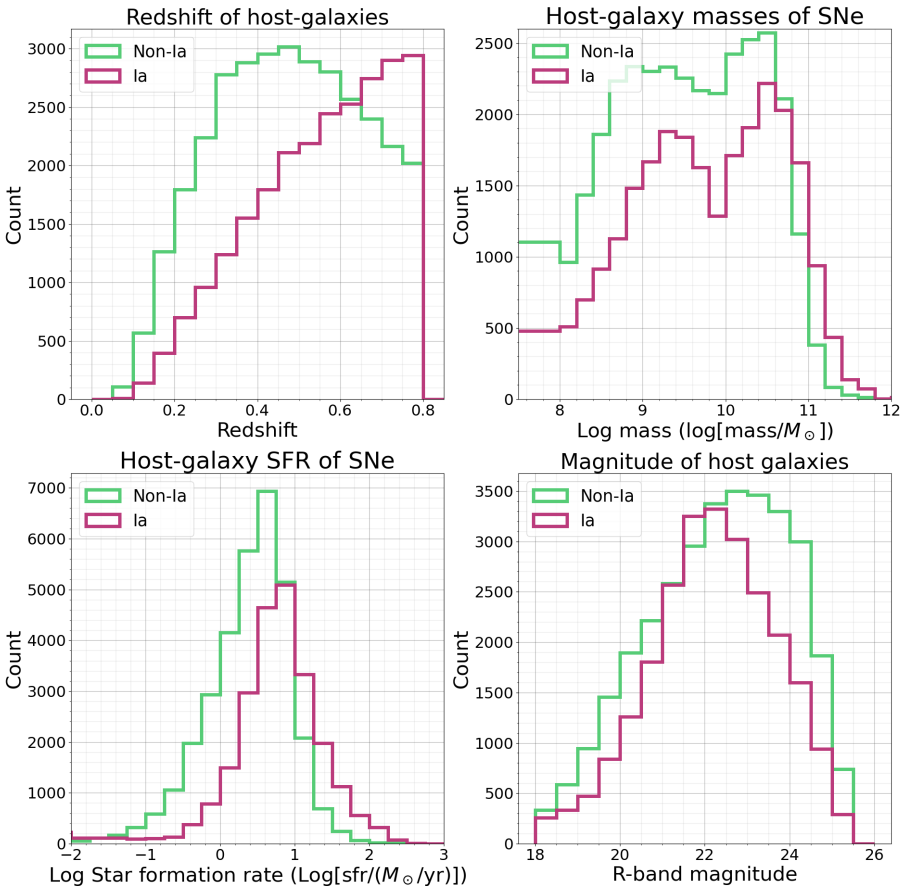

Figure 4.2: An exploration of the dataset for the classes Ia and non-Ia after our cuts. Top left: a histogram of the distribution of host-galaxy redshifts. Top right: a histogram of the distribution of host-galaxy M_*. Bottom left: a histogram of the distribution of host-galaxy SFRs. Bottom right: a histogram of the distribution of host-galaxy r-band peak magnitudes.

of two different calculations. One calculation is used below $10^{10} M_\odot$ and another calculation is used at values grater than 10^{10}. As expected, our mass and SFR distributions match those shown in (Vincenzi et al., 2021). More details about the simulation used to calculate the galaxy properties can be found in (Vincenzi et al., 2021).

We explored classifying SNe using two methods. The first method used a lookup table to calculate the probability of a host-galaxy hosting a SN Ia or a non-Ia. The second method used a recurrent neural network to assign SNe a classification of either SN Ia or non-Ia. The following subsections cover the methods carried out with both classification techniques.

4.1.2 Lookup table method

To understand the potential to classify SNe based on their host galaxy properties, we carried out a simple method with the data described in section 4.1.1. We use a probability calculated using a lookup table, which is simple to apply and has been used by (Foley & Mandel, 2013). We then compare our results with the machine learning method in Section 4.3.

The input data into our lookup table makes use of 10% of each class, introduced in Section 4.1.1. This is the same amount of data that will be used as the test data for the machine learning, which allows us to compare results between the two methods. We started by exploring correlations between SN classes and host-galaxy parameters, using the parameters stellar mass, star formation rate and redshift. To do this, we created a grid of values in this 3-dimensional parameter space. We first explored the number of the two SN types (Ia and non-Ia) in 2D parameter space of mass and redshift. Later we carry out our calculations in 3D, but it is simpler to visualise in 2D. We created a grid of 100 by 100 bins, equally spaced in the range of 0 to 0.9 in redshift and 6 to 14 $\log(M_{\odot})$ in mass. Each SN host-galaxy was assigned a bin in the 2-dimensional space, then the number of each SN type in a bin could be counted. The results for this can be seen in Figure 4.3. It can be seen in these figures that redshift has a larger effect on the number of each SN type than host-galaxy M_*.

We could then use the counts within each bin to work out the probability of an object whose host galaxy has given values of z and M_* being a SN Ia or non Ia. We estimated the probability by dividing the count of a SN type in a bin by the total number of SNe within a bin. We encountered an issue if there was a small number of SNe within a bin, that bin could have a probability of 1. To prevent this, we set any bins which had less than 10 SNe in them to have the average probability across all bins. We found that approximately 12% of the bins contained less than 10 SNe. The results of the probability calculation in 2D parameter space can be seen in Figure 4.4. We also explored the probability of an object appearing in 2D parameter space of SFR and redshift. In SFR space the 100 bins are equally spaced from -11 to 3 $\log(M_{\odot}yr^{-1})$. The results can be seen in Figure 4.5. It can be seen that SFR has a large effect on the probability of each SN type. It can also be seen that there is a large number of misclassifications. This is due to the threshold intentionally being set high, at 0.6, to maintain a high purity of SNe Ia. We explore in Section 4.3.3 the effect changing the threshold has on the results.

We then calculated the probability in 3D parameter space. We created a 100 by 100 by 100 cube. We used the same range for the bins in mass and redshift parameter spaces as we did for the 2D exploration. We present and discuss the results for the probability calculated in 3D in Section 4.1.4.

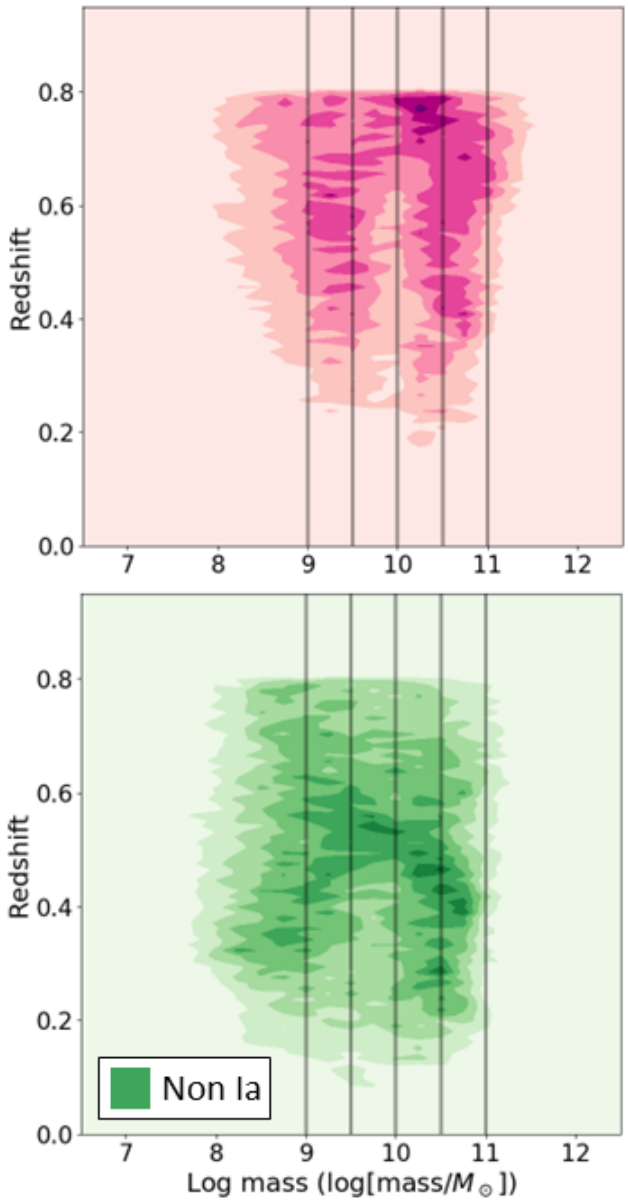

Figure 4.3: Top panel: A contour plot showing the number of SNe Ia within the mass-redshift parameter space. Bottom panel: the same as the top panel but for non-Ia SNe. The black lines are placed around the central mass values explored, at half integer values. The black lines allow easier comparison of different SNe types within the 2D parameter space.

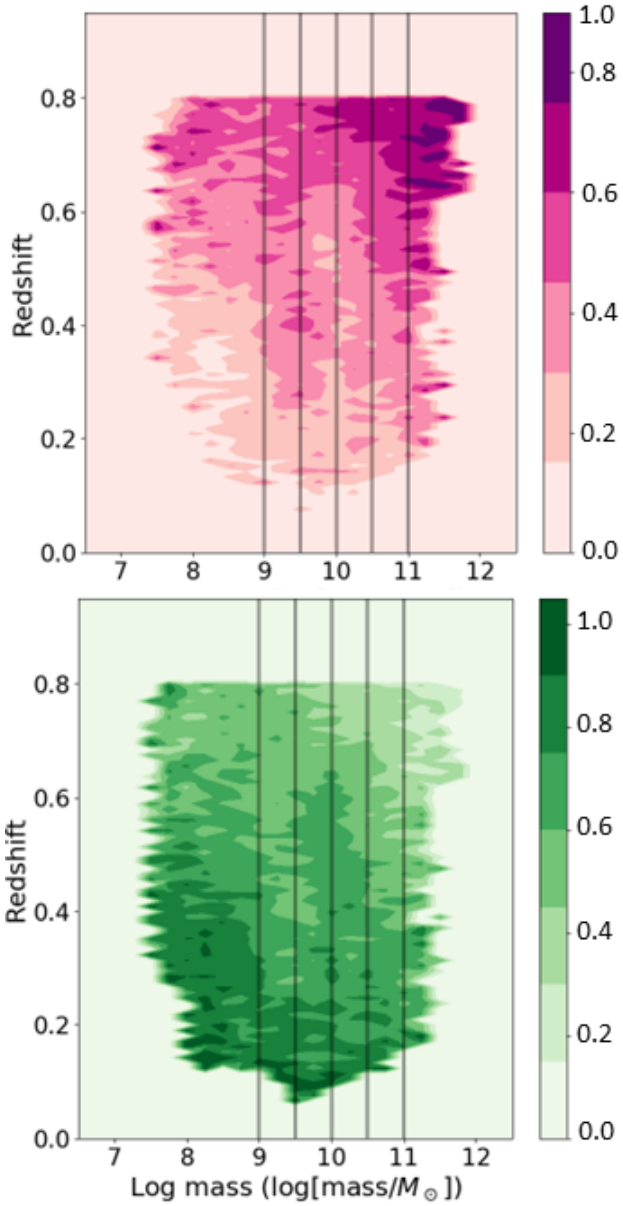

Figure 4.4: Top panel: The probability of an object being a SNe Ia within a bin in host-galaxy mass-redshift space. Bottom panel: the same as the top panel but for non-Ia SNe. Black lines are placed around the central mass values to improve the ease of comparison of the different SNe types. It can be seen that redshift has a large effect on the probability on the probability.

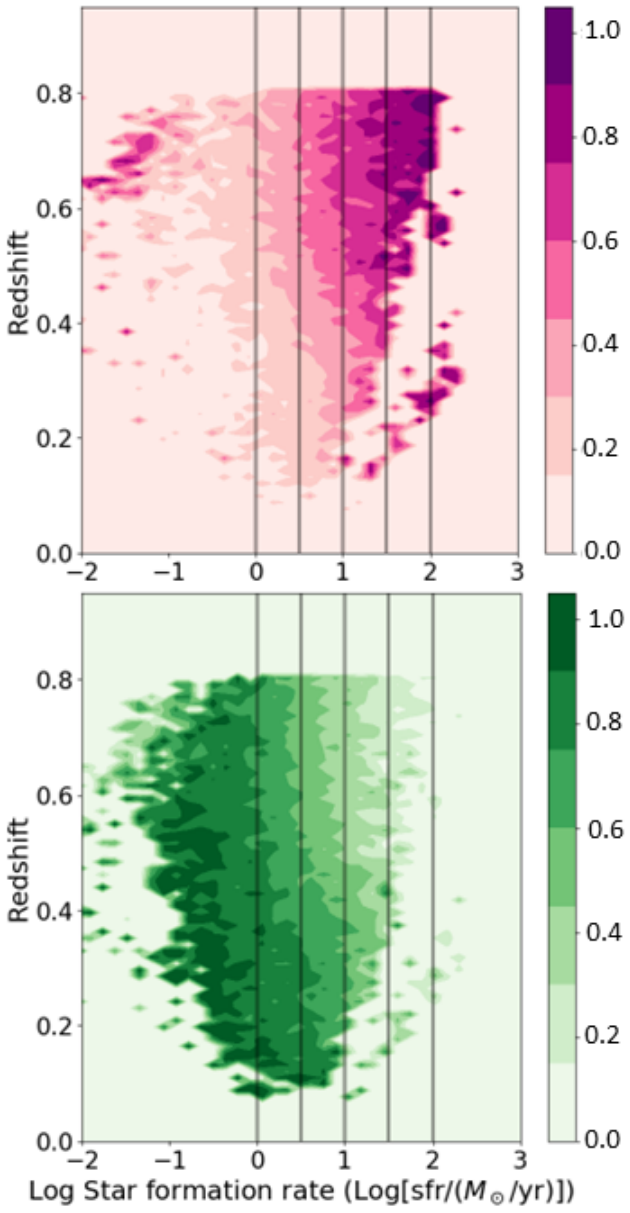

Figure 4.5: Top panel: The probability of an object being a SNe Ia within a bin in host-galaxy SFR-redshift space (instantaneous SFR). Bottom panel: the same as the top panel but for non-Ia SNe. Black lines are placed around the central SFR values to improve the ease of comparison of the different SNe types. It can be seen that SFR has a large effect on the probability.

4.1.3 Success metrics

In order to evaluate the improvements to the machine learning, we used several metrics. We briefly covered success metrics in Section 2.1.1, however we will cover the most relevant metrics to our work here. The first metric used was to measure the precision of each model, also known as purity. This compares the number of true positive (TP) classifications with the total number of objects classified as positive. It is defined by the following equation:

$$Precision = \frac{TP}{TP \ + \ FP} \tag{4.1}$$

where, FP is the number of False Positive classifications. The precision value is calculated to be between 0 and 1. A value of 1 means all of the SNe labelled as Ia are correctly assigned the classification of Ia. A value of 0 means that all labelled SNe Ia are incorrectly classified. By aiming to have a value as close to 1 as possible we can create a purer SN Ia sample.

Precision is the metric which we are most concerned about as we wish to reduce contamination of a SN Ia sample, to allow better cosmology measurements. However, there are other metrics we can use to check the performance of each model. It should be noted that it is very difficult for a machine learning model to perform well in every metric. The next metric we can use to check the performance of our models is recall, also known as efficiency. Recall is a metric which represents the accuracy of true classifications of a given class. It can be calculated as follows:

$$Recall = \frac{TP}{TP \ + \ FN} \tag{4.2}$$

where FN is the number of False Negative classifications. A recall value of 1 means that all labelled SNe Ia have been classified correctly. A value of 0 means that all labelled SNe Ia have been classified incorrectly. This is a useful metric to our work as it will represent the portion of true SNe Ia which are being discarded. There is a compromise between precision and recall (e.g. Möller et al. 2016; Dai et al. 2018). Typically, enhancing precision results in a reduction in recall, and vice versa. Adjusting the classification threshold is the most common method to enhance either metric. Increasing the threshold decreases false positive SNe Ia, raising precision but lowering recall due to increased false negatives. Conversely, reducing the threshold enhances recall while decreasing precision due to fewer false negatives or more false positives.

Another metric we will use is the F1-score. This is a metric which combines both precision and recall. It can be calculated as follows:

$$F1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \tag{4.3}$$

when precision and recall have similar values, F1 will be close to their mean. When precision and recall differ then F1 will be closest to the lowest value. This means if the value is closer to 1 our model has classified a pure sample of SNe Ia, with few false negatives. A value of 0 can stem from two scenarios: either one metric is high while the other is low, or both metrics are low. The situations can be distinguished using the precision and recall values.

Another metric which is powerful at showing how well a model can classify data is the receiver operating characteristic (ROC) curve. This metric was discussed in Section 2.1.1, but as a reminder the ROC curve is a plot of false positive rate against true positive rate. For a model which classifies between two classes, a diagonal line of $y = x$ represents the performance of the model if it classified objects at random. The further a model's ROC curve is above the diagonal line, the better the model's performance. The ROC curve can be quantified by calculating the area under the curve (AUC). An AUC value of 1 represents a perfect model, whilst a value of 0 represents a model which misclassifies every object.

A final way to view a model's performance is to use a confusion matrix. It works by visually showing the number of true positive, false positive, true negative and false negative objects. A confusion matrix is an easy way to visually compare multiple models. An ideal model will have high true positive and true negative values, meaning the diagonal will have the largest values. On the confusion models that we present this diagonal goes from the bottom left corner to the upper right corner. All of the metrics we have discussed can be calculated using the *Python* module *scikit-learn*.

4.1.4 Lookup table results

We found that using the lookup table method could be used to classify transients into the two given classes, SNe Ia and non-Ia. A confusion matrix of the results can be seen in Figure 4.6. It can be seen that the lookup table method is able to produce a SNe Ia sample with a precision of 0.788. It can also be seen that the model is able to classify non-Ia objects effectively, with a low false positive rate. However, the lookup table struggles to classify a significant number of SNe Ia resulting in a large false negative rate. The values we obtain here will be a

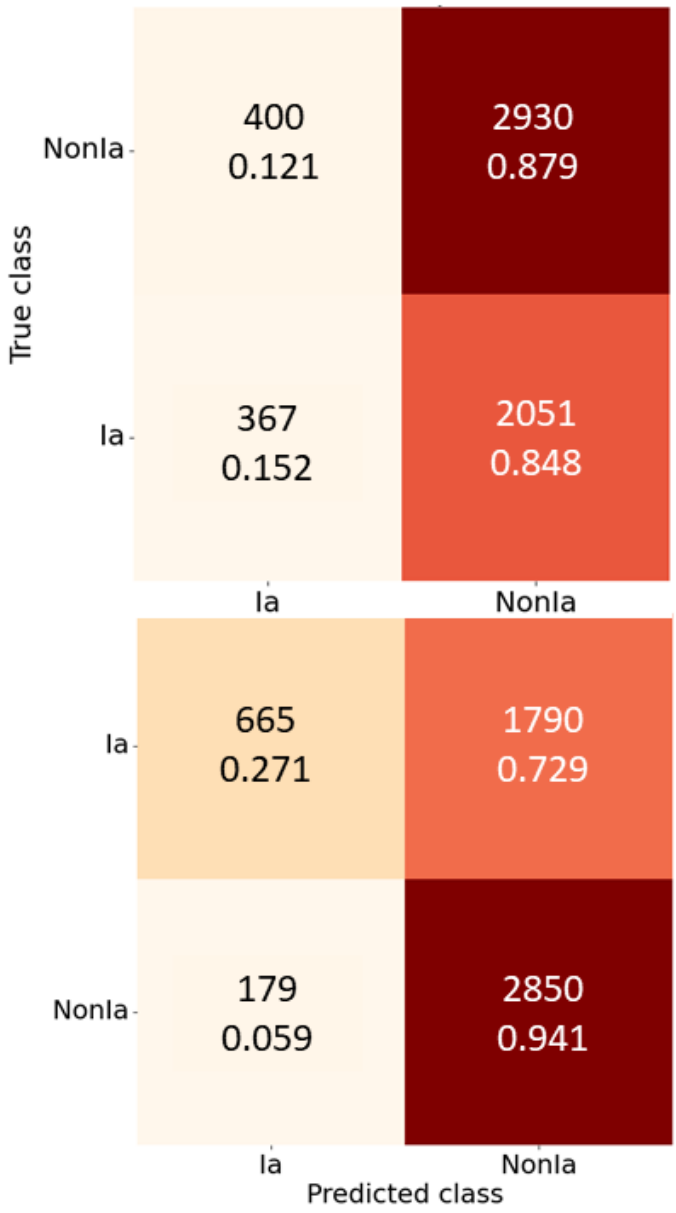

Figure 4.6: Two confusion matrices showing the probability a galaxy with given properties hosts a SNe Ia. The top panel shows the results for a 2D parameter space (redshift and M_*), whilst the bottom panel shows the results for a 3D parameter space (redshift, M_* and SFR). The numbers shown are the number of objects falling into each square. The colours represent how much of the data is within a box, the darker the red the more objects whilst the lighter the colour the fewer objects. It can be seen that the 3D parameter classification outperforms the 2D classification, which is to be expected. The 3D lookup table is able to produce a SNe Ia sample with a purity of 0.788, although this is dependent on the threshold value. However, a large quantity of SNe Ia are misclassified. The number of SNe shown are less than what was stated in Section 4.1.1 as the confusion matrix was made using the test data which will be used for the ML method, which is 10% of each class, in order to allow the comparison of results.

baseline which we can compare to our ML results to.

4.1.5 The machine learning method - Using SuperNNova

We now explore a more sophisticated approach to incorporate host-galaxy properties into SN classification. As explored in the introduction, a wide range of machine learning models are available to classify SNe. However, we wished to use a model that could be used to classify SNe with both photometric light curves and host-galaxy properties. We chose to use SuperNNova which can classify SNe light curves, and be expanded to take host galaxy information as an input. We were able adapt the model to take host-galaxy mass and Star-formation rate as input parameters.

SuperNNova was created by Möller & de Boissière (2019b) and uses a Recurrent Neural Network with Long Short term memory (LSTM) to classify SNe, where LSTM is a modification to a RNN to prevent issues with gradient descent when large quantities of data are processed. SuperNNova is also able to incorporate SNe host galaxy redshift. Möller & de Boissière found that the addition of host-galaxy redshifts enhanced the accuracy of the model by approximately 2.63%. SuperNNova is available as a random forest model or a RNN. We chose to use the RNN version as Möller & de Boissière showed it to perform better at classification than the random forest model. SuperNNova also makes use of Bayesian statistics applied to the results calculated by the RNN. However, this is only used by Möller & de Boissière when classifying between 3 or more classes. As we are only using 2 classes, we will not cover this.

To ensure the best performance of the ML model we carried out hyper parameter training. We found the optimal learning rate to be 0.0001, by exploring between 0.1 and 0.000001. The optimal batch size was found to be 128, from a series of values between 32 and 512. We found the optimal number of epochs was 270, after letting the model run for 450 epochs. As can be seen in Figure 4.7 the model overfits after 270 epochs. To improve the speed of training, SuperNNova was designed to use cyclic learning. Cyclic learning alters the learning rate in a cyclic fashion within reasonable limits (Smith, 2015). SuperNNova sets the maximum learning rate as the defined value. The minimum learning rate is set as a tenth of the maximum. Then when the model is run, with each epoch the learning rate is varied between the minimum and maximum values. We experimented with using cyclic learning, but ran into an issue with training. As can be seen in Figure 4.8, at around 20 epochs the loss would increase dramatically. Loss is the penalty for a bad prediction, loss is minimised. We found that it was caused by the host-galaxy properties not being normalised, but were unable to fix the problem. However, this is issue doesn't change the results. Instead we chose to use the

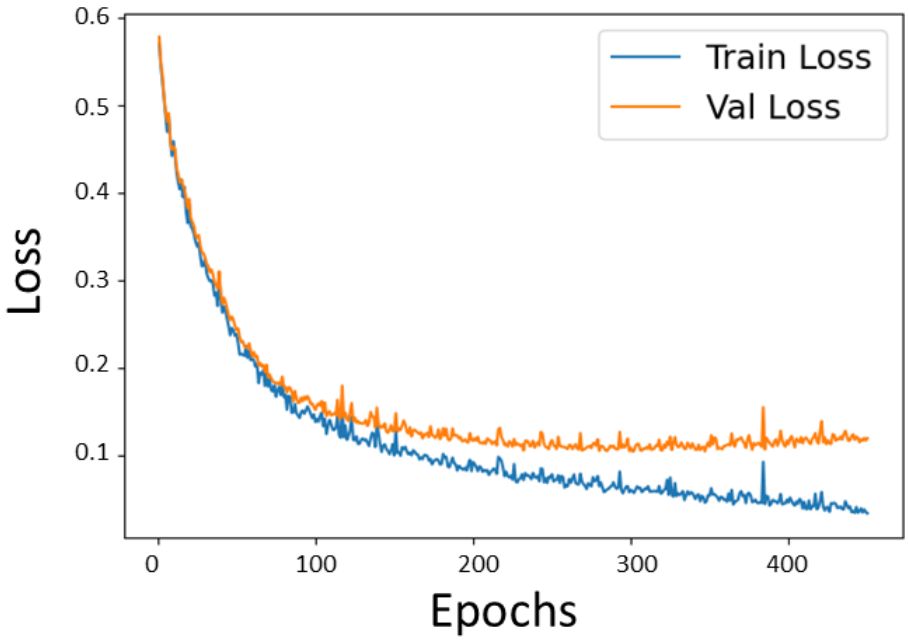

Figure 4.7: The loss after each epoch during training and validation. Loss is the penalty for a bad prediction, during training loss is minimised. We ran the model for 450 epochs to see how long it took for the model to converge. It can be seen that after 270 epochs the validation loss curve reached the turning point, this is the point of convergence. After 270 epochs the model began to overfit.

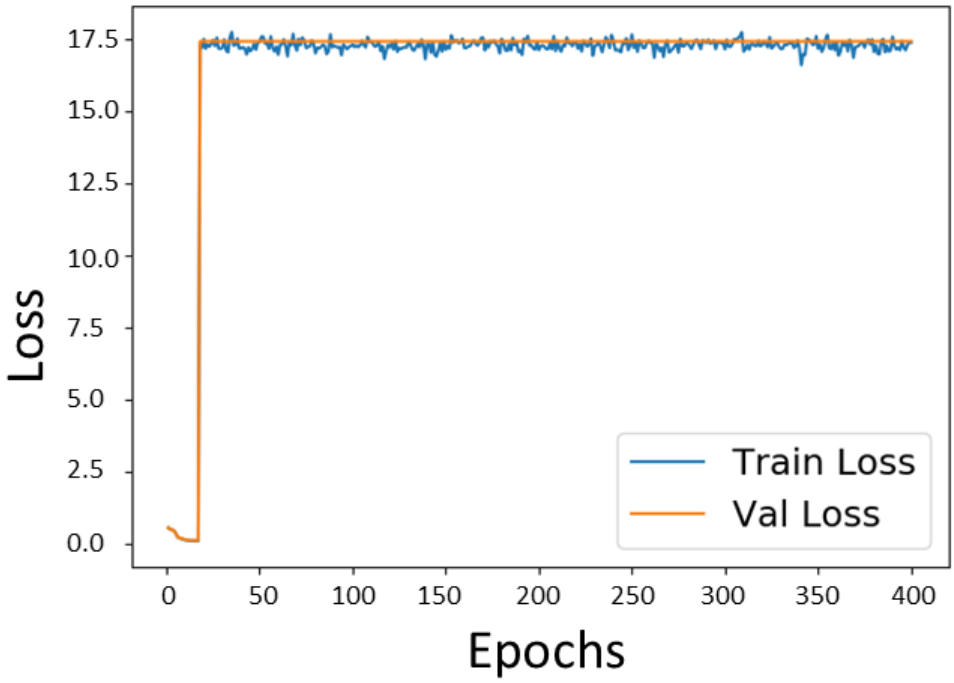

Figure 4.8: The loss after each epoch during training and validation for cyclic learning. Until epoch 20 the validation and training loss are converging. However, the loss then increases significantly after the 20th epoch. We tried to solve this issue, but were unable to find the cause of it. We used the non-cyclic method to collect all results.

non-cyclic approach since, the speed of convergence was not a limiting factor for this work. Future versions of this work could investigate the cyclic learning issue. It should be noted that SuperNNova automatically ensures equal numbers of all classes during training by randomly removing objects from the larger class. This prevents biases from arising in the model, due to disproportionate numbers of SN within each class.

We wished to use SuperNNova to test different input features, and see the effect on cosmology. Our goal is to create a larger and purer SNe Ia dataset, which can be used to improve constraints on w. We modified the SN dataset described in Section 4.1.1 to create 6 different versions. We presented each version of the data as an input to a separate but identical SuperNNova model. Model 1 used only the supernova light curves for classification. Model 2 used the the supernova light curves and photometric redshift. Model 3 used the supernova light curves and true spectroscopic redshifts. For models 4 and 5 we combined the simulated SNe with our previous work measuring host-galaxy property uncertainties when observed by 4MOST and Rubin (Dumayne et al., 2023). For model 4 we assigned the phot uncertainties for M_* and SFR. For model 5 we assigned the phot + 4MOST uncertainties for M_* and SFR. Our final model, Model 6, used the light curves, spectroscopic redshift and host-galaxy properties (M_* and Star formation rate) with no uncertainties. All redshift values provided are for the host galaxy of the SN. Model 6 represents a perfect observation of the SN host-galaxies. All of the models used an 80/10/10 split for training, validation and testing, respectively. The data is sampled randomly every time the machine learning model is run. All of the values we give in Section 4.2 used a threshold value of 0.9 to be able to compare our results with Möller & de Boissière. We discuss the effect of varying the threshold value in Section 4.3.

When assigning uncertainties, we could not simply add the uncertainty as extra columns in the data table because the ML model would not associate the uncertainty to the correct property. Instead we chose to simulate uncertainty measurements by sampling using a Gaussian distribution with the true value as the mean, and the size of the uncertainty as σ. For the SFR property, some of the minimum uncertainty values from FAST had a value of -99, meaning that FAST could not find a value. As the true SFR value had no lower limit, it meant we did not have an accurate value. To account for this, any SNe assigned a lower limit of -99 had its SFR value set to -99.

In our previous work we calculated uncertainties for an Elliptical and Sc galaxy. Our simulated dataset does not specify the type of each host-galaxy. Therefore we assigned a host-galaxy type at random, weighted by the rates at which SNe Ia appear in elliptical and Sc galaxies, as provided by Hakobyan et al. (2012). This study gives a proportion of 0.54 SNe Ia being hosted in elliptical, and the

remainder being hosted in spiral galaxies.

We ran each of the six models 10 times. Every time models 4 and 5 were run we resampled the galaxy properties. We ran each model 10 times to get an estimate of the effect of the randomness produced by machine learning models. 10 runs of each model is sufficient to understand the differences between each of the 6 different models. We were able to speed up the process by running multiple models in parallel. SuperNNova does have the ability to be run on a GPU. However, none of the computers we had access to used dedicated graphics cards. This means the use of the GPU processing would not have created much of an improvement to computing time (if any).

4.2 Machine learning results

We present the average precision values obtained by the six different models used for machine learning in Figure 4.9 and Table 4.1. Our research revealed that incorporating 4MOST redshift into the analysis of photometric transients, for which 4MOST had observed the host galaxy but not a live transient, led to enhanced precision in the classification of supernovae. We also found that the addition of spectroscopic host-galaxy M_* and SFR observations improved the performance precision value of our model.

The results show that when SNe lightcurves are used with photometric host information (model 4) the precision is lowest. However an average precision of 0.958 is still very high. The lower precision value may be due to the large uncertainties associated with each galaxy property. The classification carried out with only SNe lightcurves achieved a slightly improved precision (model 1, 0.961). The results show that the addition of photometric redshift improves the average precision value (model 2, 0.962), and the addition of spectroscopic redshift improves the average precision further (model 3, 0.977).

The reduction of host-galaxy M_* and SFR uncertainties used as an input for model 5 corresponding to observations with with phot + 4MOST, boosted the average precision of the model (0.980) compared with model 3 which used SNe light curves and spec redshift. Finally, the model incorporating SNe light curves, spectroscopic redshift, and precisely measured host-galaxy M_* and SFR attained an average precision similar to that of model 5 (model 6, 0.983), with higher minimum and maximum precision values. The result of model 6 is our perfect case. It shows that a reduction of the host-galaxy M_* and SFR uncertainties would not improve the precision of the models further.

In addition to the precision scores, we calculated the average recall, average F1-score and average AUC for each model. These values are presented in Table 4.1

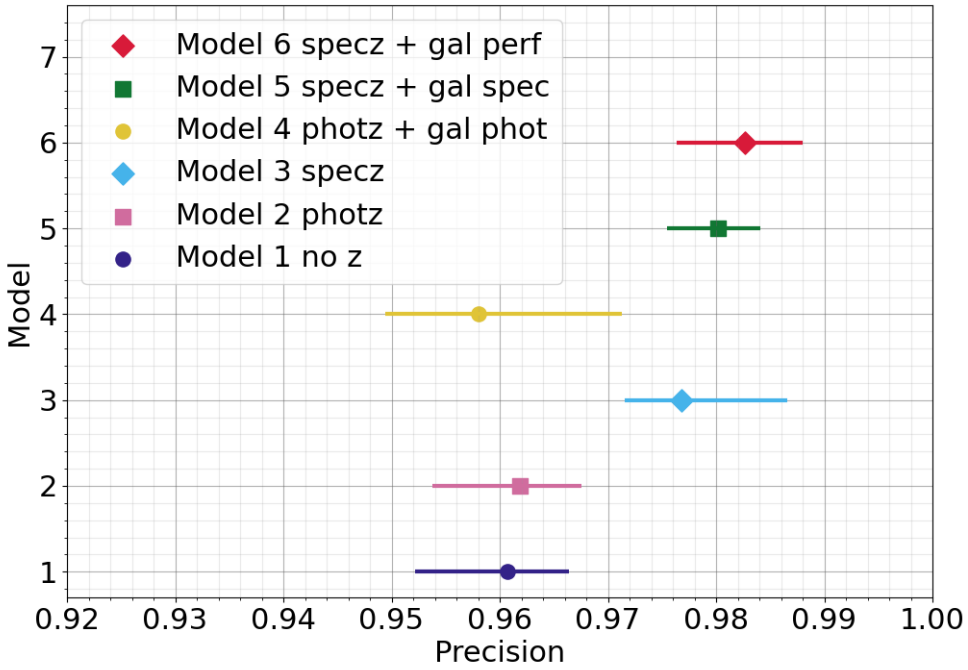

Figure 4.9: A comparison of the precision values of all of the different models. Model 1 is with only SN lightcurves (purple circle). Model 2 uses lightcurves and photometric redshift (pink square). Model 3 uses lightcurves and spectroscopic redshift (blue diamond). Model 4 uses lightcurves, phot redshift and host-galaxy mass and SFR with phot uncertainties (gold circle). Model 5 uses light curves, spec redshift and host-galaxy mass and SFR with phot + 4MOST uncertainties (green square). Model 6 uses light curves, spec redshift and perfectly measured host-galaxy mass and SFR (red diamond). All redshift values are for the host galaxy. We ran each model 10 times. Here we show the minimum, mean and maximum precision values achieved by each model. It can be seen Model 4 which used photometric uncertainties for each of the host-galaxy properties produced the lowest average precision value. The highest average precision value was obtained by models 5 and 6. This result shows the addition of host-galaxy properties can aid classification if the host-galaxy properties have small uncertainties.

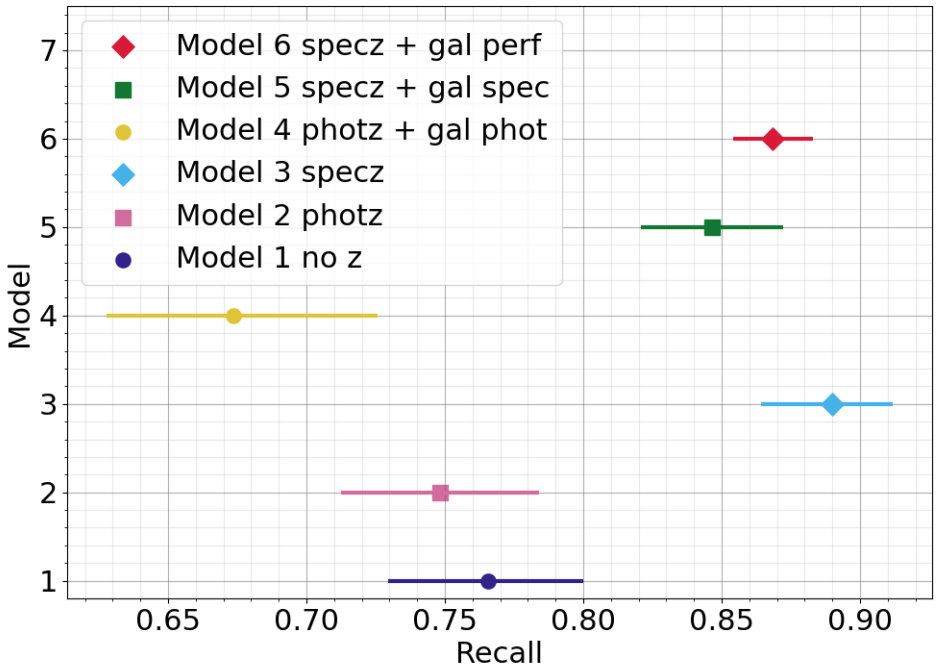

Figure 4.10: The minimum, mean and maximum recall values achieved by each model. It can be seen that all models have lower recall values than their precision values shown in Figure 4.9. This is to be expected. The high threshold value used creates more false negative classifications, bringing down the recall values. Models 3, 5 and 6 are able to achieve higher values than the other 3 models.

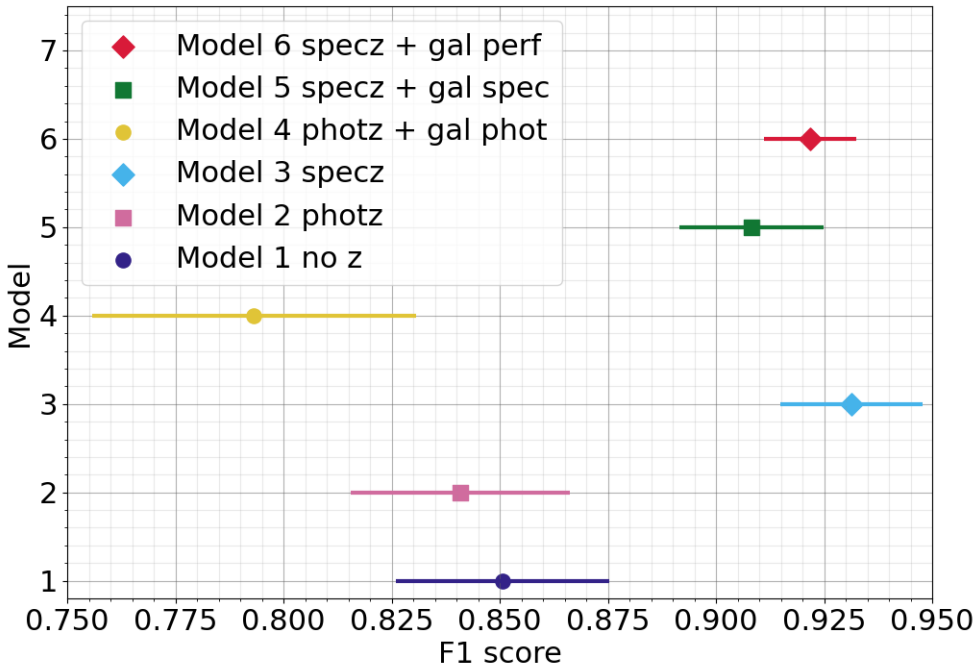

Figure 4.11: The minimum, mean and maximum F1 values achieved by each model. It can be seen that models 1, 2 and 4 produce low F1 values whilst models 3, 5 and 6 produce high values. Models 1, 2 and 4 perform worse due to misclassifying more SNe Ia than the other models. The F1 values show the addition of spectroscopic data and the addition of host-galaxy properties, with models 5 and 6, creates a well-rounded classifier. However, the addition of spectroscopic redshift (model 3) also adds a significant improvement to the F1 value.

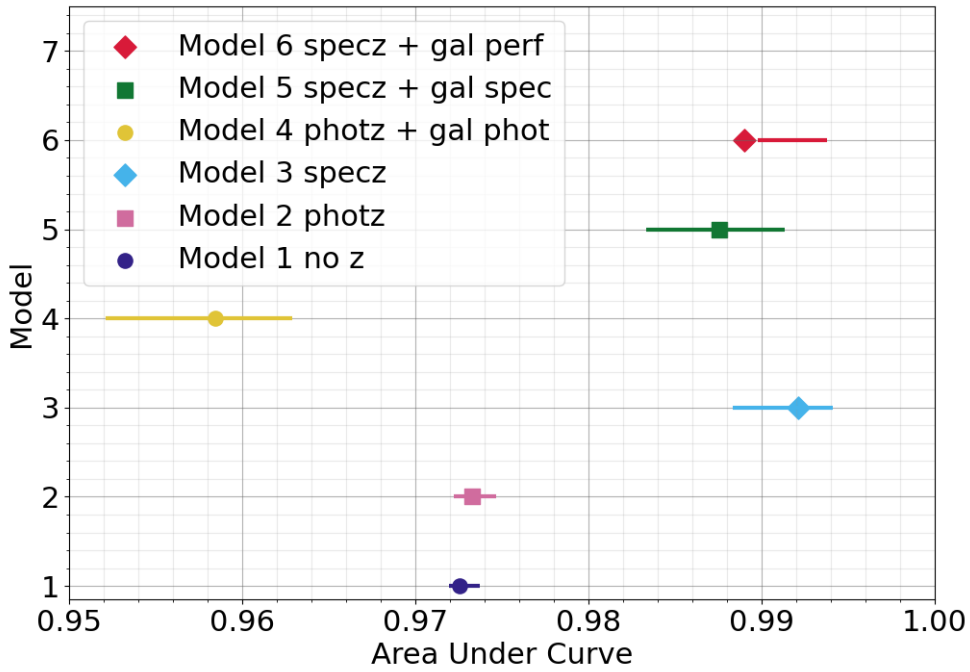

Figure 4.12: The minimum, mean and maximum AUC values calculated for each model. It can be seen that all the models produce high AUC values. These values represent high-performing classification models, which is reflected in the other metrics we have evaluated. Model 4 produces the lowest AUC values, but even its values are very high. An example ROC curve for each model can be seen in Figure 4.13.

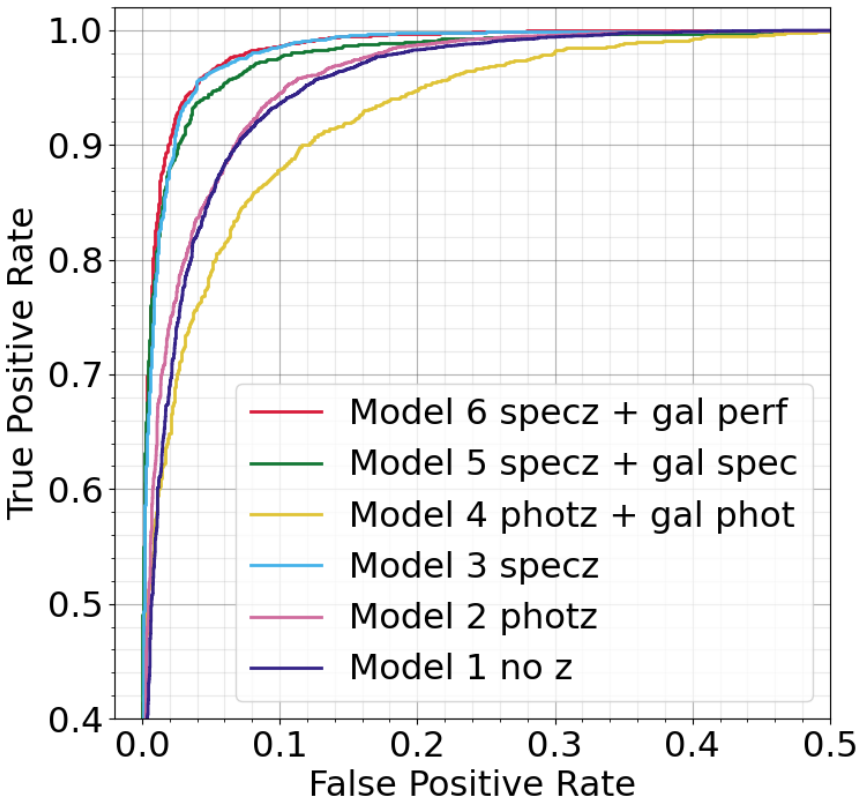

Figure 4.13: A comparison of an example ROC curve produced by each of the 6 models. Model 1 is shown in purple. Model 2 is shown by the pink line. Model 3 is shown in blue. Model 4 is represented by the gold line. Model 5 is shown in green. Finally, Model 6 is shown in red. It can be seen that all 6 models have ROC curves in the upper left quadrant of the graph, representing a good classification model. Model 4 is lower than all of the other models. This is a reflection of the low precision, recall and F1 values. There were also issues with this model training more slowly than the other models. We discuss the slower training issue in Section 4.3. The AUC values for each of the models can be seen in Table 4.1 and in Figure 4.12.

Model	Average precision	Average recall	Average F1-score	AUC
Model 1 no z	$0.961^{0.966}_{0.952}$	$0.766^{0.800}_{0.729}$	$0.851^{0.875}_{0.826}$	$0.973^{0.974}_{0.972}$
Model 2 photz	$0.962^{0.968}_{0.953}$	$0.741^{0.784}_{0.712}$	$0.841^{0.866}_{0.816}$	$0.973^{0.975}_{0.972}$
Model 3 specz	$0.977^{0.987}_{0.972}$	$0.890^{0.912}_{0.864}$	$0.931^{0.947}_{0.915}$	$0.992^{0.994}_{0.988}$
Model 4 photz + gal phot	$0.958^{0.971}_{0.949}$	$0.674^{0.726}_{0.628}$	$0.793^{0.831}_{0.756}$	$0.958^{0.963}_{0.952}$
Model 5 specz + gal spec	$0.980^{0.984}_{0.975}$	$0.847^{0.872}_{0.821}$	$0.908^{0.925}_{0.989}$	$0.988^{0.991}_{0.983}$
Model 6 specz + gal perf	$0.983^{0.988}_{0.976}$	$0.868^{0.882}_{0.854}$	$0.922^{0.932}_{0.911}$	$0.989^{0.994}_{0.990}$

Table 4.1: The average precision value obtained by each of our 6 models, after 10 runs. We also present the average recall, average F1-score and average AUC for each model. The maximum and minimum value for each metric is also given. All the values presented are also shown in Figures 4.9, 4.10, 4.11 and 4.12 with the minimum and maximum values obtained. It can be seen that Models 5 and 6 obtain the best average precision values. Model 4 achieves the worst average precision value. This is probably due to the large uncertainties on the host-galaxy properties creating difficulties for the machine learning model.

and in Figures 4.10, 4.11 and 4.12. We have plotted an example ROC curve of each model, which can be seen in Figure 4.13. As expected the average recall values are lower than the average precision values for each model. This is due to the use of a high threshold value of 0.9, which results in a larger number of false negative classifications. However, when the average F1-score is calculated it can be seen that for models 3, 5 and 6 we have reached a good compromise between precision and recall. Models 3, 5 and 6 all use spectroscopic redshift, whilst models 5 and 6 use host-galaxy M_* and SFR with spectroscopic uncertainties. The average F1 values show that the addition of spectroscopic data and host-galaxy properties create a well-rounded classifier, able to classify between the two classes effectively. Similarly, the high average AUC values for models 3, 5 and 6 show this too. However, the other models also achieve high average AUC values due to their capabilities of classifying SNe Ia.

4.3 Discussion

As discussed in Section 4.1.2 we explored using probability to test the capability of using host-galaxy properties to classify SNe. Each SN was assigned a bin in

3D (host-galaxy redshift, M_* and SFR), then for each bin in the parameter space the probability that a galaxy would host a SN Ia or non-Ia could be calculated. We attempted this with spectroscopic redshift and host-galaxy properties with no uncertainty. The results had a low precision value (0.788). This precision value is significantly lower than any of the machine learning models. A confusion matrix of the results was shown in Figure 4.6. It can be seen that using the probability method fails to classify 65% of SNe Ia. The lookup table precision value is better than if a SN's type was sampled randomly between the two possible classes, Ia and non-Ia.

To further compare the lookup table with the ML model, we combined the probability for an object being a SNe Ia calculated by the lookup table with the probability calculated by SuperNNova model 1 (the light curve only model). The two probabilities were combined by multiplication. As both values are decimals, it is unlikely that the combined probability would have a maximum of 1 or a minimum of 0. To overcome this we normalised the values, by setting the new values to be between 0 and 1. We then set a threshold of 0.6 and produce a confusion matrix to evaluate the classification, seen in the top panel of Figure 4.14. This modification to the lookup table approach produces a precision value of 0.987. This precision value is larger than any of the machine learning models, and is able to produce a larger sample of SNe Ia than the basic lookup table approach. However, 71% of true SNe Ia are still being misclassified as non-Ia. This value is affected by the threshold value. The threshold value of 0.6 was chosen as it maximised both precision and recall. We explore the effect threshold has on the results in Section 4.3.3. We compare this method with our ML method which had no uncertainties on the galaxy properties (model 6) in Figure 4.14.

The initial test of using the lookup table to use probability in the 3D parameter space produced a small, contaminated dataset. If the results had been better we would have explored the effect of adding phot and phot + 4MOST uncertainties to the host-galaxy values. Due to time constraints we prioritiesd the work exploring the machine learning method.

To test the effect different host-galaxy properties had on the machine learning classification, we trained models 4 (phot z + gal phot) and 5 (spec z + gal spec) with only M_*, and a separate case of each model with only SFR. We ran each of the new versions of models 4 and 5 10 times. The results for this can be seen in Figure 4.15. Our results show that using only one galaxy property lowered the precision value. This confirms the results we found previously when exploring the parameter space in 2D, as was shown in Figures 4.4 and 4.5. In particular we wished to understand the effect on our machine learning model if SFR was not used. The fact that removing SFR results in lower precision implies that the model is correctly using SFR, even when some values are nonphysical, i.e. the ML model

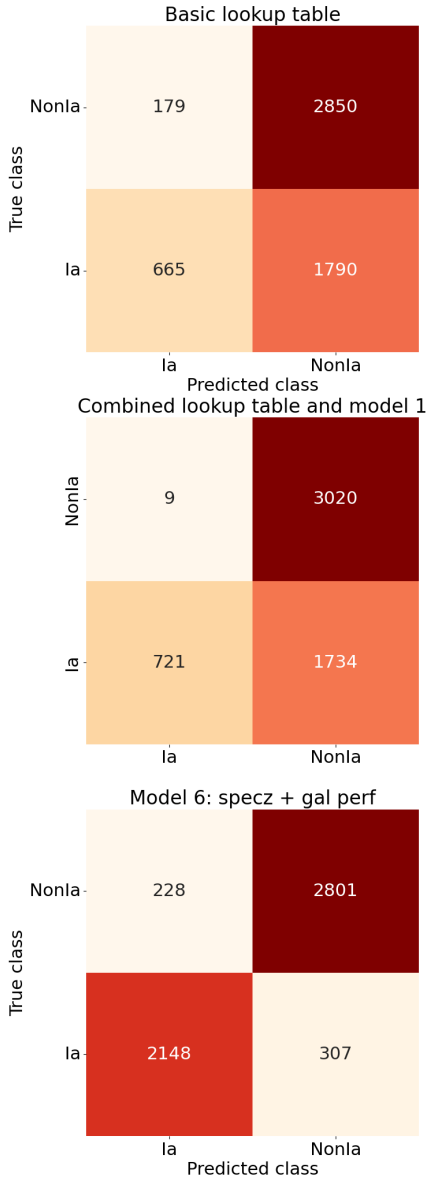

Figure 4.14: Top panel: As shown in Figure 4.5, a confusion matrix of the lookup table. This allows a comparison with the other approaches. Middle panel: A confusion matrix showing the result of combining the probability of an object from the lookup table method with the probability of the same object calculated by SuperNNova model 1 (light curve only). After multiplying the probabilities together the combined probability is normalised. The number show the number of objects falling into each square. The colours represent how much of the data is within a box, the darker the red the more objects whilst the lighter the colour the fewer objects. It can be seen that this variant of the lookup table is able to produce a larger sample of SNe Ia. However, 64% of true SNe Ia are still being misclassified as non-Ia. Bottom panel: The confusion matrix for model 6 using SuperNNova. The combined lookup table and model 1 achieves a precision of 0.987, whilst the machine learning model achieves an average precision value of 0.983.

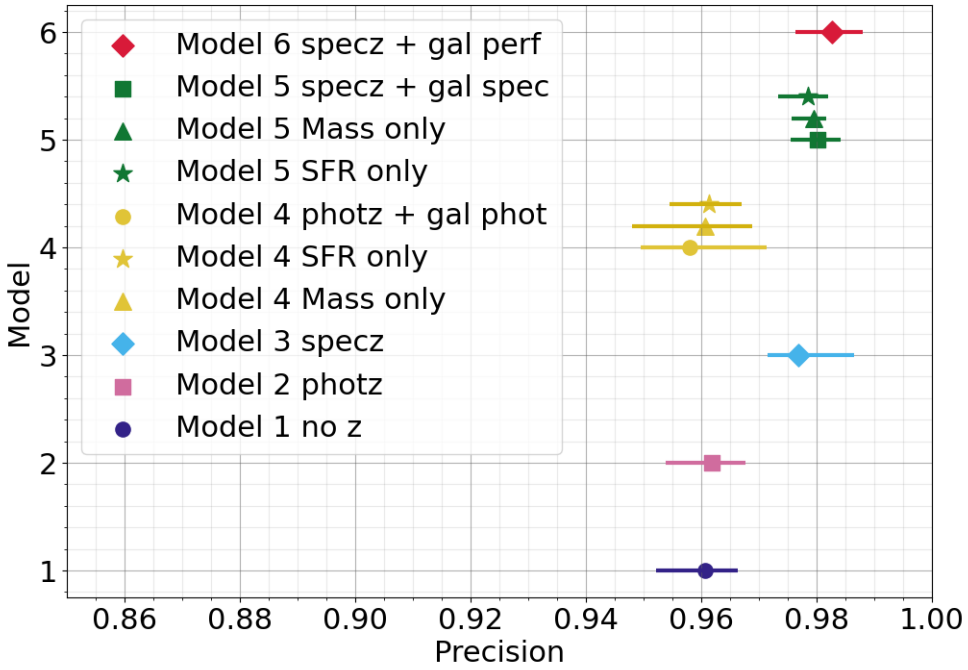

Figure 4.15: The same as Figure 4.9, but with a variant of models 4 (gold triangle) and 5 (green triangle) using only M_*. A separate variant of models 4 (gold star) and 5 (green star) used only SFR. It can be seen that removing SFR or mass for models 4 and 5 had a small effect on the precision.

has learned to recognise -99 and effectively ignores these entries.

4.3.1 Training issues

Upon analysis of our results we have noticed that model 4 didn't train at the same rate as other models. This is unusual as models 5 and 6, which also used host-galaxy properties, trained at the same rate as models 1, 2 and 3. This would explain why model 4's ROC curve is much lower than any of the other models. Model 4 has underfit to the data. We present the training curve of model 4 in Figure 4.16 (top panel), alongside the training of model 6 (bottom panel). We also noticed that when the version of model 4 was trained without host-galaxy M_* the training rate was lower than the regular model 4's training rate. We present the results for the training of model 4 without M_* in Figure 4.16 (middle panel).

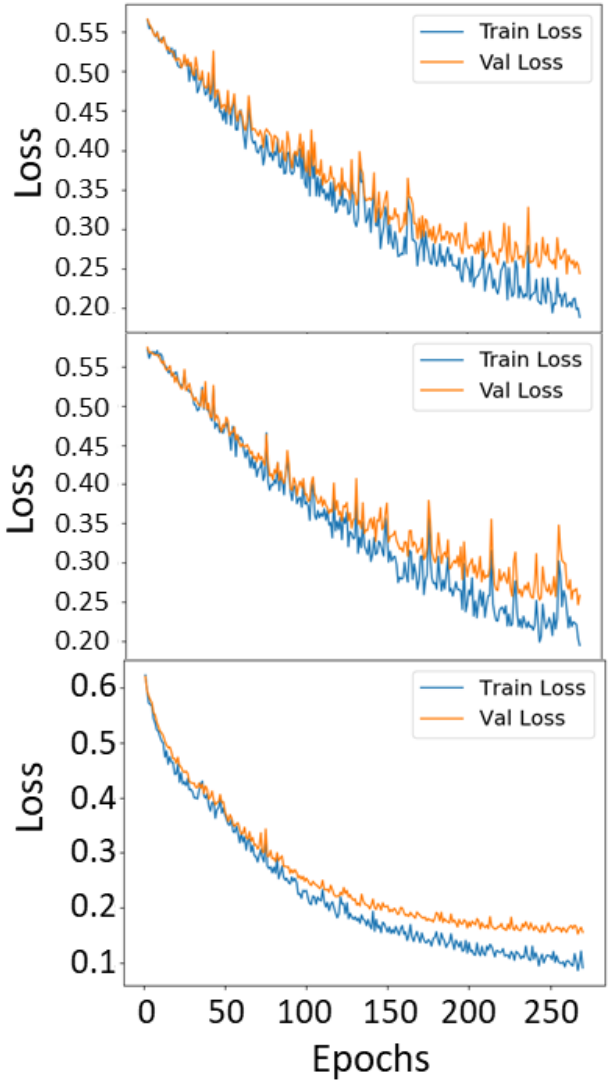

Figure 4.16: Top panel: The training of model 4. Middle panel: The training of model 4 without M_*. Bottom panel: The training of model 6. It can be seen that model 4 took took more epochs to achieve the same loss as model 6 (model 6's training is a good representation of models 1-3 and 5). This explains the lower ROC curve seen for model 4 in Figure 4.13. It can also be seen that the modified version of model 4 which used only SFR trained even slower than the regular model 4.

4.3.2 Comparison with the literature

We find our precision scores for model 1 is similar to the same model carried out by Möller & de Boissière, however our work uses different data to Möller & de Boissière. They obtained a precision value of 0.954 ± 0.005, whilst we achieved a value of 0.961 ± 0.009. Our uncertainty comes from the range of values achieved across the 10 runs. However, their results which use photometric redshift (0.9949 ± 0.0005) outperform our perfect model (0.983 ± 0.007). We can also compare our obtained recall values. We find our models achieve worse recall values than their models. When using only lightcurves, they achieved a recall value of 0.973 ± 0.004, whilst we obtained a value of 0.766 ± 0.037. When redshift is included they found a recall value of 0.9961 ± 0.0009, model 6 achieved a value of 0.868 ± 0.014.

There are a number of reasons why our results differ from those achieved by Möller & de Boissière. They used a different dataset from us. Their data was comprised of simulated light curves using SNANA. However, the exact parameters used for the simulation are not given, nor is the redshift range they use. They were able to normalise their features, which we were not able to. We discuss normalisation in Section 4.4. Möller & de Boissière used larger numbers of SNe Ia and non-Ia (912,691 for each class). Since the default split of the data into train and test would have been optimised for their dataset, future versions of this research would need to explore changing the percentage size of the train and test data sets.

We can compare the results we obtained with other studies which have used host-galaxy information to classify SNe. It should be noted that each of the mentioned studies used different datasets to our work. Foley & Mandel (2013) used a probabilistic method to classify SNe (between Ia, Ibc and II) using only host-galaxy data. For a threshold of 0.9, they achieve a precision of approximately 0.9 and a recall of approximately 0.36. This shows our model improves over theirs. However this is to be expected as we only use two classes and they highlight that their method could be improved by combining their results with a light curve method. Baldeschi et al. (2020) make use of a random forest to classify between SNe Ia and core collapse SNe using light curves and host-galaxy SFR. They achieve comparable precision as Foley & Mandel, which is surprising. This could be due to Baldeschi et al. using a small dataset containing only 249 SNe Ia and 268 core-collapse SNe. Gagliano et al. (2021) also used a random forest to classify between SNe Ia and core-collapse SNe. They achieve a precision of 0.79 and an AUC of 0.72. These values are worse than our worst performing machine learning model (model 4, precision: 0.958, AUC: 0.958). One possible reason for such a difference between their results and ours is they provide their model with 24 different input parameters. Gagliano et al. show that some of these parameters are not correlated with SNe type. Therefore it could be possible that some of the input features are

reducing performance. Additionally, their model was tuned to produce the highest possible accuracy value (the number of correct classifications) not precision, like ours was.

Vincenzi et al. (2023) carried out a study into the effect of SNe Ia sample purity on the measurement of cosmological parameters, using BEAMS with Bias Corrections to model the effects of contamination. BEAMS is a framework designed to include SN classification probabilities and reduce the effect of contamination (Kunz et al., 2007). BEAMS with Bias Corrections is a framework by Kessler & Scolnic (2017) which expands upon BEAMS by using a novel approach for the correction of selection biases and contamination from core-collapse SNe. Vincenzi et al. found that a contamination of $0.8 - 3.5\%$ results in a bias on w of < 0.008, a bias on w_0 of < 0.009, and a bias on w_a of < 0.108. We see from classification with model 1 (no z) to classification with model 5 (specz + gal spec) an improvement in purity of 1.9%. Therefore we could obtain a similar upper limit on the improvement to the bias on the w parameter as Vincenzi et al.. However, Vincenzi et al. include SNe at $z > 1$. A study by Jones et al. (2017) found that a contamination of 2% could introduce a bias of 5% on w. Another study by Campbell et al. (2013) found w to be -0.96 ± 0.1 from a SN Ia sample which had a contamination of 3.9%. However, they found that the contamination had a negligible effect when estimating constraints on w. Further research would need to be carried out to determine the specific improvements our work could achieve to the measurement of w.

4.3.3 Changing threshold value

We explored the effect changing the classification threshold had on our results. We began by observing the effects on the results obtained by the lookup table. We present the results in Figure 4.17 of the recall plotted against precision at different thresholds. It can be seen that whilst the precision value increases, the recall value drops significantly. This plot shows that a large number of SNe Ia being classified correctly is only possible when the confidence of that object being an SNe Ia is low. As the threshold increases, the low recall value reflects more SNe Ia being discarded. As the threshold increases the lookup table is able to produce a purer SNe Ia dataset, at the cost of a smaller dataset. Whilst the ML method is able to produce a larger dataset with comparable purity, with a high threshold. It is currently unknown if a purer or a larger dataset could be used to constrain w or the time varying cosmology parameters (w_0 and w_a). Currently simulations do not contain correlations between w_0 and w_a and the size and/or purity of the SNe Ia sample. This would need to be explored using simulations. Alternatively, the real data from 4MOST will contain these correlations, and will be available in 2024.

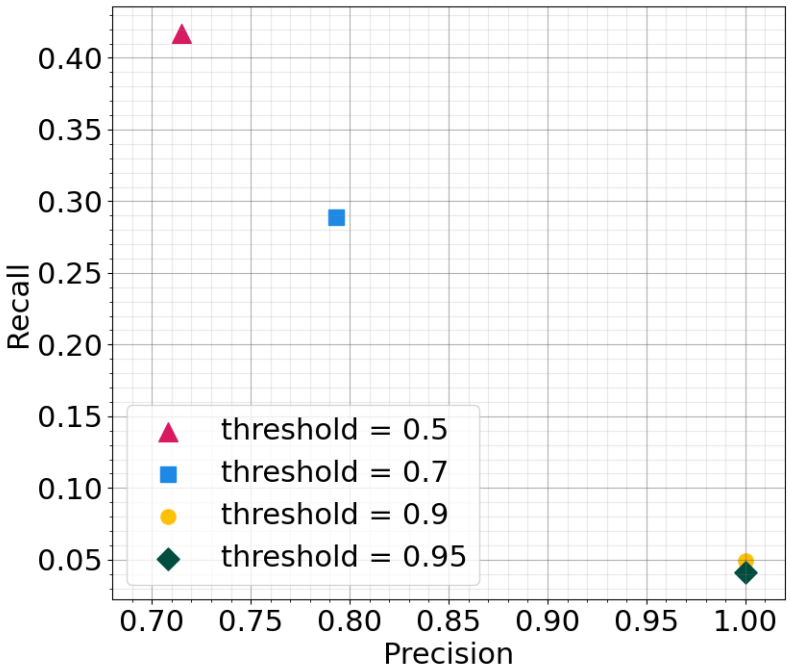

Figure 4.17: The precision plotted against recall values for the lookup table approach for differing thresholds. It can be seen that at thresholds 0.9 and 0.95 a completely pure SNe Ia sample is achieved. However, majority of the SNe Ia are being discarded.

Next we explored the effect changing the classification threshold has on our evaluation metrics produced by SuperNNova: precision, recall and F1-score. The results can be seen in Figures 4.18, 4.19 and 4.20. It can be seen that the precision values increase as the threshold increases, for all models. This is due to the number of false positive objects being reduced. However, the higher threshold leads to a higher number of false negatives which brings the recall value down. As discussed previously, it is extremely difficult to have both high precision and high recall values for a model. The figure shows that the recall value increases for all models as the threshold value decreases. We also observed the effect on the F1-score as threshold increases. Generally, the F1-score decreases for all models as the threshold increases. However, the models which perform worst at classification are affected more (models 1, 2 and 4) than the models which are better at classifying SNe (models 3, 5 and 6). Model 3, 5 and 6 having a higher F1 value at higher thresholds show that the models are capable of creating a pure SNe Ia sample without too many SNe Ia being discarded. We note that the AUC measures a models classification quality regardless of the specific threshold value. Therefore AUC will not change with the threshold value.

4.3.4 Effect on cosmology

To understand the impact of the precision of our ML results on cosmology we carried out an investigation to quantify the effect on the Hubble residual. In order to do this, we carried out the same method used in Figure 3.12. This made use of a sample of Type Ia SNe to calculate the Hubble residual. To simulate the effect of an impure dataset we randomly made a selection of the SNe fainter, by increasing the magnitude value (m_b). For the objects which were made fainter, the magnitude value was increased by a random amount between 0 and 2 magnitudes. The number of objects assigned a fainter magnitude was equivalent to 1 - the average precision measured for each model. We could then measure the RMS. To simplify this investigation, we assumed no mass correction was applied. This meant we could compare our results with the results shown in Table 3.6.

We show our results in 4.2. We found that all precision values increased the RMS, which was to be expected. These results show that even a small amount of contamination will have an effect on the measured uncertainty of w. These results also show that contamination has a much larger effect on the uncertainty of w than our work in the previous Chapter, therefore the effects of contamination need to be investigated further.

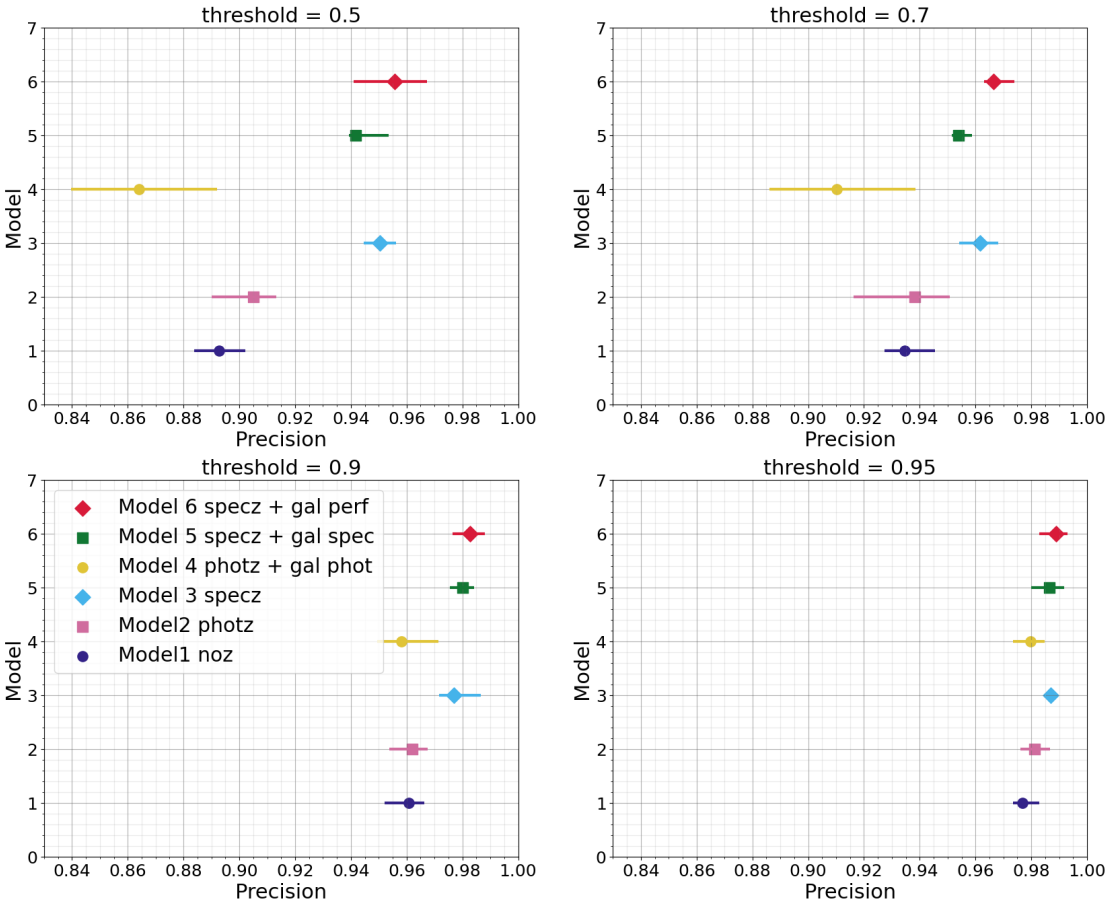

Figure 4.18: The precision value for each model as it changes with the threshold. It can be seen that as the threshold increases the precision value for all models improve. This is due to a decrease in false positives SNe Ia.

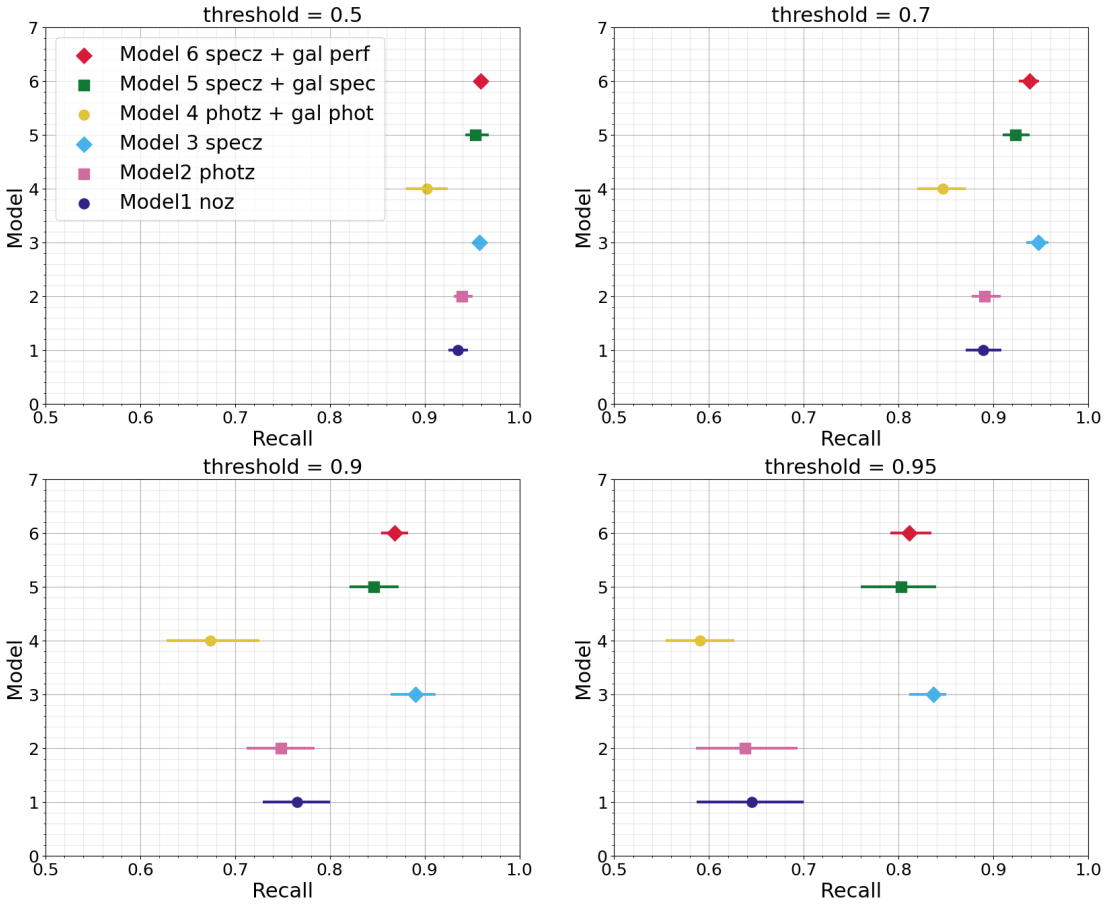

Figure 4.19: The recall value for each model as it changes with the threshold. It can be seen that as as the threshold increases the recall value decreases for all models. This is to be expected, as there is a trade-off between precision and recall when the threshold increases. As the threshold increases the number of false negatives increases.

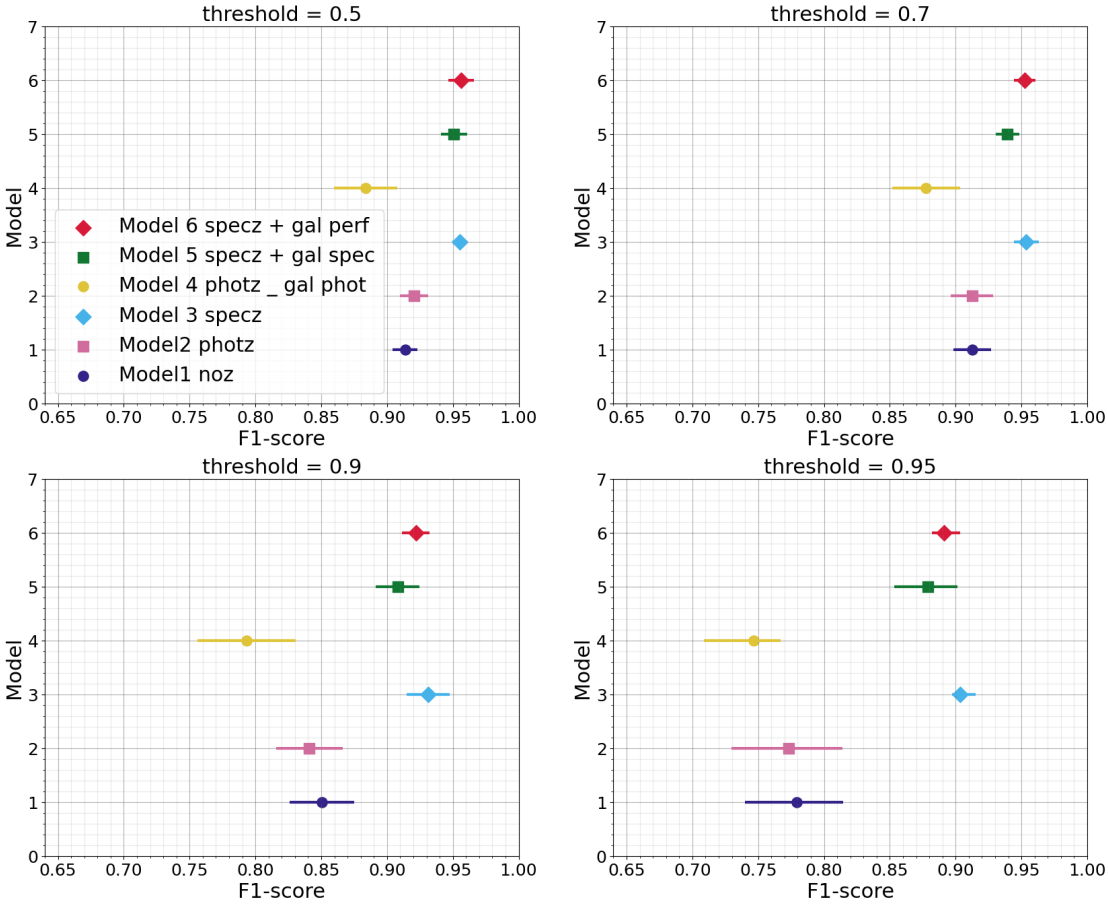

Figure 4.20: The F1-score for each model as it changes with the threshold. It can be seen that the F1-score changes differently with threshold for different models. For the worst models for classification (models 1, 2 and 4) the F1-score decreases drastically as the threshold increases. The models which classify SNe best (models 3, 5 and 6) see less of a decline in F1-score as the threshold increases.

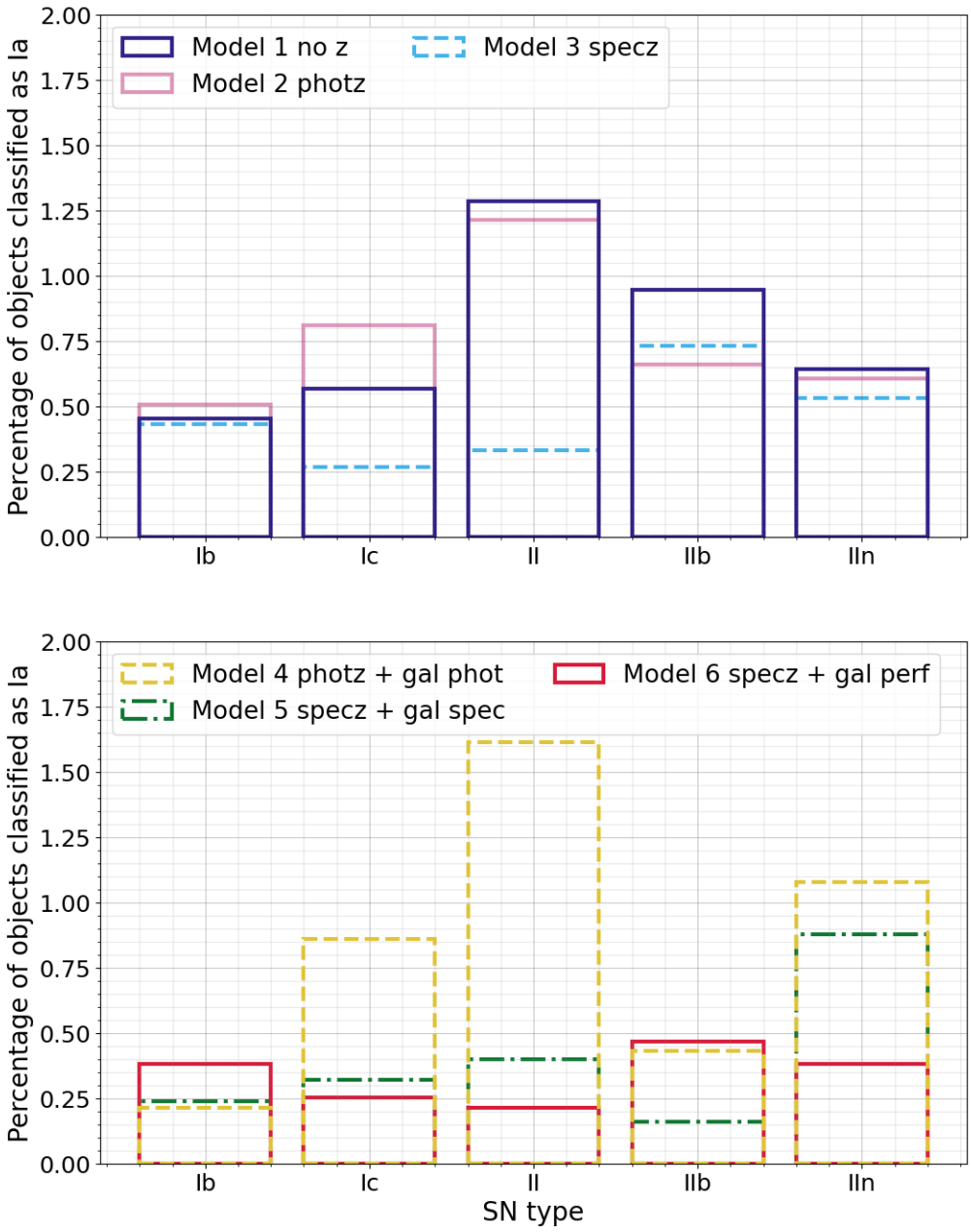

Figure 4.21: A comparison of the objects contaminating SNe Ia samples for each of the ML models. The figures shown are as percentages of the total number of objects classified as a SNe Ia. Top panel: The contaminants for models 1, 2 and 3. Bottom panel: The contaminants for models 4, 5 and 6. For every model, the total of each contaminant SNe in comparison to the SNe Ia sample size is small. However, in order to improve precision the number of contaminants needs to be reduced further. The contamination for the lookup + model 1 was not included in this figure as the contamination amount is too small to draw any meaningful insights.

Correction	RMS Hubble residual (mag)
Perfectly pure	0.1752
Model 1 no z	0.2577
Model 2 photo$z$	0.2349
Model 3 spec$z$	0.2292
Model 4 photo$z$ + gal phot	0.2720
Model 5 spec$z$ + gal spec	0.2114
Model 6 spec$z$ + gal perf	0.2041

Table 4.2: We used the average precision value for each ML to randomly adjust the magnitude of a sample of SNe Ia, to simulate contamination. We could then compare our results with those found in Table 3.6. All of the impurity values increased the RMS for the Hubble residual, which is to be expected.

4.3.5 Contamination exploration

To understand the contamination of our SNe Ia samples further we explored the object types found in our samples. a visualisation of this can be found in Figure 4.21. It can be seen that all machine learning variants struggle with different SNe types. We have included the lookup table combined with ML model 1 in the figure for completion. However, with only 9 objects contaminating the SNe Ia sample it is difficult to draw any meaningful conclusions from the data for this classification approach. Generally, models 1, 2 and 3 have the most of each type of the SN contaminants. As model 5 makes use of 4MOST observations and model 6 used perfect galaxy properties, these are the two we are most interested in improving. Surprisingly, model 5 and model 6 have large contaminants of different SNe. Model 5's largest contaminant is type IIn SNe, whilst model 6's largest contaminant is type IIb. It should be noted that all of the SNe objects are in small numbers compared with the total SNe Ia sample. However, a reduction in any of these SNe classes would reduce contamination. Some contaminants could be removed by raising the classification threshold but this would reduce the number of SNe Ia in the sample too. Another possible way to reduce contamination would be to try a three way classification, between the classes Ia, Ibc (Ib and Ic) and II (II, IIb, IIn). Möller & de Boissière report that using three way classification produced a precision of 0.986 ± 0.002, which improved upon their two way classification precision (0.954 ± 0.005).

4.4 Conclusion

We have shown that using host-galaxy properties with uncertainties from photometry and 4MOST improves the purity of SN Ia samples classified by SuperNNova, compared with using only SN light curves or host-galaxy properties with phot uncertainties. Although 4MOST will not be able to observe as many objects as Rubin, this work has proven that 4MOST will be able to aid in the classification of SNe observed by Rubin and other photometric surveys. Our results are summarised as follows:

- We found that the lookup table combined with the ML performed better at creating a pure SN Ia dataset than using just the machine learning approach, producing a dataset with only 1.2% contamination. However, the machine learning was capable of producing a larger dataset with a comparable contamination (1.9%).
- We found using machine learning the contamination of a SNe Ia sample classified using only SN light curves could be reduced, from 3.9% to 2.3%, with the addition of host-galaxy spectroscopic redshift used during classification. Similarly, a SNe Ia sample classified using light curves and photometric redshift would have the contamination reduced, from 3.8% to 2.3%, when spectroscopic redshift is used instead of photometric redshift. Each of these datasets were larger than the dataset produced by the lookup table.
- We found the addition of host-galaxy M_* and SFR resulted in a reduction of contamination in SNe Ia dataset, from 2.3% to 1.9%, compared with spectroscopic redshift with light curves being used for classification.
- We see similar improvements to the recall, AUC and F1-score values achieved by our models.
- We expect our purer datasets to lead to a reduction in the bias on w. However, further work would need to be carried out to find the exact impact.

There is room for improvement in our work, if the research were to be continued. Future versions of the research should incorporate normalisation of the input parameters. Normalisation would convert all values for each input to be scaled to be between 0 and 1. The normalisation of the host-galaxy properties could prove effective in increasing the precision values. However, normalisation of the SFR values would need to be carried out carefully to ensure the flagged values of -99 were still maintained as a flag value. SuperNNova includes an option to normalise SN light curves, but not the other parameters. The host-galaxy properties would need to be normalised using *scikit-learn*. We did not carry out normalisation due

to time constraints.

All of our work has utilised simulations of observations. As Rubin and 4MOST come online, another next step would be to repeat the work with real data. Similarly, it would be essential to incorporate Euclid and Roman observations as the space telescopes come online. The Dark Energy Survey instrument (DESI) is about to begin carrying out the DESI-II survey which will have 5,000 fibres. DESI covers large sections of the northern hemisphere, which will complement 4MOST. Machine learning algorithms are constantly evolving and improving. It would be imperative that future versions of this research stay up to date with any developments, especially regarding the results of the SN classification challenge PLAsTiCC.

Chapter 5

Conclusion

In this thesis we have presented research quantifying the improvements 4MOST will enable. We find that 4MOST will be able to improve the precision of derived galaxy properties, when 4MOST is combined with LSST observations, compared with galaxy properties derived from photometry alone. In a separate research project we found that the improved galaxy property measurements allow a machine learning model to more accurately classify SNe Ia, resulting in purer data samples. In this Section, we summarise the two main research projects carried out in this thesis.

5.1 Measuring host galaxy properties

Our research has confirmed that 4MOST can effectively be used in synergy with imaging telescopes, such as the Vera Rubin Observatory, to achieve a more precise measurement of galaxy properties compared to those obtained solely through photometric methods. This outcome, which aligns with our expectations, has been quantified in our study. Our results are summarised as follows:

- For elliptical galaxies with brighter magnitudes ($r \leq 22$) the uncertainty for a galaxy's log (M_*) is 49 – 95% smaller when spectroscopy is used with photometry, compared with that derived from photometry alone. The range of improvements depends on the magnitude and redshift.

- At fainter magnitudes ($r \geq 23$) the M_* uncertainty is 24 – 71% smaller for phot + 4MOST, compared to photometry alone. The range depends on magnitude and redshift, with more improvements seen for brighter galaxies.

- We see similar improvements to the precision of Sc host-galaxy masses, when adding 4MOST spectroscopy.
- We applied our uncertainties derived from FAST to real SNe host-galaxy masses. The smaller uncertainties produced when adding 4MOST spectroscopy make it easier to distinguish which side of the 10^{10} ($M_{\odot}$) line a host galaxy falls. This has implications for applying the correct mass step corrections in cosmological analysis.
- Other galaxy properties see a significant improvement in uncertainties, including: star formation rate, age, V-band extinction, metallicity, specific star formation rate and star formation timescale.
- Whilst there is not as significant an improvement to the precision of a galaxy's metallicity, there is still an improvement at most simulated magnitudes and redshifts.
- Our improved precision of host-galaxy mass would improve the mass-correction applied to standardise SNe Ia. We calculate that this increased precision of correction would result in a reduction in the uncertainties of w of 2%

Recently there have been discussions within the 4MOST community around the exposure time carried out at each pointing. When our work was carried out it was set to be 1 hour. The current discussion seems to indicate that the exposure time will be much shorter. If multiple observations are carried out for a galaxy, then a longer exposure time will be possible. However, the observing strategy has not yet been decided. The improved precision for each galaxy property we measure would be effected by a change in exposure time. The exposure time discussion needs to be followed closely, and any changes implemented in future research.

The correlations between SNe Ia peak brightness and host-galaxy properties are one of the main systematic effects in SNe Ia cosmology. The improved galaxy property measurements from 4MOST and Rubin we calculated have the potential to improve the corrections used in SN cosmology. We have chosen to focus on the host-galaxy mass and the implications to the mass-step, but a deeper investigation could be carried out with all of the galaxy properties. Simulations do not currently contain correlations between the additional properties and SN type. However, the data observed by 4MOST will, which will be available in 2024. The galaxy-property precision improvements come from the fact that spectra contain significantly more information than photometry, which leads to the breaking of degeneracies.

Our work has explored the effect we expect 4MOST's measurement of galaxy properties to have on w. However, the time varying model parameters w_0 and w_a are likely to be more sensitive to systematic effects from galaxy properties.

5.2 Machine learning classification of SNe Ia

Our research demonstrates that by incorporating host-galaxy properties, simulated with uncertainties from phot + 4MOST, we can enhance the purity of SN Ia samples classified by SuperNNova. This improvement is particularly notable when compared to relying solely on SN light curves or host-galaxy properties with photometric uncertainties. While we acknowledge that 4MOST may not have the observational capacity of Rubin for a large number of objects, our findings prove its valuable potential in contributing to the classification of supernovae observed by Rubin and other photometric surveys. Our results are summarised as follows:

- We found that the lookup table combined with the ML performed better at creating a pure SN Ia dataset than using just the machine learning approach, producing a dataset with only 1.2% contamination. However, the machine learning was capable of producing a larger dataset with a comparable contamination (1.9%).
- We found using machine learning the contamination of a SNe Ia sample classified using only SN light curves could be reduced, from 3.9% to 2.3%, with the addition of host-galaxy spectroscopic redshift used during classification. Similarly, a SNe Ia sample classified using light curves and photometric redshift would have the contamination reduced, from 3.8% to 2.3%, when spectroscopic redshift is used instead of photometric redshift. Each of these datasets were larger than the dataset produced by the lookup table.
- We found the addition of host-galaxy M_* and SFR resulted in a reduction of contamination in SNe Ia dataset, from 2.3% to 1.9%, compared with spectroscopic redshift with light curves being used for classification.
- We see similar improvements to the recall, AUC and F1-score values achieved by our models.
- We expect our purer datasets to lead to a reduction in the bias on w. However, further work would need to be carried out to find the exact impact.

The next stage of the research would be to expand the wavelength ranges of the galaxy observations, by including additional observations. To reduce the uncertainty of each of the galaxy properties we could incorporate infrared data from Euclid. We expect the reduced uncertainty from the addition of infrared data would improve the results of the machine learning classification. However, models with reduced host-galaxy property uncertainties would not be able to improve upon the precision value found for model 6 which has no uncertainty. We chose to focus on the effect of M_* and SFR on SNe classification. However, our previous

work in Section 3 measured uncertainties for other galaxy properties. A deeper investigation could be carried out with all of the galaxy properties.

The most exciting way in which the research could be improved would be to incorporate real data as 4MOST and Rubin come online. We would carry out an in-depth study to determine the effect of contamination on the ability to measure cosmological parameters. We would carry out a full simulation, which includes the effect of 4MOST's observing strategy. This simulation would allow us to understand if the cosmological parameters are better constrained by a small but pure sample of SNe Ia, or better constrained by a large sample. This work would make use of the Beams Bias Corrections framework (Kessler & Scolnic, 2017). Finally, we explored the latest machine learning models to carry out this research. Each month there are new machine learning models being created and old models being updated. It will be exciting to see where SN machine learning research goes in the near future to improve SN classification.

5.3 Summary

We have quantified some of the capabilities of 4MOST in preparation for next generation surveys. TiDES will utilise 4MOST for transient follow up of a selection of LSST transients. The synergy between these two surveys will make them both powerful tools for cosmology. We have demonstrated the improved precision which LSST and 4MOST will be able to measure SN host-galaxy properties, both for elliptical and Sc galaxies. We proved that the improved measurements of stellar mass will lead to a more accurate correction applied to standardise SNe Ia.

We then applied the improved host-galaxy property measurements to a simulated SNe sample to prove that SNe host-galaxy properties could be used to classify SNe using a lookup table. We then took this a step further by adapting a deep-learning model to use SNe light curves and host-galaxy stellar mass and star formation rate to produce a purer SNe Ia sample. We have demonstrated the ability for 4MOST to aid in the classification of SNe for which live transients were not taken, but observations of the host-galaxy were taken. We have also shown that SN host-galaxy properties can be a useful tool to aid in the machine learning classification of SNe.

We have shown original research conducted by utilising data science and deep learning techniques, to provide a robust foundation in quantifying the capabilities of 4MOST. The implications of this thesis work extend beyond mere analysis, we hope our work will catalyze an improvement on the constraints of the cosmological parameters. Improved measurements of the dark energy equation of state parameter, w, will improve our understanding of the expansion of the Universe.

References

Abbott T. M. C., et al., 2019, ApJ, 872, L30

Addison G. E., Hinshaw G., Halpern M., 2013, MNRAS, 436, 1674

Ade P., et al., 2019, Journal of Cosmology and Astroparticle Physics, 2019, 056

Akeson R., et al., 2019, The Wide Field Infrared Survey Telescope: 100 Hubbles for the 2020s (arXiv:1902.05569)

Alam S., et al., 2015, The Astrophysical Journal Supplement Series, 219, 12

Alam S., et al., 2021, Physical Review D, 103, 083533

Altman N. S., 1992, The American Statistician, 46, 175

Amendola L., et al., 2013, Living Reviews in Relativity, 16, 6

Amendola L., et al., 2018, Living Reviews in Relativity, 21, 2

Angwin J., Larson J., Mattu S., Kirchner L., 2016, Propublica

Arnett W. D., Bahcall J. N., Kirshner R. P., Woosley S. E., 1989, Annual Review of Astronomy and Astrophysics, 27, 629

Astier P., et al., 2014, Astronomy & Astrophysics, 572, A80

Aubourg É., et al., 2015, Physical Review D, 92, 123516

Bailey A. C., Vincenzi M., Scolnic D., Cuillandre J. C., Rhodes J., Hook I., Peterson E. R., Popovic B., 2023, MNRAS, 524, 5432

Baldeschi A., Miller A., Stroh M., Margutti R., Coppejans D. L., 2020, ApJ, 902, 60

Balland C., et al., 2009, Astronomy and Astrophysics, 507, 85

Barbon R., Ciatti F., Rosino L., 1979, Astronomy and Astrophysics, 72, 287

Barris B. J., Tonry J. L., 2004, ApJ, 613, L21

Bastian N., Covey K. R., Meyer M. R., 2010, Annual Review of Astronomy and Astrophysics, 48, 339

Bell E. F., de Jong R. S., 2000, MNRAS, 312, 497

Bell E. F., de Jong R. S., 2001, ApJ, 550, 212

Ben-Ami S., et al., 2022, Proceedings of the SPIE, 12181

Bentz M. C., et al., 2013, ApJ, 767, 149

Bethe H. A., Brown G. E., Applegate J., Lattimer J. M., 1979, Nuclear Physics A, 324, 487

Blondin S., Tonry J. L., 2007, ApJ, 666, 1024

Bohlin R. C., Gordon K. D., Tremblay P.-E., 2014, Publications of the Astronomical Society of the Pacific, 126, 711

Bohlin R. C., Hubeny I., Rauch T., 2020, The Astrophysical Journal, 160, 21

Boone K., 2019, ApJ, 158, 257

Boone K., et al., 2021a, The Astrophysical Journal, 912, 70

Boone K., et al., 2021b, The Astrophysical Journal, 912, 71

Botticella M. T., et al., 2017, Astronomy and Astrophysics, 598, A50

Branch D., Tammann G. A., 1992, Annual review of astronomy and astrophysics, 30, 359

Branch D., Livio M., Yungelson L. R., Boffi F. R., Baron E., 1995, PASP, 107, 1019

Briday M., et al., 2022, Astronomy and Astrophysics, 657, A22

Brout D., Scolnic D., 2021, ApJ, 909, 26

Brout D., et al., 2019, ApJ, 874, 150

Brout D., et al., 2022, ApJ, 938, 110

Brush S. G., 1967, Rev. Mod. Phys., 39, 883

Bruzual G., Charlot S., 2003, MNRAS, 344, 1000

Buolamwini J., Gebru T., 2018, in Friedler S. A., Wilson C., eds, Proceedings of Machine Learning Research Vol. 81, Proceedings of the 1st Conference on Fairness, Accountability and Transparency. PMLR, pp 77–91, https://proceedings.mlr.press/v81/buolamwini18a.html

Caldwell R. R., Dave R., Steinhardt P. J., 1998, Physical Review Letters, 80, 1582

Campbell H., et al., 2013, ApJ, 763, 88

Carrick J. E., Hook I. M., Swann E., Boone K., Frohmaier C., Kim A. G., Sullivan M., LSST Dark Energy Science Collaboration 2021, MNRAS, 508, 1

Carroll S. M., 2001, Living Reviews in Relativity, 4

Chabrier G., 2003, PASP, 115, 763

Chandrasekhar S., 1931, ApJ, 74, 81

Chen R., et al., 2022, ApJ, 938, 62

Chevallier M., Polarski D., 2001, International Journal of Modern Physics D, 10, 213

Childress M., et al., 2013, ApJ, 770, 107

Chung C., Yoon S.-J., Park S., An S., Son J., Cho H., Lee Y.-W., 2023, arXiv e-prints, p. arXiv:2310.06011

Conley A., et al., 2011, ApJS, 192, 1

Conroy C., 2013, Annual Review of Astronomy and Astrophysics, 51, 393

Cortes C., Vapnik V., 1995, Machine learning, 20, 273

Dai M., Kuhlmann S., Wang Y., Kovacs E., 2018, MNRAS, 477, 4142

Desert F. X., Boulanger F., Puget J. L., 1990, Astronomy and Astrophysics, 237, 215

Dhawan S., Thorp S., Mandel K. S., Ward S. M., Narayan G., Jha S. W., Chant T., 2022a, MNRAS, p. arXiv:2211.07657

Dhawan S., et al., 2022b, MNRAS, 510, 2228

Dumayne J., et al., 2023, RAS Techniques and Instruments, 2, 453

Eifler T., et al., 2021, MNRAS, 507, 1514

Euclid Collaboration et al., 2022, Astronomy & Astrophysics, 662, A112

Feast M., 1999, PASP, 111, 775

Filippenko A. V., 1997, Annual Review of Astronomy and Astrophysics, 35, 309

Filippenko A. V., Matheson T., Ho L. C., 1993, ApJ, 415, L103

Foley R. J., Mandel K., 2013, ApJ, 778, 167

Freedman W. L., 1998, Proceedings of the National Academy of Sciences, 95, 2

Freedman W. L., et al., 2019, ApJ, 882, 34

Friedman A., 1922, Zeitschrift für Physik, 31, 1999

Friedman J. H., 2002, Computational Statistics & Data Analysis, 38, 367

Frieman J. A., et al., 2008, ApJ, 135, 338

Fukushima K., 1975, Biological Cybernetics, 20, 121

Gagliano A., Narayan G., Engel A., Carrasco Kind M., LSST Dark Energy Science Collaboration 2021, ApJ, 908, 170

Galbany L., et al., 2014, Astronomy & Astrophysics, 572, A38

Galbany L., et al., 2022, Astronomy and Astrophysics, 659, A89

Gallagher J. S., Garnavich P. M., Caldwell N., Kirshner R. P., Jha S. W., Li W., Ganeshalingam M., Filippenko A. V., 2008, ApJ, 685, 752

Garnavich P. M., et al., 1998, ApJ, 509, 74

Giallongo E., D'Odorico S., Fontana A., Cristiani S., Egami E., Hu E., McMahon R. G., 1998, ApJ, 115, 2169

González-Gaitán S., de Jaeger T., Galbany L., Mourão A., Paulino-Afonso A., Filippenko A. V., 2021, MNRAS, 508, 4656

Graur O., Bianco F. B., Huang S., Modjaz M., Shivvers I., Filippenko A. V., Li W., Eldridge J. J., 2017a, ApJ, 837, 120

Graur O., Bianco F. B., Modjaz M., Shivvers I., Filippenko A. V., Li W., Smith N., 2017b, ApJ, 837, 121

Grillo C., Lombardi M., Bertin G., 2008, Astronomy and Astrophysics, 477, 397

Grossen L., Kronig L., Araujo R., Caseiro S., Sayres C., Sánchez-Gallego J., Bouri M., Kneib J. P., 2020, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series. p. 114478P, doi:10.1117/12.2562545

Groves B., Dopita M. A., Sutherland R. S., Kewley L. J., Fischera J., Leitherer C., Brandl B., van Breugel W., 2008, ApJS, 176, 438

Guiglion G., et al., 2019a, The Messenger, 175, 17

Guiglion G., et al., 2019b, Published in The Messenger vol. 175, pp. 17-21, March 2019.

Gupta R. R., et al., 2011, ApJ, 740, 92

Guy J., et al., 2007, Astronomy and Astrophysics, 466, 11

Guy J., et al., 2010, Astronomy and Astrophysics, 523, A7

Hakobyan A. A., Adibekyan, V. Zh. Aramyan, L. S. Petrosian, A. R. Gomes, J. M. Mamon, G. A. Kunth, D. Turatto, M. 2012, Astronomy and Astrophysics, 544, A81

Hamuy M., Trager S. C., Pinto P. A., Phillips M. M., Schommer R. A., Ivanov V., Suntzeff N. B., 2000, The Astrophysical Journal, 120, 1479

Hamuy M., Trager S. C., Pinto P. A., Phillips M. M., Schommer R. A., Ivanov V., Suntzeff N. B., 2001, ApJ, 122, 3506

Harkness R. P., et al., 1987, ApJ, 317, 355

Hazumi M., et al., 2020, in Lystrup M., Perrin M. D., eds, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 11443, Space Telescopes and Instrumentation 2020: Optical, Infrared, and Millimeter Wave. p. 114432F (arXiv:2101.12449), doi:10.1117/12.2563050

Hložek R., et al., 2023, ApJS, 267, 25

Hochreiter S., Bengio Y., 2001. https://api.semanticscholar.org/CorpusID:17278462

Hochreiter S., Schmidhuber J., 1997, Neural Computation, 9, 1735

Holz D. E., Hughes S. A., 2005, ApJ, 629, 15

Homeier N. L., 2005, ApJ, 620, 12

Hounsell R., et al., 2018, ApJ, 867, 23

Hoyle F., Fowler W. A., 1960, ApJ, 132, 565

Hubble E., 1929, ApJ, 15, 168

Huppenkothen D., et al., 2023, arXiv e-prints, p. arXiv:2310.12528

Ivezić Ž.., the LSST Science Collaboration 2013, LSST Science Requirements Document, `http://ls.st/LPM-17`

Ivezić Ž., et al., 2019, ApJ, 873, 111

Jeong H., et al., 2009, Monthly Notices of the Royal Astronomical Society, 398, 2028

Jha S., Riess A. G., Kirshner R. P., 2007, ApJ, 659, 122

Johnson J., Khoshgoftaar T., 2019, Journal of Big Data, 6, 27

Jones D. O., et al., 2017, ApJ, 843, 6

Jones G. T., Stanway E. R., Carnall A. C., 2022, MNRAS, 514, 5706

Karachentsev I. D., Karachentseva V. E., Huchtmeier W. K., Makarov D. I., 2004, ApJ, 127, 2031

Karttunen H., 2007, Springer

Kaspi S., 2001, in Peterson B. M., Pogge R. W., Polidan R. S., eds, Astronomical Society of the Pacific Conference Series Vol. 224, Probing the Physics of Active Galactic Nuclei. p. 347 (`arXiv:astro-ph/0009488`), doi:10.48550/arXiv.astro-ph/0009488

Kauffmann G., et al., 2003, MNRAS, 341, 33

Kelly P. L., Hicken M., Burke D. L., Mandel K. S., Kirshner R. P., 2010, ApJ, 715, 743

Kennicutt R. C., 1998, Annual Review of Astronomy and Astrophysics, 36, 189

Kessler R., Scolnic D., 2017, ApJ, 836, 56

Kessler R., et al., 2009, PASP, 121, 1028

Kinney A. L., Calzetti D., Bohlin R. C., McQuade K., Storchi-Bergmann T., Schmitt H. R., 1996, ApJ, 467, 38

Kokusho T., Kaneda H., Bureau M., Suzuki T., Murata K., Kondo A., Yamagishi M., 2017, Astronomy and Astrophysics, 605, A74

Kowalski M., et al., 2008, ApJ, 686, 749

Kriek M., Conroy C., 2013, ApJ, 775, L16

Kriek M., van Dokkum P. G., Labbé I., Franx M., Illingworth G. D., Marchesini D., Quadri R. F., 2009, ApJ, 700, 221

Kunz M., Bassett B. A., Hlozek R. A., 2007, Phys. Rev. D, 75, 103508

Lampeitl H., et al., 2010, ApJ, 722, 566

Laureijs R., et al., 2011, arXiv e-prints, p. arXiv:1110.3193

Leavitt H. S., Pickering E. C., 1912, Harvard College Observatory Circular, 173, 1

Li M., et al., 2022, arXiv e-prints, p. arXiv:2211.01382

Lidman C., et al., 2020, MNRAS, 496, 19

Linder E. V., 2003, Physical Review Letters, 90, 091301

Lokken M., et al., 2023, Monthly Notices of the Royal Astronomical Society, 520, 2887

Lower S., Narayanan D., Leja J., Johnson B. D., Conroy C., Davé R., 2020, ApJ, 904, 33

MacArthur L. A., Courteau S., Holtzman J. A., 2003, ApJ, 582, 689

Macri L. M., Riess A. G., 2009, in Guzik J. A., Bradley P. A., eds, American Institute of Physics Conference Series Vol. 1170, Stellar Pulsation: Challenges for Theory and Observation. pp 23–25, doi:10.1063/1.3246452

Mannucci Della Valle, M. Panagia, N. Cappellaro, E. Cresci, G. Maiolino, R. Petrosian, A. Turatto, M. 2005, Astronomy and Astrophysics, 433, 807

Maoz D., Mannucci F., Nelemans G., 2014, Annual Review of Astronomy and Astrophysics, 52, 107

Marshall P., et al., 2017, Lsst Science Collaborations Observing Strategy White Paper: "Science-Driven Optimization Of The Lsst Observing Strategy", doi:10.5281/ZENODO.842713, https://zenodo.org/record/842713

Minkowski R., 1941, PASP, 53, 224

Modjaz M., et al., 2014a, ApJ, 147, 99

Modjaz M., et al., 2014b, ApJ, 147, 99

Modjaz M., et al., 2020, ApJ, 892, 153

Möller A., et al., 2016, Journal of Cosmology and Astroparticle Physics, 2016, 008

Muthukrishna D., Narayan G., Mandel K. S., Biswas R., Hložek R., 2019a, PASP, 131, 118002

Muthukrishna D., Parkinson D., Tucker B. E., 2019b, ApJ, 885, 85

Möller A., de Boissiè re T., 2019a, Monthly Notices of the Royal Astronomical Society, 491, 4277

Möller A., de Boissière T., 2019b, MNRAS, 491, 4277

Narayan G., ELAsTiCC Team 2023, in American Astronomical Society Meeting Abstracts. p. 117.01

Nyholm A., et al., 2020, Astronomy and Astrophysics, 637, A73

O'Connell R. W., 1980, ApJ, 236, 430

Oemler A. J., Tinsley B. M., 1979, Astronomical Journal, 84, 985

Pacifici C., et al., 2023, ApJ, 944, 141

Papovich C., Dickinson M., Ferguson H. C., 2001, ApJ, 559, 620

Peebles P. J. E., 1968, ApJ, 153, 1

Peebles P. J. E., Ratra B., 2003, Reviews of Modern Physics, 75, 559

Perlmutter S., et al., 1999, ApJ, 517, 565

Peterson B. M., et al., 2014, ApJ, 795, 149

Phillips M. M., 1993, ApJ, 413, L105

Phillips M. M., Lira P., Suntzeff N. B., Schommer R. A., Hamuy M., Maza J., 1999, ApJ, 118, 1766

Plackett R. L., 1972, Biometrika, 59, 239

Planck Collaboration et al., 2020, Astronomy and Astrophysics, 641, A6

Qu H., Sako M., 2022, ApJ, 163, 57

Rabin D., 1982, ApJ, 261, 85

Ratra B., Peebles P. J. E., 1988, Phys. Rev. D, 37, 3406

Revsbech E. A., Trotta R., van Dyk D. A., 2017, MNRAS, 473, 3969

Riess A. G., Press W. H., Kirshner R. P., 1996, ApJ, 473, 88

Riess A. G., et al., 1998, The Astronomical Journal, 116, 1009

Riess A. G., et al., 2004b, ApJ, 607, 665

Riess A. G., et al., 2004a, ApJ, 607, 665

Riess A. G., et al., 2007, ApJ, 659, 98

Riess A. G., et al., 2022, ApJ, 934, L7

Rigault M., et al., 2020, Astronomy and Astrophysics, 644, A176

Rose B. M., Rubin D., Strolger L., Garnavich P. M., 2021, ApJ, 909, 28

Rosenblatt F., 1960, Proceedings of the IRE, 48, 301

Sako M., et al., 2008, ApJ, 135, 348

Salim S., Narayanan D., 2020, Annual Review of Astronomy and Astrophysics, 58, 529

Salim S., et al., 2007, ApJS, 173, 267

Sawicki M., Schade D., Mallén-Ornelas G., Durand D., Barrientos L. F., 1999, in Weymann R., Storrie-Lombardi L., Sawicki M., Brunner R., eds, Astronomical Society of the Pacific Conference Series Vol. 191, Photometric Redshifts and the Detection of High Redshift Galaxies. p. 74

Sayres C., et al., 2021, The Astronomical Journal, 161, 92

Schlafly E. F., Finkbeiner D. P., 2011, The Astrophysical Journal, 737, 103

Schlafly E. F., Finkbeiner D. P., Schlegel D. J., Jurić M., Željko Ivezić Gibson R. R., Knapp G. R., Weaver B. A., 2010, The Astrophysical Journal, 725, 1175

Schlegel E. M., 1990, MNRAS, 244, 269

Schlegel D. J., Finkbeiner D. P., Davis M., 1998, The Astrophysical Journal, 500, 525

Scolnic D. M., et al., 2018, ApJ, 859, 101

Scolnic D., et al., 2022, ApJ, 938, 113

Serra P., Trager S. C., 2007, MNRAS, 374, 769

Shivvers I., et al., 2017, PASP, 129, 054201

Silva L., Granato G. L., Bressan A., Danese L., 1998, ApJ, 509, 103

Smith L. N., 2015, arXiv e-prints, p. arXiv:1506.01186

Smith M., et al., 2020, MNRAS, 494, 4426

Speagle J. S., Steinhardt C. L., Capak P. L., Silverman J. D., 2014, ApJS, 214, 15

Spergel D., et al., 2015, arXiv e-prints, p. arXiv:1503.03757

Spinrad H., 1972, ApJ, 171, 463

Sullivan M., Balland C., 2008, The Messenger, 133, 42

Sullivan M., et al., 2006, ApJ, 648, 868

Sullivan M., et al., 2010, MNRAS, 406, 782

Swann E., et al., 2019, The Messenger, 175, 58

Tempel E., et al., 2020, MNRAS, 497, 4626

The Dark Energy Survey Collaboration 2005, arXiv e-prints, pp astro–ph/0510346

The LSST Dark Energy Science Collaboration et al., 2018a, arXiv e-prints, p. arXiv:1809.01669

The LSST Dark Energy Science Collaboration et al., 2018b, arXiv e-prints, p. arXiv:1809.01669

The PLAsTiCC team et al., 2018, The Photometric LSST Astronomical Time-series Classification Challenge (PLAsTiCC): Data set (arXiv:1810.00001)

Tripp R., 1998, Astronomy and Astrophysics, 331, 815

Uddin S. A., Mould J., Lidman C., Ruhlmann-Kleider V., Zhang B. R., 2017, ApJ, 848, 56

Vilarino R., Vicente R., 2021, An experiment on the mechanisms of racial bias in ML-based credit scoring in Brazil (arXiv:2011.09865)

Vincenzi M., 2022, Student Thesis: Doctoral Thesis. University of Porstmouth.

Vincenzi M., et al., 2021, MNRAS, 505, 2819

Vincenzi M., et al., 2023, MNRAS, 518, 1106

Walcher J., Groves B., Budavári T., Dale D., 2011, Astrophysics & Space Science, 331, 1

Werbos P., 1974, Ph.D. Thesis, Harvard University, Cambridge

Wheeler J. C., Harkness R. P., 1986, in Madore B. F., Tully R. B., eds, NATO Advanced Study Institute (ASI) Series C Vol. 180, Galaxy Distances and Deviations from Universal Expansion. pp 45–54

Wiseman P., et al., 2020, MNRAS, 498, 2575

Wiseman P., et al., 2021, MNRAS, 506, 3330

Wiseman P., Sullivan M., Smith M., Popovic B., 2023, MNRAS, 520, 6214

Wolf R. C., et al., 2016, ApJ, 821, 115

Wood-Vasey W. M., et al., 2008, ApJ, 689, 377

Worthey G., 1994, ApJS, 95, 107

York D. G., et al., 2000, ApJ, 120, 1579

Yungelson L., Livio M., 1998, ApJ, 497, 168

chul Lee H., Worthey G., Trager S. C., Faber S. M., 2007, The Astrophysical Journal, 664, 215

de Barros S., Schaerer D., Stark D. P., 2014, Astronomy and Astrophysics, 563, A81

de Jong R. S., et al., 2016, in Evans C. J., Simard L., Takami H., eds, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 9908, Ground-based and Airborne Instrumentation for Astronomy VI. p. 99081O, doi:10.1117/12.2232832

de Jong R. S., et al., 2019, The Messenger, 175, 3