UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

COSMOLOGY WITH TYPE IA SUPERNOVAE: ROBUST TECHNIQUES FOR DATA ANALYSIS AND PHOTOMETRIC CLASSIFICATION

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

MI DAI
Norman, Oklahoma
2017

COSMOLOGY WITH TYPE IA SUPERNOVAE:
ROBUST TECHNIQUES FOR DATA ANALYSIS
AND PHOTOMETRIC CLASSIFICATION

A DISSERTATION APPROVED FOR THE
HOMER L. DODGE DEPARTMENT OF PHYSICS AND ASTRONOMY

BY

Dr. Yun Wang, Chair

Dr. Xinyu Dai

Dr. Ronald Kantowski

Dr. Mukremin Kilic

Dr. S. Lakshmivarahan

To Yiming, and my parents, for their love and support.

Acknowledgements

First and foremost, I would like to thank my advisor, Yun Wang, for all her guidance and support along the way of finishing this dissertation. She has been a great mentor, encouraging me to think and work independently, and guiding me to be always on track of this PhD journey. I would not be able to achieve so much without her help.

I would like to thank Steve Kuhlmann, for providing me with an opportunity to work at the Argonne National Lab in Summer 2016, and offering me a lot of advice and help during my time there. I would also like to thank the Argonne supernova group members, Eve Kovacs, Ravi Gupta and Hal Spinka, for all the help and discussions.

I thank my committee members, Xinyu Dai, Ronald Kantowski, Mukremin Kilic, and S. Lakshmivarahan, for their time and support.

Some of the computing for this project was performed at the OU Supercomputing Center for Education & Research (OSCER) at the University of Oklahoma (OU). This work was supported in part by Professor Yun Wang's DOE grant DE-SC0009956 (2011-2014) and her Carl T. Bush Professorship (2012-2015) at University of Oklahoma, and by DOE contract DE-AC02-06CH11357 at Argonne National Laboratory (summer 2016).

Table of Contents

List of Tables

List of Figures

Abstract

Since the accelerating expansion of the Universe was discovered by observing distant Type Ia supernovae (SNe Ia), SNe Ia have been playing an important role in constraining the unknown cause behind the observed cosmic acceleration, or what we refer to as dark energy.

In order to obtain robust cosmological constraints from SN Ia data, we have applied Markov Chain Monte Carlo (MCMC) to SN Ia light curve fitting. We develop a method for sampling the resultant probability density distributions (pdf) of the SN Ia light curve parameters in the MCMC likelihood analysis to constrain cosmological parameters, and validate it using simulated data sets. Applying this method to the Joint Lightcurve Analysis (JLA) data set of SNe Ia, we find that sampling the SN Ia light curve parameter pdf's leads to cosmological parameters closer to that of a flat Universe with a cosmological constant, compared to the usual practice of using only the best fit values of the SN Ia light curve parameters.

On the other front, SN classification and redshift estimation using photometric data only have become very important for the Large Synoptic Survey Telescope (LSST), given the large number of SNe that LSST will observe and the impossibility of spectroscopically following up all the SNe. We investigate the performance of a SN classifier that uses SN colors to classify LSST SNe with the Random Forest classification algorithm. Our classifier results in an AUC of 0.98 which represents excellent classification. We are able to obtain a photometric SN sample containing 99% SNe Ia by choosing a probability threshold. We estimate the photometric redshifts (photo-z) of SNe in our

sample by fitting the SN light curves using the SALT2 model with nested sampling. We obtain a mean bias ($\left\langle z_{\rm phot} - z_{\rm spec} \right\rangle$) of 0.012 with $\sigma \left(\frac{z_{\rm phot} - z_{\rm spec}}{1+z_{\rm spec}} \right) = 0.0294$ without using a host-galaxy photo-z prior, and a mean bias ($\left\langle z_{\rm phot} - z_{\rm spec} \right\rangle$) of 0.0017 with $\sigma \left(\frac{z_{\rm phot} - z_{\rm spec}}{1+z_{\rm spec}} \right) = 0.0116$ using a host-galaxy photo-z prior. Assuming a flat ΛCDM model with $\Omega_m = 0.3$, we obtain Ω_m of 0.305 ± 0.008 (statistical errors only), using the simulated LSST sample of photometric SNe Ia (with intrinsic scatter $\sigma_{\rm int} = 0.11$) derived using our methodology without using host-galaxy photo-z prior.

The methodology developed in this dissertation will be useful in the use of SN Ia data for precision cosmology and help boost the power of SNe from the LSST as cosmological probes.

Chapter 1

Introduction

1.1 Cosmology

In 1929, Edwin Hubble discovered that our Universe is expanding through observations of distant galaxies (Hubble, 1929). Hubble's law expresses the linear relation between our distance D to a galaxy and its velocity v relative to the observer:

$$v = H_0 D \tag{1.1}$$

where H_0 is the Hubble constant.

As the Universe expands, the wavelength of a photon is stretched and becomes longer than when it is emitted. We define the cosmological redshift as:

$$z = \frac{\lambda_{\text{obs}}}{\lambda_{\text{emit}}} - 1 \tag{1.2}$$

According to the cosmological principle, our Universe is homogeneous and isotropic on large scales. Our Universe can be described using the Robertson-Walker metric:

$$ds^2 = -c^2 dt^2 + a^2(t) \left[\frac{dr^2}{1 - kr^2} + r^2 d\theta^2 + r^2 \sin^2 \theta d\phi^2 \right] \tag{1.3}$$

where c is the speed of light, t is the cosmic time, $a(t)$ is the cosmic scale factor at time t which evolves as the Universe expands, and k is the curvature constant that describes our Universe as "open", "flat", and "close" when k is -1, 0, 1, respectively. (r, θ, ϕ) are the spacial coordinates in spherical coordinates.

By definition,

$$\frac{a_0}{a(t)} = \frac{\lambda_{\text{obs}}}{\lambda_{\text{emit}}} = 1 + z \tag{1.4}$$

where a_0 is the scale factor today.

We can also define a comoving distance $r(z)$ of an object at redshift z in the comoving frame, which is related to the luminosity distance D_L of that object:

$$D_L = (1 + z)r(z) \tag{1.5}$$

The luminosity distance D_L (in parsecs) is defined in the following expression that relates the apparent magnitude m and the absolute magnitude M of an object:

$$m - M = 5 \log_{10} \left(\frac{D_L}{10pc} \right) = 5 \log_{10} D_L - 5 \tag{1.6}$$

$m - M$ is often called the distance modulus μ.

We define the Hubble parameter $H(t)$ as the expansion rate of the Universe at time t:

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

The comoving distance of an object at redshift z becomes:

$$r(z) = cH_0^{-1} |\Omega_k|^{-\frac{1}{2}} \text{sinn} \left[|\Omega_k|^{\frac{1}{2}} \Gamma(z) \right] \tag{1.8}$$

where

$$\Gamma(z) = \int_0^z \frac{\mathrm{d}z'}{E(z')} \tag{1.9}$$

$$E(z) = H(z)/H_0 \tag{1.10}$$

and

$$\text{sinn}(x) = \begin{cases} \sin(x), & \Omega_k < 0 \\ x, & \Omega_k = 0 \\ \sinh(x), & \Omega_k > 0 \end{cases} \tag{1.11}$$

Given the Roberson-Walker metric and Einstein's equation, we can derive the following expression:

$$H^2(z) = H_0^2[\Omega_m(1+z)^3 + \Omega_r(1+z)^4 + \Omega_k(1+z)^2 + \Omega_{DE}(1+z)^{3(1+w)}] \quad (1.12)$$

where H_0 is the Hubble constant as in Eq. (1.1), and equals to the value of the Hubble parameter today, Ω_m, Ω_r, Ω_k, Ω_{DE} are the density fraction of the matter, radiation, curvature and dark energy component, with the density fraction defined as the density of each component, divided by the critical density (which is defined as the density of the Universe assuming the Universe is flat: $\rho_c = 3H^2/8\pi G$). The dark energy component is introduced in order to match the observational data, as described in the next section. Here we assume that dark energy has a constant equation of state, w (defined as the ratio of pressure to density). When $w = -1$, the dark energy terms becomes simply a cosmological constant. The radiation term is often negligible.

1.2 The dark energy problem

The dark energy problem refers to the unknown cause of the cosmic acceleration. The cosmic acceleration was first discovered by two teams independently via data analysis of distant Type Ia supernovae (SNe Ia). The use of SNe Ia as calibrated standard candles gives the measurement of the distance modulus. Together with the measurement of redshift through spectroscopy, a distance-redshift relation can be shown using the Hubble diagram, which plots the distance modulus vs the redshift in a 2-D diagram. The Hubble diagram can provide information about the expanding history of the Universe, and can be fitted to a cosmological model to obtain constraints on the cosmological

parameters. The measurement of the cosmological parameters and the deceleration rate (as people believed at that time that our Universe was matter dominated so that it was undergoing a deceleration process) relies on observations of high redshift SNe Ia. Efforts have been put into obtaining high quality SNe Ia data at higher redshifts.

The two teams – the High-z Supernova Search Team (Riess et al., 1998) and the Supernova Cosmology Project (Perlmutter et al., 1999), both aimed at obtaining high redshift SNe Ia and using them to measure the cosmological parameters by fitting the Hubble diagram. Both using the 4 meter telescope at Cerro Tololo Inter-American Observatory (CTIO), the High-z Supernova Search Team obtained 16 high-redshift SNe Ia with redshift from 0.16 to 0.62, and the Supernova Cosmology Project obtained light curves of 42 high redshift SNe Ia in a redshift range from 0.18 to 0.83. Combining with a sample of nearby SNe Ia, both the teams concluded that the resultant cosmological constant was non-zero and positive with 99% confidence, after carefully examined the possible systematic effects, which led to the discovery that the Universe is accelerating. Fig. 1.1 and Fig. 1.2 show the Hubble Diagrams created by the two teams. The discovery won the Nobel prize in 2011.

1.3 Type Ia supernovae as distance indicators

Type Ia supernova (SN Ia) is one type of supernova explosion that has no Hydrogen lines but strong silicon lines in its spectrum. It is commonly accepted that this type of supernova is the consequence of a white dwarf star accreting mass from its companion star and resulting in a powerful explosion after reaching the Chandrasekhar mass limit

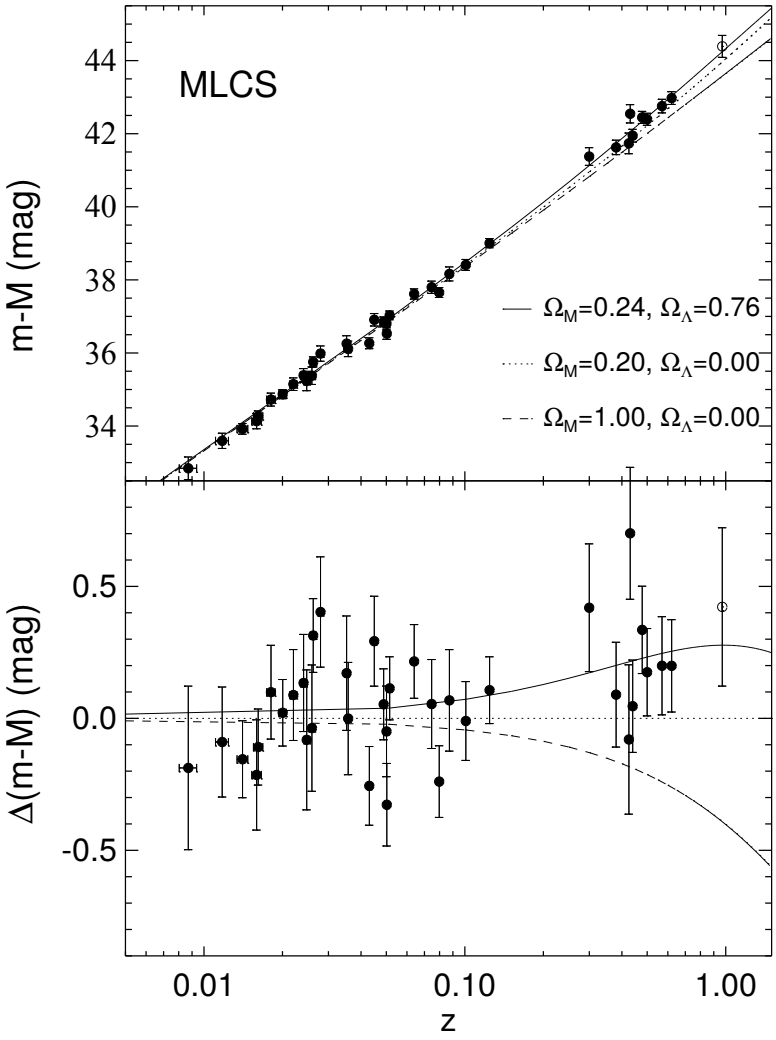

Figure 1.1: The Hubble Diagram from the High-z Supernova Search Team, Fig. 4 of Riess et al. (1998)

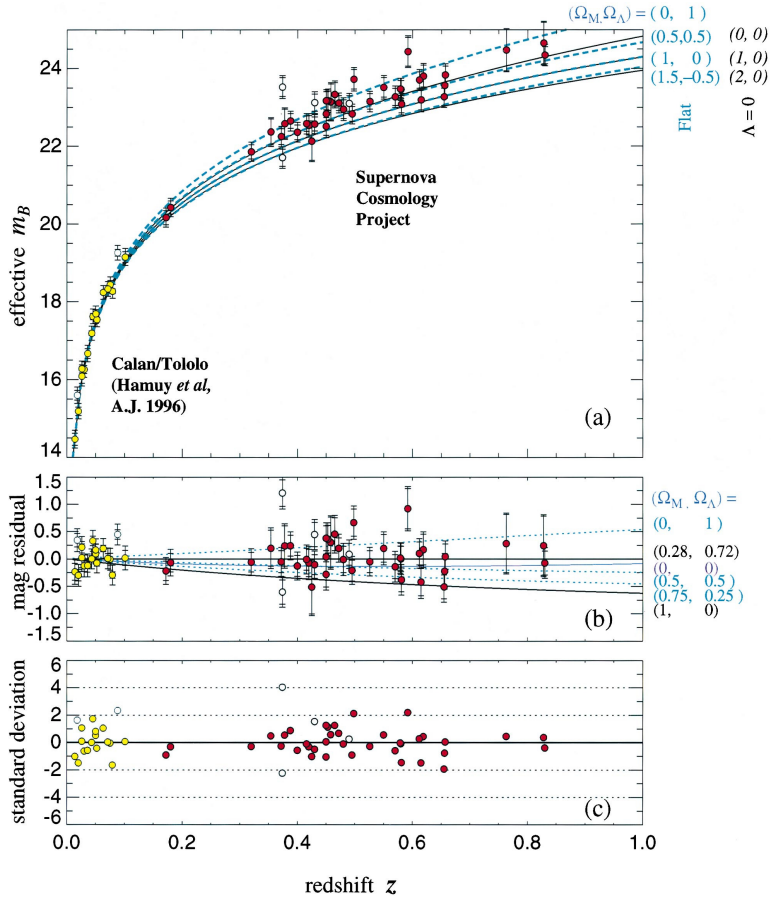

Figure 1.2: The Hubble Diagram from the Supernova Cosmology Project, Fig. 2 of Perlmutter et al. (1999).

of 1.4 $M_{\odot}$. Due to this unique physical mechanism, SN Ia explosions occur at the same mass, which produces the same luminosity, making them a well-known distance indicator.

SN Ia explosion produces tons of energy, making it shine in the sky even if it's located in a galaxy far away. The observation of supernovae can be traced back to the ancient times. As technology improved, observation of high redshift supernovae became possible and revealed a big mystery of our Universe in the late 1990s. The Nobel prize winning discovery was made by two independent groups of researchers who showed that our Universe is accelerating, by examining the distant Type Ia supernovae (Riess et al., 1998; Perlmutter et al., 1999).

The assumption that SNe Ia are "standard candles" with the same intrinsic luminosity is, however, not exactly true. Research show that SNe Ia have a scatter in their peak magnitude of ~2.5 mags. In the investigation of the possible source of the SN Ia's intrinsic scatter, Phillips (1993) found that peak magnitudes of the SNe Ia are correlated with their light curve decline rate, and that brighter SNe have broader light curves, while dimmer SNe have narrower ones (Fig. 1.3). On the other hand, Hamuy et al. (1996a) also showed that the SN peak magnitudes are also correlated with the colors of the SNe, and that brighter SNe appear bluer while fainter SNe are redder (Fig. 1.4). Riess et al. (1996) incorporated the two relations and developed a method that utilized the muticolor light curves of the SNe to standardize the SN peak magnitudes. The method leads to a significant reduction in the SN velocity dispersion with $\sigma_v = 0.12$, comparing to the uncorrected $\sigma_v = 0.52$. Tripp (1998) proposed to correct the SN peak brightness using a linear relation with the two parameters representing the SN decline

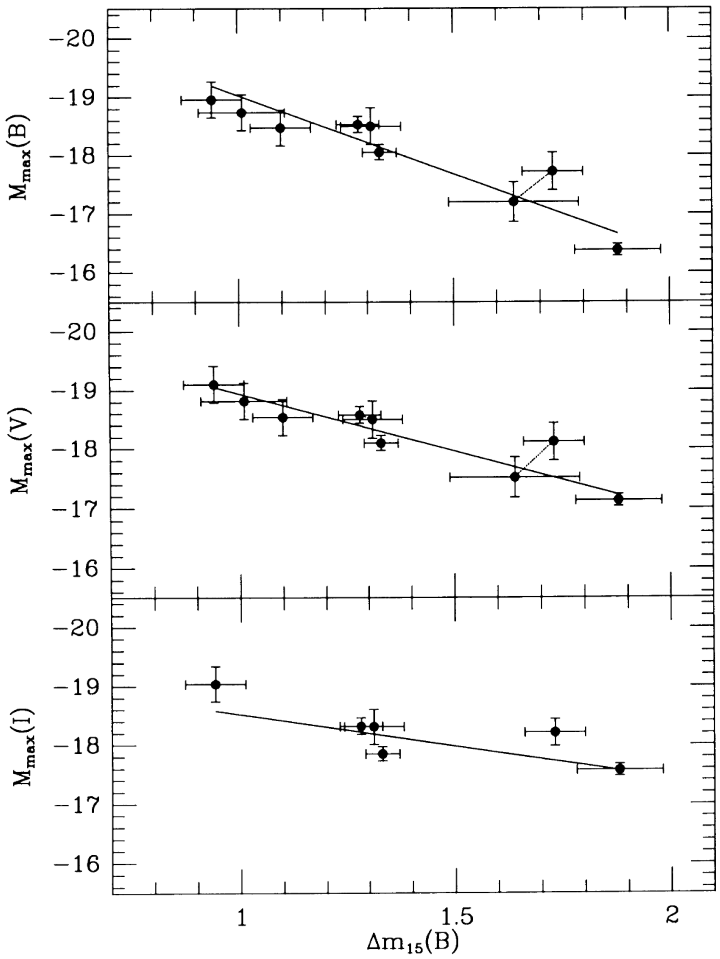

Figure 1.3: SN peak magnitude in each band vs the shape parameter Δm_{15} – the magnitude change after 15 days from the date of maximum, Fig. 1 of Phillips (1993)

rate (Δm_{15}) and color (B-V). SNe Ia are therefore "calibrated standard candles" that can be used in cosmological analysis.

1.4 SN Ia light curve models

Since the physics behind SNe Ia explosion is still not completely understood, we rely on empirical models to standardize the SN peak brightness. A lot of effort has been put

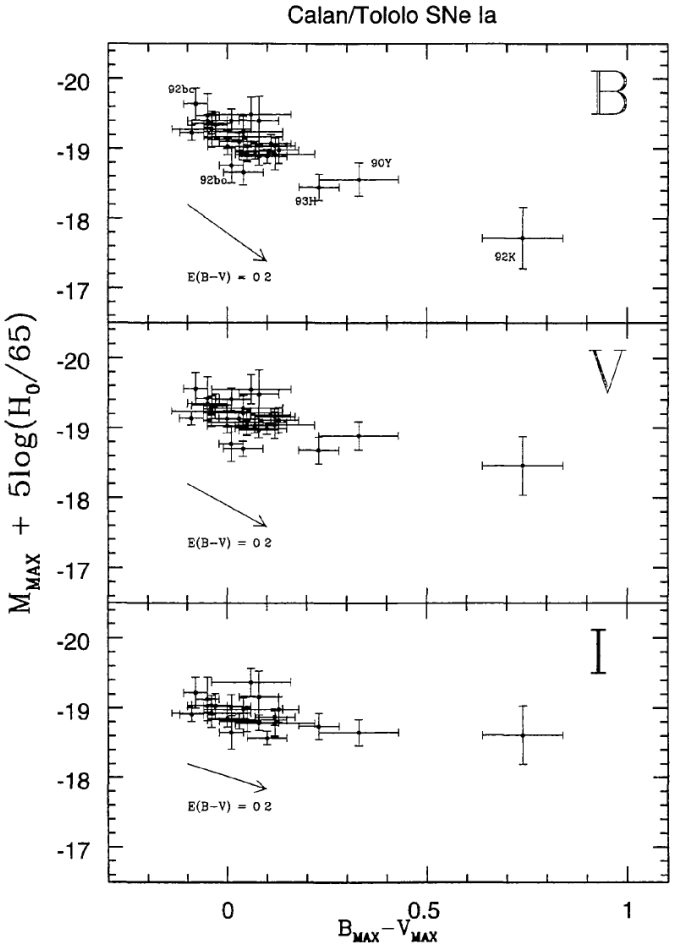

Figure 1.4: SN peak magnitude in each band vs SN color (B-V), Fig. 2 in Hamuy et al. (1996a)

in developing models for using SNe Ia as standard candles. In this section, we describe models that have been used in the analysis of SNe Ia for cosmology.

The SALT2 model A commonly used model by the SN cosmology community is the SALT2 (Spectral Adaptive Lightcurve Template 2) model (Guy et al., 2007, 2010), which is an observer frame model that provides an average spectral sequence that can be integrated over the observing bandpass to obtain flux at a given epoch. A color variation model (a.k.a. color law) is also provided. The model is generated using a training set of SNe consisting of both low-z and high-z (z up to 1) SNe Ia. The use of high-z SNe Ia in the training process provides better estimation of the model in the UV range, which is essential for measuring the distance for higher redshift SNe ($z \sim 1$).

In order to estimate the SN distances from fitting the observed SN light curves using some empirical model, the SN fluxes are translated from an observer-frame bandpass to a rest-frame bandpass, in which a K-correction is needed to account for the different shapes between the two bandpasses. In SALT2, the K-correction is naturally built into the model, as the flux is calculated through integration of the spectral sequence over an observer-frame bandpass.

The SALT2 model describes the SN light curve using 4 parameters, the date at the maximum flux t_0, the amplitude of the light curve x_0, the shape parameter x_1, and the color parameter c. The model flux is defined in the following form:

$$\frac{dF}{d\lambda}(p,\lambda) = x_0 \times [M_0(p,\lambda) + x_1 M_1(p,\lambda) + \ldots] \times \exp[c\, CL(\lambda)] \tag{1.13}$$

where $M_0(p,\lambda)$, $M_1(p,\lambda)$, and $CL(\lambda)$ are provided by the model and are derived from the training procedure. $M_0(p,\lambda)$ is the average spectral template, $M_1(p,\lambda)$ is the

first order variation of the template, and $CL(\lambda)$ is the color variation law.

An integration over a given bandpass with known transmission functions is needed to obtain the flux in that band.

We describe the model in more detail in the next chapter (Section 2.2.1) in order to implement it in our own analysis.

MLCS2k2 Another commonly used SN light curve model is the MLCS2k2 (Multicolor LightCurve Shape 2k2, Jha et al. (2007)), which is a rest frame model that describes the SN light curves using five parameters, the date at maximum light t_0, the true distance modulus μ_0, the shape parameter Δ, the host-galaxy extinction parameter R_V and A_V^0. For a given band X, the light curves are first corrected for K-correction, Galactic extinction and time dilation, and then fitted in the following form:

$$\mathbf{m}_X(t - t_0) = \mathbf{M}_X^0 + \mu_0 + \zeta_X \left(\alpha_X + \frac{\beta_X}{R_V}\right) A_V^0 + \mathbf{P}_X \Delta + \mathbf{Q}_X \Delta^2 \tag{1.14}$$

where $\mathbf{M}_X^0$ represents the absolute magnitude of the SN Ia, $\mathbf{P}_X$ and $\mathbf{Q}_X$ are parameters that scale the corrections to the SN light curve according to their shape parameter Δ. $\mathbf{M}_X^0$, $\mathbf{P}_X$ and $\mathbf{Q}_X$ are obtained using a training set of SNe Ia. ζ_X is a function of time that characterized the extinction variation with time. The extinction at any time for any band is calculated using ζ_X and fixed parameter α_X and β_X.

1.5 Supernova classification

Different types of SNe are classified according to different features in their spectra. Like SNe Ia, SNe with no hydrogen lines are classified as Type I, while those with

hydrogen lines in their spectra are classified as Type II. Within each type, the SNe are further classified into subtypes according to other spectral features. Type I SNe with silicon features are classified as Type Ia; those with no silicon feature, but strong helium lines are classified as Type Ib; and those with no silicon feature or helium lines are classified as Type Ic. Type II SNe are usually classified into subtypes according to both their spectra and light curves. There are generally four subtypes for Type II's: Type IIP, Type IIL, Type IIn and Type IIb. For Type II SNe, those with a relatively longer decline time and showing a "plateau" feature in their light curves in the declining phase are classified as Type IIP ("P" stand for "plateau"), whereas those with a relatively shorter decline time and a linear decline rate are classified as Type IIL ("L" stand for "linear"). Type IIn ("n" stand for narrow emission lines) SNe show narrow emission features in their spectra, rather than the typical absorption features. Another type that has been mentioned in the literature is Type IIb, which is considered to evolve into Type Ib, as it has the Type II features in the early time spectra, while looks like Type Ib/c in the late time spectra.

Among different types of SNe, SNe Ia stand out as they are believed to result from thermal-nuclear explosions near the Chandrasekhar mass of 1.4 $M_{\odot}$, and can be used as standard candles to constrain cosmology. All other types of SNe are considered to be consequences of the core collapse of massive stars, and are referred to as core-collapse (CC) SNe.

While there has been studies on constraining cosmology using Type II SNe, SNe Ia remain the main object of interest in cosmological analysis. The standard way of picking out the SNe Ia from all types of supernova explosions is looking at their

spectra. The spectroscopically classified SN Ia sample is considered to be pure with very low core-collapse contaminations. Accurate redshifts are also obtained through spectroscopy. So it is ideal to classify SNe and obtain accurate redshift information using spectroscopy. However, doing spectroscopy is quite expensive, especially for the high-redshift SNe, which play an important role in the cosmological analysis. Our spectroscopically confirmed SN Ia sample for cosmology is already limited by telescope resources. So it becomes important to make use of the large photometric sample of SNe and develop methods for photometric classification and analysis methods for cosmology using photometric data.

1.6 SN surveys

Since the discovery of dark energy by the High-z Supernova Search Team and the Supernova Cosmology Project, many SN surveys have been carried out in order to obtain SN samples in a variety of ranges of redshifts with high data quality. Here we summarize several major SN surveys, completed, ongoing and planned, which play important roles in SN cosmology.

ESSENCE The ESSENCE (Equation of State: SupErNovae trace Cosmic Expansion) supernova survey (Miknaitis et al., 2007; Wood-Vasey et al., 2007) was conducted from 2002 to 2008 using the 4-meter Blanco Telescope at CTIO, aiming at measuring the dark energy equation of state w with 10% precision. The survey obtained SN light curves in R and I bands, with a redshift range of 0.1 to 0.8. Wood-Vasey et al. (2007) presented the analysis using the data set obtained in the first 4 years of the project,

which contains 60 SNe Ia. Assuming a constant w and a flat Universe, the data results in a measurement of $w = 1.05^{+0.13}_{-0.12}(stat) \pm 0.13(sys)$. Narayan et al. (2016) presented the final SN Ia sample with 213 spectroscopically identified SNe Ia. The data is recalibrated by Narayan et al. (2016) and the estimated systematic uncertainty from photometry is $\sim 1\%$.

SNLS The SNLS (SuperNova Legacy Survey) survey (Astier et al., 2006; Guy et al., 2010) is part of the Canada-France-Hawaii Telescope (CFHT) Legacy Survey using the MegaPrime/MegaCam instrument at CFHT. It is a 5-year survey starting in 2003 that images the sky every 3 or 4 nights in four one square degree fields. The images are taken in 4 bands (griz) to obtain multi-color light curves with high data quality. The survey is designed to improve the discovery and follow-up strategy, and to reduce the systematics from photometry. In the first 3 year of the survey, 252 SNe Ia ($0.15 < z < 1.1$) were obtained and analysed to constrain cosmology (Guy et al., 2010). Using SN data alone, Guy et al. (2010) measured $\Omega_m = 0.211 \pm 0.034(stat) \pm 0.069(sys)$ by fitting a flat ΛCDM model, and claimed that the largest systematic uncertainty came from the photometric calibration of the data. The survey have utilized two independent pipelines in identifying SN Ia candidates. The real-time pipeline searches for SN Ia candidate and follows up with spectroscopy. The deferred pipeline uses SN photometry to identify possible SN Ia candidates. Möller et al. (2016) presents the SNLS 3-year results from the deferred pipeline, which obtained a sample of 486 photometrically identified SNe Ia, twice the size of the sample from the real-time pipeline.

SDSS The SDSS-II (Sloan Digital Sky Survey-II) SN Survey (Frieman et al., 2008) is an intermediate redshift SN survey using the SDSS camera at Apache Point Observatory (APO). The survey was conducted from 2005 to 2007. It scans part of the sky in the SDSS Stripe 82 in an average of every 4 nights using the 5 SDSS filters – ugriz. The total survey area is 300 square degrees. The survey aims at addressing (i) the lack of intermediate redshift sample from previous surveys, (ii) reducing the systematic uncertainties. The survey can provide a uniform sample of SNe imaged from the same telescope, which can minimize the systematic uncertainty arising from cross-calibration of different surveys. With the redshift ranging from 0.05 to 0.35, the survey bridges the redshift dessert of SNe for cosmology. Sako et al. (2014) presented the data from the full 3-year survey, including 1443 spectroscopically identified SNe Ia and also 677 possible photometrically identified SNe Ia.

Pan-STARRS The PAN-STARRS1 (PS1) medium deep field survey (Kaiser et al., 2010) is designed for searching transient events including SNe. The survey images ten 7-square-degree fields using the PAN-STARRS system in Hawaii. The images are taken in 5 bands – grizy, with an average cadence of 6 images every 10 days. The cosmological constraints derived from the data of the first 1.5 years of the survey is presented in Rest et al. (2014). Using 113 spectroscopically confirmed SNe Ia ($0.03 < z < 0.65$), Rest et al. (2014) measured $\Omega_m = 0.226^{+0.057}_{-0.061}$ by assuming $w = -1$ and a flat Universe. Jones et al. (2016) presented the cosmological constraint using the full PS1 survey data including 143 spectroscopically confirmed SNe Ia and 969 photometrically identified SNe Ia. By utilizing the BEAMS framework (Kunz et al., 2007; Hlozek et al., 2012)

to take into account probabilities of SNe Ia from classification, they found that the cosmological constraints were consistent with Rest et al. (2014). They also claimed that the cosmological analysis was very sensitive to the core-collapse contaminations.

DES The Dark Energy Survey[1] (DES, The Dark Energy Survey Collaboration (2005)) is a 5-year survey starting in 2013, aiming at probing dark energy with different probes, including SNe. The survey camera – DECam, is installed on the 4-meter telescope at CTIO, which images the sky in 4 different bands – griz. The DES-SN survey (Bernstein et al., 2012) includes 10 SN fields covering a total of 30 square degrees of the sky. The survey is expected to discover ~4000 SNe Ia in a redshift range of 0.05 to 1.2, and plans to obtain follow-up spectra for 20% of the SNe discovered. Host galaxy spectra are planned to be obtained for the majority of the rest of the sample in order to obtain host-galaxy spectroscopic redshift. A homogeneous sample with a large number of SNe Ia are expected to reduce systematic effects. And the survey is expected to achieve a significant improvement in the color measurement for the high redshift SNe, benefiting from an increased red efficiency of the camera.

LSST The Large Synoptic Survey Telescope[2] (LSST, LSST Science Collaboration et al. (2009)) is a next-generation wide-field survey telescope which is under construction in Chile, with first light planned for 2019. The LSST has many science goals including probing dark energy. The survey is designed to scan a large area of sky (20,000 square degrees) in a deep, wide, fast mode. Several deep drilling field are

[1] https://www.darkenergysurvey.org/
[2] www.lsst.org

planned to be observed in greater depth. The survey will take images of the sky in 6 bands – ugrizy. While the LSST observing strategy is still under determination, we expect to observe hundreds of thousands of SNe Ia in 10 years of the survey operation; the challenge is to utilize those photometric SNe in a new era of SN cosmology.

WFIRST The Wide Field Infrared Survey Telescope (WFIRST, Spergel et al. (2015)) is a NASA space mission planned for launch in the mid-2020's; one of its primary goals to constrain dark energy through multiple probes. The WFIRST SN Survey is proposed to be a 2-year survey of a single field. The telescope will discover SNe Ia and obtain SN Ia light curves in 6 near-infrared filters, with a 5-day cadence. The survey strategy is to observe SNe in three tiers – a shallow field with $z < 0.4$, an intermediate field with redshift from 0.4 to 0.8, and a deep field with redshift up to 1.7. A uniform infrared SN Ia sample observed from space has the ability to tighten the dark energy constraint and reduce the systematic effects.

Euclid Another relevant space mission is the ESA led Euclid (Laureijs et al., 2011), a dedicated dark energy space mission, scheduled for launch in 2020. The Euclid mission could also contribute significantly to the high-z SN sample with redshift ~1.5.

In addition to the high-z ($z > 0.1$) samples mentioned above, a number of low-z ($z < 0.1$) samples also play important roles in generating better light curve models, understanding the SN magnitude dispersions, and anchoring the Hubble Diagram. These low-z surveys includes the Calan-Tololo sample (Hamuy et al., 1996b), the CfA sample (Riess et al., 1999; Jha et al., 2006), the CfA3 sample (Hicken et al., 2009), the Carnegie

Supernova Survey (CSP, Folatelli et al. 2010; Contreras et al. 2010; Stritzinger et al. 2011), the Lick Observatory Supernova Search (LOSS, Ganeshalingam et al. 2013), etc. The highest redshift sample today consists of SNe Ia observed by several groups using the Hubble Space Telescope (Riess et al., 2007; Suzuki et al., 2012).

1.7 Tools for SN cosmology

In this dissertation, we have utilized several softwares that have been developed for the needs of simulations and analyses of SNe for cosmology. We provide information for two very useful softwares for SN cosmology in this section.

SNANA SNANA[3] (Kessler et al., 2009) is designed to meet the needs for using SNe Ia in cosmological analysis. It can be used to simulate SN light curves, fit the SN light curve using different light curve models, and fit for cosmological parameters using the derived SN light curve parameters. Rather than simulating the image for a given survey, SNANA simulates SN light curves basing on actual survey conditions, such as cadence, sky noise, CCD noise, etc. Those survey conditions are written in a simulation library (SIMLIB), and are generated by someone who has the information of the telescope and the survey itself. A SIMLIB for future surveys (such as LSST) are generated basing on current proposed survey strategy and information from previous surveys at a similar site. Using SNANA, it is possible to generate many light curves efficiently and accurately, which is useful in characterizing the systematic effects and optimizing future surveys. The software can simulate both SNe Ia and Core-collapse

[3] http://snana.uchicago.edu/

SNe, using an SN Ia model and different CC templates. The number of SNe generated can be determined by providing the SN rate measured previously as an input.

We have utilized the software to generate realistic simulations for LSST in Chapter 3. The software can also perform light curve fitting, SN typing, and cosmological fit in the same framework.

SNCosmo SNCosmo[4] (Barbary et al., 2016) is a python package that does similar jobs as SNANA, including simulations and light curve fitting. Instead of providing several executives for the users to run, it provides a library of functions and classes so that users can write their own programs by utilizing the APIs and modify their own codes as needed. It contains the popular SN models and can perform the light curve fitting using either a maximum likelihood method or a Bayesian inference method such as MCMC and nested sampling. We have used SNCosmo to fit for light curve parameters with the nested sampling method in Chapter 3.

1.8 Summary

SN cosmology remains an active and exciting field and is facing more challenges as the number and quality of SNe increases dramatically in the next decades. In this dissertation, we focus on two aspects in using SNe Ia as cosmological probes.

In Chapter 2, we focus on developing methods for the analysis of the spectroscopic SN Ia sample using robust statistical methods. We apply the Markov Chain Monte Carlo (MCMC) method in the SN light curve fitting to obtain well-sampled probability

[4]http://sncosmo.readthedocs.org

distributions (pdf's) of the SN light curve parameters and develop a method to sample the resultant SN light curve parameter pdf's to obtain cosmological constraints.

In Chapter 3, we focus on constraining cosmology using the photometric SN sample from LSST. We simulate realistic SN light curves using SNANA, and develop methods for photometric classification and estimation of the photometric redshifts (photo-z's) of SNe. We show the result of our cosmological analysis using our photometric sample with photo-z's.

In Chapter 4, we summarize our results and discuss directions for future work.

Chapter 2

Analysis of current spectroscopic SN Ia sample

2.1 Introduction

As described in Section 1.4, we can expect a dramatic increase in the quantity and quality of SN Ia data in the next decade and beyond. For $z < 1$, thousands of SNe Ia are expected from DES, and hundreds of thousands of SNe Ia are expected from LSST. WFIRST will observe thousands of SNe Ia at $z > 1$. In order to use the SN Ia data in precision cosmology, it is important that we develop robust analysis techniques that can be applied to the observed SN Ia light curves.

In this work, we apply Markov Chain Monte Carlo (MCMC) to SN Ia light curve fitting. We develop a method for sampling the resultant probability density distributions (pdf) of the SN Ia light curve parameters in the MCMC likelihood analysis to constrain cosmological parameters. Both of these are new approaches in SN Ia cosmology. We present our methodology in Section 2.2, and results in Section 2.3. We conclude with a summary and discussion in Section 2.4.

2.2 Methodology

As the mechanism of SNe Ia explosion is still unclear, empirical models are used for fitting SN Ia light curves, see Section 1.3 and Section 1.4. The light curve of a SN Ia can be characterized by its shape and color, which can be corrected to reduce the intrinsic dispersion in SN Ia peak magnitudes. The shape correction utilizes the correlation

between SN Ia peak brightness and decline time (brighter SNe Ia decline more slowly in brightness) (Pskovskii, 1977; Branch, 1981; Phillips, 1993; Phillips et al., 1999). The color correction models the variation of SN Ia colors and dust extinction.

Following Betoule et al. (2014), we use the SALT2 model (first proposed in Guy et al. (2007), and updated in Guy et al. (2010) and Betoule et al. (2014)) for SN Ia light curve fitting, as described in Section 1.4. SALT2 provides an average spectral sequence and color dispersion from the training of a subset of SN Ia light curves. For SALT2, The four SN Ia light curve parameters are: date of maximum light in the rest-frame B band, the amplitude of the spectral sequence (which is described conventionally by the peak magnitude in the rest-frame B band), light curve shape, and color. The date of maximum light is a nuisance parameter, since the distance-indicator properties of a SN Ia only depend on its peak magnitude, light curve shape, and color. The default fitting procedure is minimizing a χ^2 to find the best fit light curve parameters from a 4D grid of parameter values. This grid method has its limitations: it may result in a local minimum, and the error estimate is sensitive to the choice of the grid size and the spacing of the grid points.

In this work, we use Markov Chain Monte Carlo (MCMC) method to fit for light curve parameters, using the SALT2 model in Betoule et al. (2014). The MCMC method has the following advantages over a grid-based method:

1. The posterior is drawn randomly from the proposed distribution; the true probability distribution is recovered given enough number of realizations.

2. The statistics can be calculated using multiple chains that have converged, leading

to robust error estimates.

3. The resultant probability density functions (pdf) are smooth and can be utilized in the subsequent cosmological analysis.

2.2.1 The SALT2 model

The SALT2 model (Guy et al., 2007, 2010) is an empirical light curve model that provides an average spectral sequence and its first order variation. The model flux is defined as:

$$\frac{dF}{d\lambda}(p,\lambda) = x_0 \times [M_0(p,\lambda) + x_1 M_1(p,\lambda) + \ldots] \times \exp[c\, CL(\lambda)] \tag{2.1}$$

Where p is the phase from the maximum light in the rest frame, λ is the rest-frame wavelength, M_0 and M_1 are the average spectral sequence and its first order variation, CL is the color law which is independent of phase. The light curve parameters x_0, x_1, and c are determined by light curve fitting. To find the best fit parameters, the default method finds the minimum χ^2 using a grid-based method. The χ^2 is expressed as:

$$\chi^2 = (F_{model} - F_{obs})\, C_{SN}^{-1}\, (F_{model} - F_{obs}), \tag{2.2}$$

where F_{obs} is the observed flux after calibration, F_{model} is the model flux integrated over the observing filter $T\,(\lambda(1+z))$:

$$F_{model} = (1+z) \int \lambda \frac{dF}{d\lambda}(p,\lambda) T\,(\lambda(1+z))\, d\lambda. \tag{2.3}$$

The covariance matrix C_{SN} consists of three parts: a diagonal term of the model errors, a regular matrix (non-diagonal) term of color dispersions (including K-correction errors),

an error matrix (for SNLS) or a diagonal term of the observational flux errors (for SN samples other than SNLS):

$$C_{SN} = D_{model} + C_{model} + C_{obs}. \tag{2.4}$$

The model parts of the covariance matrix are dependent on the light curve parameters; they are varied to minimize the χ^2 for a given set of light curve parameters.

We will now discuss in detail the uncertainties in the model fluxes (D_{model} term in Eq. (2.4)). The model fluxes are calculated using Eq. (2.1) and Eq. (2.3). The model uncertainties are calculated as

$$\sigma_{model} = (f_0/f_{total}) \times S \times (V_0 + x_1^2 V_1 + 2x_1 V_{01})^{1/2} \times F_{model}, \tag{2.5}$$

where S, V_0, V_1 and V_{01} are provided as part of the SALT2 model, and are dependent on phase and wavelength. Basically, V_0 is the variance in M_0, V_1 is the variance in M_1, V_{01} is the covariance between M_0 and M_1. S is a scale factor. f_0 is the main component of the model flux, which is M_0 integrated over the observing filter; f_{total} is the total model flux including the M_1 component (not including the color term). When $f_{total} \leq 0$, the model uncertainty is set to be 100 times that of the model flux. Note that in the SALT2 code, $(V_0 + x_1^2 V_1 + 2x_1 V_{01})$ is set to 0.0001 when it becomes less then zero.

2.2.2 MCMC light curve fitting

We use CosmoMC as a generic sampler to generate multiple chains using the Metropolis-Hastings algorithm. We assume convergence when $R - 1 < 0.01$ using the Gelman and Rubin "R-1" statistic (Brooks & Gelman, 1998). For a detailed discussion on CosmoMC, see Lewis & Bridle (2002).

To implement SALT2 light curve fitting using MCMC, we first need to understand how the SALT2 default grid-based method works. In particular, how the covariance matrix from Eq.(2.4) is handled, since the contribution from model uncertainties depends on the values of the light curve parameters (which are being fitted). There is an optional "update weights" feature in the public SALT2 code, but it seems not to be used by either Conley et al. (2011) or Betoule et al. (2014); we are only able to reproduce their results without using this option (see Section 2.3). However, by not updating the weights for the covariance matrix in the SALT2 code, it does *not* mean keeping the covariance matrix fixed. It means deriving a converged covariance matrix by doing the following:

1. Initial fit with x_1 fixed to be 0, without including model uncertainty.

2. Second fit with x_1 allowed to vary, without including model uncertainty.

3. Iterations of fits with model uncertainty included, until the changes in all parameters are less than 0.1 times the errors in the parameters.

At the beginning of each iteration in step 3, the covariance matrix is recalculated using the parameters from the previous step or iteration. And the covariance matrix is kept fixed during this iteration of the fit. In MCMC we cannot do the fit by steps as in the grid-based method, so we calculate the covariance matrix using fiducial light curve parameter values and keep it unchanged during the MCMC lighcurve fitting. We assume that likelihood $\propto \exp(-\chi^2/2)$, with χ^2 defined in Eq.(2.2), and carry out the following steps:

1. Perform a grid-based fit to obtain a set of fiducial values of light curve parameters (which are the best fit values from this fit).

2. The covariance matrix C_{SN} is calculated using the fiducial values of the model and light curve parameters.

3. An MCMC likelihood analysis is performed to obtain the light curve parameters, while fixing the covariance matrix to be that calculated using the fiducial light curve parameters.

A compelling reason for us to fix the covariance matrix for MCMC light curve fitting to be that calculated at fiducial values of the light curve parameters is as follows. The model uncertainty from Eq. (2.5) is dependent on the light curve parameters (especially x_1 and time of maximum flux), so it is varied as the light curve parameters are vared. For example, if x_1 is a very large number, it is possible that f_{total} will become negative, which means that the model uncertainty is set to 100 times that of the model value, leading to a very small χ^2 — a numerical artifact that can bias the light curve fitting. We have found that simply updating the covariance matrix as the light curve parameters vary during the MCMC steps can lead to unreasonable results for the light curve parameters, since the model errors can be much larger when the light curve parameters fall out of the reasonable ranges.

To assess the implications of our choice of fixing the covariance matrix for MCMC light curve fitting, we have compared the light curve parameter results using two different fiducial values — one set contains the best fit results from a grid-based fit, the other set the best fit results from a grid-based fit without model errors. The light

curve parameters we get from MCMC using the two different covariance matrices are very similar to each other (with a mean difference $\sim 10^{-4}$ for all three parameters used in a cosmological fit, i.e. m_B, x_1, and c), and lead to almost identical cosmological constraints. This is not surprising, since the cosmological constraints are not very sensitive to the light curve parameters, as long as no erroneous lighthcurve parameter values are used. We conclude that it is reasonable to fix the covariance matrix at fiducial values of the light curve parameters in MCMC light curve fitting.

2.2.3 Cosmological analysis

Having derived SN Ia light curve parameters using MCMC, we can use them to derive cosmological constraints. We follow the definition for the model magnitude in Conley et al. (2011):

$$m_{\rm mod} = 5 \log_{10} \mathcal{D}_L - \alpha\, x_1 + \beta\, c + \mathcal{M}, \tag{2.6}$$

where $\mathcal{D}_L$ is a scaled luminosity distance that is independent of the Hubble constant, α, β and $\mathcal{M}$ are nuisance parameters which describe the shape and color corrections of the light curve, and the SN absolute magnitude in combination with the Hubble constant. In order to model the dependence of SN Ia intrinsic brightness on the host galaxy mass, $\mathcal{M}$ is defined as a function of the host galaxy stellar mass M_{host} (in units of solar masses):

$$\mathcal{M} = \begin{cases} \mathcal{M}_1 & \text{for } \log_{10} M_{host} < 10 \\ \mathcal{M}_2 & \text{for } \log_{10} M_{host} > 10 \end{cases} \tag{2.7}$$

The χ^2 is then

$$\chi^2 = \Delta \mathbf{m}^T \cdot \mathbf{C}^{-1} \cdot \Delta \mathbf{m}, \tag{2.8}$$

where $\Delta\mathbf{m} = \mathbf{m}_B - \mathbf{m}_{\rm mod}$, and m_B is calculated by using $m_B = -2.5\log_{10}(x_0) + 10.635$ (Mosher et al., 2014). $\mathbf{C}$ is the covariance matrix. We use the same covariance matrix in our cosmological analysis as that used by Betoule et al. (2014). Note that $m_{\rm mod}$ is defined in Eq. (2.6).

In this work we assume a flat Universe with constant dark energy equation of state w, since SN Ia data alone do not provide meaningful constraints on additional cosmological parameters. The Hubble constant free luminosity distance $\mathcal{D}_L$ is defined as:

$$\mathcal{D}_L \equiv c^{-1} H_0 (1 + z_{\rm hel})\, r(z), \tag{2.9}$$

where $z_{\rm hel}$ is the heliocentric redshift, z is the CMB-frame redshift (i.e., the cosmological redshift), $r(z)$ is the comoving distance:

$$r(z) = c H_0^{-1} \Gamma(z) \tag{2.10}$$

$$\Gamma(z) = \int_0^z \frac{\mathrm{d}z'}{E(z')} \tag{2.11}$$

$$E(z) = H(z)/H_0 \tag{2.12}$$

With the assumption of flat Universe and constant w,

$$H(z) \equiv \frac{\dot{a}}{a} = H_0 \sqrt{\Omega_m (1+z)^3 + \Omega_{DE} (1+z)^{3(1+w)}} \tag{2.13}$$

and $\Omega_m + \Omega_{DE} = 1$ (the radiation contribution is negligible here).

2.2.4 pdf sampling

The MCMC analysis gives the marginalized pdf's of the light curve parameters. Those pdf's contain the distribution and error information of the light curve parameters and

can be used in a cosmological analysis. We have developed a method to sample the pdf's, and derive cosmological results by combining results from different sets of light curve parameters drawn from the pdf's, as described below.

For each SN, we choose N points with equal probability intervals from the pdf of each light curve parameter, x_0, x_1 and c, with probabilities equal to P_1, P_2, ... , P_N. This gives N^3 sets of light curve parameters, with each set having the same SNe with different values of light curve parameters. We then use these sets of light curve parameters to fit cosmology, resulting in N^3 sets of cosmological parameters. We combine the results in the following way to get the combined cosmological parameters:

$$\mathbf{s} = \frac{\sum_{i,j,k} P_i P_j P_k \, \mathbf{s}_{ijk}}{\sum_{i,j,k} P_i P_j P_k}, \tag{2.14}$$

where $\mathbf{s}$ are the cosmological parameters from sampling the pdf's of the light curve parameters, $\mathbf{s}_{ijk}$ are the cosmological parameters derived from the data set with a given set of light curve parameters drawn from the pdf's of the light curve parameters. The cosmological parameters from different data sets are weighted by the product of the relative probabilities of the three light curve parameters, P_i, P_j, and P_k. Using only light curve parameters that correspond to the peaks of the pdf's gives $P_i = P_j = P_k = 1$; this is similar to the usual practice of using only the best fit light curve parameters in the grid-based method for SALT2 (which is equivalent to using the peaks of the mean likelihood distributions from MCMC).

It is not practical to densely sample the pdf's of the light curve parameters. In order to gauge how the cosmological parameter constraints depend on the sampling density of the light curve parameter pdf's, we study the following cases:

1. **N=3**. We choose three points from the pdf, with the probabilities $P_1 = 1$ and $P_2 = P_3 = 1/2$. This results in 3 points on the pdf: the peak, and the half height point on either side of the peak.

2. **N=7**. We choose seven points on the pdf, $P_1 = 1$ (the peak), $P_2 = P_3 = 3/4$, $P_4 = P_5 = 1/2$, and $P_6 = P_7 = 1/4$. This divides the pdf in 1/4 segments in height, resulting in three points on either side of the peak.

3. **N=15**. In this case 15 points are chosen from the pdf, the probabilities are 1, 7/8, 3/4, 5/8, 1/2, 3/8, 1/4, 1/8.

4. **N=19**. When dividing the pdf into 10 equal probability intervals with probabilities $P_i = i/10 (i = 1, 2, ..., 10)$, we get 19 points on each pdf.

We will show that pdf sampling converges with increased N in Section 2.3.4.

2.3 Results

We have applied our methodology for SN Ia light curve fitting using MCMC and constraining cosmology with sampling the pdf's of SN Ia light curve parameters to the "Joint light curve Analysis" (JLA) data set of SNe Ia from Betoule et al. (2014), which combines the SNLS and SDSS data of SNe Ia in a consistent, well-calibrated manner. We do our MCMC light curve fitting using the calibrated photometric data provided by Betoule et al. (2014). The JLA sample, as an extension to the C11 Compilation (Conley et al., 2011), contains a combination of data sets of 740 spectroscopically confirmed SNe Ia from several low-z samples ($z < 0.1$) (mostly Hicken et al., 2009; Contreras

et al., 2010; Hamuy et al., 1996b; Jha et al., 2006; Riess et al., 1999), the full three-year SDSS-II supernova survey ($0.05 < z < 0.4$) (Sako et al., 2014), the first three years data of the SNLS survey ($0.2 < z < 1$) (Guy et al., 2010; Conley et al., 2011) and a couple of high redshift HST SNe ($0.7 < z < 1.4$) (Riess et al., 2007). The photometry of SDSS and SNLS is recalibrated in this data combination, and the SALT2 model is retrained using the joint data set.

2.3.1 Definitions

We will show comparative results on cosmological constraints using different sets of light curve parameters we have obtained in different approaches. These are:

1. SALT2 : This is the published light curve parameter set from Betoule et al. (2014), we use it directly in the cosmological analysis as the base of comparison to other sets.

2. GRID-SALT2 : We obtain this set of light curve parameters by running the published version of the SALT2 code[1], adding in the bias correction term to the peak magnitude and the uncertainties in redshift, lensing and intrinsic dispersion to the magnitude uncertainties. All the values of the terms above are the same as those used in the JLA set as described in Betoule et al. (2014). This is supposed to reproduce the results of SALT2. All other sets of light curve parameters described in the following are processed as described here.

3. GRID : We use our own grid-based code using the SALT2 model to calculate

[1] http://supernovae.in2p3.fr/salt/doku.php

the χ^2; this is an important cross-check, to ensure that we understand all the nuances of the public SALT2 code and its output. We will use this as the grid method to compare with our MCMC analysis. We obtain this GRID set of light curve parameters by minimizing the χ^2 using exactly the same approach as in the SALT2 code (i.e. using the function minimization package called MINUIT[2]). We expect to get the same values for the light curve parameters within numerical errors.

4. MCMC-LIKE : The MCMC chains can be used to calculate mean likelihood distributions; its maximum corresponds to the least χ^2 value from the grid method. If the grid method ever falls into a local minimum, the maximum likelihood value of the MCMC chains would instead give the correct global minimum upon convergence.

5. MCMC-MARGE : The MCMC chains can be used to obtain marginalized one-dimensional distributions of the fitted parameters; these give the standard error distribution information from an MCMC analysis. For this MCMC-MARGE set we use the means of the marginalized pdf's as light curve parameters for fitting cosmology (with no pdf sampling). We will discuss the pdf sampling results in Section 2.3.4.

In an MCMC analysis, the differences of the marginalized and the mean likelihood distributions indicate non-Gaussianity, although it is possible to have a non-Gaussian distribution where both curves are the same (Lewis & Bridle, 2002). In general, the

[2]http://seal.web.cern.ch/seal/snapshot/work-packages/mathlibs/minuit

marginalized pdf differs more from the mean likelihood for parameters that are less well constrained by the data. Mean likelihood shows how good a fit you could expect if you drew a random sample from the marginalized distribution. It is customary to quote marginalized constraints in a cosmological analysis. We will follow this practice regarding the light curve parameters, except when we need to make a direct comparison with the grid-based method (which gives results that are equivalent to the mean likelihood). However, when showing the cosmological constraints, we give the mean likelihood pdf's instead, as our tests with simulated data sets show that the peaks of the mean likelihood pdf's of the cosmological parameters are less biased than the marginalized means. (For more details, see section 2.3.4.)

2.3.2 Reproducing the SALT2 results

We first compare results of the three grid sets – SALT2, GRID-SALT2 and GRID, to verify that our grid-based code is correct. In this cross-check exercise, we found that we are only able to reproduce the Betoule et al. (2014) results without using the "update weights" option in the public SALT2 code, but instead carrying out the steps as described in Section 2.2.2.

The light curve parameters from the three sets are very similar. Fig. 2.1 shows a comparison of the cosmological constraints from SALT2, GRID-SALT2, and GRID. As expected, all three approaches give nearly the same constraints on the cosmological and SN nuisance parameters.

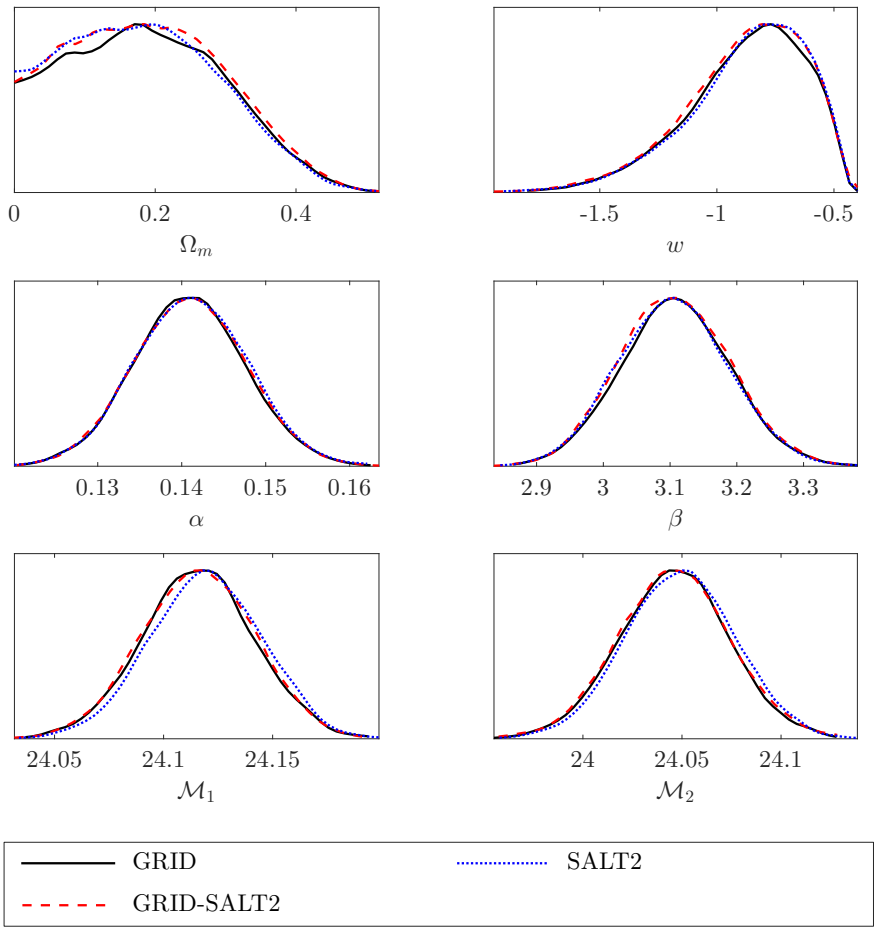

Figure 2.1: Mean likelihood pdf's of the parameters for the cosmological fit for the grid sets, using all 740 SNe from the JLA data set. Black solid lines represent the GRID set; red dashed lines represent the GRID-SALT2 set; blue dotted lines are for the SALT2 set.

2.3.3 MCMC vs GRID comparison

We now compare the results of MCMC and GRID. We find that the light curve parameters from the MCMC-LIKE set are also consistent with that from the GRID set. However, the light curve parameters of the MCMC-MARGE set are offset from the MCMC-LIKE (or the GRID) set for some SNe. The cosmological results are shown in Fig. 2.2. The MCMC-LIKE set gives similar cosmological constraint as the GRID set, as expected. The MCMC-MARGE set shows pdf's shifted from the MCMC-LIKE (or the GRID) set in Ω_m and w; this is not surprising given the difference between the marginalized and the mean likelihood in an MCMC analysis (see the discussion near the beginning of Section 2.3). The nuisance parameters are well constrained and have similar constraints in all sets.

The peaks of the mean likelihood pdf's and their 68% confident intervals of the parameters are listed in Table 2.1. We find the approximate 68% confidence levels by finding the parameter values where the probability has dropped by a factor of $e^{-1/2}$. Note that for some sets there is only one upper limit in Ω_m, because the pdf's are truncated at $\Omega_m = 0$ before the probability has dropped by a factor of $e^{-1/2}$. The grid methods (SALT2, GRID-SALT2, GRID) show similar values and errors. The MCMC-LIKE has consistent values with the grid methods but has slightly higher errors in Ω_m and w. The MCMC-MARGE set show even higher errors in w, but Ω_m and w are closer to the "concordance model" of $\Omega_m = 0.27$ and $w = -1$.

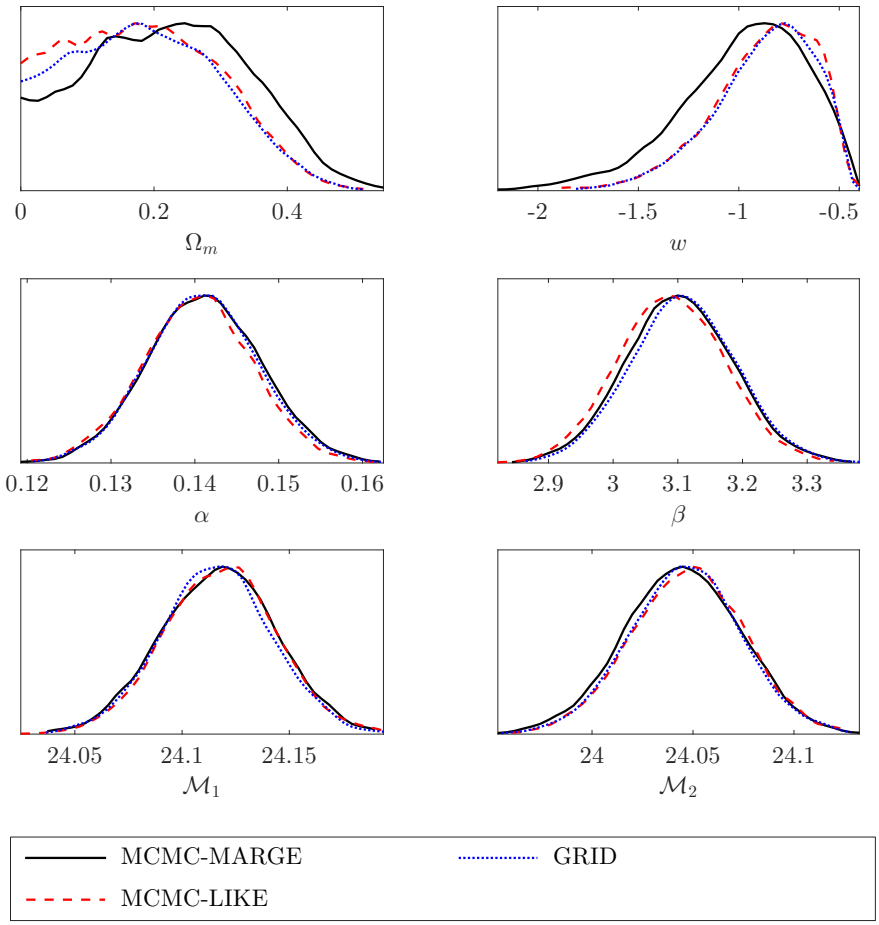

Figure 2.2: Mean likelihood pdf's of the parameters from the cosmological fit for the MCMC sets comparing with the GRID set, using all 740 SNe from the JLA data set. Black solid lines represent the MCMC-MARGE set; red dashed lines represent the MCMC-LIKE set; blue dotted lines represent the GRID set.

Table 2.1: Parameters from the cosmology fit using all 740 JLA SNe

	Ω_m	w	α	β	$\mathcal{M}_1$	$\mathcal{M}_2$
SALT2	$0.196^{+0.116}$	$-0.795^{+0.261}_{-0.236}$	$0.141^{+0.008}_{-0.007}$	$3.107^{+0.078}_{-0.088}$	$24.118^{+0.027}_{-0.024}$	$24.050^{+0.026}_{-0.029}$
GRID-SALT2	$0.185^{+0.139}$	$-0.806^{+0.268}_{-0.267}$	$0.141^{+0.006}_{-0.007}$	$3.109^{+0.086}_{-0.090}$	$24.118^{+0.026}_{-0.029}$	$24.045^{+0.029}_{-0.028}$
GRID	$0.171^{+0.145}$	$-0.769^{+0.229}_{-0.282}$	$0.141^{+0.006}_{-0.007}$	$3.105^{+0.085}_{-0.076}$	$24.117^{+0.024}_{-0.027}$	$24.044^{+0.030}_{-0.025}$
MCMC-LIKE	$0.171^{+0.157}$	$-0.788^{+0.257}_{-0.282}$	$0.141^{+0.006}_{-0.007}$	$3.088^{+0.087}_{-0.082}$	$24.123^{+0.021}_{-0.032}$	$24.050^{+0.027}_{-0.030}$
MCMC-MARGE†	$0.246^{+0.125}_{-0.181}$	$-0.871^{+0.293}_{-0.338}$	$0.141^{+0.007}_{-0.007}$	$3.100^{+0.086}_{-0.083}$	$24.119^{+0.025}_{-0.029}$	$24.045^{+0.029}_{-0.030}$

† The MCMC-MARGE set uses the marginalized means of the light curve parameter pdf's.

2.3.4 Cosmological constraints from sampling the pdf's of SN Ia light curve parameters

We now present cosmological constraints derived from sampling the pdf's of SN Ia light curve parameters. For this work, we have to limit our analysis to SNe Ia with well-behaved light curve parameter pdf's. For most SNe, the pdf's of their light curve parameters are well-behaved, single-peaked smooth bell curves. However, there are several exceptions with multi-peak pdf profiles for the light curve parameters. We have tracked the multi-peak profiles to data quality issues: some SNe have no data after the maximum light, some SNe have light curves with too few data points, or very noisy data. We exclude those problematic, multi-peak SNe from our cosmological analysis. We also exclude other SNe that don't have any data in any bandpass after the peak magnitude as a quality cut. This results in a set of 729 SNe Ia; we will use only this set in our analysis from this point on.

The effects of excluding the 11 problematic SNe (listed in Table 2.2) from the cosmological analysis are shown in Fig. 2.3. The constraints on Ω_m and w are noticeably shifted by excluding these 11 SNe, indicating that cosmological results could be biased by including poor quality data. We observe the same effect both using the original JLA data and using our MCMC-fitted light curve parameters.

Tests with simulations Before showing the results of pdf sampling applied to the JLA dataset, we first show results from the simulated datasets to illustrate that pdf sampling gives less biased cosmological results than the usual practice without pdf sampling. To generate the simulated datasets, we replace the bias-corrected B-band

Table 2.2: Problematic SNe

	SN name
1	Lancaster
2	Patuxent
3	SDSS11206
4	SDSS14318
5	SDSS16619
6	SDSS16737
7	SDSS16793
8	SDSS19067
9	SDSS21510
10	SDSS21669
11	Torngasek

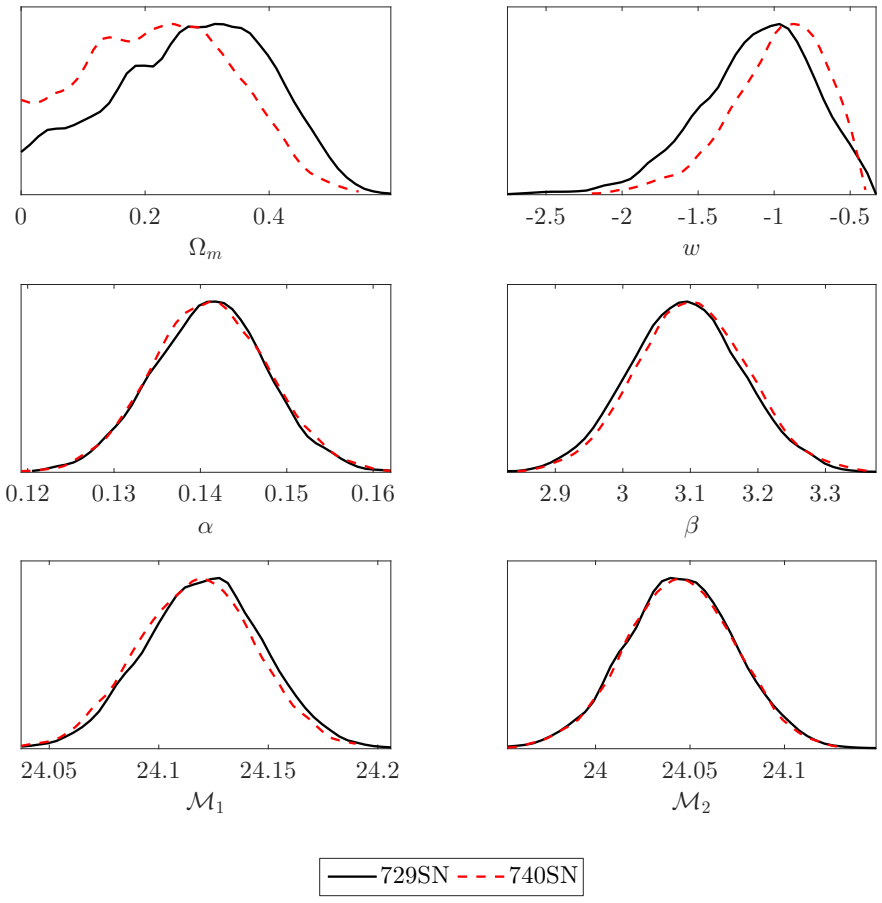

Figure 2.3: Mean likelihood pdf's of the parameters from the cosmological fit for different numbers of SNe Ia using the MCMC-MARGE set. Black solid lines show the results from 729 SNe excluding the 11 problematic SNe; red dashed lines represent the whole SNe sample including all 740 SNe from the JLA data set.

peak magnitude (m_B) in the JLA dataset with true peak magnitude calculated from a fiducial cosmological model, added with ${x_1}^{random}$ and c^{random} randomly drawn from the light curve parameter pdf's we obtain using our MCMC light curve fitter, and a Gaussian scatter ($\mu = 0, \sigma^2 = 0.1^2$):

$$m_B = \mu(\Omega_m, w) - \alpha \times {x_1}^{random} + \beta \times c^{random} + \mathcal{M} + \mathcal{N}(0, 0.1^2) \tag{2.15}$$

The nuisance parameters α, β and $\mathcal{M}$ are defined in Eq. (2.6), and are fixed with fiducial values.

We generate 1000 sets of simulated data and perform cosmological analysis to each set of data, with and without pdf sampling. When applying pdf sampling to the simulated dataset, we only sample the x_1 and c parameter, and only 3 points are chosen from each pdf to speed up the process. We list the input parameters and the means and standard deviations of the 1000 sets of resultant cosmological parameters in Table 2.3. We have shown both the peak values of the mean likelihood pdf and the means of the marginalized pdf. The two values and their standard deviations show differences in the Ω_m and w parameter, which is due to the differences in the meaning of the two kind of pdf's. We have briefly discussed the differences between the mean likelihood pdf and the marginalized pdf in Section 2.3.1. For more detailed discussions, see Lewis & Bridle (2002). Since the peaks of the mean likelihood pdf are less biased in general, we only show the mean likelihood pdf's for cosmological constraints in this work. When quoting the peak values of the mean likelihood pdf's, comparing to not using pdf sampling, applying pdf sampling gives Ω_m and w that are closer to the input parameters, which is expected since pdf sampling utilizes more information from the light curve

parameter pdf's. However, the standard deviations are larger in these two parameters when applying pdf sampling, as the values spread in larger ranges.

Pdf sampling with the JLA dataset We now proceed to implement pdf sampling of the light curve parameters in our cosmological analysis, as described in Section 2.2.4, to the set of 729 SNe Ia, as described below:

1. PDF-COMBINED-3 : We draw $3^3 = 27$ sets of light curve parameters from the pdf's as described in Section 2.2.4. We also apply the same bias correction as used in the JLA set to each individual set of light curve parameters. The cosmological results of the individual sets are combined as described in Section 2.2.4 to obtain the cosmological results with pdf sampling of light curve parameters.

2. PDF-COMBINED-7 : Similarly, $7^3 = 343$ sets of light curve parameters are drawn from the pdf's and the cosmological results are combined.

3. PDF-COMBINED-15 & PDF-COMBINED-19: $15^3 = 3375$ and $19^3 = 6859$ sets of light curve parameters are drawn respectively. These ensure that the combined results have converged with increasing N. (See Fig. 2.4)

We compare the results from using pdf sampling of light curve parameters with the GRID set and MCMC-MARGE (no light curve parameter pdf sampling), shown in Fig. 2.5. Since we have already shown that the pdf sampling results are converged with increasing N, we only show one set of results from using pdf sampling – PDF-COMBINED-15.

Table 2.3: Test with simulated data

	Ω_m	w	α	β	$\mathcal{M}_1$	$\mathcal{M}_2$
input values	0.3	-1.0	0.14	3.1	24.11	24.04
w/o pdf (like[1])	0.205±0.098	-0.928±0.196	0.152±0.006	3.543±0.088	24.114±0.019	24.034±0.021
w/ pdf (like[1])	0.258±0.111	-0.990±0.288	0.152±0.007	3.541±0.089	24.123±0.032	24.038±0.036
w/o pdf (marge[2])	0.247±0.057	-1.038±0.163	0.152±0.006	3.548±0.088	24.112±0.018	24.032±0.020
w/ pdf (marge[2])	0.287±0.055	-1.121±0.181	0.148±0.006	3.545±0.089	24.111±0.018	24.029±0.020

[1] The means and standard deviations of cosmological parameters of the 1000 simulated datasets, by quoting the peak values of the mean likelihood pdf's of the cosmological parameters;

[2] The means and standard deviations of cosmological parameters of the 1000 simulated datasets, by quoting the means of the marginalized pdf's of the cosmological parameters.

Note that the pdf's for Ω_m and w are shifted, compared to the results using the GRID set (grid-based method with no pdf sampling). When comparing with the results using only the marginalized mean values of the light curve parameters (without pdf sampling), the pdf's are also shifted, and pdf-sampling gives a little tighter constraints. The peak of the mean likelihood pdf's and their 68% confidence intervals of the cosmological analysis using 729 SNe excluding the problematic ones are shown in Table 2.4. When excluding the problematic SNe, we get slightly larger Ω_m values, and the w values are closer to -1. In Fig. 2.6, we show the corresponding 2D mean likelihood contours of the fitted parameters of PDF-COMBINED-15, GRID and MCMC-MARGE. It is interesting to note that pdf sampling of SN Ia light curve parameters, and using the means of the light curve parameter pdf's, leads to cosmological constraints closer to a flat Universe with a cosmological constant.

2.4 Summary and discussion

We have developed a method to utilize the probability density distribution (pdf) of SN Ia light curve parameters in cosmological analysis using SNe Ia. First, we have applied Markov Chain Monte Carlo (MCMC) to SN Ia light curve fitting, in order to obtain smooth and well-behaved pdf's of SN Ia light curve parameters. Then we derived cosmological constraints with sampling of the pdf's of the SN Ia light curve parameters. For a complementary approach of sampling the underlying SN Ia population in cosmological model fitting, see March et al. (2011).

In order to validate our method, we applied it to 1000 sets of simulated SN Ia data.

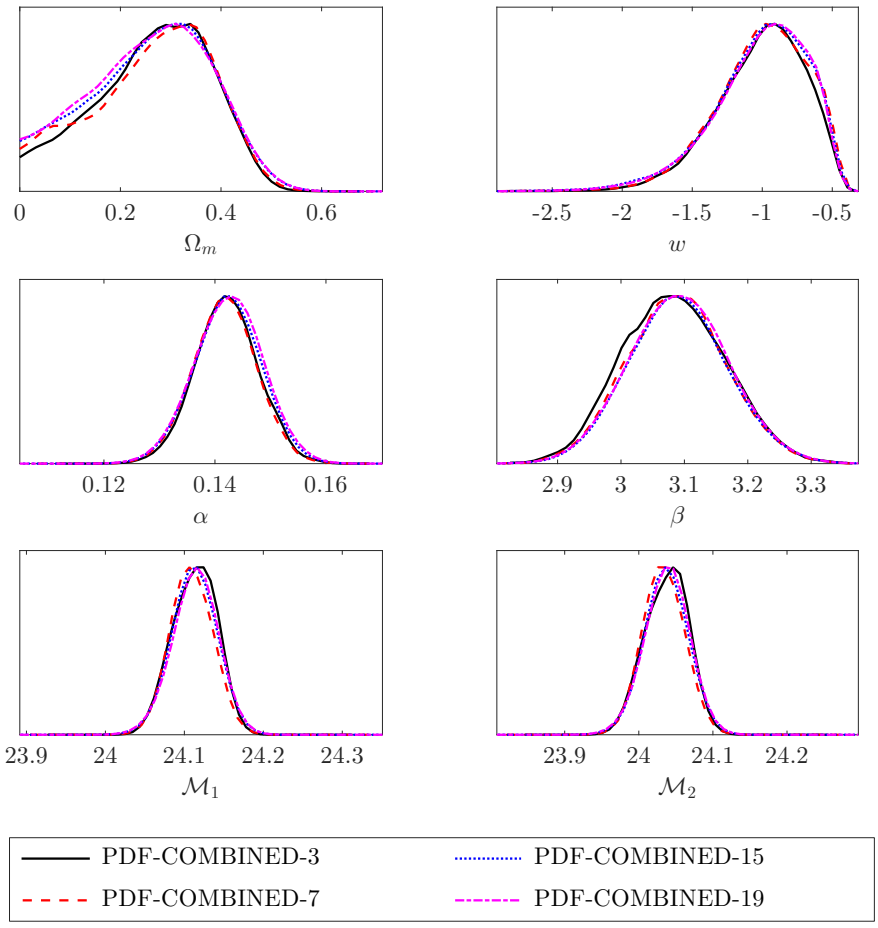

Figure 2.4: Mean likelihood pdf's of the parameters from the cosmological fit with sampling of the SN Ia light curve parameter pdf's, using 729 SNe Ia from the JLA data set (excluding 11 problematic ones). Black solid lines are results from sampling 3 points on each pdf (PDF-COMBINED-3); red dashed lines are results from sampling 7 points on each pdf (PDF-COMBINED-7); blue dotted lines are results from sampling 15 points on each pdf (PDF-COMBINED-15); magenta dash-dotted lines are results from sampling 19 points on each pdf (PDF-COMBINED-19).

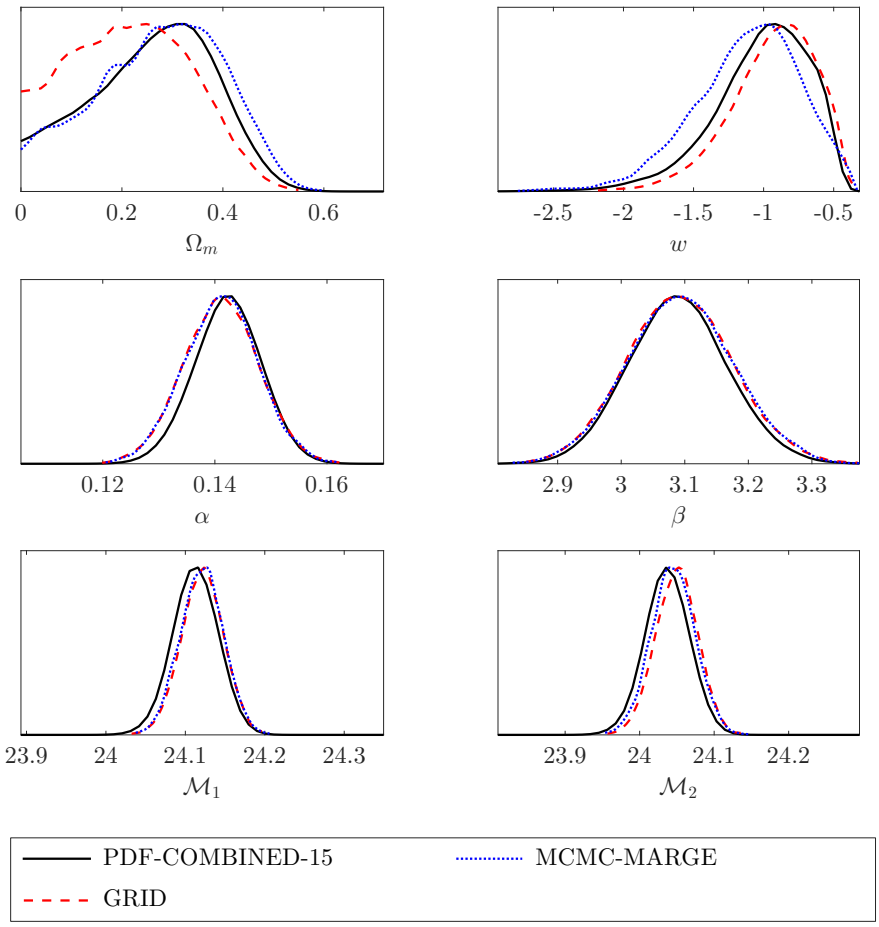

Figure 2.5: Mean likelihood pdf's of the parameters from the cosmological fit with and without sampling of the SN Ia light curve parameter pdf's, using 729 SNe Ia from the JLA data set (excluding 11 problematic ones). Black solid lines are results from sampling 15 points on each pdf (PDF-COMBINED-15); red dashed lines are results from using the GRID set; blue dotted lines are results from using the MCMC-MARGE set, without pdf sampling.

Table 2.4: Parameters from the cosmology fit using 729 JLA SNe (excluding 11 problematic ones)

	Ω_m	w	α	β	$\mathcal{M}_1$	$\mathcal{M}_2$
SALT2	$0.238^{+0.115}_{-0.180}$	$-0.848^{+0.283}_{-0.270}$	$0.140^{+0.007}_{-0.006}$	$3.092^{+0.082}_{-0.092}$	$24.123^{+0.029}_{-0.024}$	$24.054^{+0.028}_{-0.030}$
GRID-SALT2	$0.258^{+0.104}_{-0.228}$	$-0.866^{+0.320}_{-0.306}$	$0.141^{+0.007}_{-0.006}$	$3.090^{+0.087}_{-0.083}$	$24.123^{+0.026}_{-0.028}$	$24.051^{+0.027}_{-0.030}$
GRID	$0.249^{+0.119}$	$-0.830^{+0.291}_{-0.336}$	$0.141^{+0.007}_{-0.006}$	$3.093^{+0.083}_{-0.087}$	$24.124^{+0.024}_{-0.028}$	$24.053^{+0.026}_{-0.031}$
MCMC-LIKE	$0.263^{+0.120}_{-0.205}$	$-0.859^{+0.317}_{-0.359}$	$0.141^{+0.007}_{-0.007}$	$3.071^{+0.083}_{-0.080}$	$24.126^{+0.025}_{-0.030}$	$24.049^{+0.028}_{-0.031}$
MCMC-MARGE†	$0.312^{+0.118}_{-0.158}$	$-0.965^{+0.266}_{-0.420}$	$0.142^{+0.006}_{-0.007}$	$3.096^{+0.080}_{-0.087}$	$24.128^{+0.021}_{-0.033}$	$24.040^{+0.035}_{-0.024}$
PDF-COMBINED-3	$0.339^{+0.065}_{-0.161}$	$-0.910^{+0.288}_{-0.339}$	$0.142^{+0.006}_{-0.005}$	$3.076^{+0.092}_{-0.085}$	$24.124^{+0.023}_{-0.042}$	$24.046^{+0.025}_{-0.039}$
PDF-COMBINED-7	$0.335^{+0.072}_{-0.145}$	$-0.975^{+0.417}_{-0.288}$	$0.142^{+0.005}_{-0.006}$	$3.098^{+0.066}_{-0.093}$	$24.107^{+0.031}_{-0.026}$	$24.026^{+0.037}_{-0.023}$
PDF-COMBINED-15	$0.325^{+0.081}_{-0.162}$	$-0.916^{+0.351}_{-0.344}$	$0.143^{+0.005}_{-0.007}$	$3.084^{+0.079}_{-0.072}$	$24.116^{+0.026}_{-0.032}$	$24.036^{+0.031}_{-0.028}$
PDF-COMBINED-19	$0.309^{+0.098}_{-0.162}$	$-0.912^{+0.343}_{-0.330}$	$0.143^{+0.006}_{-0.006}$	$3.096^{+0.075}_{-0.086}$	$24.116^{+0.028}_{-0.029}$	$24.035^{+0.034}_{-0.026}$

† The MCMC-MARGE set uses the marginalized means of the light curve parameter pdf's.

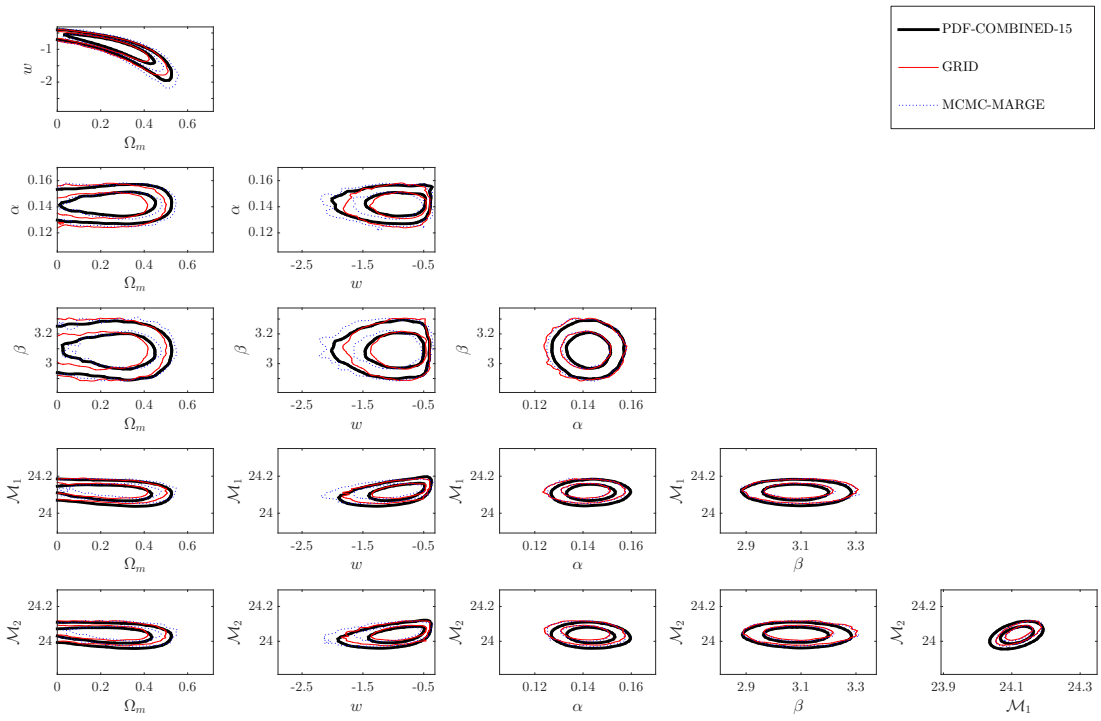

Figure 2.6: Joint confidence level contour plots from the cosmological fit with and without sampling of the SN Ia light curve parameter pdf's, using 729 SNe Ia from the JLA data set (excluding 11 problematic ones). The contours are 68% and 95% confidence levels. Thick black solid contours are are results from sampling 15 points on each pdf (PDF-COMBINED-15); thin red solid contours are results from using the GRID set; blue dotted contours are results from using the MCMC-MARGE set, without pdf sampling.

We found that compared to not using pdf sampling, applying pdf sampling gives Ω_m and w that are closer to the input parameters, which is expected as pdf sampling utilizes more information from the light curve parameter pdf's. We also found that the peak values of the mean likelihood pdf's are closer to the input values than the marginalized means.

Our method differs from the usual approach in two ways: (1) We use MCMC, instead of a grid-based method, in fitting the SN Ia light curve parameters; (2) We sampled the pdf's of the SN Ia light curve parameters in the cosmological analysis, instead of just using the peaks of the pdf's.

We have applied our method to the Joint Lightcurve Analysis (JLA) data set of SNe Ia derived by Betoule et al. (2014), which combines the SNe Ia from SDSS and SNLS in a consistent, well-calibrated manner. Interestingly, we find that the resultant cosmological constraints are closer to that of a flat Universe with a cosmological constant, compared to the usual practice of using only the best fit values of the SN Ia light curve parameters.

The JLA set has a bias correction term which is determined using simulations generated by the SNANA software (Kessler et al., 2009), by comparing the reconstructed distance using SALT2-fitted light curve parameters to the simulation inputs (Betoule et al., 2014). We use this bias correction in our analysis, assuming that it is independent of the fitting technique. Ideally we should perform our own bias calculation by fitting the simulations with our MCMC light curve fitter. We will leave this for future work. We also note that the significantly smaller marginalized errors (compared to likelihood errors) from the simulated datasets could be due to the simulated datasets not being

sufficiently realistic. We will investigate this further in future work.

As SN Ia data increases in both quantity and quality, our method will be useful in the quest to illuminate the nature of dark energy using cosmological data.

Chapter 3

Photometric classification and redshift estimation of the LSST Supernovae

3.1 Introduction

The Supernovae (SNe) we observe need to be correctly typed and accurate redshift information needs to be obtained, before the SNe Ia can be used to constrain cosmological models. With a sample size of < 1000, it is possible to obtain correct types and redshifts of SNe via spectroscopy. Ongoing and planned surveys such as the Dark Energy Survey (DES, Bernstein et al. 2012), and the Large Synoptic Survey Telescope (LSST, LSST Science Collaboration et al. 2009), will observe a dramatically increased number of SNe, making it difficult to spectroscopically follow up all the SNe. SN cosmology will rely on photometric typing and redshift estimation. It is important to derive methods for reliable and accurate typing and redshift estimation using the photometric data of the SNe only.

In this work, we use realistic LSST SN simulations to study the performance of SN classification with the Random Forest classification algorithm, using SN colors as features for the first time, together with parameters from a general function fit of the light curves. We are able to obtain a photometrically classified SN Ia sample that is 99% pure. We derive the photometric redshifts of the SNe by fitting their light curves using the SALT2 model with a nested sampling algorithm, and study the performance of this photometrically classified sample with photometric redshifts in constraining cosmology.

We describe our simulations in Section 3.2, our classification method in Section 3.3, our photo-z estimator in Section 3.4. The construction of a photometric Hubble Diagram is described in Section 3.5, followed by a summary and discussion in Section 3.6.

3.2 SN simulations

We use the SNANA[1] (Kessler et al., 2009) software to generate realistic SN light curve simulations; it provides a simulation library that contains the survey condition information, and the LSST filter transmissions. Our simulation is based on 10 years of operation on 10 deep drilling fields of LSST. The actual number and location of the deep drilling fields, as well as the observing cadence, are still under study. The observing cadence will affect the final number of SNe that we can observe and the quality of the light curves. We generate mixed-type SNe by using SN rates for both Type Ia and non-Ia (core-collapse (CC) SN). The SN rate is parametrized in the form of $\alpha(1+z)^{\beta}$. For SNe Ia, we use the rate measured by Dilday et al. (2008), with $\alpha_{\rm Ia} = 2.6 \times 10^{-5}\,{\rm SNe\,Mpc^{-3}}\,h_{70}^{3}\,{\rm yr^{-1}}$ and $\beta_{\rm Ia} = 1.5$; for the core-collapse SNe, we assume $\alpha_{\rm CC} = 6.8 \times 10^{-5}\,{\rm SNe\,Mpc^{-3}}\,h_{70}^{3}\,{\rm yr^{-1}}$ and $\beta_{\rm CC} = 3.6$, following Bernstein et al. (2012) and LSST Science Collaboration et al. (2009). The light curves are generated using the extended SALT2 model, which extends the standard SALT2 model on both ends of the spectra. The use of this extended model is essential in generating light curves in all six LSST bands, since with the standard SALT2 model some bands with certain redshifts will not be generated if they fall out of the wavelength range of the standard model. However, since there is little flux out of the wavelength range of

[1] http://snana.uchicago.edu/

the standard SALT2 model, the light curves generated in those bands are basically constant noise. Section 3.4.1 gives a brief review of the SALT2 model. We make a quality cut during the simulation by requiring that the photometry in at least 3 bands have a maximum signal-to-noise-ratio (SNR) greater than 5. A total of 144246 SNe are obtained, including 62147 Ia, 67631 II, and 14468 Ibc. The redshift range of our simulation is from 0.01 to 1.2. We also apply other quality cuts on this sample in different stages, which are discussed in section 3.3.3.

3.3 SN classification

3.3.1 Using SN colors for classification

In Wang et al. (2015), SN colors are used to build an analytic photo-z estimator with very good performance. Inspired by the ability of using SN colors only to estimate the SN photo-z, we expand the usage of SN colors in SN classification. The SN light curves are first fitted into a general functional form that is described in detail in Section 3.3.2, so that a relatively accurate peak for each band is obtained. The peak magnitudes (converted from peak fluxes) are then calculated using the fitted parameters (Eq. (3.3) and (3.4)), and the colors are calculated as the difference between the peak mag of the two adjacent bands.

$$c_{ij} = m_{p,i} - m_{p,j} \tag{3.1}$$

where i and j represents two of the adjacent bands of the LSST filters (ugrizY).

We find that the SN colors can be used in SN classification with very good performance using the machine learning classification algorithms when they are used together

with the two parameters from the general parametrization that describes the rising and falling time of the SN light curves.

3.3.2 General parametrization of SN light curves

The SN light curves (both Type Ia and non-Ia) are found to be well fitted using a general functional form; it has no specific physical motivation, but describes the lightcurve shapes in a model-independent manner. Following Bazin et al. (2009), we use the following equation to fit the SN light curves (we refer to this as the "Bazin function"):

$$f(t) = A\frac{\exp^{-(t-t_0)/t_{\rm fall}}}{1+\exp^{-(t-t_0)/t_{\rm rise}}} + B \tag{3.2}$$

In this equation, $t_{\rm fall}$ and $t_{\rm rise}$ measure the declining and rising time of the light curve, A is the normalization constant, and B is a constant term. We should note that t_0 is not exactly the date of maximum flux (date at the peak of the light curve), and similarly A is not exactly the maximum flux. By calculating the derivative of the function we are able to obtain the functional form of date at the maximum flux and the value of the maximum flux:

$$t_{\rm max} = t_0 + t_{\rm rise}\ln(\frac{t_{\rm fall}}{t_{\rm rise}} - 1) \tag{3.3}$$

$$f(t_{\rm max}) = Ax^x(1-x)^{1-x} + B, x = \frac{t_{\rm rise}}{t_{\rm fall}} \tag{3.4}$$

The equation above also indicates that we should set $\frac{t_{\rm rise}}{t_{\rm fall}} < 1$ in order to have a meaningful $t_{\rm max}$. This constraint is useful in excluding some of the bad fits. More details on the quality cuts are described in Section 3.3.3.

For higher redshift SNe, there is little flux in the u, g band. At the other end of the spectra, the Y band data are usually noisy. We notice that the Bazin function does not

fit well for these bands with low SNR. So we fit the light curve with different forms depending on the SNR of the band being fitted. For SNR > 5, the light curve is fitted using Eq. (3.2), otherwise the light curve is fitted to a constant $f(t) = B$.

We utilize the curvefit procedure in python scipy[2]. It is necessary to set initial values and limits for the parameters being fitted. To achieve better results, we do the fit in two steps with different initial conditions and parameter limits. We list the initial values and parameter limits in Table 3.1. The initial values of the 2nd fit are calculated using the results of the 1st fit. We define a "successful fit" as satisfying the following conditions: $t_{\rm fall} > t_{\rm rise} > 1$. The successfully fitted bands in the 1st fit are kept and the median of the parameters of such bands are used as the initial value in the 2nd fit. A 2nd fit is also performed when the constant fitting returns a B value larger than 5; this usually happens for the Y band where the SNR is lower than 5 but there is indeed signal and can be fitted with Eq. (3.2).

The peak magnitude in each band is calculated as:

$$m_p = -2.5 \log_{10}(f_{\rm max}) + \text{zero-point} \tag{3.5}$$

where $f_{\rm max}$ is calculated using Eq. (3.4). For any band that is fitted to a constant, $t_{\rm fall}$, $t_{\rm rise}$, and m_p are set to 0. The colors are calculated using Eq. (3.1), regardless of the value of m_p. We notice that the colors can be 0, or the opposite of the peak magnitude in one band, instead of actual colors, when the zero peak magnitude is used in calculating the colors. We treat this as a property of our sample and pass it to the classifier.

[2]https://www.scipy.org/

Table 3.1: 2-step general parametrization fit: initial conditions and parameter limits

	1st step		2nd step	
	initial value	limits	initial value	limits
A	flux at peak	[0, inf]	flux at peak	[0, inf]
t_0	time at peak	[-inf, inf]	median(t_0)	fixed
$t_{\rm fall}$	15	[0, inf]	median($t_{\rm fall}$)	[1, inf]
$t_{\rm fall}$	5	[0, inf]	median($t_{\rm rise}$)	[1, inf]
B	0	[-inf, inf]	0	[-inf, inf]

3.3.3 Quality cuts

In order to obtain a high quality sample, we apply several quality cuts, both before and after the general function fit. Before the fitting, we require that the max SNR is greater than 5 for at least 3 bands, and that at least 3 bands have 1 point before the peak and 2 points after the peak, including at least one of the SNR > 5 bands. These cuts ensure that the light curve has a well-defined peak and not too noisy in at least one band so that at least one successful fit is achieved in the first step described above.

After all the pre-fitting quality cuts are applied, the fitting program is able to return a set of parameter values for most of the SNe, although some of the values are not in a reasonable range. So we make a series of cuts based on the parameter distributions. The cuts we used are listed below:

- $t_{\rm rise} > 1$, and $t_{\rm rise}$ not close to 1 with tolerance = 0.01
- $-20 < B < 20$

- $\chi^2/d.o.f < 10$
- $t_{\rm fall} < 150$
- $t_{\rm rise} < t_{\rm fall}$
- $A < 5000$
- $A_{\rm err} < 100$
- $t_{0,\rm err} < 50$
- $t_{\rm fall,err} < 100$
- $t_{\rm rise,err} < 50$
- $A(Y), A(u) < 1000$

These cuts also serve to exclude some of the non-Ia's, especially type II's which have a rather larger $t_{\rm fall}$ value. The remaining fractions after cuts for each of the three SN types in the simulations are 55%, 38% and 15%, for Ia, Ibc and II respectively.

This after-cut sample is used to test our classification algorithm, which contains 68% Ia's, 11% Ibc's and 20% II's. We notice that the χ^2 cut eliminates almost all of the low-z ($z < 0.3$) SNe, since the low-z SNe usually have very high SNR and have a second peak in the redder bands, which cannot be well-fitted using Eq. (3.2), and thus result in very large χ^2 per d.o.f.

Detailed lists of the remaining number of SNe after each cut are shown in Appendix A.

3.3.4 SN classification with random forest algorithm

Machine learning algorithms are used in SN classification recently (Lochner et al., 2016; Möller et al., 2016), and have excellent performance when the features used in the classifier are carefully selected. Here we choose the Random Forest algorithm to demonstrate the performance of classification using SN colors. We adopt a code similar to the one used in host galaxy identification by Gupta et al. (2016), and modified it to suit our needs. For details about the algorithm, see Breiman (2001). There are also many other machine learning algorithms that can be used, most of which are very easy to implement. The comparison of performance for several commonly used machine learning algorithms can be found in Lochner et al. (2016).

As described in Section 3.3.1, a total number of 17 features are passed to our classifier. The features are: 12 Bazin-fit parameters $t_{\rm fall}(u)$, $t_{\rm rise}(u)$, $t_{\rm fall}(g)$, $t_{\rm rise}(g)$, $t_{\rm fall}(r)$, $t_{\rm rise}(r)$, $t_{\rm fall}(i)$, $t_{\rm rise}(i)$, $t_{\rm fall}(z)$, $t_{\rm rise}(z)$, $t_{\rm fall}(Y)$, $t_{\rm rise}(Y)$, and 5 colors c_{ug}, c_{gr}, c_{ri}, c_{iz}, c_{zY}.

We now summarize the concepts that are commonly used in presenting the classification results.

Confusion Matrix. For a binary classification problem, a confusion matrix is defined in Table 3.2. In our case, the two classes are 'Ia' (Yes) and 'non-Ia' (No).

Receiver Operating Characteristic (ROC) Curves. We define the true positive rate (TPR) and the false positive rate (FPR) as the following (according to the confusion

Table 3.2: Confusion matrix for a binary classification

		Predicted Class	
		Yes	No
Actual Class	Yes	Ture Positive (TP)	False Negative (FN)
	No	False Positive (FP)	True Negative (TN)

matrix):

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3.6}$$

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}} \tag{3.7}$$

By varying the probability threshold within a classifier in determining the class, different values of TPR and FPR are returned. The ROC curve is defined as TPR vs FPR, since we would expect an excellent classifier to have high TPR with low FPR. Another value that is often used in comparing classification results is the area-under-the-curve (AUC) of a ROC curve. For perfect classification, AUC = 1, the ROC curve behaves as a step function, while for a random classification, AUC = 0.5, the ROC curve behaves as a diagonal line. An AUC that is larger than 0.9 usually represents excellent classification.

Efficiency and Purity We can also define the efficiency and purity using the confusion matrix:

$$\text{efficiency} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3.8}$$

$$\text{purity} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{3.9}$$

To achieve higher purity usually means sacrificing the efficiency, and vice versa, given that the probability threshold is varied.

3.3.5 Training sample size determination

The classification algorithms rely on a training sample with known types to predict types for the test sample. There are generally two ways that a training sample can be obtained: one is to use a spectroscopic sample from the same survey, but the sample size can be relatively small, compared to the large number of SNe that LSST can observe; the other is simply using realistic simulations, so the sample size can be as large as we need, representing good statistics of the test sample.

We compare the effects on the classification results by varying the training sample size as a fraction of 0.05, 0.1, 0.3, 0.5 and 0.9 of the total sample. We find that a training set of 0.05 or 0.1 fraction of the whole sample results in an AUC of 0.96, while a fraction greater or equal to 0.3 results in an AUC of 0.98, which indicates that a large enough training sample size is required to best represent the sample and thus leads to better classification performance. However, as our sample size after all the quality cuts is $\sim$ 39000 including all types, a 0.1 fraction with $\sim$ 3900 SNe is already larger than current spectroscopically confirmed data sets. This means that a spectroscopic training set for classification will be challenging to obtain. We also conclude that a training sample with comparable size to the test sample will result in best performance. In this work, we show the classification result with the 0.3-fraction training sample. We use the same sample for the subsequent analysis, in order to obtain as large a sample size as we can for the cosmological analysis. When dealing with real data in the future, a simulated

sample with the same size as the real sample can be used for the best performance.

3.3.6 Classification results

We now present the classification results using the concepts defined in Section 3.3.4. Fig. 3.1 shows the ROC curve for our classification, with an AUC of 0.98, indicating that we reach excellent classification by using the features we described in Section 3.3.1. This AUC value is comparable to recent studies (Lochner et al., 2016; Möller et al., 2016) with different data sample or simulations.

Fig. 3.2 shows the purity and efficiency curves as the threshold probability varies. High purities can be obtained by sacrificing some efficiency. We notice that a 90% - 95% purity can be easily achieved with efficiency larger than 90%. While we aim at a purity of 99% for the following cosmological analysis, the efficiency dropped to 74%. Note that this efficiency is the classification only efficiency, not including the quality cuts through all the procedures. Our analysis results in a photometric sample with 13744 SNe with 99% purity.

3.4 SN photometric redshift

Accurate redshift information is essential in constructing the Hubble Diagram and constraining cosmology. Analyses of past and ongoing surveys rely on spectroscopic redshifts – either from the SN spectra or the host-galaxy spectra. With LSST, it is impossible for us to obtain spectroscopic redshifts for all the SNe; it is also unclear whether we will be able to obtain spectroscopic redshifts for the host galaxies of all the SNe; thus it is useful to develop methods for SN redshift estimations using the

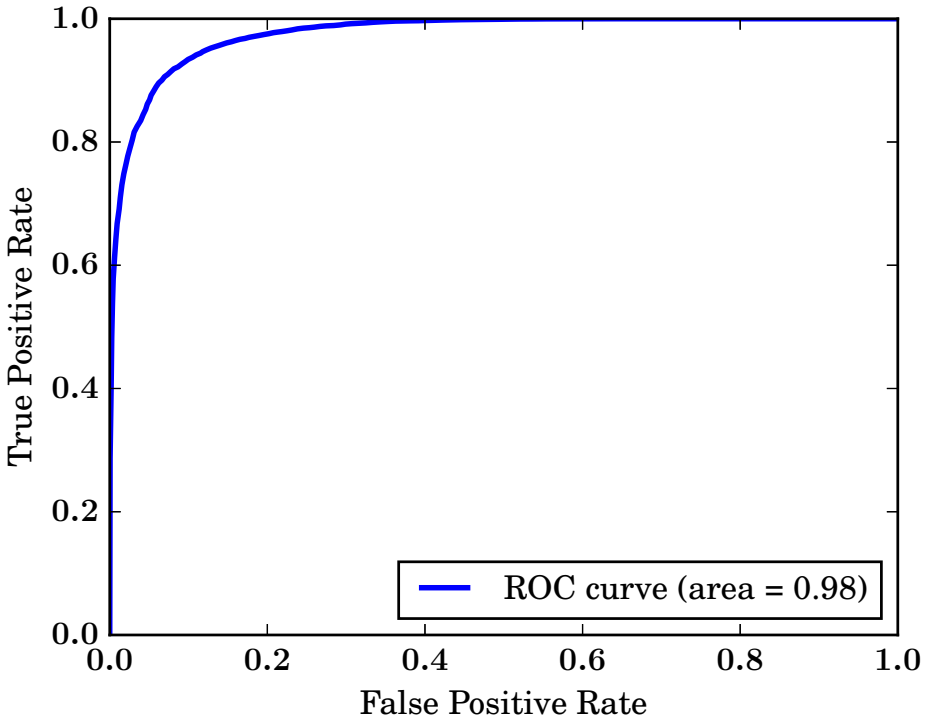

Figure 3.1: ROC curve for our classification result.

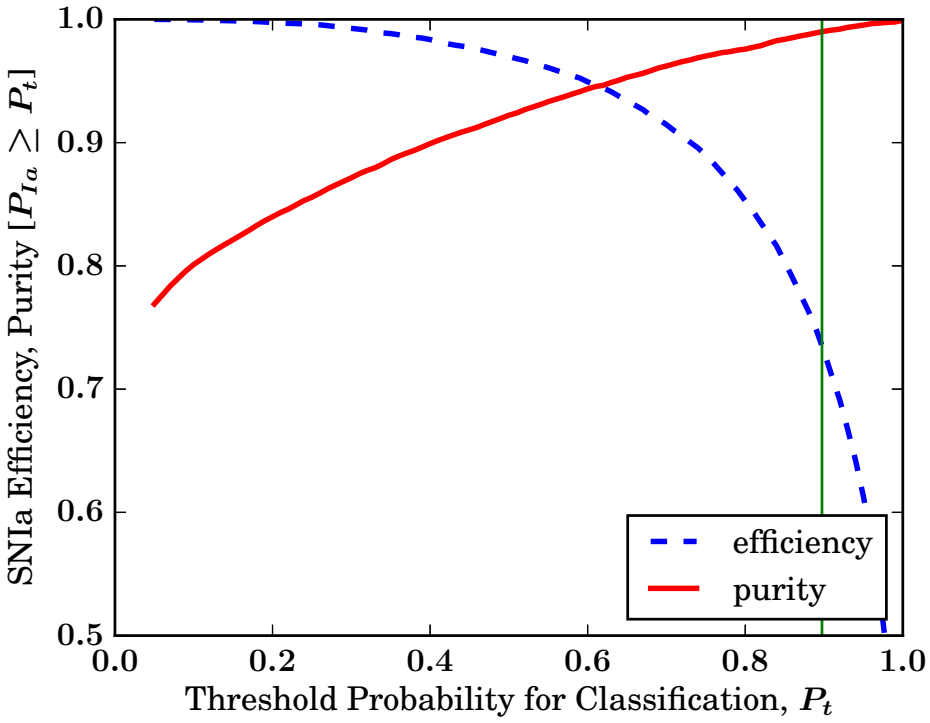

Figure 3.2: Purity and efficiency curves for our classification result, red solid line shows the purity curve with respect to the threshold probability chosen, blue dashed curve shows the efficiency curve. The vertical line indicates the threshold probability for 99% purity.

photometric data. Currently two kinds of approaches are proposed for SN photo-z estimation: one is analytic utilizing the multi-band SN colors (Wang et al., 2007; Wang, 2007; Wang et al., 2015), the other is template-based by fitting the light curves into SN Ia models (Kessler et al., 2010; Palanque-Delabrouille et al., 2010). We adopt the template approach in this work, by fitting the SN light curves using the commonly used SALT2 model, using the nested-sampling method. Our fits are performed using the SNCosmo[3] package. We also investigate the performance of our photo-z estimator using a host-galaxy photo-z prior.

3.4.1 The SALT2 model

The SALT2 model (Guy et al., 2007, 2010) provides an average spectral sequence and its higher order variations, as well as a color variation law, which can be used to fit SN Ia light curves using several parameters. The model flux is given as:

$$\frac{dF}{d\lambda}(p,\lambda) = x_0 \times [M_0(p,\lambda) + x_1 M_1(p,\lambda) + \ldots] \times \exp[c\, CL(\lambda)] \tag{3.10}$$

where p is the phase (rest-frame time from the maximum light), λ is the rest-frame wavelength, M_0 and M_1 are the spectral sequence and its first order variation, CL is the color law. x_0, x_1 and c are light curve parameters that describe the amplitude, stretch and color of the light curve; they are fitted in a fitting process in which the date of maximum t_0 and the redshift z can also be fitted simultaneously.

In this work, we use an extended model that covers wider wavelength ranges than the standard SALT2 model.

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

3.4.2 2-stage fit using nested sampling

We choose to use the nested sampling method for light curve fitting using SALT2. We find that nested sampling results in better photo-z estimates, compared to the normal maximum likelihood method using *MINUIT*.

The model is not well characterized in the UV region with a dramatically increased uncertainty, which can lead to a wrong fit if the parameter limits are not set correctly.

We take advantage of a 2-stage fit again in the SALT2 fit. Now we describe the two stages in detail:

For the initial fit, we aim at locating the right ranges of the parameters. The model covariance is not used in this fit, only the statistical errors from the photometry are included. We set the parameter limits as follows: $0.01 < z < 1.2$, $|x_1| < 5$, $|c| < 0.5$, $t_{\rm min} - 15 < t_0 < t_{\rm max} + 15$; and the bound of the amplitude parameter, x_0, is determined internally by SNCosmo.

The second fit is limited to a smaller range determined by the 1st fit, with x_1, c, t_0 limited to a 3-sigma range from the mean value of the 1st fit, while x_0 bounds are still "guessed" by *SNCosmo*. We limit the redshift to a rather larger range: $z_{\rm ini} \pm 10\sigma_z$, in order to better estimate the final uncertainty of z. We find that the 3σ limit for the other three parameters is necessary for obtaining a good fit, since the fit can easily be trapped in an unreasonable parameter region where the χ^2 is very small due to a large value of uncertainty in the UV bands. This fitting deficiency has been observed in another analysis (Dai & Wang 2016, see Chapter 2), which uses Markov Chain Monte Carlo (MCMC) method to fit light curves to the SALT2 model. In Dai & Wang (2016),

the model covariance is kept fixed to mimic and reproduce the original SALT2 result. When we are simultaneously fitting the redshift, fixing the covariance is not applicable. We also use another condition in the fitting that helps to reduce this fitting bias: for SNe with redshifts (from the initial fit) that are less than 0.65, all 6 bands are used in the fitting; for those that have redshifts greater than 0.65, the u and g bands are excluded from the fitting. The 0.65 line is determined by observing the fitted results using all 6 bands and determining where the bias starts to occur. A better characterized model in the UV band can be very useful in all aspects.

By applying the 2-step fit described above, we obtain a set of photo-z's with an accuracy $\sigma\left(\frac{z_{\rm phot}-z_{\rm spec}}{1+z_{\rm spec}}\right) = 0.0294$, and a mean bias ($\left\langle z_{\rm phot} - z_{\rm spec}\right\rangle$) of 0.0120, after applying a cut on the reduced χ^2 of the SALT2 fit ($\chi^2_{\rm red} < 1.5$). The results are shown in Fig. 3.3. The outlier ($\left|\frac{z_{\rm phot}-z_{\rm spec}}{1+z_{\rm spec}}\right| > 0.1$) fraction is 1.12%. We also show that our method results in accurate photo-z errors, as illustrated in Fig. 3.4. In Fig. 3.4, the histograms of $(z_{\rm phot} - z_{\rm spec})/z_{\rm err}$ are plotted in different redshift ranges, where the $z_{\rm err}$ is the error in the photo-z which is output from the SALT2 fitting. The histograms are fitted with a Gaussian function. With the fitted σ close to 1, we conclude that photo-z error estimation from the SALT2 fitting is accurate. Such a fit can only be achieved when the SALT2 model covariance is included in the fitting and the parameter limits are carefully chosen.

3.4.3 Effect of host galaxy priors

We investigate the effect of using a host-galaxy photo-z prior in the SALT2 fitting. We apply a Gaussian prior with mean and sigma values set as the host-galaxy photo-z and

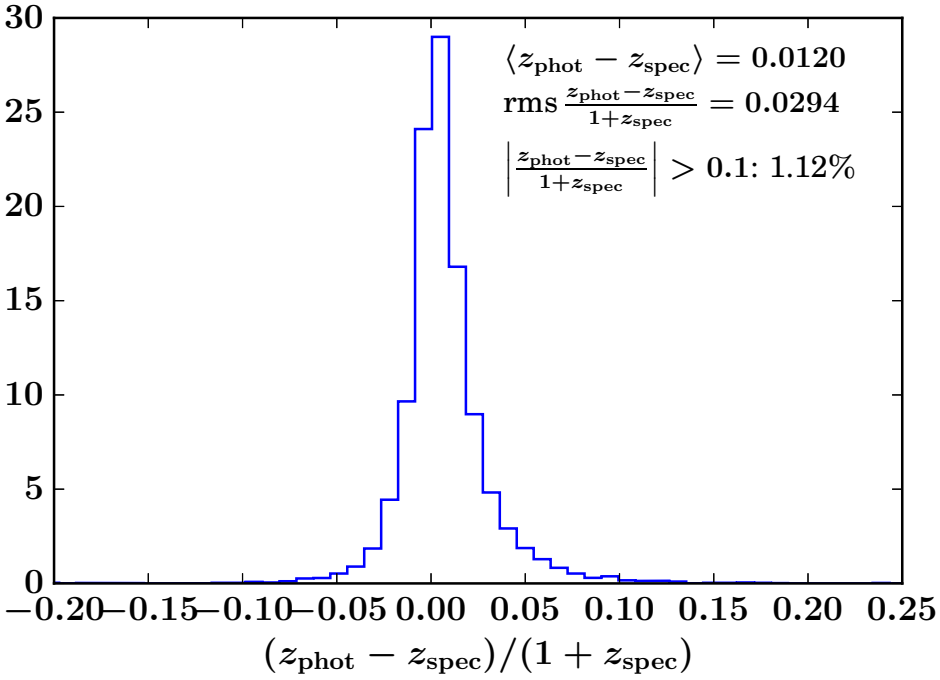

Figure 3.3: Distribution of $(z_{\rm phot} - z_{\rm spec})/(1 + z_{\rm spec})$, with no host-galaxy photo-z prior.

error from a simulated host-galaxy library for LSST. The SNe with host-galaxy photo-z smaller than 0.01 or greater than 1.2 are dropped. Using this host-galaxy photo-z prior, we obtain a set of photo-z's with an accuracy $\sigma\left(\frac{z_{\rm phot}-z_{\rm spec}}{1+z_{\rm spec}}\right) = 0.0116$, and a mean bias ($\left\langle z_{\rm phot} - z_{\rm spec}\right\rangle$) of 0.0017, after applying a cut on the reduced χ^2 of the SALT2 fit ($\chi^2_{\rm red} < 1.5$). The outlier ($\left|\frac{z_{\rm phot}-z_{\rm spec}}{1+z_{\rm spec}}\right| > 0.1$) fraction is 0.16%. The results are shown in Fig. 3.5. Using a host galaxy photo-z prior leads to significant improvement in the photo-z estimation, although the currently available LSST host galaxy library that we have used may have optimistic host galaxy photo-z errors. We will re-evaluate the performance of our photo-z estimator with host-galaxy priors when a more realistic LSST host galaxy library becomes available.

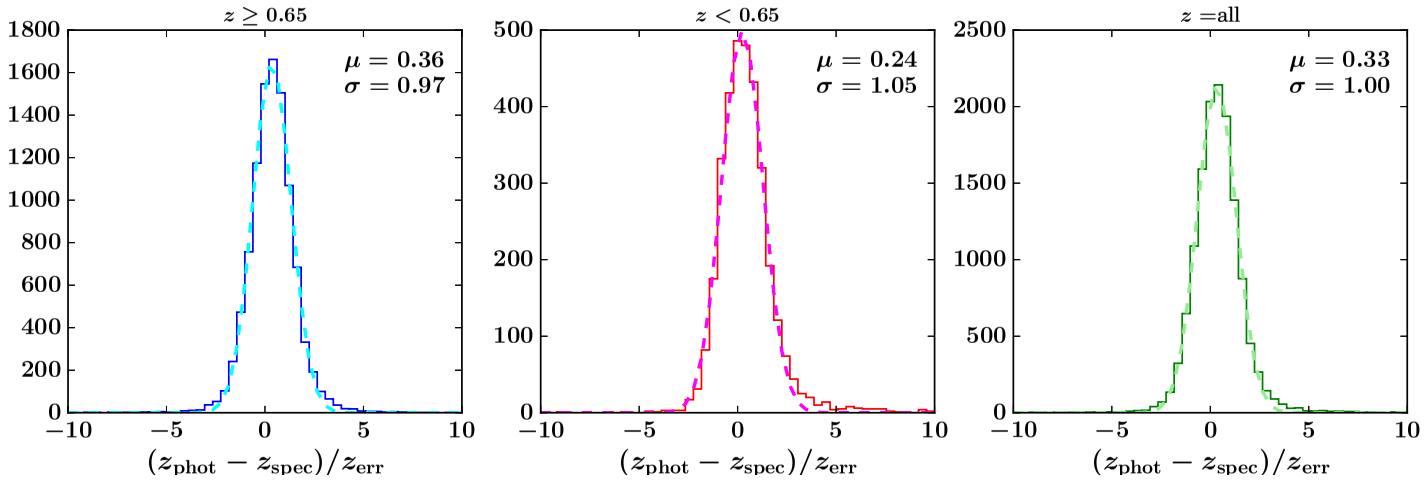

Figure 3.4: Distribution of $(z_{\rm phot} - z_{\rm spec})/z_{\rm err}$ in different redshift ranges. Left: $z \geq 0.65$, middle: $z < 0.65$, right: z in the whole sample range. Dashed lines are from Gaussian fits with best-fit value shown in the right corner.

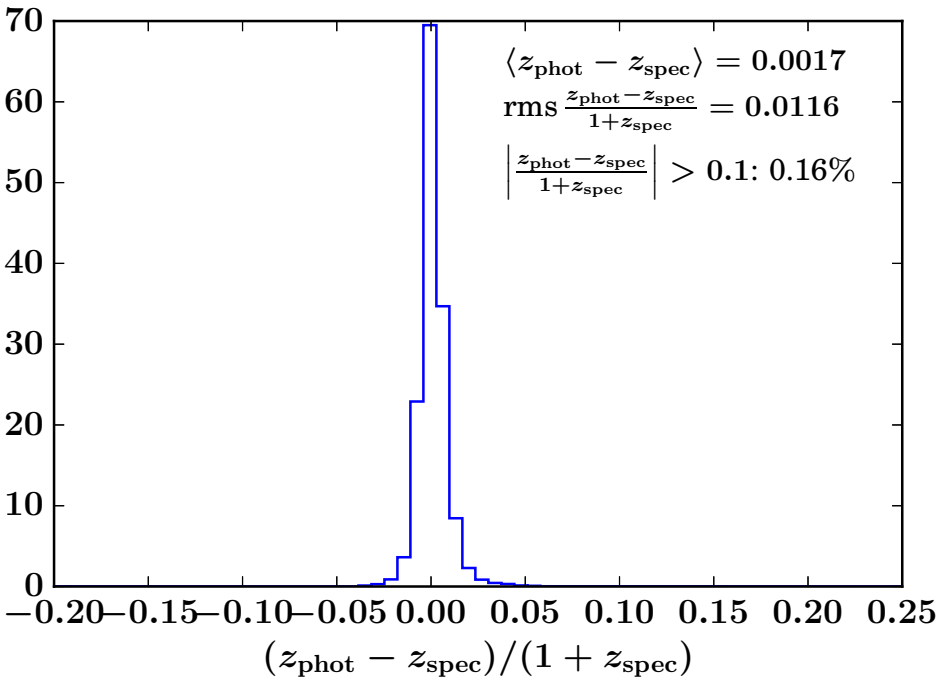

Figure 3.5: Distribution of $(z_{\rm phot} - z_{\rm spec})/(1 + z_{\rm spec})$, with host-galaxy photo-z prior.

3.5 Fitting cosmology

Our final step is to examine the performance of using our photometric-only SN Ia sample (with CC contamination) in constraining cosmology. We use the SALT2 parameters (x_0, x_1, c) obtained from Section 3.4.2 to calculate the distance modulus μ:

$$\mu = m_B + \alpha\, x_1 - \beta\, c + \mathcal{M}, \tag{3.11}$$

where α, β and $\mathcal{M}$ are nuisance parameters, and $m_B = -2.5 \log_{10}(x_0)$.

We have simulated the data assuming the ΛCDM model with $\Omega_m = 0.3$ and a flat Universe. The χ^2 is calculated as:

$$\chi^2 = \sum_i \frac{(\mu_i - \mu_{\text{model},i})^2}{\sigma_i^2} \tag{3.12}$$

where

$$\begin{aligned}\sigma^2 = \sigma_{m_B}^2 + \alpha^2\sigma_{x_1}^2 + \beta^2\sigma_c^2 + \left(\frac{\partial\mu_{\text{model}}}{\partial z}\right)^2 \sigma_z^2 + \sigma_{\text{int}}^2 \\ + 2\alpha\text{Cov}_{m_B,x_1} - 2\beta\text{Cov}_{m_B,c} - 2\left(\frac{\partial\mu_{\text{model}}}{\partial z}\right)\text{Cov}_{m_B,z} - 2\alpha\beta\text{Cov}_{x_1,c} \\ - 2\alpha\left(\frac{\partial\mu_{\text{model}}}{\partial z}\right)\text{Cov}_{x_1,z} + 2\beta\left(\frac{\partial\mu_{\text{model}}}{\partial z}\right)\text{Cov}_{c,z}\end{aligned} \tag{3.13}$$

The $\partial\mu_{\text{model}}/\partial z$ term is calculated numerically by:

$$\frac{\partial\mu_{\text{model}}}{\partial z} = \frac{5}{\log 10}\left(\frac{1}{1+z} + \frac{1}{r(z)}\frac{\partial r(z)}{\partial z}\right) \tag{3.14}$$

where $r(z)$ is the comoving distance as described in Eq. (1.8)

We set σ_{int} to be a constant value and vary it with different values to see whether it affects the fitting outcome. Note that our simulation is generated using a more complicated intrinsic scatter model, so using this constant σ_{int} may introduce a bias.

Only the statistical uncertainties are considered in this fit. The fit is performed using the CosmoMC (Lewis & Bridle, 2002)[4] software.

Before fitting to cosmological models, a bias correction term is calculated and applied to the distance modulus μ in Eq. (3.11). The bias correction term is determined using a separate "bias correction" set of simulation generated assuming a different cosmology (with $\Omega_m = 0.27$ and $w = -1$), with a sample size similar to our original data set. We apply our methodology for photometric classification and photo-z estimation to this separate data set, including the same cuts, and calculate the bias in 20 redshift bins by taking the mean of the differences between the fitted distance moduli, $\mu_{\rm fit}$, and the true distance moduli, $\mu_{\rm true}$ of the SNe in each bin:

$$\Delta\mu(z_i) = \langle \mu_{\rm fit} - \mu_{\rm true} \rangle_{z_i} \tag{3.15}$$

where $\mu_{\rm true}$ is calculated from the input cosmological model, and $\mu_{\rm fit}$ is calculated from the fitted SALT2 parameters, with α and β set to be the simulated values.

The correction for each SN is then obtained using linear interpolation. So the bias-corrected distance of each SN becomes:

$$\mu_{\rm SN} = \mu_{\rm fit,SN} - \Delta\mu_{\rm SN} \tag{3.16}$$

3.5.1 Ellipse cut and other quality cuts

We utilize an ellipse cut (Fig. 3.6) to exclude the SNe with extreme values in the x_1-c plane. We adopt a similar cut as in Bazin et al. (2011) and Campbell et al. (2013). The ellipse we draw has semi-axes $a_{\rm x_1} = 3$, and $a_{\rm c} = 0.25$, centered at $(x_1, c) = (-0.2, 0)$.

[4] http://cosmologist.info/cosmomc/

Note that this cut excludes a higher fraction of non-Ia's than Ia's in our sample, and thus imporves the purity of the final sample (99.7%).

We also require the photo-z to be greater than 0.2, since the low-z SNe are already excluded in the general-function-fit step before classification; we require the photo-z error to be less than 0.1.

Our final sample for cosmological analysis has a total number of 12618 SNe including 12586 (99.7%) SNe Ia and 32 (0.3%) Core-collapse SNe. We show the marginalized means of the parameters in Table 3.3. The Hubble Diagram is shown in Fig. 3.7.

3.6 Summary and discussion

We have developed a method for SN classification using their colors and parameters from a general function fit of their light curves, utilizing the Random Forest classification algorithm. Our method is independent of the SN models, and make no use of redshift information of the SN or its host galaxy. We have achieved performance comparable to other photometric classification methods. By varying the probability threshold we are able to obtain samples with different purity as needed. A sample with 99% purity is chosen for our cosmological constraints study in this work.

We have obtained photo-z's for our photometric sample by fitting the SN light curves using the extended SALT2 model, with the nested sampling method. We show that initial conditions and parameter limits need to be set very carefully in order to obtain good results, especially when no host-galaxy prior is given. Our photometric

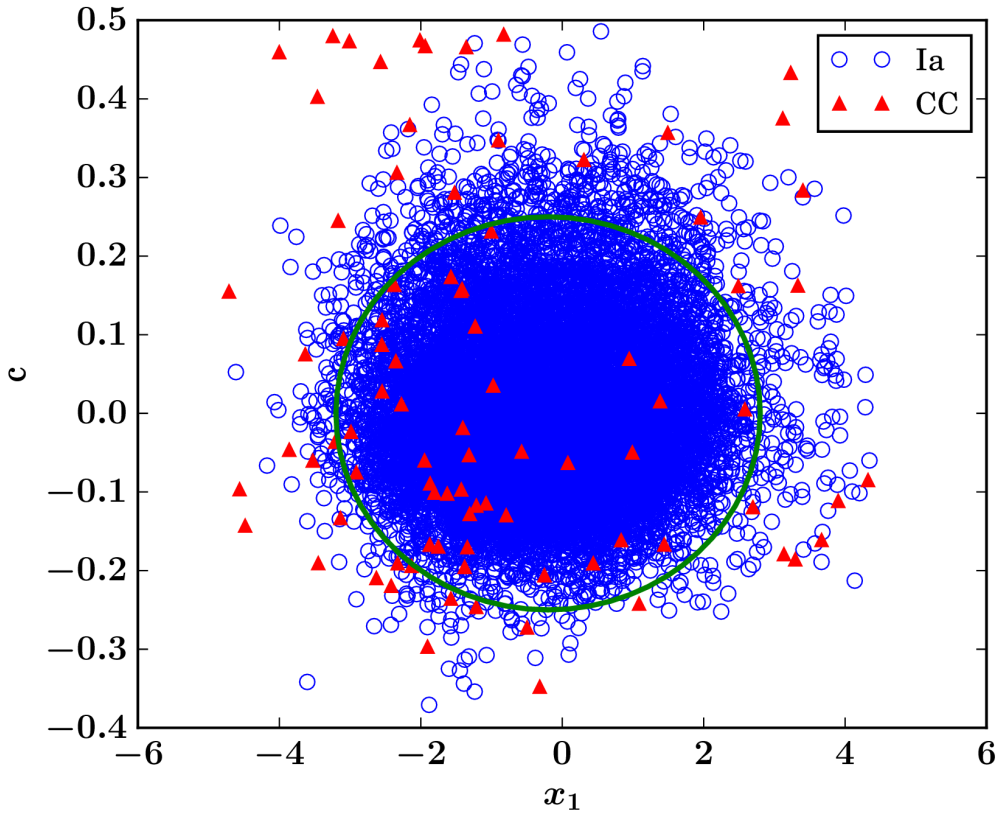

Figure 3.6: Ellipse cut for SALT2 parameters x_1 and c.

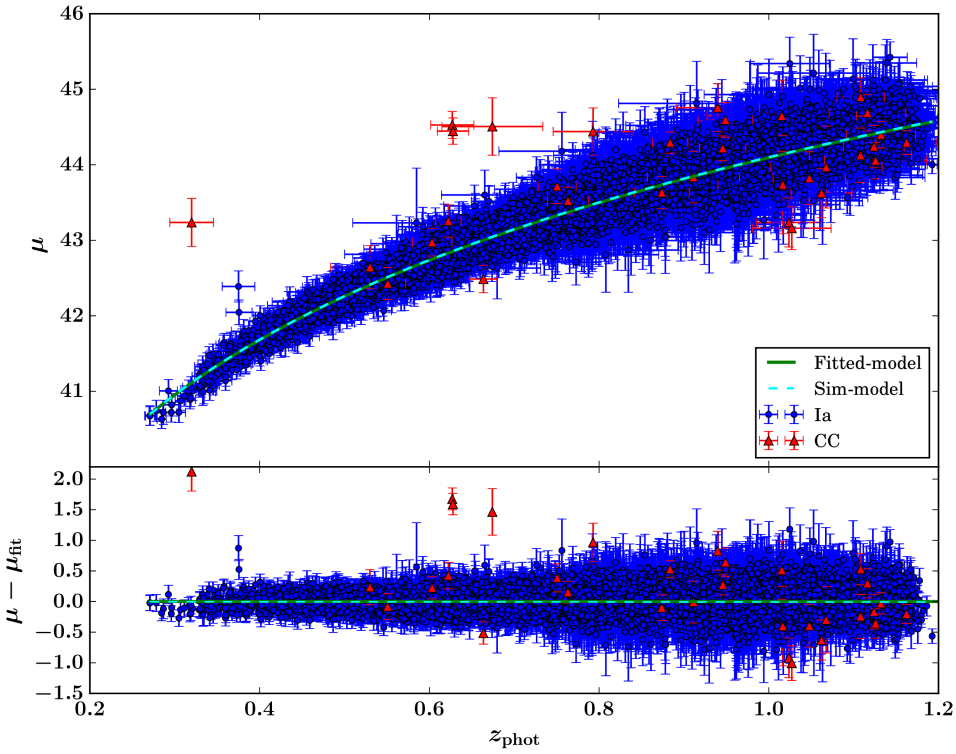

Figure 3.7: Hubble Diagram of our photometric SN Ia sample derived without using host-galaxy photo-z prior. Blue dots are true SNe Ia, red triangles are Core-collapse SNe that are classified as Ia's. Green solid line is the fitted cosmology and cyan dashed line is the simulated cosmology.

Table 3.3: Simulation input parameter values and the marginalized means of the cosmological parameters obtained using the photometric SN Ia sample derived without using host-galaxy photo-z prior.

	Ω_m	α	β	$\mathcal{M}$	χ^2	d.o.f
Sim input	0.30	0.135	3.19	13.43	-	-
$\sigma_{\rm int} = 0.1$	0.299 ± 0.008	0.119 ± 0.001	3.32 ± 0.02	13.43 ± 0.01	13795.1	12614
$\sigma_{\rm int} = 0.11$	0.305 ± 0.008	0.117 ± 0.001	3.30 ± 0.02	13.43 ± 0.01	12668.9	12614
$\sigma_{\rm int} = 0.12$	0.312 ± 0.009	0.116 ± 0.001	3.28 ± 0.02	13.44 ± 0.01	11663.3	12614

redshifts have a mean bias ($\left\langle z_{\rm phot} - z_{\rm spec} \right\rangle$) of 0.0120 with $\sigma \left(\frac{z_{\rm phot} - z_{\rm spec}}{1+z_{\rm spec}} \right) = 0.0294$ without using a host-galaxy photo-z prior, and a mean bias ($\left\langle z_{\rm phot} - z_{\rm spec} \right\rangle$) of 0.0017 with $\sigma \left(\frac{z_{\rm phot} - z_{\rm spec}}{1+z_{\rm spec}} \right) = 0.0116$ using a host-galaxy photo-z prior.

We obtained a final photometric sample with further cuts on the photo-z errors and the light curve parameters from the SALT2 fit. Using our final photometric SN Ia sample derived without using host-galaxy photo-z prior, and assuming a flat ΛCDM model, we obtain a measurement of Ω_m of 0.305 ± 0.008 after bias correction, with statistical errors only and the intrinsic scatter set to $\sigma_{\rm int} = 0.11$. The fitted Ω_m is consistent with our simulations ($\Omega_m = 0.3$). The fitted value varies a little with different choice of the intrinsic scatter terms. We do not expect the bias correction to be sensitive to the cosmological model assumed for the "bias correction" simulation set, based on the fact that the bias correction calculated from the separate "bias correction" set assuming a different cosmology from our original simulation set works well on the original simulation set. We will investigate this in further detail in future work. With the small statistical uncertainty due to the large sample size, the study of the systematic effects becomes more important. Here we aim at showing the capability of constraining cosmology using the photometric sample. We will leave the systematic studies for future work.

Chapter 4

Conclusion

With the ongoing and planned surveys, SN cosmology is entering a new era, where the data sample increases dramatically in size and quality, and spreads in a wider redshift range. The systematic uncertaianties will become dominant over the statistical uncertainties. It becomes more and more important to better characterize the systematic effects in SN cosmology. On the other hand, the study on better utilizing the photometric information in the absence of spectroscopic follow ups is also crucial. This dissertation has focused on both developing analysis method in order to better characterize the systematics in SN cosmology, and on investigating the use of photometric SN data only to determine the cosmological constraints.

In Chapter 2, we have applied the MCMC method in the SN light curve fitting and developed a method for sampling the SN light curve parameter pdf's in cosmological analysis. Our method shows that the cosmological constraints are shifted when using the sampled pdf's in the cosmological fitting compared to using the best-fit parameters from a maximum likelihood method. Our results also indicate that the cosmological constraint could be biased by including the problematic SNe with multiple peak pdf's. It would be useful to further investigate the cause of the multiple peak pdf's and develop better selection cuts to exclude them from a cosmological analysis. It will also be useful to utilize advanced statistical techniques to include the whole SN light curve pdf in the cosmological analysis.

In Chapter 3, we have shown that we can obtain a photometric SN sample with

99% purity using Machine Learning algorithms to classify the mixed type SNe from LSST, utilizing the SN colors and parameters from a general function fit of the SN light curves. We have also demonstrated that we can obtain photo-z estimation with high precision, with the correct error estimation. We have also shown that cosmological parameters can be measured from our photometric sample with high precision. In this dissertation, we have focused on photometric classification and photo-z estimation, and only considered statistical errors in fitting the Hubble Diagram. It is important to study the systematic effects for the photometric sample with contaminations, and develop methods to take into account the systematics due to both contaminations and light curve fitting without spectroscopic redshifts (either from SN spectra or host-galaxy spectra). We have seen that the light curve model uncertainty in SALT2 is not well-modeled in some wavelength ranges, which will bias the light curve fitting and the cosmological analysis from it. As more SNe with improved photometry and spectra are obtained, it would be useful to develop new light curve models to suit the needs for future analysis.

In addition to the two aspects that we have discussed in this dissertation, it is also important to study the intrinsic magnitude dispersions of the SNe Ia, which will rely on better understanding of the physics behind the SN Ia explosion.

SN cosmology remains an exciting field and there is a lot of challenge ahead to uncover the mystery of dark energy in the coming decades. I plan to continue my quest for using SNe Ia as dark energy probes, which will involve utilizing advanced statistical techniques and machine learning in the cosmological analysis, developing new SN models, and understanding the intrinsic differences in SNe Ia.

References

Astier, P., Guy, J., Regnault, N., et al. 2006, A&A, 447, 31

Barbary, K., Barclay, T., Biswas, R., et al. 2016, SNCosmo: Python library for supernova cosmology, Astrophysics Source Code Library, ascl:1611.017

Bazin, G., Palanque-Delabrouille, N., Rich, J., et al. 2009, A&A, 499, 653

Bazin, G., Ruhlmann-Kleider, V., Palanque-Delabrouille, N., et al. 2011, A&A, 534, A43

Bernstein, J. P., Kessler, R., Kuhlmann, S., et al. 2012, ApJ, 753, 152

Betoule, M., Kessler, R., Guy, J., et al. 2014, A&A, 568, A22

Branch, D. 1981, ApJ, 248, 1076

Breiman, L. 2001, Machine Learning, 45, 5

Brooks, S. P., & Gelman, A. 1998, Journal of computational and graphical statistics, 7, 434

Campbell, H., D'Andrea, C. B., Nichol, R. C., et al. 2013, ApJ, 763, 88

Conley, A., Guy, J., Sullivan, M., et al. 2011, ApJS, 192, 1

Contreras, C., Hamuy, M., Phillips, M. M., et al. 2010, AJ, 139, 519

Dai, M., & Wang, Y. 2016, MNRAS, 459, 1819

Dilday, B., Kessler, R., Frieman, J. A., et al. 2008, ApJ, 682, 262

Folatelli, G., Phillips, M. M., Burns, C. R., et al. 2010, AJ, 139, 120

Frieman, J. A., Turner, M. S., & Huterer, D. 2008, ARA&A, 46, 385

Ganeshalingam, M., Li, W., & Filippenko, A. V. 2013, MNRAS, 433, 2240

Gupta, R. R., Kuhlmann, S., Kovacs, E., et al. 2016, AJ, 152, 154

Guy, J., Astier, P., Baumont, S., et al. 2007, A&A, 466, 11

Guy, J., Sullivan, M., Conley, A., et al. 2010, A&A, 523, A7

Hamuy, M., Phillips, M. M., Suntzeff, N. B., et al. 1996a, AJ, 112, 2398

—. 1996b, AJ, 112, 2408

Hicken, M., Wood-Vasey, W. M., Blondin, S., et al. 2009, ApJ, 700, 1097

Hlozek, R., Kunz, M., Bassett, B., et al. 2012, ApJ, 752, 79

Hubble, E. 1929, Proceedings of the National Academy of Science, 15, 168

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

Jha, S., Kirshner, R. P., Challis, P., et al. 2006, AJ, 131, 527

Jones, D. O., Scolnic, D. M., Riess, A. G., et al. 2016, ArXiv e-prints

Kaiser, N., Burgett, W., Chambers, K., et al. 2010, in Proc. SPIE, Vol. 7733, Ground-based and Airborne Telescopes III, 77330E

Kessler, R., Bernstein, J. P., Cinabro, D., et al. 2009, PASP, 121, 1028

Kessler, R., Cinabro, D., Bassett, B., et al. 2010, ApJ, 717, 40

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

Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, ArXiv e-prints

Lewis, A., & Bridle, S. 2002, Phys. Rev. D, 66, 103511

Lochner, M., McEwen, J. D., Peiris, H. V., Lahav, O., & Winter, M. K. 2016, ApJS, 225, 31

LSST Science Collaboration, Abell, P. A., Allison, J., et al. 2009, ArXiv e-prints

March, M. C., Trotta, R., Berkes, P., Starkman, G. D., & Vaudrevange, P. M. 2011, MNRAS, 418, 2308

Miknaitis, G., Pignata, G., Rest, A., et al. 2007, ApJ, 666, 674

Möller, A., Ruhlmann-Kleider, V., Leloup, C., et al. 2016, J. Cosmology Astropart. Phys., 12, 008

Mosher, J., Guy, J., Kessler, R., et al. 2014, ApJ, 793, 16

Narayan, G., Rest, A., Tucker, B. E., et al. 2016, ApJS, 224, 3

Palanque-Delabrouille, N., Ruhlmann-Kleider, V., Pascal, S., et al. 2010, A&A, 514, A63

Perlmutter, S., Aldering, G., Goldhaber, G., et al. 1999, ApJ, 517, 565

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

Phillips, M. M., Lira, P., Suntzeff, N. B., et al. 1999, AJ, 118, 1766

Pskovskii, I. P. 1977, sovast, 21, 675

Rest, A., Scolnic, D., Foley, R. J., et al. 2014, ApJ, 795, 44

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

Riess, A. G., Filippenko, A. V., Challis, P., et al. 1998, AJ, 116, 1009

Riess, A. G., Kirshner, R. P., Schmidt, B. P., et al. 1999, AJ, 117, 707

Riess, A. G., Strolger, L.-G., Casertano, S., et al. 2007, ApJ, 659, 98

Sako, M., Bassett, B., Becker, A. C., et al. 2014, ArXiv e-prints

Spergel, D., Gehrels, N., Baltay, C., et al. 2015, ArXiv e-prints

Stritzinger, M. D., Phillips, M. M., Boldt, L. N., et al. 2011, AJ, 142, 156

Suzuki, N., Rubin, D., Lidman, C., et al. 2012, ApJ, 746, 85

The Dark Energy Survey Collaboration. 2005, ArXiv Astrophysics e-prints, astro-ph/0510346

Tripp, R. 1998, A&A, 331, 815

Wang, Y. 2007, ApJ, 654, L123

Wang, Y., Gjergo, E., & Kuhlmann, S. 2015, MNRAS, 451, 1955

Wang, Y., Narayan, G., & Wood-Vasey, M. 2007, MNRAS, 382, 377

Wood-Vasey, W. M., Miknaitis, G., Stubbs, C. W., et al. 2007, ApJ, 666, 694

Appendix A

Details of the quality cuts for SN classification

We present the numbers of SNe remaining after each quality cut in Section 3.3.3 in detail in Table A.1 and A.2.

Table A.1: Summary of number of SNe remaining for each type after each quality cut

	Ia	II	Ibc
Total number of SNe before any cuts	199400 (1)	1941000 (1)	
Max SNR > 5 for 3 bands	62147 (0.31)	67631 (0.035)	14468 (0.007)
1 point before and 2 after the peak[a] for 3 bands, 1 of which has max SNR> 5	48298 (0.24)	54900 (0.028)	11468 (0.006)
Bazin fit success[b] (all 6 bands)	48159 (0.24)	52342 (0.027)	11311 (0.006)
Bazin parameter cuts[c]	26616 (0.13)	7960 (0.004)	4354 (0.002)
Final fraction[d]	0.684	0.204	0.112

[a] Here peak refers to the highest flux point in the raw light curve whose SNR is greater than the median SNR of that band.

[b] Here success refers to any fit that returns a set of values (does not return a failure by the curvefit program), whether they are in a reasonable range or not.

[c] Detailed in Table A.2.

[d] Fraction of types in the final sample (added up to 1).

[e] Numbers in parentheses indicate the fractions.

Table A.2: Number of SNe remaining for each type after each Bazin parameter cut

	Ia	II	Ibc
After pre-fit cuts	48159 (1)	52342 (1)	11311 (1)
$t_{\text{rise}} > 1$, and t_{rise} not close to 1 with tolerance = 0.01	40337 (0.84)	34162 (0.65)	8320 (0.74)
$-20 < B < 20$	38317 (0.80)	17439 (0.33)	6228 (0.55)
$\chi^2/d.o.f < 10$	36041 (0.75)	16308 (0.31)	5995 (0.53)
$t_{\text{fall}} < 150$	33580 (0.70)	13038 (0.25)	5742 (0.51)
$t_{\text{rise}} < t_{\text{fall}}$	31248 (0.65)	11201 (0.21)	5219 (0.46)
$A < 5000$	31191 (0.65)	11172 (0.21)	5207 (0.46)
$A_{\text{err}} < 100$	27806 (0.58)	9612 (0.18)	4627 (0.41)
$t_{0,\text{err}} < 50$	27545 (0.57)	9121 (0.17)	4567 (0.40)
$t_{\text{fall,err}} < 100$	26826 (0.56)	8078 (0.15)	4390 (0.39)
$t_{\text{rise,err}} < 50$	26616 (0.55)	7961 (0.15)	4354 (0.38)
$A(Y), A(u) < 1000$	26616 (0.55)	7960 (0.15)	4354 (0.38)

† Numbers in parentheses indicate the fractions.