**Title**
Improving Cosmological Distance Measurements with Type Ia Supernovae: From Pixels to Dark Energy



**Author**
Boone, Kyle Robert

**Publication Date**
2019

Peer reviewed|Thesis/dissertation

# Improving Cosmological Distance Measurements with Type Ia Supernovae: From Pixels to Dark Energy

by

Kyle Boone

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Physics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Saul Perlmutter, Chair
Professor Joshua Bloom
Associate Professor Daniel Kasen

Summer 2019

Improving Cosmological Distance Measurements with Type Ia Supernovae: From Pixels to Dark Energy

## Abstract

Improving Cosmological Distance Measurements with Type Ia Supernovae: From Pixels to Dark Energy

by

Kyle Boone

Doctor of Philosophy in Physics

University of California, Berkeley

Professor Saul Perlmutter, Chair

In the late 1990s, precise distance measurements with Type Ia supernovae (SNe Ia) were used to show that the expansion of the universe is accelerating. One possibility is that this accelerated expansion is due to an additional form of energy referred to as "dark energy" which contributes roughly 70% of the total energy in the present day universe. The properties of dark energy are not currently well-constrained, and a wide range of different cosmological probes are currently being designed to explore the fundamental nature of the accelerated expansion of the universe. SNe Ia have remained one of the strongest cosmological probes, and upcoming experiments such as the Large Synoptic Survey Telescope (LSST) are expected to discover over 100,000 SNe Ia that can be used for cosmology. The uncertainties on cosmological parameters derived from these large samples of SNe Ia will be entirely dominated by the systematic uncertainties of distance measurements to SNe Ia. In this dissertation, we discuss several different methods of improving the systematic uncertainties in distance measurements to SNe Ia.

This dissertation is split into three main chapters each discussing how to improve a different aspect of distance measurements to SNe Ia. In Chapter 2, we examine how instrumental calibration can affect these distance measurements, and discuss a new anomalous behaviour of CCD readout electronics related to the binary encoding of pixel values that affects most astronomical instruments currently in use. For the Nearby Supernova Factory, this anomaly introduces a dispersion in the measured B-band/U-band magnitudes of 0.11 mag/0.51 mag for the faintest 20% of measurements.

Another major source of systematic uncertainty in distance measurements to SNe Ia is intrinsic variation of the SNe Ia. In Chapter 3, we develop a new method of parametrizing SNe Ia using manifold learning to generate a non-linear decomposition of the intrinsic diversity of their spectra near maximum light. We identify regions of the parameter space of SNe Ia where previous standardization methods such as SALT2 have biases of up to 0.3 mag, and show how correlations between host galaxy properties and distance estimates are greatly reduced when standardizing SNe Ia using our new parametrization.

Finally, in Chapter 4, we discuss how upcoming surveys such as LSST will need to rely on photometric classification to identify the majority of the transients that they discover, which means that samples of SNe Ia used for cosmology will be contaminated with other types of transients. We developed a set of techniques for photometric classification to address the fact that spectroscopic subsamples used for training classifiers are typically highly biased compared to the full samples of transients and variables that will be discovered. Using these techniques, we built a photometric classifier that won the PLAsTiCC photometric classification challenge out of 1,094 competing teams.

*To my parents, Tom and Renée,*
*and my wife, Amy,*
*for always supporting me.*

# Contents

**Contents** **ii**

**List of Figures** **iv**

**List of Tables** **vii**

**1 Introduction** **1**
1.1 Cosmology and the Evolution of the Universe . . . . . 1
1.2 The Hubble Constant Discrepancy . . . . . 2
1.3 Beyond ΛCDM . . . . . 4
1.4 Measurements of Cosmological Parameters with Standard Candles . . . . . 4
1.5 Type Ia Supernovae as Standardizable Candles . . . . . 5
1.6 Calibrating observations of Type Ia supernovae . . . . . 7
1.7 Photometric classification of SNe Ia . . . . . 9
1.8 Gaussian process regression . . . . . 11

**2 The Binary Offset Effect** **13**
2.1 Overview . . . . . 13
2.2 Identifying the Binary Offset Effect . . . . . 14
2.3 Modeling and Correcting the Offset for Active Instruments . . . . . 25
2.4 Impact of the Binary Offset Effect on Scientific Results . . . . . 28
2.5 Conclusions . . . . . 31
2.6 Appendix: Modeling the Binary Offset Effect in SNIFS Data . . . . . 32
2.7 Appendix: Code used to identify the Binary Offset Effect in arbitrary instruments. . . . . 36

**3 Mapping the Diversity of Type Ia Supernovae with Manifold Learning** **43**
3.1 Overview . . . . . 43
3.2 Dataset . . . . . 44
3.3 Estimating the Spectra of SNe Ia at Maximum Light . . . . . 49
3.4 Reading Between the Lines . . . . . 63
3.5 Manifold Learning on Spectra of SNe Ia . . . . . 68

3.6 Standardization of Type Ia Supernovae With Manifold Learning . . . . . 85
3.7 Discussion . . . . . 98

**4 Photometric Classification of Astronomical Transients with Gaussian Process Augmentation 112**
4.1 Overview . . . . . 112
4.2 Dataset . . . . . 113
4.3 Methods . . . . . 120
4.4 Results . . . . . 131
4.5 Discussion . . . . . 138
4.6 Appendix: Implementation of the training set augmentation . . . . . 142

**5 Conclusions 148**

**6 Acknowledgements 150**

**Bibliography 154**

# List of Figures

2.1 Count deficits for short exposures with the SNIFS instrument . . . . . 15
2.2 Example of the binary offset effect on the SNIFS blue channel . . . . . 16
2.3 Alternative example of the binary offset effect on the SNIFS blue channel . . . . 18
2.4 Deficit seen for stacked cosmic rays on the SNIFS blue channel . . . . . 19
2.5 "Blue step" on the SNIFS blue channel . . . . . 21
2.6 Examples of the binary offset effect in instruments other than SNIFS . . . . . 24
2.7 Results of the binary offset effect model for short exposures with the SNIFS instrument . . . . . 27
2.8 Example of the corrected binary offset effect . . . . . 28
2.9 Example of how the binary offset effect corretion removes the "blue step" for the SNIFS blue channel . . . . . 29

3.1 SALT2 residual outliers as a function of SALT2 phase . . . . . 46
3.2 SALT2 residual outliers as a function of the statistical S/N in the bluest 500 Å of the spectrum . . . . . 47
3.3 Comparison of the spectral evolution of different SNe Ia near maximum light . . 50
3.4 Estimated spectrum of PTF13ayw at maximum light . . . . . 53
3.5 Estimated differences between the observed spectra of PTF13ayw and its spectrum at maximum light . . . . . 54
3.6 Residuals of the spectral evolution of different SNe Ia near maximum light, after model correction . . . . . 55
3.7 Estimated spectrum of SN2004gc at maximum light . . . . . 56
3.8 Model of the differential spectral evolution near maximum light . . . . . 57
3.9 Model of the spectral evolution near maximum light for different supernova stretch values . . . . . 58
3.10 Recovered gray offsets as a function of the phase of a spectrum . . . . . 59
3.11 Model uncertainties for the interpolation of spectra to maximum light . . . . . 61
3.12 Uncertainties on the estimated spectra at maximum light for a selection of supernovae . . . . . 62
3.13 Comparison of the diversity of spectra at maximum light before and after dereddening . . . . . 66
3.14 Intrinsic dispersion of SNe Ia at maximum light . . . . . 67

3.15 Comparison of the uncertainties on the spectra at maximum light to the recovered intrinsic dispersion of SNe Ia . . . . . 69
3.16 Explained variance for each of the Isomap components . . . . . 72
3.17 Recovery of the F15 spectral "twinness" for different numbers of Isomap components . . . . . 74
3.18 Fraction of non-twins confused as twins for different numbers of Isomap components . . . . . 75
3.19 Effect of different Isomap components on the spectra of SNe Ia at maximum light 76
3.20 Effect of different Isomap components on the spectra of SNe Ia at maximum light, between 3300 and 4500 Å, highlighting the Ca II H&K feature. . . . . 77
3.21 Effect of different Isomap components on the spectra of SNe Ia at maximum light, between 4900 and 6700 Å, highlighting the Si II 5972 Å and 6355 Å features. . . 78
3.22 Effect of different Isomap components on the spectra of SNe Ia at maximum light, between 7200 and 8600 Å, highlighting the Ca II IR triplet features. . . . . 79
3.23 Branch classifications for our spectra of SNe Ia at maximum light . . . . . 80
3.24 Comparison of the Branch classifications to the Isomap embedding . . . . . 81
3.25 Comparison of spectra of Core Normal supernovae that are separated by the Isomap embedding. . . . . 82
3.26 Comparison of the SALT2 $x_1$ parameter to the Isomap embedding . . . . . 83
3.27 Predictions of the SALT2 $x_1$ parameter from the Isomap embedding . . . . . 84
3.28 Spectra of different Type Iax supernovae that are outliers in the Isomap embedding 85
3.29 Residual magnitudes estimated by the Reading Between the Lines algorithm . . 88
3.30 RBTL residual magnitudes as a function of the Isomap components 1 and 2 . . 89
3.31 RBTL residual magnitudes as a function of the Isomap components 1 and 3 . . 90
3.32 RBTL residual magnitudes as a function of the Isomap components 2 and 3 . . 90
3.33 GP predictions for the RBTL residual magnitudes as a function of the Isomap components 1 and 2 . . . . . 92
3.34 GP predictions for the RBTL residual magnitudes as a function of the Isomap components 1 and 3 . . . . . 93
3.35 GP predictions for the RBTL residual magnitudes as a function of the Isomap components 2 and 3 . . . . . 93
3.36 Difference in brightness for pairs of SNe Ia as a function of their twinness . . . . 94
3.37 Comparison of SALT2 corrected residual magnitudes to the first Isomap component 97
3.38 GP predictions for the SALT2 residual magnitudes as a function of the Isomap components 1 and 2 . . . . . 99
3.39 Differences between the SALT2 + Isomap standardization and traditional SALT2 standardization . . . . . 100
3.40 Reproduction of the host correlations seen in Rigault et al. (2018) for the SNe Ia in our analysis . . . . . 104
3.41 Host correlations for SALT2 + $x_1$ standardization for the subset of SNe Ia in common to all of our analyses . . . . . 106

3.42 Host correlations for SALT2 + Isomap standardization for the subset of SNe Ia in common to all of our analyses . . . . . 107
3.43 Host correlations for RBTL + Isomap standardization for the subset of SNe Ia in common to all of our analyses . . . . . 108

4.1 Comparison of the redshift distributions for the training and test sets in the PLAsTiCC dataset . . . . . 116
4.2 Comparison of the fraction of objects belonging to each extragalactic object type as a function of redshift . . . . . 117
4.3 Examples of GP models for SN Ia light curves . . . . . 123
4.4 Example of a feature that probes the fit quality . . . . . 129
4.5 Relative importance of each feature for classification . . . . . 132
4.6 Confusion matrix for the flat-weighted classifier. . . . . 133
4.7 Completeness of the sample of SNe Ia in the test set as a function of redshift for a fixed overall sample purity of 95%. . . . . 134
4.8 Completeness of the sample of SNe Ia in the WFD sample of the test set as a function of redshift for a fixed overall sample purity of 95% when different biases have been introduced in the training set. . . . . 137
4.9 Examples of augmented light curves for a Mira variable . . . . . 145
4.10 Examples of augmented light curves for a single lens microlensing event. . . . . 146
4.11 Examples of augmented light curves for a superluminous supernova light curve. 147

# List of Tables

2.1 Size of the binary offset effect in different instruments . . . . . 23
2.2 Parameters of the binary offset effect model for the different SNIFS amplifiers . 34

3.1 SNfactory synthetic filter definitions . . . . . 45
3.2 Sample selection requirements for the manifold learning analysis . . . . . 48
3.3 Effect of different components of the model of spectral evolution near maximum light . . . . . 63
3.4 Gaussian process standardization parameters . . . . . 91
3.5 SALT2 standardization parameters . . . . . 96
3.6 Hybrid Isomap and SALT2 standardization parameters . . . . . 98
3.7 Comparison of standardization results for different methods and samples . . . . 101
3.8 Measured host-galaxy property step sizes for different standardization methods . 105

4.1 Summary of the object types included in the PLAsTiCC simulations . . . . . 114
4.2 Confusion matrix for classification of a single object type out of a larger sample of other object types . . . . . 120
4.3 Overview of features used for classification . . . . . 125
4.4 Optimized hyperparameter values used for the LightGBM model. . . . . 130
4.5 Classifier performance on various metrics . . . . . 135

## Acknowledgments

Working as part of the Supernova Cosmology Project for the past six years has been an incredibly rewarding experience. I am deeply grateful to my advisor **Saul Perlmutter**, whose unrivaled enthusiasm for science motivated me to enter the field of cosmology that I was totally unfamiliar with at the start of my graduate student career. Saul's guidance has been invaluable my development as a scientist, and I am very thankful to have had the opportunity to explore and participate in a wide range of different projects with him.

I owe a lot to **Brian Hayden** who was a close mentor to me for the crucial first few years of graduate school. I'll never forget the time that we spent building and running the SeeChange pipeline together, and the many hours that we spent discussing anything and everything. **Greg Aldering** has been a constant source of inspiration, and he taught me most of what I now know about astronomy. Thank you, Greg, for putting up with all of my naive questions during those long nights of observing!

The work described in this dissertation wouldn't have been possible without the expertise and help of a large number of people. Thank you to **Sam Dixon** who contributed in major ways to almost everything that I worked on. Much of the work in this dissertation wouldn't have been possible without **Kara Ponder**'s tireless work on the SNfactory, and she has also been a great officemate! **Ravi Gupta** has provided detailed feedback on almost everything that I have ever written, and he has taught me a lot about how to communicate scientifically. Thank you to **David Rubin** for introducing me to Bayesian statistics (and insisting that I always use Stan) and for the many discussions about crazy cosmology ideas that we have had over the years. I am very grateful to **Xiaosheng Huang** and **Mark Strovink** who both provided essential mentorship and guidance throughout my graduate student career. Many thanks to the rest of the group, **Alex Kim**, **Clare Saunders**, **Maria Vincenzi**, **Kyle Barbary**, **Jakob Nordin**, **Caroline Sofiatti**, **Mike Yen**, and **Kyle Luther** among others who constantly provided helpful suggestions and useful feedback for my work.

Thank you to everyone on my qualifying exam and dissertation committees, **Saul Perlmutter**, **Dan Kasen**, **Josh Bloom**, and **Bill Holzapfel**, who provided lots of feedback and guidance throughout my graduate studies. I very much appreciated having **Tony Spadafora** around, he always seems to know what to do when something is going wrong! **Bryan Abshier** and **Jeff Anderson** kept all of our computing machines running smoothly, and were quick to respond whenever I needed help or figured out how to break a server.

My parents, **Tom** and **Renée**, and my siblings, **Spencer** and **Jocelyne**, have all been incredibly supportive of me over the years, and pushed me to pursue my passions. Finally, graduate school wouldn't have been the same without the support of my amazing wife **Amy Tresenrider**. She has been a wonderful partner, keeping me sane throughout the ups and downs of graduate school while also working on a Ph.D. herself. Amy, you are my best friend, and I don't know how I would have finished this Ph.D. without you.

# Chapter 1

# Introduction

## 1.1 Cosmology and the Evolution of the Universe

One of the main goals of cosmology is to understand how the size of the universe evolves over time. In an isotropic and homogeneous universe, the relative size of the universe can be parameterized using the scale factor $a(t)$. This scale factor is defined to be one at the present time, and has increased over time from an initial value after the Big Bang approaching zero. The expansion of the universe can be seen directly by measuring the recessional velocities of extragalactic objects. For not-too-distant objects, these recessional velocities are proportional to the distances to these objects, and the constant of proportionality is labeled the "Hubble parameter". The Hubble parameter can be written in terms of the scale factor as $H = \dot{a}/a$ where $\dot{a}$ is the derivative of $a$ with respect to time. The time evolution of the Hubble parameter is uniquely determined using the Friedmann equation derived from General Relativity if the energy density of the universe $\rho$ and the curvature $k$ are known:

$$H^2 = \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G\rho}{3} - \frac{kc^2}{a^2} \tag{1.1}$$

Here $G$ is Newton's gravitational constant, and $c$ is the speed of light.

In this equation, the time evolution of the Hubble parameter depends on how the energy density $\rho$ evolves. This, in turn, depends on the composition of the universe. The energy density of relativistic particles or radiation scales as $a^{-4}$, while the energy density of matter scales as $a^{-3}$. A cosmological constant introduces a constant energy density regardless of the scale factor. For a universe consisting of only radiation, matter, curvature and a cosmological constant, the Friedmann equation can then be rewritten by defining the "critical density":

$$\rho_c = \frac{3H^2}{8\pi G} \tag{1.2}$$

and by defining the density parameter for each component of the universe:

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

We also define the Hubble constant $H_0$ to be the value of the Hubble parameter at the present time ($t = 0$), resulting in the following formula:

$$\frac{H^2}{H_0^2} = \Omega_{R,0}a^{-4} + \Omega_{M,0}a^{-3} + \Omega_{k,0}a^{-2} + \Omega_{\Lambda,0} \tag{1.4}$$

The density parameters $\Omega_{R,0}$, $\Omega_{M,0}$, $\Omega_{k,0}$, and $\Omega_{\Lambda,0}$ are effectively the fractional energy compositions of the universe at the present time from radiation, matter, curvature, and a cosmological constant, respectively, satisfying:

$$\Omega_{R,0} + \Omega_{M,0} + \Omega_{k,0} + \Omega_{\Lambda,0} = 1 \tag{1.5}$$

With the separation of matter into baryonic and dark matter components, the introduction of initial density fluctuations (both of which do not affect the universe at large scales), and the assumption that $\Omega_{k,0} = 0$, this model is referred to as the ΛCDM model or the "standard model" of cosmology. Here Λ refers to the presence of dark energy in the form of a cosmological constant, and CDM refers to the presence of cold dark matter. The ΛCDM model has been shown to be a very good description of the universe on large scales. The same set of parameters can explain the behavior of many different cosmological probes such as the brightnesses of Type Ia supernovae, the length scales of baryon acoustic oscillations, the power spectrum of the cosmic microwave background, and the abundance of galaxy clusters (Planck Collaboration et al., 2016).

## 1.2 The Hubble Constant Discrepancy

One current source of discrepancy in cosmology is in measurements of the Hubble constant ($H_0$) that is effectively a measure of the expansion rate of the universe at the present time. A local measurement of $H_0$ can be made using distances to individual objects in the local universe. Currently, direct distance measurements can only be made for relatively nearby objects. For these objects, the peculiar velocities of galaxies overwhelm the signal of the expansion rate of the universe. To address this, we use the direct measurements to nearby objects to calibrate a sequence of increasingly distant indirect distance measurements that extend out into the "Hubble flow" where the expansion rate of the universe can be accurately measured. This procedure is referred to as the "cosmic distance ladder".

The current best local measurements of $H_0$ (Riess et al., 2016, 2019) first use a variety of direct distance estimates to nearby galaxies. These include using detached eclipsing binaries to measure the distance to the Large Magellanic Cloud to 1.2% (Pietrzyński et al., 2019), modeling masers to obtain a geometric distance to NGC 4258 accurate to 3.4% (Humphreys et al., 2013), and obtaining geometric distances to individual objects in the Milky Way using parallax. These measurements are then used to calibrate the period-luminosity relationship of Cepheid variables. The absolute luminosity of Type Ia supernovae (SNe Ia) is then calibrated by measuring the distances to Cepheid variables in a variety of galaxies that

hosted SNe Ia. Finally, SNe Ia can be observed well out into the Hubble flow, giving a measurement of $H_0$. This procedure results in a measurement of $H_0$ of $74.03 \pm 1.42$ km s$^{-1}$ Mpc$^{-1}$ (Riess et al., 2019).

Alternatively, the Hubble constant can be estimated by measuring cosmological parameters in the early universe from the cosmic microwave background (CMB). Under the assumption that the ΛCDM model holds, we can then infer what the current value of $H_0$ should be. Using measurements of the CMB from the Planck mission, this results in an inferred value for $H_0$ of $67.36 \pm 0.54$ km s$^{-1}$ Mpc$^{-1}$ (Planck Collaboration et al., 2018). The Planck results can be combined with various measurements of the scale of baryon acoustic oscillations (BAO) at lower redshifts, resulting in an inferred value for $H_0$ of $67.66 \pm 0.42$ km s$^{-1}$ Mpc$^{-1}$ (Planck Collaboration et al., 2018). These results show that there is a $4.4\sigma$ discrepancy between local measurements of $H_0$ and ones inferred from the CMB assuming ΛCDM, implying either that there is some unknown systematic issue with one of our cosmological probes, or that there is a need for additional physics beyond ΛCDM.

Several different methods have been introduced to provide alternative local measurements of $H_0$. Freedman et al. (2019) used Tip of the Red Giant Branch distances to calibrate the intrinsic luminosity of SNe Ia directly, bypassing the calibration of Cepheid variables, and find a local $H_0$ value of $69.8 \pm 0.8$ (stat.) $\pm 1.7$ (sys.) which is consistent with the Planck results to within $1.2\sigma$. Alternatively, time delays of gravitationally lensed quasars can be used to provide a geometric measurement of $H_0$ that is entirely independent of both the cosmic distance ladder and the CMB results. The latest analysis by the H0LiCOW team finds a value of $H_0 = 73.3^{+1.8}_{-1.7}$ km s$^{-1}$ Mpc$^{-1}$ (Wong et al., 2019) which is consistent with the local $H_0$ measurements. Finally, gravitational wave sources can be used as "standard sirens" to estimate their distances, providing another path to an independent measurement of $H_0$. With measurements of a single gravitational wave source GW170817, Hotokezaka et al. (2019) measured $H_0 = 70.3^{+5.3}_{-5.0}$ km s$^{-1}$ Mpc$^{-1}$, and estimated that with 15 more similar events they would have the sensitivity to distinguish between the Planck and cosmic distance ladder measurements of $H_0$.

Further work will need to be done to understand the source of the $H_0$ discrepancy and determine whether or not there is a need for additional physics beyond ΛCDM. SNe Ia are one of the keys probes for these measurements, and unknown or unmodeled systematics of SNe Ia could potentially be the source of the $H_0$ discrepancy. For example, as will be discussed in Section 1.5, the measured distances to SNe Ia are correlated with properties of their host galaxies. SNe Ia hosted in galaxies with Cepheid variables are predominately in locally star-forming environments. Rigault et al. (2015) showed that this biases the measured value of $H_0$, and find $H_0 = 70.6 \pm 2.6$ km s$^{-1}$ Mpc$^{-1}$ when this bias is taken into account. Understanding the source of the correlations of the distances to SNe Ia with host galaxy properties is essential for resolving the $H_0$ discrepancy.

## 1.3 Beyond ΛCDM

Many extensions of the ΛCDM model have been proposed that can be tested with observations. For example, Einstein's cosmological constant is only one of many different models that are consistent with our current observations of the accelerated expansion of the universe. These other models include among many others the introduction of a dynamic field, often referred to as "quintessence" (Ratra & Peebles, 1988; Zlatev et al., 1999), allowing gravity to leak into extra dimensions (Dvali et al., 2000) or modifications to general relativity (e.g. Capozziello & Fang, 2002; Carroll et al., 2004). To capture the effects of these other models of dark energy, the ΛCDM model can be extended with what is effectively a Taylor expansion of how dark energy scales with density (Chevallier & Polarski, 2001; Linder, 2003):

$$\Omega_{de} = \Omega_{de,0} a^{-3[1+w_0+w_a(1-a)]} \tag{1.6}$$

Different models of the accelerated expansion of the universe will result in different values of $w_0$ and $w_a$, with $w_0 = -1$ and $w_a = 0$ corresponding to a cosmological constant. Combining the results of analyses of the CMB, BAO, Type Ia supernovae and local $H_0$ measurements, the state-of-the-art constraints on these parameters are $w_0 = -1.007 \pm 0.089$ and $w_a = -0.222 \pm 0.407$, which remains consistent with a cosmological constant (Scolnic et al., 2018). All of these probes provide complementary constraints on these cosmological parameters, and improved measurements with all of them will be essential to constrain the properties of dark energy.

## 1.4 Measurements of Cosmological Parameters with Standard Candles

One method of probing cosmology is by using "standard candles": astronomical objects with known intrinsic luminosities. Assuming that we know the intrinsic luminosity $L$ of an astronomical object, the observed brightness $F$ can be used to infer the luminosity distance $d_L$ to that object:

$$d_L = \sqrt{\frac{L}{4\pi F}} \tag{1.7}$$

As light from the object travels through space, the wavelength of that light will be stretched, such that:

$$\lambda_{observed} = \lambda_{emitted}(1+z) \tag{1.8}$$

where the redshift $z$ is defined as:

$$a = \frac{1}{1+z} \tag{1.9}$$

If the restframe spectrum of the object is known, then we can compare the observed spectrum to the restframe spectrum to determine the redshift.

The relationship between the redshift and the luminosity distance can be calculated directly from a given set of cosmological parameters. For a flat universe ($\Omega_k = 0$), we have:

$$d_L(z) = c(1+z)\int_0^z \frac{dz'}{H(z')} \tag{1.10}$$

where $H(z')$ is given by the Friedmann equation shown in Equation 1.4. For the ΛCDM model, there is no closed form for $d_L(z)$, but it can easily be evaluated numerically. By observing many standard candles at different redshifts, we can constrain the functional form of $d_L(z)$ and thus the underlying cosmological parameters.

In practice, the observed brightnesses of astronomical objects are typically expressed in magnitudes, with the "apparent magnitude" $m$ for an object with an observed brightness $F$ given as:

$$m = -2.5\log_{10}(F) + \text{zero point} \tag{1.11}$$

where the zero point can be chosen arbitrarily. In this work we use "AB magnitudes" where the zeropoint is chosen so that an apparent magnitude of zero corresponds to 3,631 janskys. Similarly, the "absolute magnitude" $M$ is defined as the apparent magnitude for an object if it were located 10 pc away. We define the "distance modulus" $\mu$ to be the difference between an object's apparent and absolute magnitudes:

$$\mu = m - M = 5\log_{10}(d_L/10\text{ pc}) \tag{1.12}$$

## 1.5 Type Ia Supernovae as Standardizable Candles

Type Ia supernovae (SNe Ia) have proven to be one of the strongest probes of cosmology because of their ability to be used as standard candles. Using supernovae as standard candles for cosmological measurements has been suggested since the earliest studies of supernovae (Baade, 1938). However, due to the challenge of reliably discovering and following up SNe Ia with high quality observations, it took until the 1990s for light curves of SNe Ia to be obtained that were useful for cosmology (Branch & Tammann, 1992; Branch, 1998) and at high redshifts (Perlmutter et al., 1995). The first distance measurements with reasonably sized samples of high redshift SNe Ia led to the initial discovery of the accelerating expansion of the universe (Riess et al., 1998; Perlmutter et al., 1999). Subsequent studies have collected a sample of over 1,000 spectroscopically-confirmed SNe Ia, providing increasingly strong constraints on cosmological parameters (Knop et al., 2003; Riess et al., 2004; Astier et al., 2006; Kowalski et al., 2008; Suzuki et al., 2012; Betoule et al., 2014; Scolnic et al., 2018).

The ability to use SNe Ia as standard candles for cosmology depends on how well their brightnesses can be standardized. The brightnesses will be affected by both astronomical diversity in properties of SNe Ia, and uncertainties in the observations of SNe Ia. Unfortunately, SNe Ia are not perfect standard candles, and should instead be referred to as "standardizable candles": they have an intrinsic dispersion in their peak brightnesses in the

$B$-band of $\sim$0.4 mag. With this level of intrinsic dispersion, using Equation 1.12, the distance to an individual SN Ia can be measured with a fractional uncertainty of $\sim$20%, and the value of $H_0$ can be inferred with a similar fractional uncertainty.

To use SNe Ia as distance estimators for cosmology, several corrections need to be applied to their peak brightnesses. Phillips (1993) showed that the peak brightnesses of SNe Ia are tightly correlated with the rate of decline of their light curves, commonly referred to as the "stretch". The rate of decline, measured as the difference in magnitudes between the peak brightness of a supernova in the $B$-band and its brightness 15 days later, or $\Delta m_{15}(B)$, can therefore be used to apply a correction to distance estimates to SNe Ia. Riess et al. (1996) and Tripp (1998) showed that the color of the light curve, measured as the difference between the peak brightnesses in the B and V bands, is also highly correlated with the peak brightnesses of SNe Ia. By combining information from the stretch and color of an SN Ia, the dispersion in the corrected peak brightnesses of the SNe Ia is reduced to $\sim$0.15 mag. The SALT2 model of the spectral energy distribution of SNe Ia (Guy et al., 2007, 2010; Betoule et al., 2014) implements these two corrections, and is used to estimate distances to SNe Ia in most modern cosmological analyses. With SALT2 corrected peak brightnesses, we can measure the distances to individual SNe Ia with a fractional uncertainty of $\sim$7%.

To complicate matters, estimates of the peak brightnesses of SNe Ia corrected using only the stretch and color of the light curve have been shown to be correlated with various properties of the host galaxies of SNe Ia. These correlations are typically modeled as a "host step" where SNe Ia with a given host property below some threshold have a systematic offset in their measured distances compared to SNe Ia higher than this threshold. SALT2-standardized residual magnitudes have been shown to have "host steps" of $\sim$0.1 mag when comparing SNe Ia from host galaxies with different masses, metallicities, local colors, local star-formation rates or global star-formation rates (Kelly et al., 2010; Sullivan et al., 2010; Gupta et al., 2011; Hayden et al., 2013; Rigault et al., 2013, 2015, 2018; Roman et al., 2018). As galaxy properties evolve with redshift, correlations of the peak brightness of SNe Ia with their host galaxy properties must be well-understood to produce robust cosmological measurements. Host galaxy properties are simply a proxy for some uncaptured instrinsic variability of SNe Ia. Ideally, new standardization techniques could be developed that identify this intrinsic variability using properties of the SNe Ia themselves rather than properties of their host galaxies.

Several different techniques have been proposed to improve standardization of SNe Ia. One option is to add additional components to a linear model like SALT2. Saunders et al. (2018) built a seven component linear model (SNEMO) SNEMO that is capable of parametrizing additional diversity in the light curves of SNe Ia compared to SALT2, and standardizes SNe Ia to within $0.113 \pm 0.007$ mag. Alternatively, Léget et al. (2019, submitted to A&A) built a three-component linear model by first performing a PCA decomposition of the spectral features of SNe Ia near maximum light and then using the resulting PCA coefficients to build a linear SED model. Another option is to incorporate additional information beyond optical light curves into the standardization procedure. The brightnesses of SNe Ia in the NIR are less sensitive to the intrinsic diversity of SNe Ia and effects such as

astrophysical dust (Kasen, 2006): the corrected peak brightness of NIR light curves have intrinsic dispersions of $\sim 0.11$ mag (Krisciunas et al., 2004; Wood-Vasey et al., 2008; Mandel et al., 2011; Barone-Nugent et al., 2012; Burns et al., 2018; Stanishev et al., 2018).

Various attempts have been made to use the spectra of SNe Ia to standardize their brightnesses. Bailey et al. (2009) showed that spectral flux ratios at specific wavelengths can be used to standardize SNe Ia to within $0.125 \pm 0.011$ mag. Blondin et al. (2011) and Silverman et al. (2012) used various spectral features directly in their standardization and achieved dispersions of $0.143 \pm 0.020$ mag and $0.130 \pm 0.017$ mag, respectively. These previous methods only used specific features of the supernova spectrum for their classification. Fakhouri et al. (2015) introduced the first method to use the full spectrum simultaneously with "twin supernovae". In the twins methodology, the spectrum of a new supernova is compared to a large reference sample to find pairs of spectra with very similar spectral features that are called "twins". Standardization using twin supernovae resulted in a dispersion of $0.083 \pm 0.012$ mag for the sample of Fakhouri et al. (2015), although there is some evidence that not all twin supernovae have similar brightnesses (Foley et al., 2018).

Most previous standardization procedures have focused on either building a model of many different linear components to model the spectral energy distribution of SNe Ia, or applying linear corrections to the peak brightness based off of specific spectral features or measurements such as the light curve stretch. However, there are visible spectral sequences for SNe Ia that are highly non-linear (Nugent et al., 1995; Branch et al., 2006; Blondin et al., 2012; Siebert et al., 2019), and that cannot be captured in a low-dimensional space with linear models. Methods that do not depend on linear components, such as the twins method of Fakhouri et al. (2015), are necessary to identify these spectral sequences. Sasdelli et al. (2016) used deep learning model of the spectra of SNe Ia, and showed that a non-linear four parameter feature space (plus one dimension for time) can capture the spectral diversity of SNe Ia as well as a 15 component linear model. Rubin (2019) showed that the twins pairings from Fakhouri et al. (2015) are consistent with an underlying three to five parameter non-linear model. In Chapter 3, we expand on these techniques and develop a new non-linear parametrization of the spectra of SNe Ia at maximum light. We show that for some regions of this parameter space there are systematic biases of up to 0.3 mag for standardization methods such as SALT2. Standardization using this new non-linear parametrization of SNe Ia significantly reduces the correlations of distance residuals with host properties. Techniques such as this one are essential for reducing systematic biases in upcoming cosmological analyses with SNe Ia.

## 1.6 Calibrating observations of Type Ia supernovae

Even if SNe Ia were perfect standard candles, there are many systematic errors that can be introduced in the process of observing SNe Ia. Historically, statistical uncertainties have dominated cosmological analyses with SNe Ia due to the observed dispersion in brightness of SNe Ia (e.g. Conley et al., 2011, with 472 SNe Ia). However, recent surveys are beginning

to obtain observations of large enough numbers of SNe Ia for the statistical uncertainties to be comparable in size to the systematic uncertainties (e.g., Scolnic et al., 2018, with 1048 SNe Ia). Upcoming surveys, such as the Large Synoptic Survey Telescope (LSST; LSST Science Collaboration et al., 2009) are projected to obtain light curves for ∼100,000 SNe Ia (The LSST Dark Energy Science Collaboration et al., 2018), and the sensitivities of their cosmological analyses will be entirely dominated by systematic uncertainties.

One significant source of systematic uncertainties is at the detector level in the measurement of the observed flux levels of SNe Ia. Charge coupled devices (CCDs) have been the dominant astronomical detector for the past three decades. Ideally, light at the focal plane of the telescope is detected on a grid of pixels which each provide independent information about the incident light at their location. In practice, astronomical CCDs and the associated instruments are subject to many anomalies that introduce spurious signals, correlations between pixels or deviation from linear behavior. These anomalies will lead to errors in the derived scientific results if they are not understood and accounted for. There are several locations in the instrument where such anomalies can be introduced.

Many of the anomalies are related to imperfections in the production of the sensor. These include variations in the pixel areas, fringing due to variations in thickness of the CCD, "tree ring" patterns due to impurities in the production of the silicon wafers, manufacturing defects in the CCD electronics, and edge effects due to interactions with other components or guard rings. See Janesick (2001) or Stubbs (2014) for in-depth discussions of these effects. Localized contamination of the silicon can also lead to many undesirable effects, including hot pixels with high dark current and traps that interfere with the charge transfer of pixels read out later in the same column. Large traps can produce dead or hot columns, while smaller traps lead to charge transfer inefficiency (CTI). Baggett et al. (2012) discusses localized contamination in the HST WFC3 UVIS detectors and studies how it evolves with time.

Several anomalies arise due to normal instrument operations even with a defect-free CCD. For very bright sources, the CCD can saturate, leading to effects such as blooming where charge spills into neighboring pixels. In general, the presence of charge on the CCD distorts the nearby electric field. This distortion manifests itself as effects like the "brighter-fatter effect" where the widths of point-spread functions vary by up to 2% for bright and faint objects (Antilogus et al., 2014). It is also possible to accumulate charge on the CCD from sources other than the desired science target, such as cosmic rays which deposit charge as they pass through the CCD (Janesick, 2001), or optical reflections ("ghosts") that appear when light reflects in unintended ways off of external optical components such as filters (Brown & Lupie, 2004) or within the CCD itself.

Readout electronics can introduce another set of anomalies. Many readout systems are susceptible to pickup that appears as periodic oscillations in the measured values of pixels that are observed sequentially. This can introduce a herring-bone pattern across the CCD image (Jansen et al., 2003). Other potential artifacts of the readout electronics include undershooting after reading a bright pixel (Caldwell et al., 2010), crosstalk between the readouts from different amplifiers (Baggett et al., 2004), and biased/sticky bits in analog-to-digital converters (Robberto & Hilbert, 2005). Understanding how all of these effects interact with

scientific data is essential for proper calibration of the science output of an instrument. Techniques have been developed to correct for most of these effects, and data reduction pipelines typically endeavor to treat the ones that have a sizeable effect on the data taken by their targeted instrument.

While developing our manifold learning analysis that is discussed in Chapter 3, we discovered an artifact in the CCD electronic chain that introduced a significant amount of systematic uncertainty into our data. We identified crosstalk between the binary-coded output of the analog-to-digital converter in our CCD readout electronics and subsequently read-out pixels. We showed that this effect is not limited to our detector, and in fact is present in most CCDs currently used for astronomical observations. In Chapter 2, we discuss this CCD artifact, how to model and correct for it, and how scientific results will be impacted by it if not properly handled. This analysis was published in Boone et al. (2018).

## 1.7 Photometric classification of SNe Ia

With modern surveys, the discovery rate of SNe Ia is rapidly outpacing the growth of resources to acquire spectroscopic classifications. The Dark Energy Survey (DES; The Dark Energy Survey Collaboration, 2005) was projected to acquire spectroscopic classifications for only 20% of their sample of up to 4,000 SN Ia light curves (Bernstein et al., 2012). Similarly, the Pan-STARRS Medium Deep Survey (PS1; Kaiser et al., 2010) discovered over 5,000 likely supernovae, but only obtained spectroscopic classifications for 10% of this sample (Jones et al., 2017). Upcoming large-scale surveys such as the Large Synoptic Survey Telescope (LSST; LSST Science Collaboration et al., 2009) are projected to obtain light curves for $\sim$100,000 SNe Ia (The LSST Dark Energy Science Collaboration et al., 2018), and will almost certainly have spectroscopic classifications for a much smaller fraction of their full sample.

Cosmological analyses with photometrically-classified SNe Ia are complicated by the fact that there is contamination from other transients in the sample, such as Type Ib/c or Type II supernovae. These other transients do not have the same intrinsic luminosity as SNe Ia, and they will bias cosmological measurements if they are accidentally included in a cosmological analysis. In principle, unbiased cosmological parameters can be recovered from photometrically-classified samples of SNe Ia by using Bayesian methods to model the contamination of the non-SN Ia transients in the sample (Kunz et al., 2007; Hlozek et al., 2012; Rubin et al., 2015; Jones et al., 2017). The performance of these methods depends heavily on their ability to distinguish SNe Ia from other transients, so accurate photometric classifiers are essential to their operation.

There are several major challenges to designing a photometric classification algorithm. The light curves generated by surveys such as LSST are sparsely sampled, and the observations do not occur on regular time intervals. Observations occur in different bands, and only a subset of the bands are typically available on any given night. The uncertainties on the photometry are heteroskedastic, and some bands have much higher noise levels than others.

Currently, photometric classifiers are typically trained on a subset of light curves from the survey in question that have spectroscopic confirmation. Brighter, nearby transients are significantly easier to spectroscopically classify than fainter, more distant ones, so training sets for transient surveys will typically be highly biased towards bright, nearby objects.

To understand the performance and limitations of photometric classifiers for DES, the Supernova Photometric Classification Challenge (SNPhotCC, Kessler et al., 2010) was initiated. The organizers of this challenge produced a simulation of Type Ia, Ib, Ic and II supernovae observed with realistic DES observing conditions. The SNPhotCC dataset consists of a training set of 1,103 spectroscopically confirmed objects and a test set of 20,216 objects without spectroscopic confirmation. Participants were challenged to develop classifiers that could use the known labels of the training set to infer the types of objects in the test set.

A wide variety of models and techniques were developed for, or applied to, data from the SNPhotCC. The techniques that have been applied to photometric classification on this dataset include Bayesian template comparisons (Poznanski et al., 2007; Sako et al., 2011), diffusion maps with random forests (Richards et al., 2012), neural networks (Karpenka et al., 2013), kernel PCA with nearest neighbours (Ishida & de Souza, 2013), convolutional neural networks (Pasquet et al., 2019), and deep recurrent neural networks (Charnock & Moss, 2017). Lochner et al. (2016) compared the performance of several different machine learning algorithms on the SNPhotCC dataset, and found that fitting the SALT2 model of SNe Ia (Guy et al., 2007) to observations and training a boosted decision tree on the parameters of that model gave the best classifier performance of the methods that they tested.

The major concern with all of these photometric classification methods is that they have poor performance when the training set of objects with spectroscopically-determined types is not representative of the full dataset. One common method of evaluating the performance of a classifier is by calculating the Area under the Receiver Operator Characteristic Curve (AUC), defined in Section 4.2.2, with a value of 1 representing a perfect classifier and a value of 0.5 representing random guessing. Lochner et al. (2016) achieve an AUC for the classification of SNe Ia of 0.98 when the training set is representative of the full dataset, but an AUC of only 0.88 when training on the non-representative training set in the SNPhotCC. Revsbech et al. (2018) introduced the first effective attempt to deal with non-representative training sets in a model that they call `STACCATO`. They augment the original training data by generating new light curves from ones in the training sample to produce a new training set that is more representative of the full dataset. `STACCATO` achieves an AUC of 0.96 when trained on their augmented training set compared to 0.92 when trained on the original training set.

Most of the previously discussed classifiers were trained and evaluated on the SNPhotCC dataset. Following the success of the SNPhotCC, a new challenge was created focusing on photometric classification for the LSST. This challenge, the Photometric LSST Astronomical Time-Series Classification Challenge (Kessler et al., 2019, PLAsTiCC,), includes 18 different kinds of transients and variables, and is not limited to supernovae like the SNPhotCC was. From September 28, 2018 to December 17, 2018, a blinded version of the PLAsTiCC dataset

was provided through the Kaggle platform[1] with class labels available only for the training set of spectroscopically-classified objects. Teams were challenged to determine the types for the remainder of the dataset, and submit their predictions to the Kaggle platform where a score was assigned to their predictions. A total of 1,094 teams submitted predictions as part of this challenge.

We developed a new photometric classifier that we applied to the PLAsTiCC dataset that expands on the previously described techniques. This classifier is discussed in Chapter 4. Out of all of the models submitted during the blinded phase of the PLAsTiCC, our classifier achieved the best performance on the PLAsTiCC test set measured using the weighted log-loss metric proposed by the PLAsTiCC team (Malz et al., 2018). Fundamental to our classifier is the concept of "augmentation" where we generate a variety of simulated light curves from each given observed light curve. Using Gaussian process regression, we developed a technique to modify the redshift of the light curves in the training set. Our method does not require a representative training set to be effective for classification, and we instead suggest that future surveys relying on photometric classification focus their spectroscopic efforts on typing large numbers of well-observed, intermediate redshift transients.

## 1.8 Gaussian process regression

Several of our analyses make use of Gaussian process regression. A stochastic process $P(x)$ is a Gaussian process (GP) if for any finite set of points $x_1, x_2, ..., x_n$ the distribution $P(x_1), P(x_2), ..., P(x_n)$ is a multivariate normal distribution. A GP can be thought of as a prior over a set of functions. By conditioning the GP on observations, we obtain a posterior containing the set of functions that are consistent with the observations. For a detailed discussion of GPs and their applications, see Rasmussen & Williams (2006).

A GP is uniquely defined by its mean function:

$$\mu(x) = E[P(x)] \tag{1.13}$$

and its covariance function, or "kernel":

$$K(x_1, x_2) = E[(P(x_1) - \mu(x_1)) \times (P(x_2) - \mu(x_2))] \tag{1.14}$$

The choice of $\mu$ and $K$ determines how different functions are weighted in the prior of the GP. There are several possible choices for the kernel. One common choice in the astronomy literature is the squared-exponential kernel (e.g. Kim et al. 2013 and Lochner et al. 2016). This kernel produces infinitely differentiable functions which can often be unrealistically smooth and produce poor models of data (Stein, 1999). Instead, for our analyses, we prefer to use a Matérn 3/2 kernel whose predictions are only once differentiable:

$$K_{3/2}(x_1, x_2; \alpha, l) = \alpha^2 \left(1 + \sqrt{3\frac{(x_1 - x_2)^2}{l^2}}\right) \exp\left(-\sqrt{3\frac{(x_1 - x_2)^2}{l^2}}\right) \tag{1.15}$$

[1] `https://www.kaggle.com/c/PLAsTiCC-2018`

The parameter $\alpha$ describes the amplitude scale of the functions that will be produced by the GP, and $l$ sets the length scale over which functions vary. Although we prefer the Matérn 3/2 kernel for theoretical reasons, we do not typically notice significant differences in the performance of the GPs that we use with this kernel compared to the squared-exponential one.

# Chapter 2

# The Binary Offset Effect

## 2.1 Overview

In this Chapter, we report a newly discovered CCD electronic chain artifact which originates in the readout electronics of many commonly-used CCD electronics systems. This chapter was previously published in the Publications of the Astronomical Society of the Pacific in May 2018, and was authored by K. Boone, G. Aldering, Y. Copin, S. Dixon, R. S. Domagalski, E. Gangler, E. Pecontal, and S. Perlmutter (Boone et al., 2018).

For many instruments, we find an offset in the digitization of the CCD output voltage that depends on the binary encoding of one pixel is added to pixels that are read out one, two and/or three pixels later. One result of this effect is the introduction of a differential offset in the background when comparing regions with and without flux from science targets. Conventional data reduction methods do not correct for this offset. We find this effect in 16 of 22 instruments investigated, covering a variety of telescopes and many different front-end electronics systems. The affected instruments include LRIS and DEIMOS on the Keck telescopes, WFC3-UVIS and STIS on HST, MegaCam on CFHT, SNIFS on the UH88 telescope, GMOS on the Gemini telescopes, HSC on Subaru, and FORS on VLT. The amplitude of the introduced offset is up to 4.5 ADU per pixel, and it is not directly proportional to the measured ADU level. We have characterized this effect and modeled it with high accuracy in the SuperNova Integral Field Spectrograph (SNIFS; Lantz et al., 2004) instrument used by the Nearby Supernova Factory collaboration (SNfactory; Aldering et al., 2002) on the UH88 telescope. We call this CCD-electronics-chain artifact the "binary offset effect".

We proceed as follows. In Section 2.2 we discuss how to identify this effect in CCD data. In Section 2.3, we build a model of the effect which can predict the size of the introduced offsets given the adjacent pixel values, and we provide an example of data corrected with this model. In Section 2.4, we then discuss how scientific results can be impacted by this effect if it is not corrected for.

## 2.2 Identifying the Binary Offset Effect

### 2.2.1 Evidence of the Binary Offset Effect in SNIFS Data

The binary offset effect was first observed in data taken with SNIFS for the SNfactory (Lantz et al., 2004; Aldering et al., 2002). SNIFS contains two lenslet integral field unit (IFU) spectrographs which produce spectra over a 15×15 grid of spatial elements (spaxels). The spectrographs simultaneously cover wavelength ranges of 3200–5200 Å and 5100–10000 Å for the blue and red channels respectively. Each spectrograph uses a CCD composed of 2048×4096 15 micron pixels. The blue channel uses a thinned E2V model 44-82 CCD while the red channel uses a thinned and deep depleted E2V CCD44-82-0-A72 CCD. These CCDs are of the highest scientific grade. Each CCD is read out by two independent amplifiers using an Astronomical Research Cameras (ARC) Generation II video board (Leach et al., 1998).

The binary offset effect can be easily seen by looking at images where there is an accurate model of the light on the CCD, and by probing the residuals after this model is subtracted from the data. Here, we examine a set of 1 second dome flat exposures taken with SNIFS. We take a 300 second exposure in the same configuration, and we treat this exposure as a model of the true light on the CCD because this exposure has high count levels compared to the size of the effects that we are looking into. An example of these dome flat CCD exposures is shown in Figure 2.1, where the spectral traces due to the IFU reformatting of the dome flat are all visible. We systematically work with the output of the CCD in analog-to-digital units (ADU) throughout this chapter. The top two panels of Figure 2.1 show examples of the dome flat exposures after applying an overscan subtraction, a bias correction, and a dark correction. The bottom panel of this figure shows slices through both one of the 1 second dome flat exposures and the 300 second exposure scaled to match the exposure time of the 1 second exposure. On the right amplifier, there is a count deficit of 1–2 ADU in between each of the spaxels on the 1 second exposure relative to the 300 second one. This same deficit is not seen on the left amplifier.

To probe the cause this deficit, we subtract a scaled version of the 300 second exposure (which we use here as our reference model) from each of the 1 second exposures to obtain residual images. For each 1 second exposure, we then compare the residual values of each pixel after the model was subtracted to the raw value (number of ADU) measured in a pixel that was read out 2 pixels earlier. We call the pixel read out 2 pixels earlier the "driver pixel." We take the mean of all residuals that have the same driver pixel value, and we plot the mean residual as a function of the driver pixel value. The results of this procedure are shown in Figure 2.2.

The resulting plot gives a very jagged function: a single ADU difference in the driver pixel value can correspond to a difference of up to 2.5 ADU in the mean of the residuals of the pixel read out 2 pixels afterwards. These differences are highly statistically significant; the measurement uncertainties are less than 0.05 ADU between driver pixel values of 1260 and 1290. A careful analysis of these large differences reveals that they map out the number

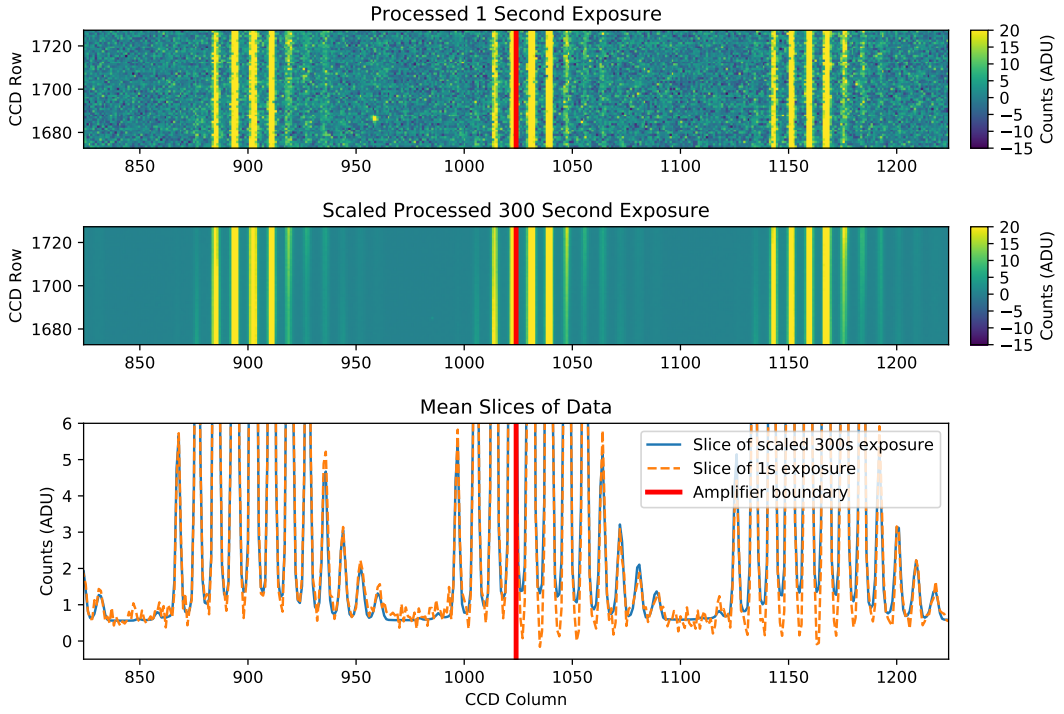



Figure 2.1: Comparison of 1 and 300 second dome flat exposures with the SNIFS instrument. Top panel: 1 second exposure. Middle panel: 300 second exposure. The CCD is read out from left to right. The vertical lines (which saturate the scale) are the traces of spectra for each of the individual spaxels. The amplifier boundary on the CCD is indicated with a vertical red line. Bottom panel: slices through the previous images showing the mean values of the pixels in each CCD column with CCD row values between 1400 and 2000. The 300 second exposure has been scaled by its exposure time to match the 1 second exposure. On the right amplifier, there is a deficit in the measured counts between each spaxel for the 1 second exposure relative to the 300 second exposure, although the background regions are in agreement. These large deficits are not present on the left amplifier.

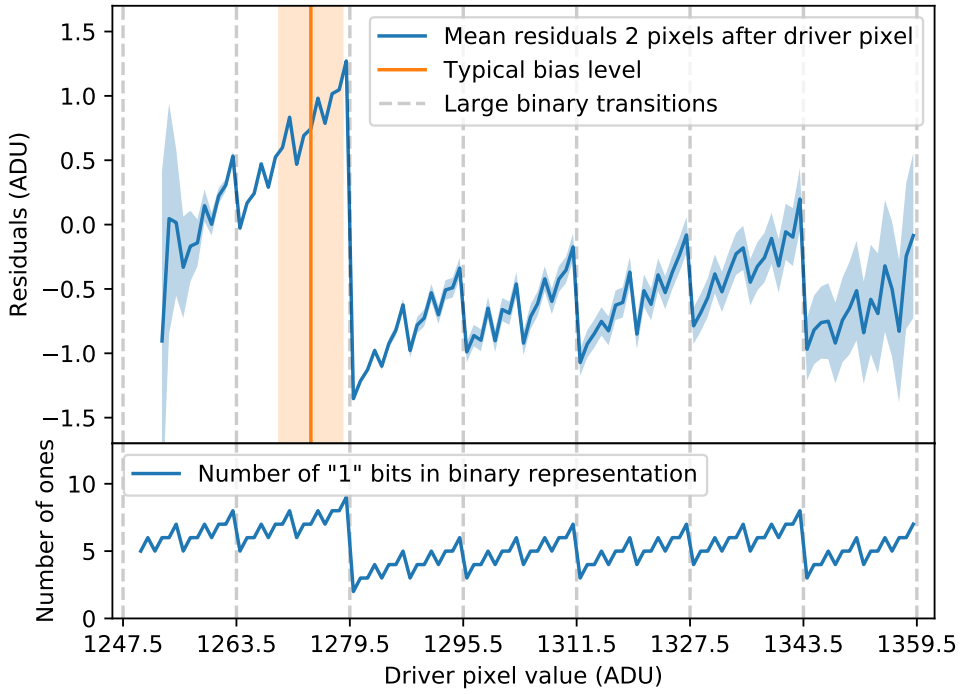



Figure 2.2: Example of the binary offset effect on the SNIFS blue channel right amplifier. See text for details on how this plot was produced. Top panel: mean residuals as a function of the raw driver pixel value read out 2 pixels earlier along with an uncertainty band. Bottom panel: number of "1" bits in the binary representation of the driver pixel value. We label transitions where more than 5 bits flip with a vertical dashed line. All of the features in the bottom panel, including the large transitions, are visible in the residual data. The mean background level and one standard deviation of read noise are illustrated with a vertical orange line and surrounding band.

of "1" bits in the binary representation of the driver pixel value. For example, Figure 2.2 shows that there is a difference of 2.5 ADU when transitioning from a driver pixel value of 1279 ADU to a driver pixel value of 1280 ADU. The binary representations of these two numbers are 100 1111 1111 and 101 0000 0000 respectively. Other large differences occur between 1311 ADU and 1312 ADU (101 0001 1111 and 101 0010 0000) and between 1343 ADU and 1344 ADU (101 0011 1111 and 101 0100 0000). The lower panel of Figure 2.2 shows the plot of the number of "1" bits in the binary representation of the driver pixel for comparison. All of the features in this plot are seen in the residual data, from the large binary transitions down to odd-even effects.

Figure 2.3 shows the same residuals from Figure 2.2 plotted directly against the number of "1" bits in the binary representation of the driver pixel value. There is a linear trend between the mean of the residuals and the number of "1" bits in the binary representation of the driver pixel value. This is consistent with this effect being mostly related to the total number of "1" bits rather than it being due to some issue with the uppermost bits. There is an offset in the zeropoint of this linear relation for values below 1280 ADU indicating that a linear model with the number of "1" bits is not a complete description of the effect. In Section 2.3 we will discuss how to properly model this effect.

### 2.2.2 Implications of the Binary Offset Effect in SNIFS Data

The binary offset effect introduces a highly non-linear signal into images taken with the CCD. The main consequence of the binary offset effect is that an offset can be introduced to CCD data that only appears in regions where sufficient flux from science targets is present on the CCD. To show how this appears in data, we examine stacks of face-on cosmic rays in dark exposures taken with the SNIFS instrument. An example of such a stack is shown in Figure 2.4. These images were originally taken in effort to measure the in-situ CTI of the SNIFS instrument (Dixon et al., 2016), and a small CTI tail can be seen on the upper part of the image. We also find a deficit in the measured count levels of pixels read out 2–3 pixels after a cosmic ray compared to the background level. We find that the size of the deficit varies from night to night, from no effect up to around 1 ADU per pixel. The size of the deficit is not the same on each amplifier. This deficit is the result of the binary offset effect, and it can be thought of as an offset that is added to the data when the count levels of previous pixels cross a specific threshold. The deficit has a roughly fixed count value that does not scale with the amount of flux in the preceding pixels, so it does not behave as a simple modification to the PSF. The binary offset effect hence effectively introduces a deficit wherever there is sufficient flux from science targets on the CCD, leading to a local offset in the measured count values for regions on the CCD with flux from science targets compared to those without any flux.

An explanation of how the binary offset effect can cause such a deficit can be seen from Figure 2.2. The raw background level (before overscan subtraction) on the dome flat images used to generate this figure is around 1274 ADU with 4.5 ADU (or 3 electrons) of read noise per unstacked pixel. This background level is illustrated with an orange line and surrounding

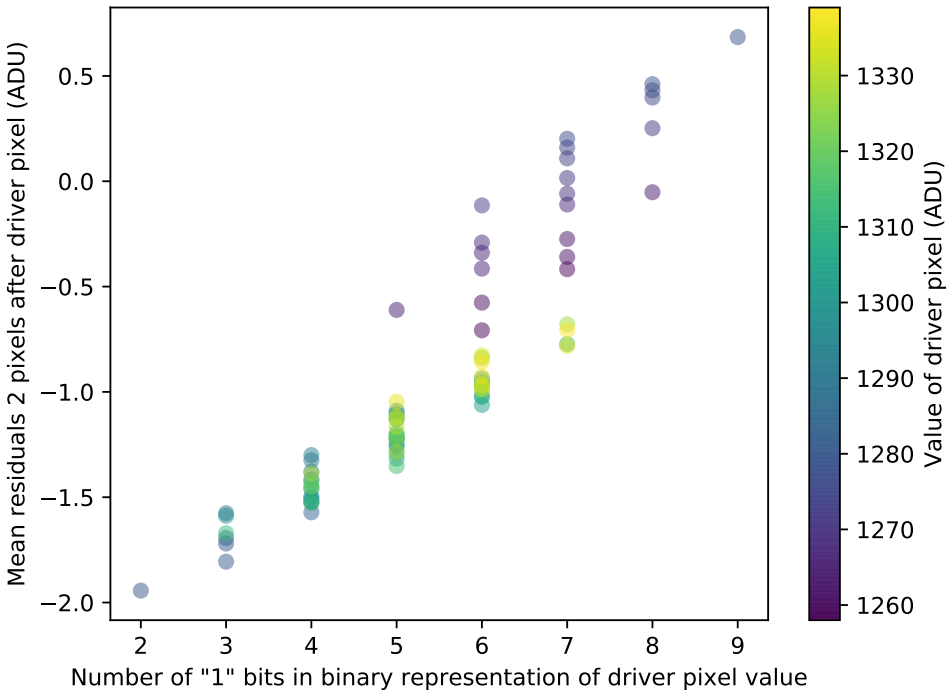



Figure 2.3: Example of the binary offset effect on the SNIFS blue channel right amplifier. This plot shows the same data as in Figure 2.2 with a cut requiring that the uncertainty on the mean of the residuals be less the 0.2 ADU. Instead of plotting the mean of the residuals directly against the driver pixel values as in Figure 2.2, we sum the number of "1" bits in the binary representation of each driver pixel value in ADU, and we plot the mean of the residuals against this number. There is a clear linear trend in the mean of the residuals with the number of "1" bits in the binary representation, at all scales. It is also apparent that there are some additional effects, such as an offset in the zeropoint of the linear relation near a driver pixel value of 1280 ADU.

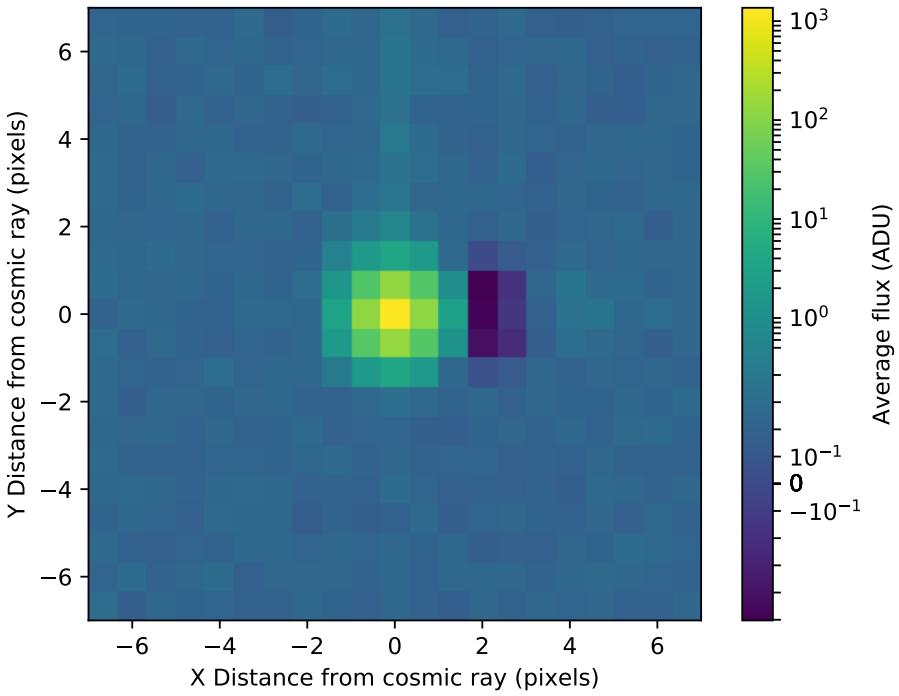



Figure 2.4: Stacked cosmic ray frames from the right amplifier of the SNIFS blue channel. This image was generated by taking a series of dark exposures on a night, identifying all face-on cosmic rays in the image, and taking the clipped mean of all pixels around cosmic rays. The color scale on this plot changes from linear to logarithmic at 0.5 ADU in order to capture the full dynamic range of the image. The serial readout reads pixels from left to right in this image. There is a visible deficit of $\sim$0.5 ADU 2–3 pixels following the cosmic ray. The faint tail above the cosmic ray is due to the well-known effect of charge transfer inefficiency.

band on the figure. Most background pixels are therefore below the threshold of 1280 ADU, where there is a large change in the binary representation of the driver pixel value. When there is at least a small amount of signal on the CCD, the pixel values are brought above the 1280 ADU threshold and a ∼2 ADU offset is introduced into subsequently read pixels compared to the background regions. The consequences of the binary offset effect for a science image can be seen in the lower panel of Figure 2.1: for the right amplifier, a deficit appears in the 1 second dome flat exposure compared to the 300 second exposure wherever the average count levels are above a few counts. On the left amplifier, the bias level is at 1285 ADU, and there are no major binary transitions in the nearby higher count levels. The binary offset effect is still present, but introduced offsets are similar for all pixels so no major difference is observed between background and science regions.

When extracting the spectra from images affected by the binary offset effect, the introduced offsets will propagate to the extracted spectra. The two amplifiers will, in general, have different bias levels, so they will be affected in different ways by the binary offset effect. Hence, systematic differences will be observed in spectra extracted from each of the two amplifiers. The SNfactory collaboration discovered these systematic differences before the cause was known, and dubbed this effect the "blue step" as it was primarily noticed at the bluest wavelengths where the expected fluxes of SNIFS's primary objects of interest (Type Ia supernovae) can be faint. To illustrate the blue step, we extract spectra for each of the IFU spaxels in one of the 1 second dome flat exposures using the standard SNfactory pipeline (Aldering et al., 2006; Scalzo et al., 2010), and we calculate the average flux in the 4000 to 4500 Å region. The results of this procedure are shown in Figure 2.5. We find a 2.2 ADU difference in the measured counts between the two amplifiers, corresponding to a 4.2% difference in the measured fluxes between the two amplifiers for this exposure. The observed difference is well-modeled by a flat difference in the background level between the different amplifiers, as expected for a signal introduced in the previously described manner. The average offset introduced by the binary offset is different in regions on the CCD with and without flux present, meaning that there is a effectively a pathological local background offset that only appears where there is flux on the CCD. Conventional background subtraction routines cannot identify or correct for the difference in background levels introduced by the binary offset effect because these methods measure the background level in locations on the CCD where there is no flux present.

### 2.2.3 Cause of the Binary Offset Effect

The discussion up to this point has shown that an offset is introduced into pixels that are read out two pixels after the driver pixel. We find a similar offset three pixels after the driver pixel for each CCD amplifier on both the blue and red SNIFS channels. There is no effect 1 pixel after or 4 or more pixels after, nor is there an offset for pixels before the driver pixel. The ARC video board used by SNIFS processes pixels one at a time, so the timeframe of 2–3 pixels between the driver and affected pixels implies that there must be some feedback in the electronics chain from somewhere after the ADC conversion. We varied the gain of the

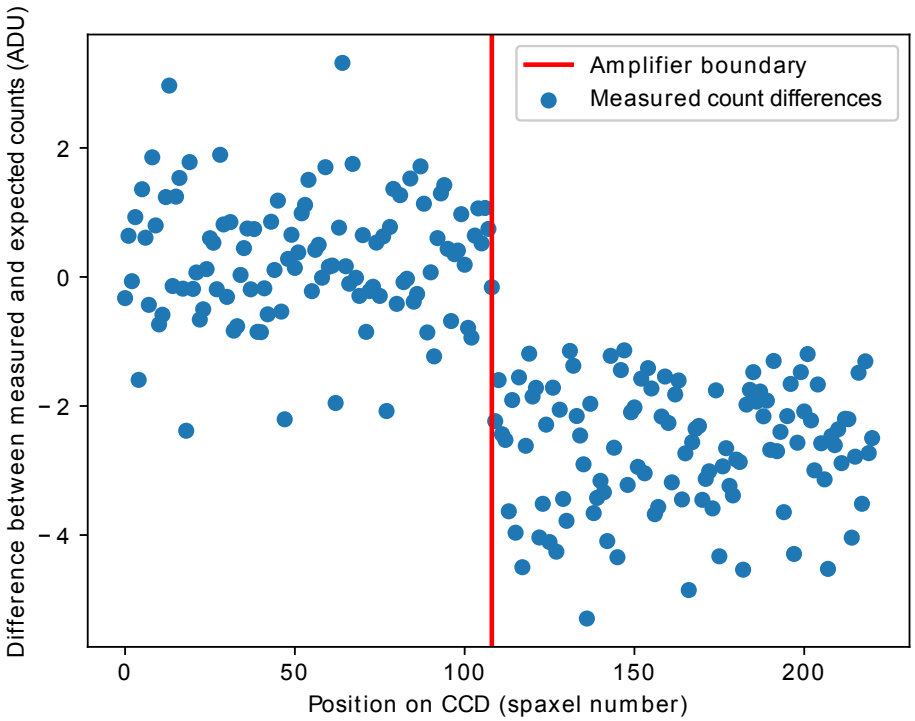



Figure 2.5: Exposure-time normalized difference between extracted 1 second and 300 second dome flats from the SNIFS blue channel, illustrating the "blue step". Each point corresponds to the average flux of a spaxel in the 4000 to 4500 Å region. We indicate the boundary where the readout switches from the left to the right amplifier with a vertical red line. We find a 2.2 ADU difference in the measured fluxes between the two amplifiers which corresponds to a 4.2% difference in the measured flux for this example.

SNIFS front-end electronics by a factor of 10, and we find that the amplitude of the effect is constant in ADU. This implies that the introduction of the feedback must occur past the gain electronics in the CCD electronics chain. We also notice that the binary offset effect shows up across amplifiers: a driver pixel will introduce an offset in the pixels that are read out 3 pixels later in the other amplifier with a similar amplitude to the offset introduced on the driver pixel's amplifier.

Combining all of these findings, we propose that the binary offset effect is being caused by feedback into the reference voltage of the ADC. The ADC is highly sensitive to small changes in its reference voltages: one ADU for a 16-bit ADC corresponds to a change of only 0.0015%. When earlier pixels are read out, the ADC outputs their binary representations and stores them temporarily. We propose that these charges (representing the ones in the binary code) then introduce a slight offset into the reference voltage when they are released

and the next pixel is read out.

### 2.2.4 The Binary Offset Effect in Other Instruments

We wrote a program that can identify the binary offset effect in CCD data from any telescope by looking for the characteristic saw-tooth shape in the mean of the residuals as a function of the driver pixel value seen in Figure 2.2. We find that the effect is present in most astronomical instruments that are currently in use across a variety of different readout systems. As we are not able to build a model using different length exposures of our own on all of these instruments, we use bias or dark images to probe the effect, and we assume that the pixel values in these images can be modeled by a smooth function. We fit for such a model using the background determined by `sep` (Barbary, 2016; Bertin & Arnouts, 1996) with $64 \times 64$ pixel boxes. We subtract this model from the original images to obtain residual images. With dark and bias images, we are limited to probing the effect over a smaller baseline of driver values (typically around 10 ADU). We are, however, able to achieve very high signal-to-noise by averaging the residuals over the full image and by examining multiple images. We estimate the amplitude of the binary offset effect by measuring the largest difference in residuals between two adjacent driver pixel values. For the SNIFS images shown in Figure 2.2, this corresponds to the size of the transition from 1279 to 1280. We note that the measurement uncertainties on the stacked residuals are around 0.01 ADU for most instruments, so this method is sensitive to effects larger than $\sim$0.05 ADU. The results are summarized in Table 2.1.

We are able to detect the binary offset effect in 16 of 22 instruments that were investigated. The amplitude of the offset varies significantly between the different instruments. We provide a `Jupyter` notebook (Kluyver et al., 2016) that contains code to probe this effect for all of the instruments listed in Table 2.1. This `Jupyter` notebook is available on GitHub[1], and the main algorithms from that notebook are reproduces in Appendix 2.7. Plots of the binary offset effect are shown for a representative sample of the different instruments in Figure 2.6.

For all of the detections in Table 2.1 from instruments on the CFHT, Gemini, Keck, UH88 and VLT telescopes (11 of the 22 instruments investigated), we see an offset in the mean of the residuals that is roughly proportional to the number of "1" bits in the binary representation of the driver pixel, similar to what is seen for SNIFS. The amplitude of the effect varies significantly between instruments, from a peak-to-peak amplitude of approximately 0.1 ADU for FORS 2 on VLT to a peak-to-peak amplitude of approximately 2.6 ADU for DEIMOS on Keck (see Figure 2.6a). We also find that the number of pixels following the driver pixel at which the binary offset effect appears varies from 1 to 3 pixels for the different instruments. All of the instruments tested using ARC Generation I or II controllers (Leach et al., 1998) show this kind of binary offset effect, although the size of the effect varies from instrument

[1] `https://www.github.com/snfactory/binaryoffset`

Table 2.1: Size of the binary offset effect in different instruments. See text and accompanying `Jupyter notebook` for details on how these measurements were made. We have attempted to determine the CCD electronics front end and ADC wherever possible for comparison purposes.

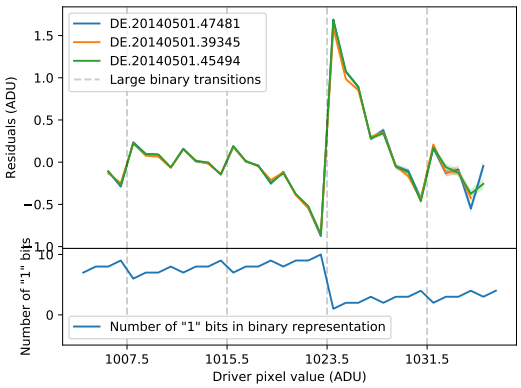
| Telescope | Instrument | Distance from Driver Pixel to Target Pixel (pixels) | Approximate Peak-to-Peak Amplitude of Binary Offset Effect (ADU) | Amplitude in Electrons | CCD Front End | ADC | Reference |
|---|---|---|---|---|---|---|---|
| Blanco | DECam | - | $< 0.05^a$ | $< 0.2^a$ | Monsoon | Analog Devices AD7674 | Castilla et al. (2010) |
| CFHT | Megacam | 1 | 0.4 | 0.6 | | Linear Technology LTC 1604 | de Kat et al. (2004) |
| Gemini | GMOS-S E2V | 2, 3 | 0.7 | 1.4 | ARC Gen. II | Datel ADS-937 | Hook et al. (2004) |
| | GMOS-S Hamamatsu | - | $< 0.05^a$ | $< 0.08^a$ | ARC Gen. III | | Gimeno et al. (2016) |
| | GMOS-N E2V | 2, 3 | 0.7 | 1.4 | ARC Gen. II | Datel ADS-937 | Hook et al. (2004) |
| | GMOS-N Hamamatsu | - | $< 0.05^a$ | $< 0.09^a$ | ARC Gen. III | | Gimeno et al. (2016) |
| HST | WFC3 UVIS | 1 | $0.05$–$0.15^b$ | $0.08$–$0.23^b$ | | | |
| | STIS (post SM4) | 1 | $0.5$–$4.5^b$ | $0.5$–$4.5^b$ | | | |
| | ACS | 1, 2, 3, 4 | $0.4$–$1.0^b$ | $0.4$–$1.0^b$ | | | |
| Keck | DEIMOS | 2 | 2.6 | 3.2 | ARC Gen. II | Datel ADS-937 | Wright et al. (2003) |
| | HIRES | 2 | 0.3 | 0.6 | ARC Gen. I | Datel ADS-937 | Kibrick et al. (1993) |
| | LRIS B | 2 | 0.15 | 0.24 | ARC Gen. I | Datel ADS-937 | McCarthy et al. (1998) |
| | LRIS R (upgraded) | 2 | 0.15 | 0.15 | ARC Gen. II | Datel ADS-937 | Rockosi et al. (2010) |
| SDSS | - | - | $< 0.05^a$ | $< 0.23^a$ | | Crystal Semiconductor CS5101A | Gunn et al. (1998) |
| Subaru | Suprime-Cam | - | $< 0.05^a$ | $< 0.15^a$ | MFront | Analogic ADC423 | Miyazaki et al. (2002) |
| | Hyper Suprime-Cam | 1 | $0.5^c$ | $1.6^c$ | MFront2 | Analog Devices AD7686C | Nakaya et al. (2012) |
| | FOCAS | 2 | $0.1^c$ | $0.2^c$ | MFront | Analogic ADC423 | Kashikawa et al. (2002) |
| UH88 | SNIFS blue channel | 2, 3 | 2.4 | 1.8 | ARC-41 Gen. II | Datel ADS-937 | Aldering et al. (2002) |
| | SNIFS red channel | 2, 3 | 1.5 | 1.1 | ARC-41 Gen. II | Datel ADS-937 | Aldering et al. (2002) |
| VLT | FORS 1 | 1 | 0.1 | 0.22 | FIERA | Analogic ADC4320A$^d$ | Beletic et al. (1998) |
| | FORS 2 | 1 | 0.1 | 0.13 | FIERA | Analogic ADC4320A$^d$ | Beletic et al. (1998) |
| | MUSE | - | $< 0.05^a$ | $< 0.06^a$ | NGC | Analog Devices AD7677$^c$ | Reiss et al. (2012) |

[a] For these instruments, the binary offset effect was not detected. We report an estimate of the sensitivity of our measurement of the binary offset effect.

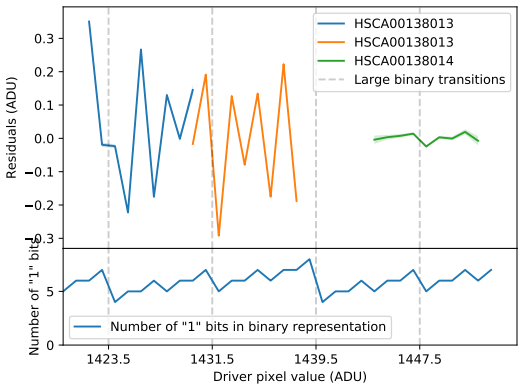
[b] For the labeled HST instruments, there is a strong trend in the mean of the residuals with the least-significant bit (odd-even) along with large offsets when higher bits change. Intermediate bits do not appear to have any effect.

[c] In Subaru Hyper Suprime-Cam and FOCAS images we find a trend in the mean of the residuals with the least-significant bit (odd-even). The higher bits do not appear to have a direct impact on the mean of the residuals.

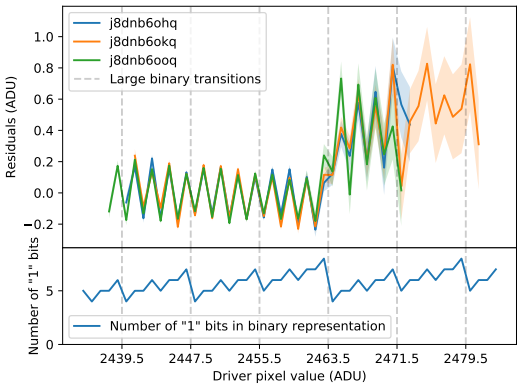
[d] We thank J. Reyes (ESO) for details of the ADCs used in the VLT instruments. (personal communication, August 2017)



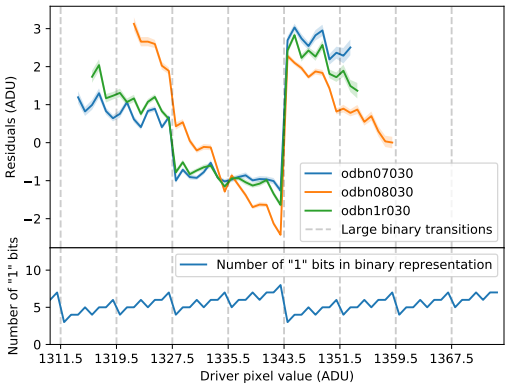

(a) DEIMOS (Keck), 2 pixels after the driver pixel. The median measurement uncertainty for the resiudals is 0.012 ADU.




(b) Hyper Suprime-Cam (Subaru), 1 pixel after the driver pixel. The median measurement uncertainty for the residuals is 0.006 ADU.




(c) ACS (HST), 1 pixel after the driver pixel. The median measurement uncertainty for the residuals is 0.030 ADU.




(d) STIS (HST), 1 pixel after the driver pixel. The median measurement uncertainty for the residuals is 0.077 ADU.

Figure 2.6: Examples of the binary offset effect in other instruments. For half of the instruments that were investigated, such as DEIMOS on Keck, we see an effect that has similar properties to the effect on SNIFS. For other instruments, such as Hyper Suprime-Cam on Subaru, ACS on HST, or STIS on HST, we find evidence of offsets related to the binary representation, but the offsets have a different dependence on the binary encoding. All plots are shown with their measurement uncertainties as a shaded contour, and the median measurement uncertainty is given in the subpanel caption. These measurement uncertainties are very small compared to the size of the observed effects, and all of these observed effects are highly statistically significant. The specific observations used to generate each plot are listed in the figure legends.

to instrument. The newer ARC Generation III controllers used in GMOS-S and GMOS-N do not show any evidence of the binary offset effect.

In the images from Hyper Suprime-Cam (HSC) on Subaru, we find offsets in the mean of the residuals, but they do not appear to be linearly related to the number of "1" bits in the binary encoding of the driver pixel value (see Figure 2.6b). The mean of the residuals shows peak-to-peak offsets of 0.5 ADU between adjacent driver pixel values with a median measurement uncertainty of 0.006 ADU, so the detection of an offset related to the driver pixel value is highly significant. There appears to be a correlation of the least-significant bit (i.e. whether the driver pixel value is even or odd) with the mean of the residuals. Unlike the previously-described instruments, large changes in the binary representation do not appear to correlate with large offsets in the mean of the residuals. The introduced offsets do still have a highly nonlinear and nonmonotonic dependence on the driver pixel values, although additional work is needed to understand these offsets. For images taken with the FOCAS instrument on Subaru we see a similar effect, although the amplitude is smaller.

Images from all of the HST instruments that were examined (WFC3 UVIS, STIS and ACS) show evidence of an effect related to the binary encoding of driver pixels but with somewhat different behavior. An example of the effect in ACS is shown in Figure 2.6c. The least-significant bit (odd-even) has a large effect on the mean of the residuals, but transitions of the next several bits do not appear to have a significant impact. However, when the sixth bit or higher is changed, a step is introduced into the mean of the residuals. For the ACS images shown in Figure 2.6c, we see an offset of 0.4 ADU at the transition from 2463 ADU to 2464 ADU (1001 1001 1111 to 1001 1010 0000).

The STIS instrument on HST displays a similar behavior to ACS. An example of the binary offset effect on STIS is shown in Figure 2.6d. For STIS, when the fifth bit or higher is changed, we notice a large offset, but there is little effect for the less significant bits. At the transition from 1343 ADU to 1344 ADU (101 0011 1111 to 101 0100 0000) there is a 4.5 ADU offset in the mean of the residuals. We find that the size of the offsets when the upper bits are changed varies dramatically, from 0.5 ADU up to the offset of 4.5 ADU previously mentioned.

For several instruments, the binary offset effect introduces artifacts that are large compared to the size of the signals being observed. On SNIFS, the blue step was identified and characterized before it was understood as being a result of the binary offset effect. It is likely that the consequences of the binary offset effect have been identified in other instruments without the root cause being fully understood.

## 2.3 Modeling and Correcting the Offset for Active Instruments

The binary offset effect is present in a large fraction of existing CCD data, as illustrated by Table 2.1, and correcting for the effect will both improve the quality of existing data and

allow previously unusable data to be recovered. As a working example of such a correction, we have derived a model that can calculate the introduced offsets in SNIFS data.

As illustrated in Figure 2.3, the amplitude of the effect is primarily proportional to the number of "1" bits in the binary encoding of the driver pixel. However, this figure illustrates that the effect is not simply a linear function of the driver pixel value, as the zeropoint of the linear relation appears to shift when the driver pixel values drop below 1280 ADU. These details of the effect are second order, but must be taken into account to generate an accurate model. We have developed a 9-parameter model which can capture the behavior of the binary offset effect in SNIFS data. We trained this model on a set of bias images covering the full history of SNIFS operations (2004–2017). The details of this model can be found in Appendix 2.6.

We find that this single set of model parameters is able to describe the behavior of an amplifier over the entire history of the SNIFS instrument. That is; the model parameters do not vary over time, and a single set of parameters per amplifier is sufficient to cover the full range of observed temperatures, bias levels and signal levels. The behavior of the binary offset effect does vary significantly between amplifiers, and we therefore use a unique set of model parameters for each amplifier. This is consistent with the idea that the digitization is the root cause, since each amplifier has its own video board. With this model, we are able to predict the amplitude of the offsets introduced by the binary offset effect to within 0.11–0.16 ADU depending on the amplifier. We can predict the amplitudes of the offsets for every pixel in an image, and we can use these predictions to build a correction image which can be subtracted from the data to remove the binary offset effect. Applying this procedure to a new image only requires the raw pixel counts of the image and a set of previously-derived model parameters for that amplifier. No additional fitting is required.

As a test of the model, we applied the derived correction to the set of 1 and 300 second dome flat exposures described in Section 2.2.1. The results of this procedure can be seen in Figure 2.7. The dome flat images were not included the dataset on which the model was trained, so this is an out-of-sample test of the model. The deficits seen on the right amplifier in the slice plot of Figure 2.1 are no longer present after correction. We applied the same algorithm used to produce Figure 2.2 to the corrected data. The results of this procedure are shown in Figure 2.8. As seen from this figure, the offsets introduced by the binary offset effect are effectively removed from the data when the correction is applied. The remaining linear slope in the residuals is a consequence of the high frequency pickup described in Appendix 2.6, and is not a feature of the binary offset effect.

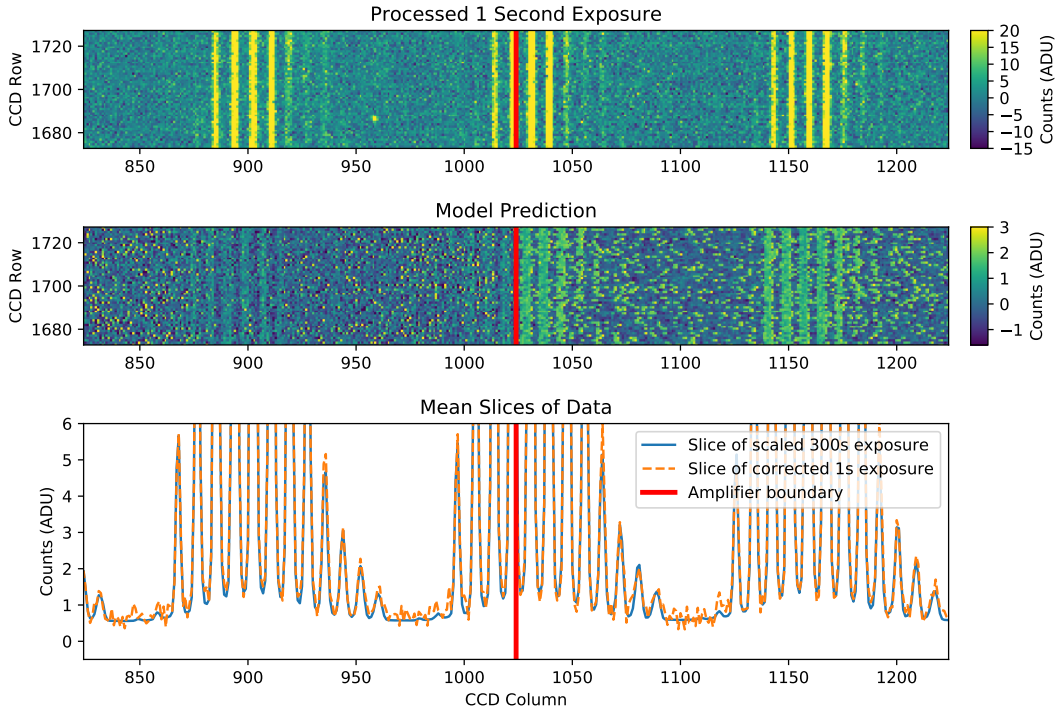



Figure 2.7: Results of the binary offset effect model for the SNIFS image shown in Figure 2.1. Top panel: original 1 second exposure. Middle panel: model prediction of the signal introduced by the binary offset effect. Bottom panel: slices through the previous images showing the mean values of the pixels in each CCD column with CCD row values between 1400 and 2000. For the right amplifier, the model predicts that the binary offset effect introduces an offset of ∼2 ADU (∼1.3 electrons) at the specific locations on the CCD where spectral traces are found relative to the background locations. When these corrections are applied to the data, the count deficits seen in Figure 2.1 for the slice of the 1 second exposure on the right amplifier are eliminated.

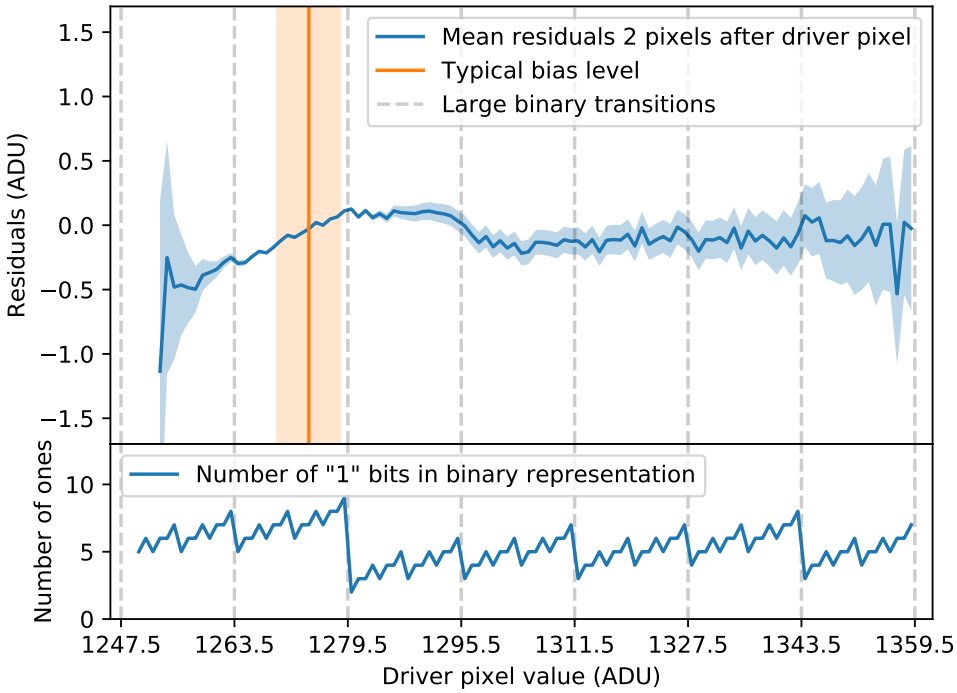



Figure 2.8: Example of data that previously exhibited the binary offset effect on the SNIFS blue channel right amplifier, after correcting for the effect. See Figure 2.2 for the same data before correction and an explanation of how this figure was produced. The residuals after correction no longer have a strong dependence on the binary encoding. There is a slight linear slope in the residuals for driver pixel values between ∼1250–1280 ADU; this slope arises due to correlations introduced by high frequency pickup, as discussed in Appendix 2.6, and is not related to the binary offset effect.

## 2.4 Impact of the Binary Offset Effect on Scientific Results

### 2.4.1 Impact of the Binary Offset Effect on SNIFS/SNfactory Data

The binary offset effect introduces a highly non-linear signal into CCD data that affects science results in several ways. As discussed in Section 2.2.2, for the SNIFS instrument, the binary offset effect can introduce an offset to the data that only appears at locations on the CCD with a significant amount of flux. This can be seen directly in the model in Figure 2.7: on the right amplifier, there is a visible offset of ∼2 ADU (∼1.3 electrons) at the specific

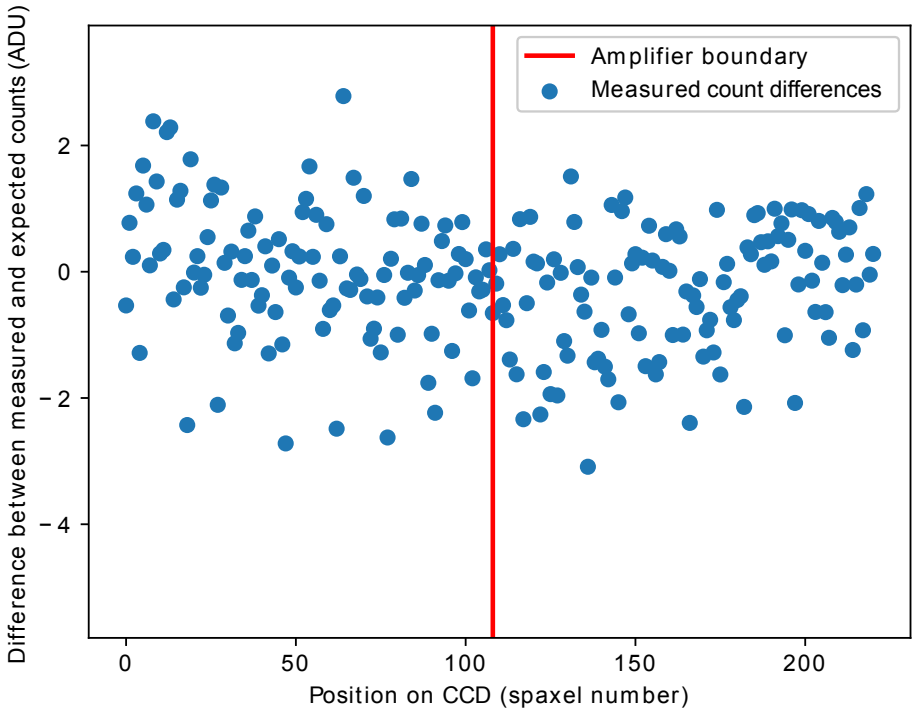



Figure 2.9: Corrected exposure-time normalized difference between extracted 1 second and 300 second dome flats from the SNIFS blue channel. See Figure 2.5 for details and comparison. We find that the offsets between the two amplifiers are consistent to within 0.2 ADU (0.13 electrons) after correction.

locations on the CCD where spectral traces are found relative to the background regions. The model predicts that the binary offset effect has little impact on the left amplifier due to the bias level being above the large binary transition at 1280.

We subtract the derived model from one of the 1 second dome flat exposures, and we extract the spectra in the corrected image using the SNfactory pipeline. The result of this procedure is shown in Figure 2.9, which can be compared to Figure 2.5 on the same data before corrections. The previous offset of 2.2 ADU has been removed and we constrain the corrected offset between the amplifiers to be less than 0.2 ADU (0.13 electrons), confirming that the binary offset effect was the cause of the "blue step" in this data and that our model is capable of removing it.

We apply the binary offset correction to a series of SNfactory observations of Type Ia supernovae that are representative of the range of measured flux levels on the CCD for scientific observations. We find that the blue channel is the most strongly affected by the binary offset effect. For the full dataset, the NMAD (normalized median absolute deviation)

of the applied corrections ranges is 0.04 mag at UV wavelengths and 0.02 mag for blue (B-band) wavelengths, while if we look at the faintest 20th percentile of data ordered by flux level, we find that the corrections in UV have a much larger NMAD of 0.51 mag and those in the blue have an NMAD of 0.11 mag. Corrections on the red channel have an NMAD of less than 0.008 mag over the full dataset, and an NMAD of less than 0.05 mag for the faintest 20 percent of observations. The binary offset correction is therefore relatively small for the majority of the dataset, but it has a disproportionately high impact on fainter regions of our spectra on the blue channel.

The affected spectra in SNfactory data are typically at late phases of the supernova lightcurves where the intrinsic supernova flux is low. In practice, however, the SNfactory collaboration had previously developed methods to identify spectra affected by the binary offset effect (before recognizing the cause), primarily by looking for anomalous behavior in the UV, where it stands out. These spectra were not included in previous published analyses. Although the corrections for the binary offset thus should not affect any previous results, we do expect that it will allow future analyses to include many spectra that were previously rejected, especially spectra at late phases.

### 2.4.2 Impact of the Binary Offset Effect on General Astronomical Data

There are several science scenarios where the binary offset effect can have large impacts on science data. The conditions that lead to the largest potential impact are:

- Low background noise.
- Low signal levels.
- A background level just below a large binary transition.

Low-flux observations with fiber-fed, lenslet, or similar spectrographs are very susceptible to this effect. In these cases, most of the CCD is not in the path of any light, so the background in those regions is entirely dominated by read noise. For modern instruments, this means that the background noise will be very low, on the order of a few ADU or less. When measuring the background level on the CCD with conventional methods, one will effectively sample the binary offset curve at the given bias level instead of measuring what the background truly is under the science data due to the small range of pixel values covered in the region used to measure the background level. Observations with ground-based slit spectrographs may be highly susceptible by the binary offset effect, although their susceptibility depends on how they are used. Observations of faint emission lines with low sky background or target continuum will be affected for the same reasons as fiber-fed and IFU spectrographs. As the binary offset effect introduces an offset that is shifted by 1–3 pixels from the data that it is applied to, the recovered wavelengths of emission lines may be biased. Higher

resolution spectrographs are likely to be more strongly affected because of the reduction in sky background per pixel leading to low background noise levels on the CCD.

Science applications using ground-based CCD imaging are not as susceptible as spectrographs to the binary offset effect. Measurements of faint targets with these detectors are often limited by the sky background, so the potential size of the offset is suppressed. Imaging applications that have low backgrounds and low signals (for example U-band observations, narrow band imaging, imaging of standard stars, or very short exposures) are more likely to be affected. An additional issue for imaging is that the local background offset will have the same shape as the science signal, but will be shifted by 1–3 pixels depending on the instrument. This can affect measurements of the shape of a galaxy and may be a concern for some weak lensing analyses. The binary offset effect will also introduce some correlations between nearby pixels on the CCD.

Science applications using detectors in space-based missions are highly susceptible to the binary offset effect, regardless of whether they are being used as an imager or any type of spectrograph. For these instruments, the sky noise is often low compared to the readout noise so conventional background subtraction routines will effectively sample from a point on the binary offset curve rather than averaging over it. The STIS spectrograph on HST is of particular interest. We find an offset of up to 4.5 ADU for some binary transitions in this instrument. This offset is per pixel, so it can easily add up to a significant fraction of the science signal.

## 2.5 Conclusions

We have discovered an anomalous behavior in the read-out of CCDs: the introduction of spurious counts into a pixel, with an amplitude that depends on the binary encoding of a pixel read out 1–3 pixels previously. One consequence of this effect is that it can introduce a local background offset that only appears where sufficient flux from science targets is present on the CCD. This background offset is not removed by conventional background subtraction procedures. In SNIFS data, the effect can cause an offset in the final measured fluxes of up to 2 ADU per pixel. The binary offset effect explains several effects previously noticed in SNIFS data, notably the "blue step" where there is a flat offset between data from different amplifiers, and a fixed count deficit in pixels following cosmic rays. We find evidence of the binary offset effect in 16 of 22 different instruments that were investigated, indicating that the effect is present in a significant amount of existing astronomical data.

In this analysis, we built a model that can predict, and thus reliably correct, the offsets introduced by this effect in affected CCD data, and we applied this model to SNIFS data. We find that the model parameters are stable in time, but that they are unique to the readout electronics of each amplifier. With this model, we can predict the amplitudes of the introduced offsets, and we use these predictions to correct for the binary offset effect. The model derived in this analysis is general, and a variant of it should be applicable to any instrument. Because of the relatively large amplitudes of the introduced offsets relative to

the noise levels, only a single bias image is typically required to characterize the effect and fit the model.

The binary offset effect can also be mitigated in hardware. Reducing opportunities for stray signals to affect the input of the ADC, e.g. using a differential input, should eliminate the binary offset effect. We find that 6 of the 22 instruments that we investigated do not show signs of this effect, as shown in Table 2.1. These can serve as a guide to design new readout electronics that are not susceptible to the binary offset effect. The size of the binary offset effect should be measured and characterized in any new CCD readout system to determine whether or not it will impact the science results. For systems where modifying the readout electronics is not practical, the bias level can at least be set to a level where there are no large binary transitions immediately above said bias level. This will minimize the potential size of the offsets although it will not eliminate them. Finally, the tests outlined in this analysis should be performed for any CCD-based instrument likely to exhibit the binary offset effect, and the correction procedure that we present in this analysis should be incorporated in the data-reduction pipeline of all instruments where the binary offset effect is found.

## 2.6 Appendix: Modeling the Binary Offset Effect in SNIFS Data

In SNIFS data, we notice that an offset is introduced into the affected pixel related to the binary encodings of the driver pixels read out 2 and 3 pixels earlier. In the following discussion, we refer to a pixel that is affected by the binary offset as the "target". Each pixel on the CCD serves as a driver pixel for several other pixels. We subtract a model of the counts on the CCD from the measured data to obtain a residual image, and we compare the value of the residual image for each pixel to the raw ADU measured on the corresponding driver pixels. We refer to driver pixels that were read out N pixels before the target pixel as T-N. There is also some crosstalk between the amplifiers: the value of the pixel that was read out 3 pixels earlier in time on the other amplifier also affects the introduced offsets. We note that the timing of the readout of the other amplifier is what matters rather than the physical location on the CCD. We refer to driver pixels that were read out N pixels before the target pixel on the other amplifier as O-N.

We find that the binary codes of the driver pixels interact with each other in a complex way to produce the offset that is applied to the target pixel. There is a baseline linear trend of the introduced offsets with the number of "1" bits in pixels T-2 and T-3. However, there is also some interaction between these pixels: a bit in pixel T-2 behaves differently if that same bit was on in pixel T-3 compared to if it was off. The largest offsets occur when a bit is "1" in both T-2 and T-3. For pixel T-2, we find that the more significant bits have a larger effect on the introduced offset than the less significant bits. For pixel T-3, we find evidence of interactions across amplifiers: turning a bit from "0" to "1" in T-3 introduces an offset that depends on how many bits were already on in T-3 and in O-3. We model this

interaction with a second order polynomial in the number of bits that are "1" in each of T-3 and O-3. We do not find any evidence of direct interactions across bits (e.g.: bit 8 does not directly affect the behavior of bit 0).

We incorporate all of these effects into a model capturing the time-invariant behavior of the electronics, which is then fit to the data. The model includes 9 parameters which are used to predict the offsets introduced by the binary offset effect as a simple function of the three pixel values T-2, T-3, and O-3. The terms can be summarized as follows: a base effect in the number of bits in T-2 are "1" that depends on whether the same bits are "1" in T-3 or not (2 parameters), larger effects for more significant bits in T-2 and T-3 (2 parameters), and a 2-dimensional polynomial in how many bits are "1" in T-3 and how many are "1" in O-3 in order to capture the cross-amplifier effects (5 parameters). The zeropoint is arbitrarily chosen such that the model predicts an offset of 0 when all of the driver pixels are 0. These terms and the effects that they capture are shown in Table 2.2.

A challenge in fitting such a model to data is that effects other than the binary offset effect can introduce correlations between the observed driver and target pixels. For example, cosmic rays will produce large signals on the detector that will show up in the residual images. To mitigate this, we mask out cosmic rays (and any other bright pixels) by identifying any pixels that are over five standard deviations above the background noise level in the residual images and flagging a region with a border of 2 pixels around them. Another challenge is that low-order spatial variations over the image can also introduce local correlations between pixel values. We perform a background subtraction on the residual images using `sep` (Barbary, 2016; Bertin & Arnouts, 1996) with $64 \times 64$ pixel boxes to reduce the potential impact of these variations. A more challenging feature of the data is a pickup signal that appears in many of the SNIFS images with a frequency of 15–20 kHz (a period of roughly 3 pixels) and an amplitude of $\sim 1$ ADU. This pickup signal is very challenging to model due to its low amplitude, and it effectively introduces a smooth variation in the mean of the target pixel residuals as a function of the driver pixel value. We note that the binary offset effect introduces large offsets into the target pixel value when a driver pixel value increases by 1 ADU while most other effects are continuous. We therefore fit for the effect of a 1 ADU change in a driver pixel on the target pixel rather than trying to fit for the offset that was added to each target pixel as a function of the driver pixel values directly.

The final fitting procedure is as follows. Given a raw image, we mask out pixels with known issues, and then perform a background subtraction to obtain a residual image. For every unique set of three driver pixel values T-2, T-3, and O-3, we find all target pixels in the raw image with those driver pixel values. We calculate the mean value of the residuals for those target pixels and we estimate the uncertainty on that mean value. Note that the mean value of the target pixel residuals is a sum of the amplitude of the binary offset effect and other effects like pickup that introduce correlated residuals. After calculating the mean of the target pixel residuals for every combination of driver pixel values, we identify sets of driver pixel values where one driver pixel value changed by 1 ADU, and we calculate the difference in the mean of the target pixel residuals associated with that change in driver pixel value and the measurement uncertainty on this difference. We fit our model to these

Table 2.2: Variables used in the model of the binary offset effect in SNIFS data. The full model involves linear corrections for each of these terms. The coefficients of these linear corrections differ for the different amplifiers, but remain stable over time. In this table we use the following notation: T-N refers to the driver pixel read N pixels previously and O-N refers to the driver pixel read on the other amplifier N pixels previously. The uncertainties shown on the parameters are the standard deviations of fits to 10 different datasets.

| Explained effect | Id | Variable | Fitted parameters for SNIFS (ADU/unit variable) | | | |
|---|---|---|---|---|---|---|
| | | | Blue channel | | Red channel | |
| | | | Left amplifier | Right amplifier | Left amplifier | Right amplifier |
| Base effect | 1 | Number of bits that are "1" in both T-2 and T-3 | $0.663 \pm 0.012$ | $0.342 \pm 0.006$ | $-0.503 \pm 0.006$ | $-0.091 \pm 0.004$ |
| | 2 | Number of bits that are both "1" in T-2 and "0" in T-3. | $0.081 \pm 0.004$ | $0.043 \pm 0.005$ | $-0.247 \pm 0.005$ | $-0.140 \pm 0.003$ |
| Larger effects for more significant bits | 3 | Sum of the indices of bits that are "1" in both T-2 and T-3. | $0.042 \pm 0.003$ | $0.034 \pm 0.002$ | $0.010 \pm 0.007$ | $0.022 \pm 0.002$ |
| | 4 | Sum of the indices of bits that are both "1" in T-2 and "0" in T-3. | $0.049 \pm 0.003$ | $0.015 \pm 0.003$ | $0.006 \pm 0.007$ | $0.021 \pm 0.002$ |
| Two-dimensional polynomial capturing interaction across amplifiers | 5 | Number of bits that are "1" in T-3. | $-0.410 \pm 0.027$ | $-0.076 \pm 0.037$ | $0.020 \pm 0.030$ | $0.107 \pm 0.023$ |
| | 6 | Number of bits that are "1" in O-3. | $0.065 \pm 0.017$ | $-0.036 \pm 0.024$ | $0.144 \pm 0.024$ | $-0.125 \pm 0.025$ |
| | 7 | Number of bits that are "1" in T-3, squared. | $0.035 \pm 0.002$ | $0.024 \pm 0.004$ | $0.019 \pm 0.003$ | $0.002 \pm 0.002$ |
| | 8 | Number of bits that are "1" in O-3, squared. | $0.037 \pm 0.002$ | $0.016 \pm 0.003$ | $-0.001 \pm 0.002$ | $0.019 \pm 0.003$ |
| | 9 | (Number of bits that are "1" in T-3) $\times$ (Number of bits that are "1" in O-3) | $-0.072 \pm 0.005$ | $-0.039 \pm 0.006$ | $-0.018 \pm 0.003$ | $-0.020 \pm 0.003$ |

differential measurements which mitigates the impact of effects like pickup.

We fit the 9-parameter model described in Table 2.2 to a subset of bias images taken roughly evenly spaced in time across the full history of the SNIFS instrument by performing a $\chi^2$ minimization. There are 98 images included in this fit for the blue channel and 99 images for the red channel. Each sample is split into 10 subsamples, and we fit the model parameters on each of these subsamples individually. We then calculate the mean model parameters from each of these fits, and we use the standard deviation of the fitted parameters across subsamples as an estimate of the systematic uncertainty associated with the model. The fitted model parameters are shown in Table 2.2.

We find that when a bit is "1" in both pixels T-2 and T-3, an offset of up to $0.663 \pm 0.012$ ADU is introduced in the target pixel. The fitted model parameters are not consistent across amplifiers: on the red channel, the left amplifier has an offset of $-0.503 \pm 0.006$ ADU per bit that is "1" in both T-2 and T-3 while the right amplifier has an offset of $-0.091 \pm 0.004$ ADU per "1" bit. The introduced offsets for the interaction across amplifiers can be up to $0.144 \pm 0.024$ ADU per "1" bit for the left amplifier on the red channel, and the fit finds a strong interaction between the bits in pixels T-3 and O-3 on the blue channel.

The model is not a perfect description of the effect: when fitting the model to the data, we find that the chi-squared per degree of freedom ($\chi^2/DoF$) of the fits is between 1.11 and 1.23 when using the measurement uncertainties estimated in the previously described procedure, while it would be 1 if the model perfectly described the data. We estimate the remaining dispersion of the data due to the binary offset effect by adding an uncertainty term in quadrature to the measurement uncertainty in order to set the $\chi^2/DoF$ to 1. For the blue channel, this requires an additional 0.156 ADU and 0.164 ADU of dispersion for the left and right amplifiers respectively. For the red channel, this requires an additional 0.126 ADU and 0.110 ADU of dispersion for the left and right amplifiers respectively.

We estimate the uncertainty in the model parameters by taking the standard deviation of the fits to the 10 different subsets. As a conservative estimate of the systematic uncertainties, we adopt the standard deviations of the fitted parameters directly rather than attempting to take out the statistical component by combining datasets. The derived uncertainties are shown in Table 2.2. These uncertainties are highly correlated, especially the terms relating higher orders in the same bit counts (eg: parameters 5 and 7) and the terms related to larger effects for more significant bits (eg: parameters 1 and 3). We estimate the impact of the variation in the model parameters on the derived corrections by repeatedly sampling realizations of model parameters from a Gaussian distribution following the full fitted covariance matrix of the model parameters. For each realization of model parameters, we calculate the implied corrections for a group of images. Across realizations, we find that the standard deviation of the implied corrections is less than 0.01 ADU for images on the blue channel, and less than 0.005 ADU for images on the red channel, so the uncertainties in the model parameters do not significantly affect the final corrections. The uncertainty of the correction is therefore dominated by the unexplained residual dispersion described previously, and is between 0.11–0.16 ADU depending on the amplifier. We do not detect any significant variation in the behavior of the binary offset effect or the model parameters over the history of

SNIFS (from 2004 to 2017).

## 2.7 Appendix: Code used to identify the Binary Offset Effect in arbitrary instruments.

A `Jupyter` notebook is available on GitHub[2] with code that can be used to identify the binary offset effect in arbitrary instruments. The main algorithms from this notebook are reproduced in this Appendix for posterity:

```
import sep
import glob
import numpy as np
from scipy.stats import sigmaclip
from scipy.signal import convolve2d
from matplotlib import pyplot as plt


def subtract_background(data):
    """Subtract the background from data.

    We assume that we are working with a bias or dark frame, so we
    don't have to worry about detecting and masking out objects.
    """
    background = sep.Background(data)

    sub_data = data - background.back()

    return sub_data


def build_cosmic_mask(data):
    """
    Build a cosmic ray mask for a residual image.

    This is a very naive algorithm. We are running on residuals, so we
    cut anything that is 5 sigma high or low. We then mask by a couple
    pixels around the cut too to capture the tail.
    """
    target_nmad = nmad(data)
    start_mask = np.abs(data - np.median(data)) > 5. * target_nmad

    num_edge_pixels = 2
    mask = ~(convolve2d(
        start_mask,
        np.ones((2*num_edge_pixels+1, 2*num_edge_pixels+1)),
        mode='same'
```

[2] https://www.github.com/snfactory/binaryoffset

```
38      ).astype(bool))
39
40      return mask
41
42
43  def clipped_mean(data):
44      clip_data, min_clip, max_clip = sigmaclip(data)
45
46      mean = np.mean(clip_data)
47      mean_err = np.std(clip_data) / np.sqrt(len(clip_data) - 1)
48
49      return mean, mean_err
50
51
52  def nmad(data, *args, **kwargs):
53      return 1.4826 * np.median(
54          np.abs(np.asarray(data) - np.median(data, *args, **kwargs)),
55          *args, **kwargs
56      )
57
58
59  _cache_num_ones = {}
60  def cached_num_binary_ones(x):
61      try:
62          return _cache_num_ones[x]
63      except KeyError:
64          num_ones = _num_binary_ones(x)
65          _cache_num_ones[x] = num_ones
66          return num_ones
67
68
69  def _num_binary_ones(x):
70      return bin(x).count('1')
71
72
73  def calc_num_binary_ones(x):
74      x = np.atleast_1d(x)
75      out = np.zeros(x.shape)
76      for i in range(len(x.flat)):
77          out.flat[i] = cached_num_binary_ones(x.flat[i])
78
79      return out
80
81
82  def calculate_mean_residual(data, offset, do_diff=True,
83          residuals=None):
84      """Calculate the mean residuals as a function of the value of a
85      reference pixel.
86
87      This will return a list of raw pixel values and the average
```

```
88      residuals for pixels separated by "offset" from pixels with those
89      values.
90
91      If do_diff is True, we look at a difference between the offset in
92      the left direction and the right direction to cancel out effects
93      like pickup. Otherwise, we only look in the direction specified
94      by offset.
95
96      The sign of offset doesn't really matter if used in diff mode since
97      the positive and negative offsets will be subtracted from each
98      other. However, if do_diff is False then they do matter. I have not
99      determined the readout direction on each instrument tested here, so
100     a positive offset may or may not be in the readout direction.
101     """
102     if residuals is None:
103         # If we weren't given residuals, just assume that the image is
104         # relatively flat and do a local background subtraction.
105         residuals = subtract_background(data.astype(float).copy())
106
107     # We look at the difference in residuals from one side to the next.
108     ref_1 = np.roll(data, offset)
109     ref_2 = np.roll(data, -offset)
110
111     # Generate a mask to filter out cosmic rays/bad pixels.
112     mask = build_cosmic_mask(data)
113     use_residuals = residuals[mask]
114     use_ref_1 = ref_1[mask]
115     use_ref_2 = ref_2[mask]
116
117     # Figure out which range of pixel values to probe.
118     min_values = np.percentile(data, 2)
119     max_values = np.percentile(data, 98)
120     pixel_values = np.arange(min_values, max_values+1)
121
122     residual_means_1 = []
123     residual_means_2 = []
124     residual_errs_1 = []
125     residual_errs_2 = []
126
127     for pixel_value in pixel_values:
128         cut_1 = use_ref_1 == pixel_value
129         cut_2 = use_ref_2 == pixel_value
130
131         # For each value that is found in at least 20 pixels we
132         # calculate a clipped mean of the residuals and estimate an
133         # error on it.
134         if np.sum(cut_1) > 20:
135             mean_1, mean_err_1 = clipped_mean(use_residuals[cut_1])
136         else:
137             mean_1 = np.nan
```

```
138             mean_err_1 = np.nan
139
140         if np.sum(cut_2) > 20:
141             mean_2, mean_err_2 = clipped_mean(use_residuals[cut_2])
142         else:
143             mean_2 = np.nan
144             mean_err_2 = np.nan
145
146         residual_means_1.append(mean_1)
147         residual_means_2.append(mean_2)
148         residual_errs_1.append(mean_err_1)
149         residual_errs_2.append(mean_err_2)
150
151     residual_means_1 = np.array(residual_means_1)
152     residual_means_2 = np.array(residual_means_2)
153     residual_errs_1 = np.array(residual_errs_1)
154     residual_errs_2 = np.array(residual_errs_2)
155
156     # Calculate the difference in residuals in each direction. This is
157     # what we will look at to see if the binary offset effect is
158     # present since it is relatively clean.
159     if do_diff:
160         residual_means = residual_means_1 - residual_means_2
161         residual_errs = np.sqrt(residual_errs_1**2
162                                 + residual_errs_2**2)
163     else:
164         residual_means = residual_means_1
165         residual_errs = residual_errs_1
166
167     return pixel_values, residual_means, residual_errs
168
169
170 def setup_figure():
171     fig, (ax1, ax2) = plt.subplots(2, 1, sharex=True,
172         gridspec_kw={'height_ratios': [2.5, 1], 'hspace': 0.})
173
174     fig_parts = (fig, ax1, ax2)
175
176     return fig_parts
177
178
179 def plot_mean_residual(data, offset, label=None, do_diff=True,
180                        residuals=None, fig_parts=None):
181     """Plot the mean residuals calculated using
182     calculate_mean_residual.
183     """
184     pixel_values, residual_means, residual_errs = \
185         calculate_mean_residual(data, offset, do_diff=do_diff,
186                                 residuals=residuals)
187
```

```
188     if fig_parts is not None:
189         fig = fig_parts[0]
190         ax = fig_parts[1]
191     else:
192         fig = plt.gcf()
193         ax = fig.add_subplot(111)
194         xlabel = 'Value of driver pixel (ADU)'
195         ax.set_xlabel(xlabel)
196 
197         if offset == 1:
198             offset_label = '1 pixel'
199         else:
200             offset_label = '%d pixels' % offset
201         plt.ylabel('Mean residuals for an offset of %s (ADU)' %
202             offset_label)
203 
204     ax.plot(pixel_values, residual_means, label=label)
205     ax.fill_between(
206         pixel_values,
207         residual_means-residual_errs,
208         residual_means+residual_errs,
209         alpha=0.2
210     )
211 
212     print('Median residual errors: %.4f' % np.median(residual_errs))
213 
214     return pixel_values
215 
216 
217 def finalize_figure(fig_parts, show_bits=4, min_bits=0, max_bits=11,
218         legend_loc=0):
219     fig, ax1, ax2 = fig_parts
220     label_bit_changes(show_bits, ax1)
221     labelled_locations = label_bit_changes(show_bits, ax2,
222                                            legend=False)
223 
224     ax1.set_ylabel('Residuals (ADU)')
225 
226     xmin, xmax = ax1.get_xlim()
227     x = np.arange(int(np.floor(xmin)), int(np.ceil(xmax)))
228 
229     x_binary_count = calc_num_binary_ones(x)
230     ax2.plot(x, x_binary_count, c='C0',
231         label='Number of "1" bits in binary representation')
232     ax2.set_ylabel('Number of "1" bits')
233     ax2.set_xlabel('Driver pixel value (ADU)')
234     ax2.set_ylim(min_bits, max_bits)
235     ax2.legend(loc=legend_loc)
236 
237     ax2.set_xticks(labelled_locations)
```

```
238
239     ax1.legend()
240
241
242 def calc_num_bits(x):
243     return bin(int(x)).count('1')
244
245
246 def label_bit_changes(min_bits=4, axis=None, legend=True):
247     if axis is None:
248         axis = plt.gca()
249
250     x_lim = plt.gca().get_xlim()
251     x_start = int(np.floor(x_lim[0]))
252     x_end = int(np.ceil(x_lim[-1]))
253
254     x_range = np.arange(x_start, x_end)
255
256     first = True
257     labelled_locations = []
258     for x in x_range:
259         num_diff_bits = calc_num_bits(x ^ (x+1))
260         if num_diff_bits >= min_bits:
261             if first and legend:
262                 label = 'Large binary transitions'
263                 first = False
264             else:
265                 label = None
266             axis.axvline(x+0.5, c='black', ls='--', alpha=0.2,
267                 label=label)
268
269             labelled_locations.append(x + 0.5)
270
271     return labelled_locations
```

The following code shows an example of how to use the previous code to process a set of dark files from the DEIMOS instrument. This code produces the plot shown in Figure 2.6a:

```
1 filenames = glob.glob('./data/deimos/*.fits.gz')
2
3 fig_parts = setup_figure()
4
5 for filename in filenames:
6     index = 4
7     print("Processing", filename)
8
9     with fits.open(filename) as fits_file:
10         data = fits_file[index].data
11         target_name = filename.split('/')[-1][:-8]
12         plot_mean_residual(data, 2, label=target_name,
13                            fig_parts=fig_parts)
```

```
14
15 finalize_figure(fig_parts, show_bits=4, legend_loc=3, min_bits=-5)
16
17 plt.savefig('./binary_offset_deimos.pdf')
```

# Chapter 3

# Mapping the Diversity of Type Ia Supernovae with Manifold Learning

## 3.1 Overview

In Fakhouri et al. (2015) (hereafter F15), we showed that pairs of "twin supernovae", identified by matching their spectral features, can be used to standardize Type Ia supernovae (SNe Ia) with a much lower dispersion in brightness than template based methods. While this standardization method is effective, it can only be applied for SNe Ia with spectrophotometry near maximum light. In this chapter, we build upon the twins analysis, and we show how we can use similar techniques to develop a new non-linear parametrization of the intrinsic diversity of SNe Ia. Standardization of SNe Ia with this parametrization results in low dispersions in brightness, similar to the twins method. With our parametrization, we can identify subsets of SNe Ia that are not correctly standardized by traditional standardization methods such as SALT2 (Betoule et al., 2014) with systematic offsets of up to 0.3 mag for some regions of the parameter space. We show that correlations between the standardized brightnesses and host galaxy properties can be greatly reduced using our parametrization of SNe Ia.

This analysis is laid out as follows. First, we discuss the dataset that we use in Section 3.2. We then describe a method to model the differential evolution of the spectra of SNe Ia and estimate their spectra at maximum light in Section 3.3. In Section 3.4, we introduce a procedure that we call "Reading Between the Lines", where we use the regions of the supernova spectra with low spectral diversity to remove the extrinsic contributions to the spectra from distance uncertainties and extinction due to dust. This analysis produces dereddened spectra that only have intrinsic variation of SNe Ia remaining. Using these dereddened spectra, we perform a manifold learning embedding in Section 3.5 and examine the recovered parameter space of SNe Ia. In Section 3.6, we show how the recovered parameter space of SNe Ia can be used to standardize SNe Ia, and evaluate the performance of our methods. Finally, in Section 3.7, we discuss future extensions to this standardization method and how it can be

applied to other datasets.

## 3.2 Dataset

For this analysis, we make use of the spectrophotometric timeseries of SNe Ia obtained by the Nearby Supernova Factory (SNfactory; Aldering et al., 2002). The spectrophotometric timeseries were collected using the Super Nova Integral Field Spectrograph (SNIFS; Lantz et al., 2004). The SNIFS spectroscopic channels consist of two lenslet integral field spectrographs (IFS, "à la TIGER"; Bacon et al., 1995, 2001), which split a fully-filled $6.''4 \times 6.''4$ field of view into a grid of $15 \times 15$ spatial elements. The two spectrographs cover the 3200–5200 Å and 5100–10000 Å wavelength ranges simultaneously. A photometric channel images the field around the IFS to monitor atmospheric transmission. The SNIFS instrument is continuously mounted on the south bent Cassegrain port of the University of Hawaii 2.2 m telescope on Mauna Kea.

The spectra from SNIFS were reduced using the SNfactory data reduction pipeline (Bacon et al., 2001; Aldering et al., 2006; Scalzo et al., 2010). The flux calibration procedure for this pipeline is described in Buton et al. (2013), and the host-galaxy subtraction procedure is presented in Bongard et al. (2011). The full SNfactory dataset at the time of this analysis consists of a total of 3951 spectra of 438 different SNe Ia between the redshifts of 0.001 to 0.14.

We fit the light curve of each of the SNe Ia in our sample using the SALT2 light curve fitter (Betoule et al., 2014) version 2.4 that is currently used for most cosmological analyses with SNe Ia (e.g., Scolnic et al., 2018). To perform these fits, we synthesize synthetic photometry in the SNfactory U, B, V, R, and I bands defined as tophat filters with wavelength ranges shown in Table 3.1. As the SALT2 model is unreliable outside of the optical bands, we fit the SALT2 model using only the B, V, and R bands, but we calculate the model residuals in all bands. We use these SALT2 fits for several different purposes. First, for this analysis, we are specifically interested in the spectra from the SNfactory dataset near maximum light. We use the SALT2 fits to determine the time of maximum light for each supernova. To ensure that we have a reasonable determination of the time of maximum, we require that each supernova have at least five spectra, and that the SALT2 day of maximum uncertainty is less than one day. We then keep all of the spectra within five restframe days of maximum light for our analysis, and reject any supernova that doesn't have such a spectrum.

We also compare the standardization of the luminosity of SNe Ia with our techniques to standardization with the SALT2 fitter in Section 3.6. For these analyses, we impose additional criteria on the SALT2 fits to ensure that the fit quality is acceptable. First, we impose phase coverage criteria similar to Guy et al. (2010), requiring at least four measurements at restframe times $t$ relative to maximum light satisfying $-10 < t < 35$ days, at least one measurement with $-10 < t < 7$ days, at least one measurement with $7 < t < 20$ days and measurements in at least two different synthetic filters with $-8 < t < 10$ days. We then require that the normalized median absolute deviation (NMAD) of the SALT2 model

Table 3.1: SNfactory synthetic filter definitions. We synthesize photometry from the SNIFS spectra in various bands using tophat filters with perfect throughput in a given wavelength range and no throughput outside of that range. The wavelength ranges for the different bands are shown in this table.

| Band name | Wavelength range (Å) |
|---|---|
| U | 3300–4102 |
| B | 4102–5100 |
| V | 5200–6289 |
| R | 6289–7607 |
| I | 7607–9200 |

residuals be less than 0.12 mag and that no more than 20% of the SALT2 model residuals have an amplitude of more than 0.2 mag.

For analyses on the spectra of SNe Ia, we shift the wavelengths of all of our spectra to the supernova's restframe, and we rebin these spectra with a common binning of 1000 km/s between 3300–8600 Å. This results in a total of 288 wavelength bins, and is the same binning used in F15 and Saunders et al. (2018). Assuming a set of fiducial cosmological parameters from Komatsu et al. (2011), we shift the brightness of each supernova so that they all appear to be at a common redshift of 0.05.

We find that low signal-to-noise (S/N) spectra from the SNIFS instrument have large systematic uncertainties, especially at the bluer ends of the spectra. These uncertainties appear to be primarily due to poor fits of the model in the extraction from the CCD, which introduces a correlated offset that gets larger towards bluer wavelengths. We do not notice strong biases in the SNIFS spectra for repeated observations of the same target. To illustrate this effect, we examined the SALT2 model residuals for all of the SNe Ia in the SNfactory sample passing the SALT2 selection criteria previously described. To probe for large systematic uncertainties, we examined how the fraction of SALT2 residuals with an amplitude of more than 0.2 mag depends on various measurement properties. First, the SALT2 model has large modeling uncertainties in the U and I bands, especially for phases away from maximum light. We plot the binned outlier fraction as a function of SALT2 phase in Figure 3.1. We find that the outlier fractions are lowest in most bands near maximum light. In Figure 3.2, we show the fraction of outliers both with and without a SALT2 phase cut between $-3$ and $+15$ days as a function of the statistical S/N in the bluest 500 Å of the spectrum (3300 to 3800 Å).

Regardless of whether or not the phase cuts are applied, the outlier fraction has a strong dependence on S/N, especially in the U-band. Note that the S/N in the U-band is roughly twice the S/N in the bluest 500 Å of the spectrum, so for a S/N of 10 in this bluest 500 Å of the spectrum, we would only expect less than 1% of U-band SALT2 residuals to have an amplitude of more than 0.2 mag due to statistical uncertainty. The observed outlier fraction of $\sim$60% is entirely due to correlated systematic uncertainties in the extraction of the spectrum from the CCD. In this analysis, we are interested in understanding the intrinsic

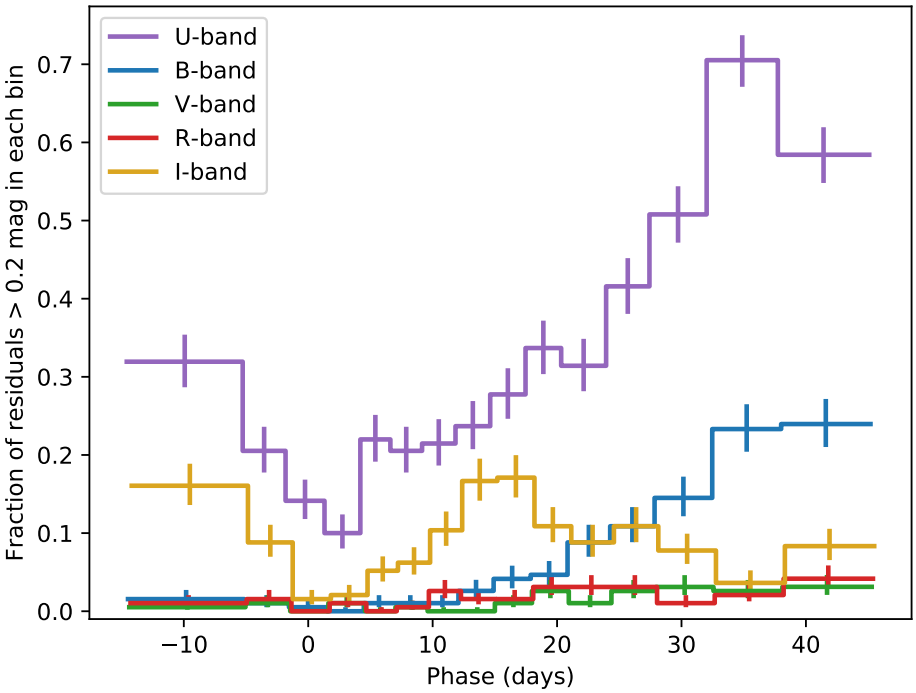



Figure 3.1: SALT2 residual outliers as a function of SALT2 phase. From the full set of valid SALT2 fits to SNfactory light curves of SNe Ia, we calculate the SALT2 residuals in each of the SNfactory bands described in Table 3.1. For each band, we calculate the fraction of spectra that have SALT2 fits residuals with an amplitude larger than 0.2 mag in bins of the SALT2 phase. The results of this procedure are shown in different colors for each band. Uncertainties on the outlier fractions in each bin are estimated using the Wilson score interval (Wilson, 1927).

spectral diversity of SNe Ia, and large instrumental sources of spectral diversity would be recovered in such an analysis and potentially confused with intrinsic spectral diversity. To avoid this issue, we require that the statistical S/N of all of the spectra used in this analysis be larger than 100 in the bluest 500 Å of the spectrum.

We also examined how the SALT2 residuals of SNIFS spectra correlate with other properties of the observations. We find that the statistical S/N in the bluest 500 Å is a better predictor of SALT2 outliers than the S/N in other redder bands because the outliers are primarily in the bluer bands: there are almost no outliers in the V or R bands, and the outliers in the I-band appear to be primarily due to SALT2 modeling uncertainties rather than S/N issues, as can be seen in Figures 3.1 and 3.2. We examined the effect of redshift, airmass, seeing, sky background level, properties of standard stars used for calibration on a

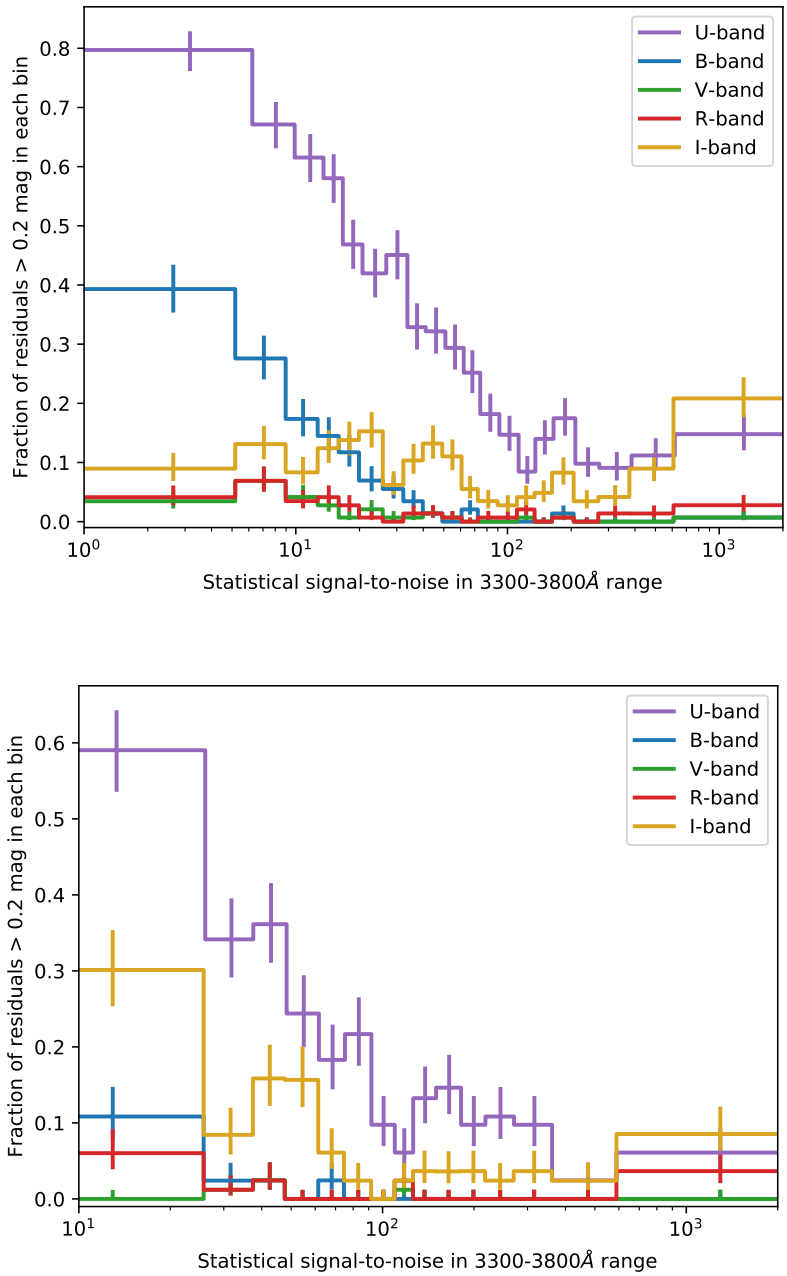



Figure 3.2: SALT2 residual outliers as a function of the statistical S/N in the bluest 500 Å of the spectrum (3300 Å to 3800 Å). See Figure 3.1 for details of the plot. Top panel: SALT2 outlier fraction in bins of S/N for all spectra. Bottom panel: SALT2 outlier fraction in bins of S/N for spectra with SALT2 phases between −3 and +15 days.

Table 3.2: Summary of sample selection requirements. The general selection requirements, listed in the first section of the table, are applied to all analyses. For other analyses described in this document, we apply different selection requirements, listed in the different sections of the table.

| Selection Requirement | Number of Supernovae passing requirement |
|---|---|
| **General selection requirements** | |
| Initial sample | 415 |
| More than 5 spectra | 280 |
| SALT2 date of maximum light uncertainty $< 1$ day | 277 |
| At least one spectrum within 5 days of maximum light | 254 |
| At least one spectrum with S/N 3300-3800 Å $> 100$ | 207 |
| **Estimation of the spectra at maximum light (Section 3.4.4)** | |
| Spectrum uncertainty $< 10\%$ of intrinsic power | 175 |
| **Valid supernova brightness requirements (Section 3.6.1)** | |
| Host galaxy redshift available | 170 |
| Host galaxy redshift above 0.02 | 146 |
| Measured $A_V < 0.5$ mag | 135 |
| Blinded training subsample | 73 |
| **Comparisons to SALT2 fits (Section 3.6.6)** | |
| SNfactory SALT2 selection requirements | 184 |
| Passes host galaxy redshift and color requirements | 152 |
| Has a valid interpolation to maximum light | 128 |
| Blinded training subsample | 66 |

given night, photometricity of the night, moon location, detector temperatures, time that the detector was on, and many other variables on the SALT2 residuals. For the latest SNfactory spectral reductions, we do not notice any relationships between the SALT2 residuals and these properties beyond S/N. A summary of the attrition for each of these steps, along with analysis specific selections that will be discussed in the following sections, are shown in Table 3.2. A total of 207 supernovae pass all of the previously described selection requirements with a total of 606 spectra within five days of maximum light passing the S/N requirements.

Note that our goal is to have a sample that covers the full diversity of SNe Ia. We include every target in the SNfactory dataset that was identified as an SN Ia in this analysis, including SNe Ia that have been labeled as "peculiar". The objective of this analysis is to understand the full diversity of the spectra of SNe Ia near maximum light, and understanding how peculiar SNe Ia are related to "normal" SNe Ia is therefore crucial to our analysis. One potential issue is that the requirement on S/N in the 3300-3800 Å wavelength range could

bias our analysis towards a specific subtype of SNe Ia. However, the variation in S/N simply due to the observed brightnesses of SNe Ia at the wide range of different redshifts considered is much larger than the intrinsic variation in brightness in this band, and is unlikely to introduce any significant selection bias into our analysis.

## 3.3 Estimating the Spectra of SNe Ia at Maximum Light

To compare different SNe Ia to each other, their spectral time series must be interpolated so that their spectra are compared at the same times relative to maximum light (their "phases"). Even though we have only included supernovae with spectra within five days of maximum light for this analysis, we could still potentially be comparing spectra with phases that differ by as much as ten days to each other if we do not perform any sort of interpolation in phase.

In F15, we performed this interpolation using Gaussian Process (GP) regression to generate models of the spectral timeseries for each supernova. This method is very effective when the time series is well-sampled, but each supernova is fit completely independently of all other supernovae, so the GP predictions have large uncertainties when interpolating or extrapolating a spectrum more than a couple days from observed data. As a result, 11 of the 66 supernovae that could otherwise have been used in the F15 analysis were rejected due to having unreliable GP predictions, a somewhat subjective cut that relied on human judgement about whether the GP predictions seemed like a reasonable model of the light curve.

In this work, we build a new interpolation method that simultaneously models the time evolution of all SNe Ia in a sample. With such a model, if we know the time of maximum for a supernova, then we can estimate the spectrum of that supernova at maximum light even if we have only a single spectrum of that supernova and a reliable estimate of the time of maximum. We do this by modeling the differential evolution of supernovae rather than the supernova spectra themselves. We find that despite the large differences in supernova spectra near maximum light, the spectra of SNe Ia evolve in time in a remarkably similar way. An example of this is seen in Figure 3.3.

### 3.3.1 Modeling the Differential Time Evolution of SNe Ia

We take advantage of the similarity in the time evolution of SNe Ia to build a differential model that can be applied to supernovae to both interpolate their spectra to maximum light and to produce an estimate of the uncertainty of the interpolated spectra. We assume that the differential evolution of a supernova from maximum light to some nearby phase can be modeled in each wavelength bin using a quadratic function in wavelength. We also include a correction to this function that depends on the stretch of the supernova measured as SALT2's $x_1$ parameter:

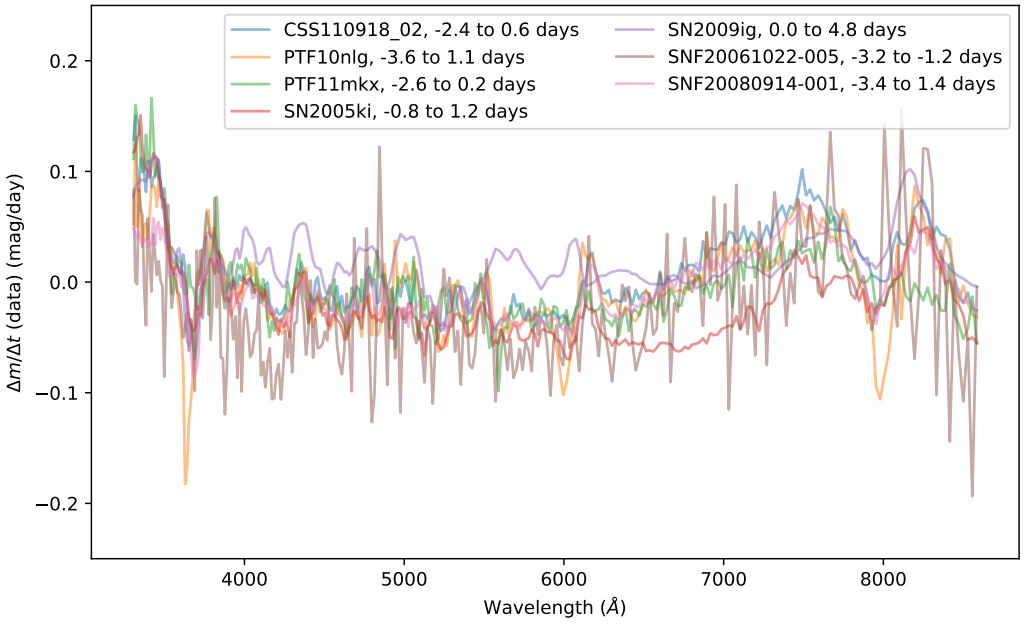



Figure 3.3: Comparison of the spectral evolution of different SNe Ia near maximum light for a randomly chosen subset of supernovae in the sample. For each supernova, we take its spectra at two different phases, and we calculate the difference in brightness in magnitudes per day between the two spectra. Despite the fact that these supernovae have a wide range of different properties (for example, SALT2 $x_1$ values range from $-2.3$ to 1.2), they show remarkable similarity in their spectral evolution in time.

$$m_{true}(p;\lambda) = m_{max}(\lambda) + p \cdot c_1(\lambda) + p^2 \cdot c_2(\lambda) + p \cdot x_1 \cdot c_3(\lambda) + p^2 \cdot x_1 \cdot c_4(\lambda) \tag{3.1}$$

where $m_{max}(\lambda)$ is the flux of the supernova at maximum light in magnitudes, $m_{true}(p;\lambda)$ is the flux of the supernova at phase $p$ in days away from maximum light, $x_1$ is the SALT2 $x_1$ parameter for that supernova, and $c_1(\lambda)$ through $c_4(\lambda)$ are arbitrary functions of wavelength that are fixed for all SNe Ia and that represent the time evolution of supernovae near maximum light. Note that because we are only modeling the differential evolution of the supernova's spectrum, any static extrinsic effects such as the distance to the supernova or reddening due to static interstellar dust will be absorbed into $m_{max}(\lambda)$ and have no impact on the model.

This simple model will not be able to capture all of the diversity in spectral evolution of SNe Ia. To account for this, we model the dispersion of the interpolation residuals as a function of phase using a broken linear function that is fixed to zero at maximum light (where there is no interpolation uncertainty by definition) and with nodes at $-5$, $-2.5$, 2.5, and 5 days for each wavelength bin. Another potential challenge with the SNfactory dataset

is that we find "gray dispersions" in brightness of a flat ∼0.02 mag independent of wavelength for spectra of standard stars after calibration. We add these gray dispersions to our model of the observed supernova flux. Finally, we include the measurement uncertainties in our model. Combining all of these effects together, the model of an observed supernova spectrum becomes:

$$f(p;\lambda) = 10^{-0.4(m_{true}(p;\lambda)+m_{gray})} \tag{3.2}$$

$$m_{gray} \sim N(0;\sigma^2_{gray}) \tag{3.3}$$

$$\sigma^2_{obs}(p;\lambda) = \sigma^2_{meas}(\lambda) + (\epsilon_{interp}(p;\lambda)\cdot f(p;\lambda))^2 \tag{3.4}$$

$$f_{obs}(p;\lambda) \sim N(f(p;\lambda);\sigma^2_{obs}(p;\lambda)) \tag{3.5}$$

where $m_{gray}$ is the gray offset of the spectrum in magnitudes, and we assume that it is drawn from a normal distribution with a variance common to all spectra of $\sigma^2_{gray}$. $\sigma_{meas}(\lambda)$ is the measurement uncertainty of the observed spectrum. $\epsilon_{interp}(p;\lambda)$ is the previously described function representing the dispersion of the interpolation residuals, and as implemented it specifically represents the fractional uncertainty of the interpolation. Finally, $f_{obs}(p;\lambda)$ is the final observed spectrum at a given phase and wavelength, and we assume that it is drawn from a normal distribution with a mean of $f(p;\lambda)$ and a variance of $\sigma^2_{obs}(p;\lambda)$ which contains all of the previously described contributions to the variance.

We implement this model using the Stan modeling language (Carpenter et al., 2017), and fit it to our full sample of 606 spectra for 207 different SNe Ia. For each supernova, we fit for a single spectrum at maximum that combines the information from all of the different spectra of that supernova taken within 5 days of maximum light. With our spectra binned in 288 wavelength bins as described in Section 3.2, this model has a total of $288 \times 207$ parameters representing the spectra at maximum light, $288 \times 4$ parameters for the $c_i(\lambda)$ functions that represent the time evolution of spectra, $288 \times 4$ parameters for the broken linear function that we use to model the interpolation uncertainty $\epsilon_{interp}(p;\lambda)$, 606 parameters for the gray offsets $m_{gray}$ of each spectrum, and finally one parameter for the gray offset dispersion $\sigma^2_{gray}$.

We use Stan to optimize the parameters of this model to obtain the maximum a posteriori probability (MAP) estimate of the model parameters (Carpenter et al., 2017), including the predicted spectrum at the time of maximum light of each supernova. We propagate the various sources of uncertainty to obtain an estimate of the uncertainty on the spectrum at maximum light of each supernova.

This model is effectively using the supernovae that have observations at multiple phases to constrain the $c_i(\lambda)$ and $\epsilon_{interp}(p;\lambda)$ parameters that describe the spectral evolution in time of SNe Ia near maximum light. The actual spectrum at maximum light is then effectively predicted by shifting all of the different observations of the supernova to maximum light using these parameters, and taking the mean of these shifted observations weighted by the

observational and interpolation uncertainties. Supernovae that only have a single spectrum provide no constraint on the spectral evolution parameters. Instead, these parameters are simply used to shift the spectrum to maximum light, and to provide an estimate of the uncertainty on that shift.

### 3.3.2 Estimated Spectra at Maximum Light

An example of the results of our procedure to estimate the spectra of supernovae at maximum light can be seen for PTF13ayw in Figure 3.4. The difference in brightness between each spectrum of PTF13ayw and the predicted spectrum at maximum light can be seen in Figure 3.5, along with residuals of the modeled differences. The predicted models of the spectral differences relative to maximum light agree very well with the observed differences, and there is very little structure in the residuals. As described in the previous section, a single set of $c_i(\lambda)$ terms is used for all supernovae, so the modeled differences only depend on the phases of the different spectra and the SALT2 $x_1$ parameter of PTF13ayw. One subtlety with Figure 3.5 is that the spectrum at maximum light is estimated as part of the fitting procedure. This means that the magnitude zeropoint at each wavelength is effectively floating, and there are only two degrees of freedom in the model for each wavelength rather than the three that appear to be plotted.

A more direct way of illustrating the accuracy of the model's interpolation can be seen by looking at the results for direct differences of pairs of spectra from the same supernovae. In Figure 3.3, we plotted this difference for a randomly selected set of different supernovae in units of magnitudes per day. We plot the same difference after removing the estimated difference from our model in Figure 3.6. In this case, we are looking at the time evolution between two spectra directly, so the estimated spectrum at maximum light is irrelevant. The model depends only on the phases of the two spectra that are being compared, and the SALT2 $x_1$ value of that supernova. The original spectra showed differences due to spectral evolution in time of up to 0.2 mag/day. After applying our model of spectral evolution, the residuals appear to be almost entirely random noise with possibly some structure at the $< 0.05$ mag/day level near the Ca II lines around 3800 Å.

For supernovae that only have a single spectrum near maximum light, the model that is derived from supernovae with multiple spectra is effectively applied to estimate their brightness at maximum light. An example of this is shown in Figure 3.7 for the supernova SN2004gc. For this supernova, we only have a single spectrum at 3.42 days after maximum light. The model is nevertheless able to produce a robust estimate of the spectrum at maximum light along with an estimate of the uncertainty on this predicted spectrum.

### 3.3.3 Model of Spectral Evolution Near Maximum Light

The actual model of spectral evolution near maximum light for a supernova with a SALT2 $x_1$ of zero is shown in Figure 3.8. There are several interesting results in this plot. First, note that we aligned all of our supernovae to the SALT2-determined time of maximum

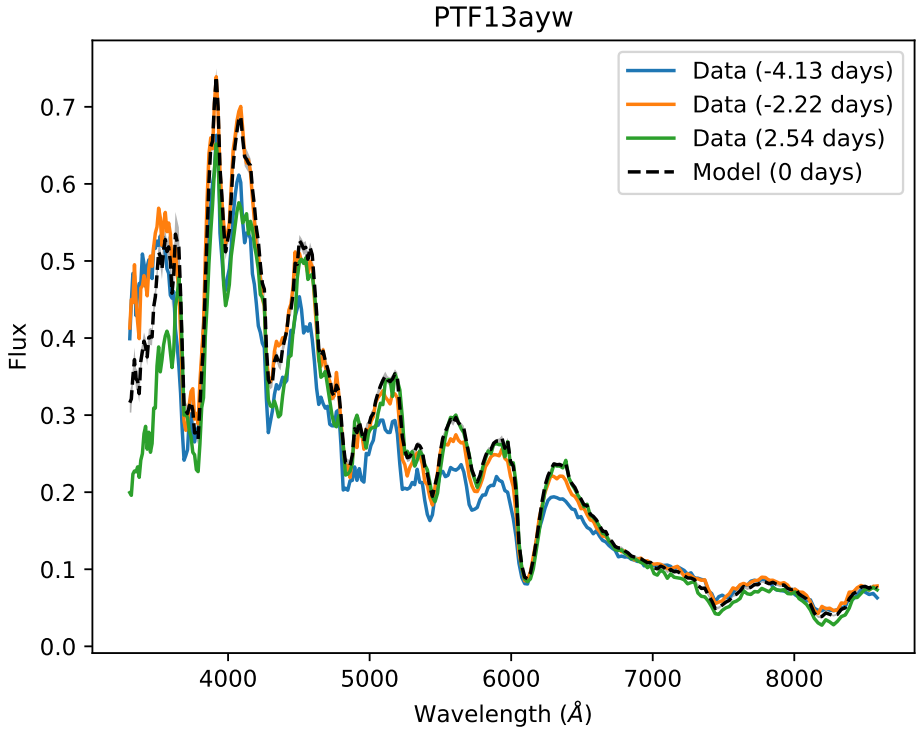



Figure 3.4: Estimated spectrum of PTF13ayw at maximum light. This supernova has three different spectra passing the selection criteria, shown in different colors. The information from all three of these spectra is used to predict the spectrum at maximum light, shown with a dashed black line. A shaded gray contour around this dashed black line shows the uncertainty on the estimate of the spectrum at maximum light, although the uncertainty is very small for this prediction so it is not easily visible.

light, which is the time of maximum light in the B-band, a filter that roughly corresponds to the integrated flux between 4000 and 5000 Å. As expected, our model predicts that the supernova gets fainter in either direction relative to maximum light in this wavelength band. We find that the time of maximum light is consistent from roughly 3900 to 6800 Å. However, for wavelengths bluer than 3900 Å or redder than 6800 Å, we find that the time of maximum light of the light curve occurs significantly earlier.

We also see very large variation in the light curve in the bluer and redder bands. For wavelengths between 3300 to 3500 Å, we find that the light curve declines by up to 0.2 mag/day, with this decline being increasingly rapid at later phases. Similarly, in the redder bands, we find that the light curve declines by up to 0.1 mag/day in two regions of the spectrum, corresponding to the OI absorption triplet and the Ca II IR triplet. If these effects are not taken into account, then the spectrum of a supernova observed 5 days after

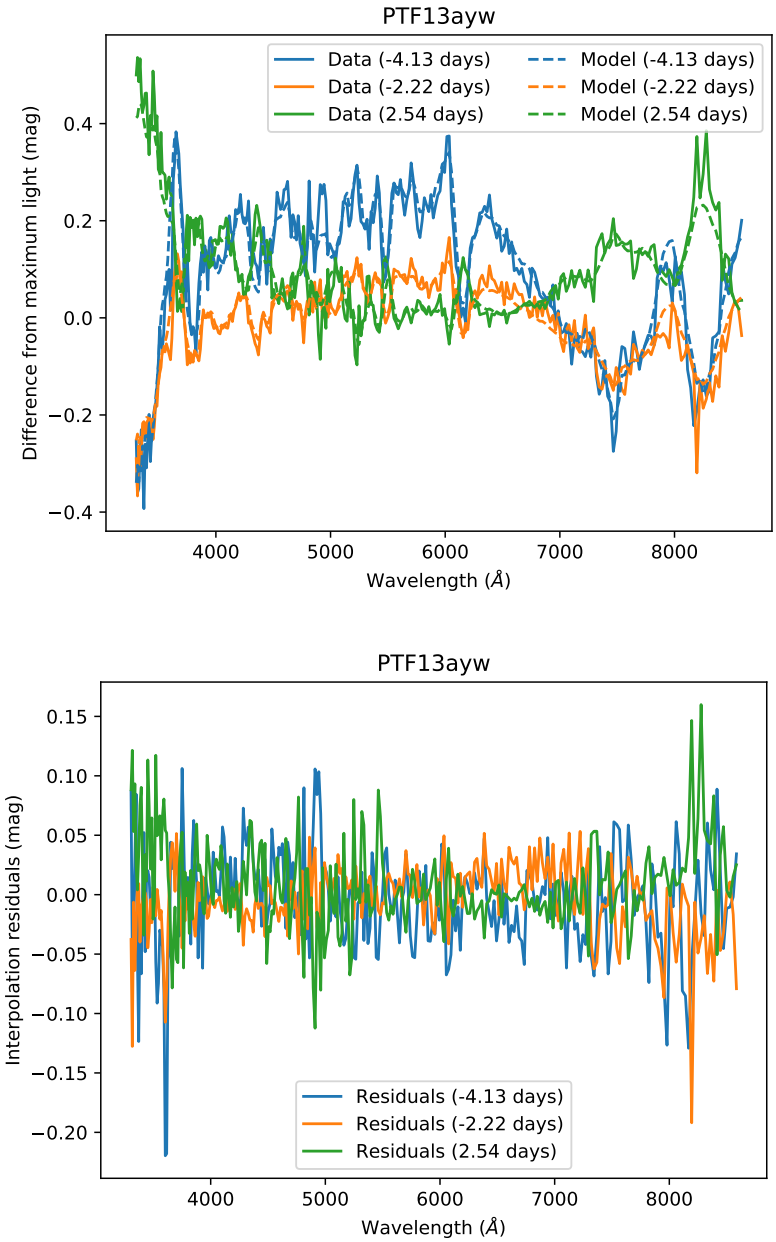



Figure 3.5: Top panel: Estimated differences between the observed spectra of PTF13ayw and its spectrum at maximum light. For each of the three spectra that we have for this supernova, we show the difference in brightness in magnitudes between that supernova and the predicted spectrum at maximum light that is produced by our model. For each spectrum, we overplot the modeled difference that is given by the $c_i(\lambda)$ terms. Bottom panel: Residuals of the measured differences compared to the modeled differences. For all three spectra there is very little structure to the residuals.

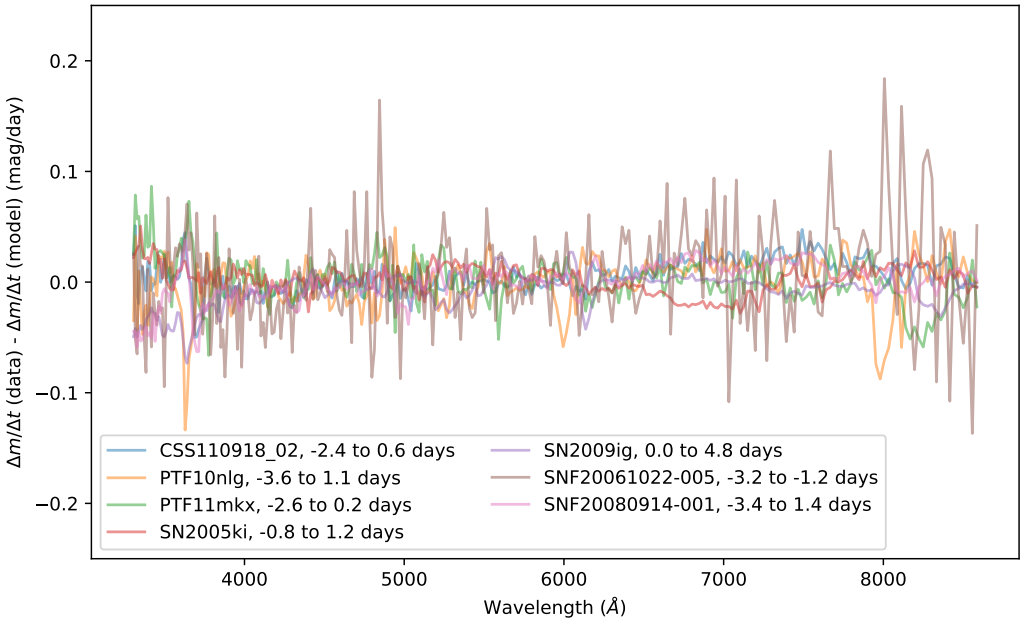



Figure 3.6: Residuals of the spectral evolution of different SNe Ia near maximum light for a randomly chosen subset of supernovae in the sample, after model correction. See Figure 3.3 for the uncorrected spectral evolution of the same set of supernovae and details of this plot. After the model correction, the residuals have very little structure at most wavelengths.

maximum light could have systematic differences of up to 0.8 mag from the true spectrum at maximum light.

Our model of spectral evolution near maximum light also takes into account the stretch of the supernova through the SALT2 $x_1$ parameter. The modeled effect of varying $x_1$ for a range of differences in phase is shown in Figure 3.9. We find that changing the SALT2 $x_1$ parameter does not have a major impact on the spectral evolution near maximum light. For differences of one day in either direction from maximum light, the effect is negligible, with an effect on the spectrum of less than 0.02 mag at all wavelengths even for extreme values of $x_1$. Before maximum light, the effect on the spectrum is relatively gray, with an offset of $\sim$0.05 mag added to the spectrum for extreme values of $x_1$ three days before maximum light, and $\sim$0.1 mag added to the spectrum five days before maximum light. After maximum light, the effect on the spectrum is highly structured. We find large effects in the bluer bands when varying $x_1$, with deviations of more than 0.3 mag for the Ca II H&K lines around 3900 Å. For our analysis, a gray offset in brightness will be absorbed by the fit to the overall brightness for a given supernova, and it will not have a major impact on the final results. Any structured residuals due to modeling errors could, however, be interpreted as real supernova variation, so the observed differences due to different stretch values for later phases could negatively impact our analysis.

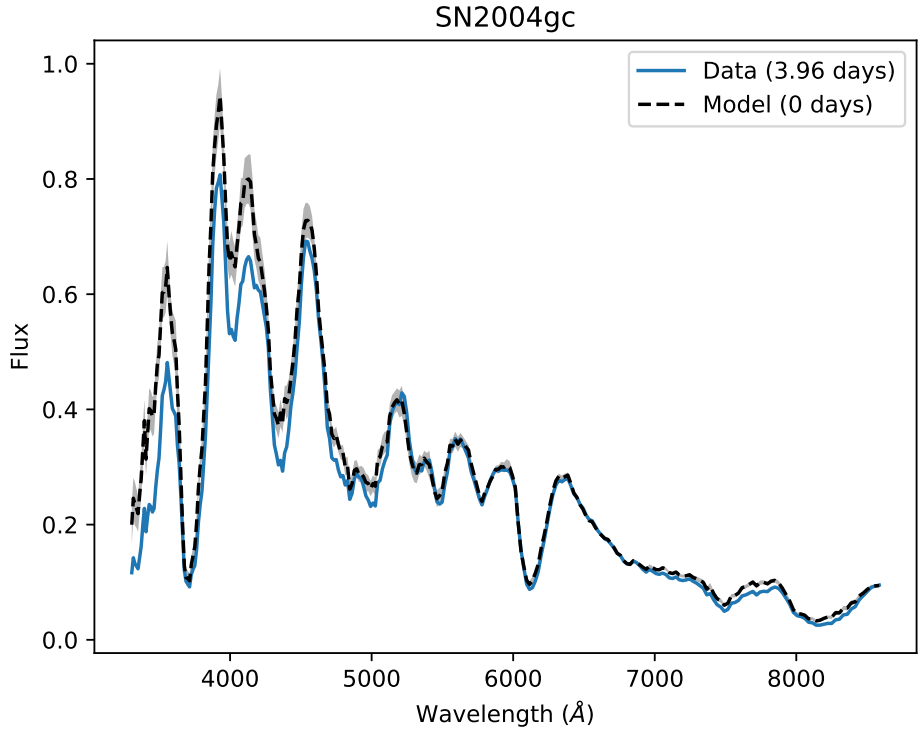



Figure 3.7: Estimated spectrum of SN2004gc at maximum light. This supernova has only a single spectrum passing our selection criteria, shown in blue. The model effectively uses information from other supernovae in the sample to predict the spectrum at maximum light, shown with a dashed black line. A shaded gray contour around this dashed black line shows the uncertainty on the estimate of the spectrum at maximum light, incorporating both observational and modeling uncertainties.

One potential issue with using the SALT2 $x_1$ parameter in the model of spectral evolution near maximum light is that the SALT2 $x_1$ parameter is not sufficient to fully describe SNe Ia and could potentially introduce artificial signals into the spectra near maximum light. To address this, we tested running the rest of the analysis described in this chapter both with and without the SALT2 $x_1$ parameter included in the model. We do not find major differences in the performance of our algorithms in either case. For SNe Ia with a spectrum within three days of maximum light, the $x_1$ value used has very little effect on the estimated spectrum at maximum light, as seen in Figure 3.9. For SNe Ia with a larger phase difference, we propagate the uncertainty in the estimate of the spectrum at maximum light. In Section 3.4.4, we will cut SNe Ia from the sample with large uncertainties, effectively removing the SNe Ia that only have spectra at phases where the stretch has a large effect on the differential evolution of the spectrum. We find that two additional spectra are cut if the $x_1$ parameter is not

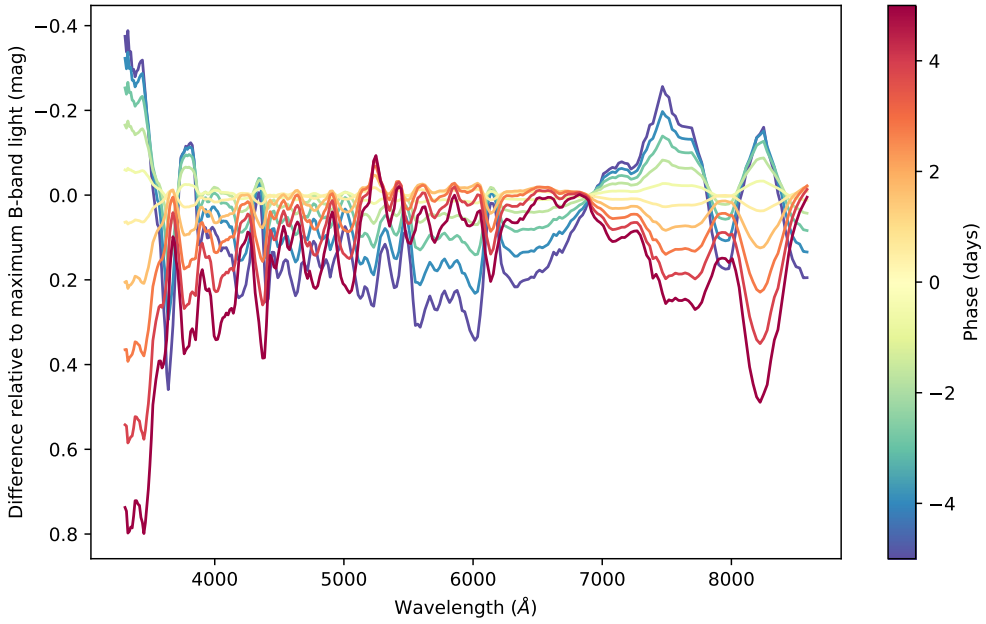



Figure 3.8: Model of the differential spectral evolution near maximum light for a supernova with a SALT2 $x_1$ of zero. The modeled differences are shown in different colors for phases within five days of maximum light with a spacing of one day. The color bar indicates which phase corresponds to which line on this plot.

included in the model of differential evolution, but the results of the rest of our analysis are otherwise unchanged.

### 3.3.4 Recovered Gray Offsets

As part of this analysis, we fit for a "gray offset" for each spectrum. For standard stars observed with SNIFS, we find that there is a residual dispersion of ∼0.02 mag after calibration that can be modeled as a flat "gray" offset that is the same at all wavelengths. In this analysis, we fit for this gray dispersion for the spectra of SNe Ia, which are at a significantly lower S/N than most of the standard stars. We recover a gray dispersion of 0.026 mag, which is consistent with what is found for standard stars. One potential issue is that this gray offset term in the model could mask actual variation in the spectra of SNe Ia. In our model, we fit for the gray offset of each supernova explicitly. To probe whether any biases are introduced, we show the recovered individual gray offsets as a function of phase for all of the supernovae in our sample in Figure 3.10.

We find that the mean of the gray offsets is consistent with zero at all wavelengths, and that the RMS of the gray offsets is flat with redshift. There is no evidence that the inclusion of gray offsets in our model is biasing the recovered spectral evolution in any way. We do,

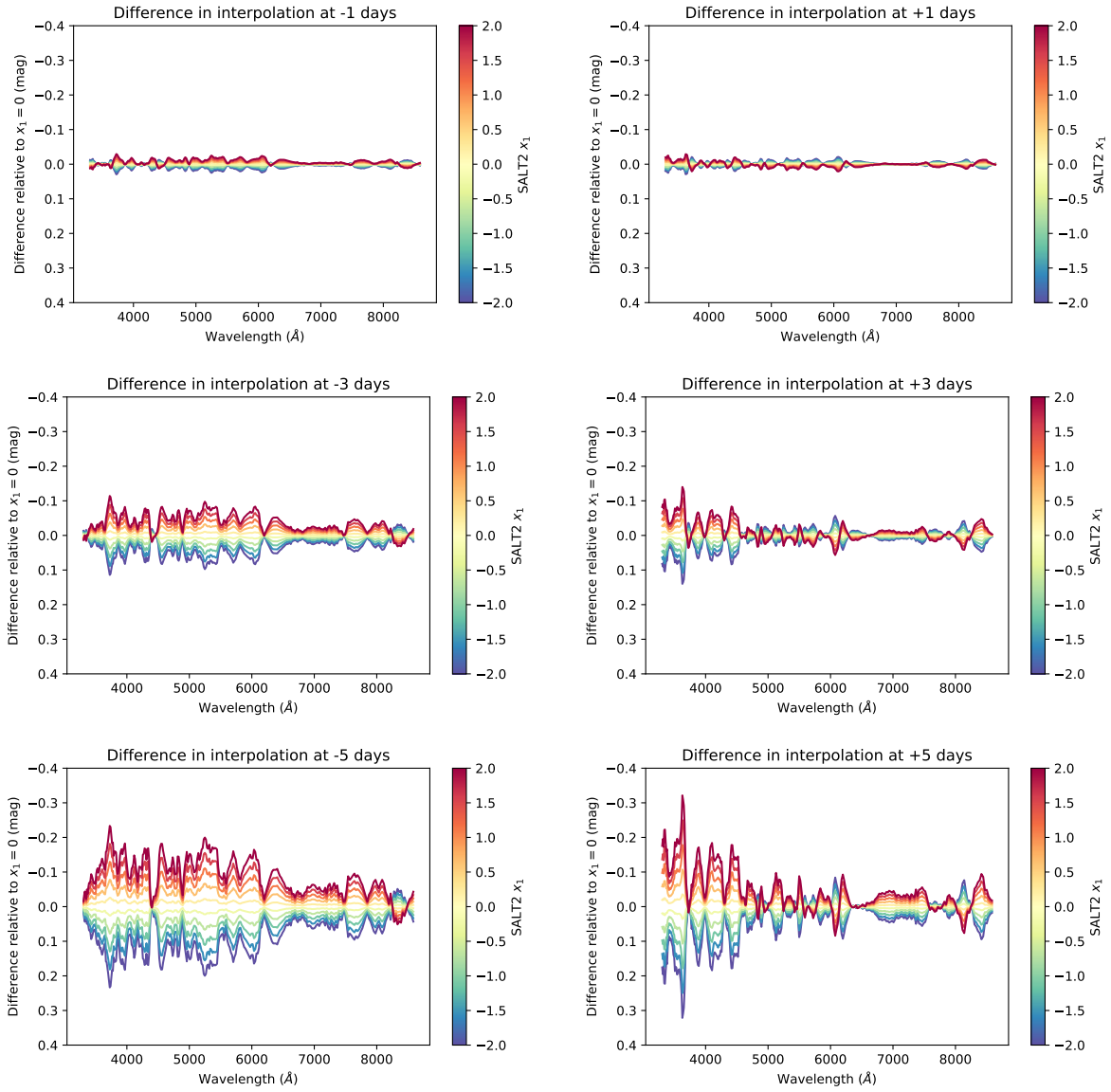



Figure 3.9: Model of the spectral evolution near maximum light for different supernova stretch values. For a given difference in phase, we show the effect of varying the SALT2 $x_1$ parameter relative to an $x_1$ value of zero. The modeled differences are shown in different colors, and the color bar indicates which SALT $x_1$ value corresponds to which line on this plot.

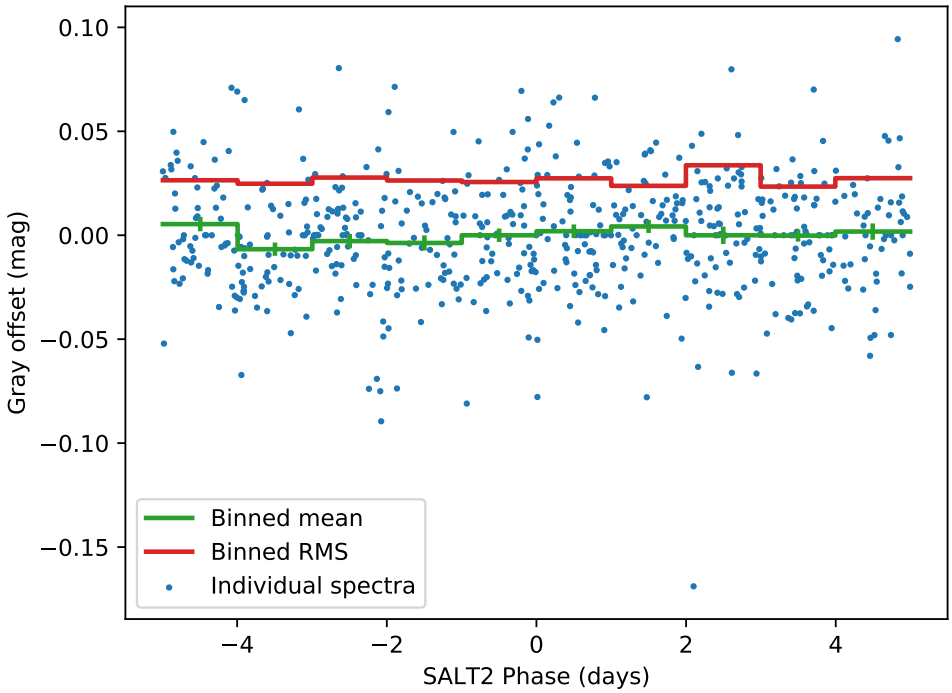



Figure 3.10: Recovered gray offsets as a function of the phase of a spectrum. The recovered gray offsets for individual spectra are shown with blue points, and the binned mean and RMS of the gray offsets are shown with green and red lines respectively. The mean of the gray offsets is consistent with zero at all phases, and there is no evidence of increased RMS for the gray offsets away from maximum light.

however, find that there are some spectra with vary large gray offsets, such as one spectrum of PTF09fox which has a recovered gray offset of $-0.17$ mag. For this supernova, we have another spectrum taken minutes earlier that has a recovered gray offset of $+0.05$ mag, and the difference between these two spectra appears to be well-modeled by a flat offset in magnitude. This large offset is due to the fact that these spectra were taken on a non-photometric night where clouds were present in the sky. For non-photometric nights, the dimming from clouds is taken into account by comparing photometry of nearby stars taken with the photometric channel to the photometry on a photometric night. The accuracy of this comparison was poor for this night, introducing a large gray offset into the calibration. While there is a large gray offset for this particular spectrum, these gray offset outliers are rare, and they are difficult to identify within our pipeline framework without another spectrum to compare to. Furthermore, when multiple spectra are available for a supernova, the effect of gray offset outliers is mitigated (PTF09fox has five spectra passing our selection requirements). As

spectra that are gray offset outliers are difficult to identify and are unlikely to significantly affect our analysis, we choose to keep them in our analysis.

### 3.3.5 Accuracy of the Model of Spectral Evolution Near Maximum Light

Along with modeling the spectral evolution of the spectra of SNe Ia, our model also fits for the residual dispersion in the spectral evolution after the model has been applied. As described in Section 3.3.1, this residual dispersion model is implemented for each wavelength bin separately as a broken linear function with nodes at phases of $-5$, $-2.5$, 2.5, and 5 days. The results of this model are shown in Figure 3.11.

We find that our model is able to estimate the spectra of supernovae 2.5 days away from maximum light in either direction with an uncertainty of less than 0.05 mag at almost all wavelengths. For spectra 5 days away from maximum, the interpolation uncertainties are around 0.05 mag at most wavelengths, but the model has a lot of difficulty modeling the spectral evolution of the Ca II H&K feature around 3900 Å. Five days before maximum light, the uncertainty of the spectral evolution of this feature is more than 0.35 mag for the worst wavelengths, indicating that there is significant additional variation in the spectral evolution at these wavelengths that is not captured by a model that only takes into account the phase of the spectrum and the SALT2 $x_1$ parameter of the supernova.

As part of our modeling procedure, we propagate the observational uncertainties and modeling uncertainties to our estimates of the spectra at maximum light. Examples of these uncertainties on the spectra at maximum light are shown in Figure 3.12. Note that the uncertainties on the estimated spectra at maximum light show large variations depending on the quality and timing of the observed spectra. For SNF20080514-002, we have six very high S/N spectra at phases between $-5.0$ and $+4.7$ days, so the spectrum at maximum light is very well constrained and the model assigns a very low uncertainty to its prediction of the spectrum at maximum light. For SNF20070424-003, on the other hand, there is only a single spectrum available at $-4.64$ days. As shown in Figure 3.11, the model cannot accurately predict the time evolution of the Ca II H&K lines at early times, so it outputs a large uncertainty on the prediction of the spectrum at maximum light at those wavelengths.

We propagate these uncertainties on the spectra at maximum light into our subsequent analyses. In Section 3.4, we incorporate the model uncertainties directly into our analysis, so poor estimates of the spectra at maximum light will be deweighted and they will not impact the analysis. As part of the analysis in Section 3.4, we estimate the intrinsic variation of SNe Ia at maximum light in each wavelength bin. We will use this estimate of the intrinsic variation to determine which features of our estimated spectra SNe Ia at maximum light are due to intrinsic variation of SNe Ia, and which are simply due to uncertainties in the interpolation to maximum light.

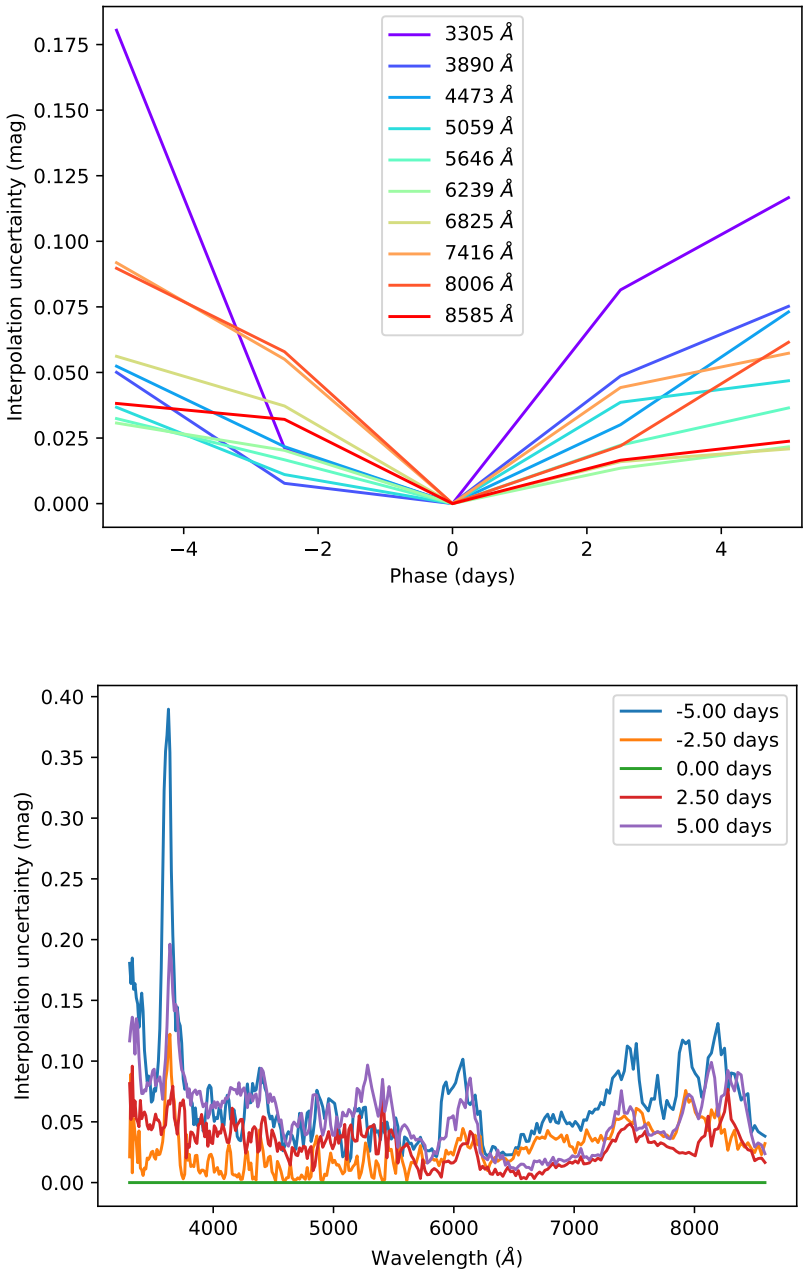



Figure 3.11: Model uncertainties for the interpolation of spectra to maximum light. Top panel: examples of the broken linear uncertainty model for a selection of different wavelengths. Bottom panel: Interpolation uncertainties for all wavelengths for the nodes of the broken linear model. Note that the interpolation uncertainty at maximum light is zero by definition.

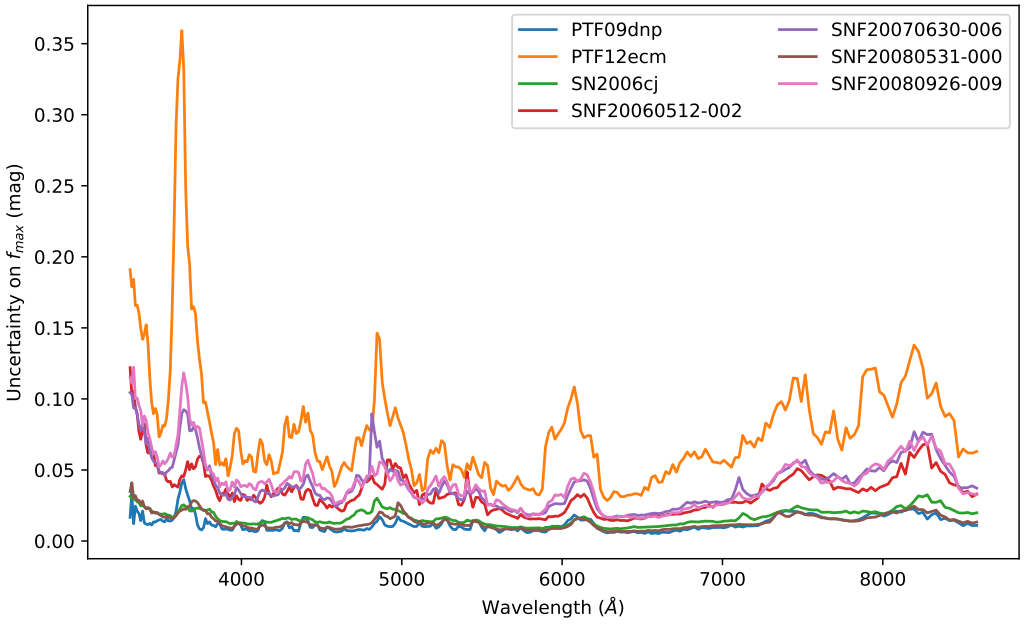



Figure 3.12: Uncertainties on the estimated spectra at maximum light for a selection of supernovae. Both the observational and modeling uncertainties are included.

### 3.3.6 Importance of Different Components of the Model of Spectral Evolution Near Maximum Light

We probe how the different components of our model impact the accuracy of predictions of spectral evolution in time. For each supernova with multiple spectra passing our selection requirements, we look at the difference in brightness in each wavelength bin between the spectrum closest to maximum light for that supernova and all of its other spectra. We then apply a subset of the components of our model of the difference in brightness for those two spectra, and calculate the RMS of the residuals over all supernovae and all wavelength bins. The results of this procedure are summarized in Table 3.3.

We find that the phase correction that is common to all supernovae has the largest impact on the RMS of the residuals, indicating that most of the spectral evolution is common among supernovae. Taking into account the gray offsets that result from the SNIFS calibration has the next largest impact on reducing the residuals. The SALT2 $x_1$ correction does not significantly affect the quality of the residuals, which is consistent with the small effects seen for most phases in Figure 3.9. Note that for this analysis we were comparing pairs of spectra within five days of maximum, so differences of up to ten days are possible. For the actual estimation of the spectrum at maximum light, the phase difference is never more than five days.

Table 3.3: Effect of different components of the model of spectral evolution near maximum light. For different pairs of spectra from the same supernova, we calculate the RMS of the differences between those pairs of spectra in magnitudes for different combinations of the components of our model.

| Components applied | RMS for pairs of spectra with a given phase difference (mag) | | | |
|---|---|---|---|---|
| | 0–1.5 days | 1.5–2.5 days | 2.5–5.5 days | >5.5 days |
| No model applied | 0.052 | 0.115 | 0.179 | 0.202 |
| Phase only | 0.046 | 0.080 | 0.095 | 0.102 |
| Phase + SALT2 $x_1$ | 0.047 | 0.078 | 0.087 | 0.103 |
| Phase + gray offset | 0.032 | 0.061 | 0.083 | 0.087 |
| Phase + gray offset + SALT2 $x_1$ | 0.032 | 0.059 | 0.074 | 0.081 |

## 3.4 Reading Between the Lines

### 3.4.1 Extrinsic Contributions to the Spectra of Type Ia Supernovae

From the analysis in Section 3.3, we have estimates of the spectra at maximum light for all of the 207 SNe Ia in our analysis, along with uncertainties on those estimates. Our objective is to decompose the variability of those spectra. There are two major components of the variability of SNe Ia that are not intrinsic to SNe Ia. First, the distance to SNe Ia introduces a flat offset to the brightness of a supernova in magnitudes. Second, interstellar dust will dim the observed brightness of a supernova and redden the observed spectrum.

As described in Section 3.2, we use a fiducial cosmology model to shift all of our supernovae to the same effective distance away. Unfortunately, this correction relies on using the redshift of a supernova to estimate the distance to it, and the measured redshift contains both contributions from the cosmological redshift and the peculiar velocity of the supernova. The peculiar velocity contributions will introduce uncertainties into the estimated distances, and thus there will be some residual gray dispersion in the observed spectra, even after correcting for their distances. Typically, these peculiar velocities are assumed to have a dispersion of $\sim$300 km/s, contributing a dispersion in brightness of 0.11 mag at $z = 0.02$, 0.07 mag at $z = 0.03$ and 0.02 mag at $z = 0.10$.

Interstellar dust is present in both the supernova host galaxy and the Milky Way. In the preprocessing of SNfactory data, we apply a correction for Milky Way dust to all of our supernovae using the dust map from Schlegel et al. (1998) with a reddening law from Cardelli et al. (1989). As shown in Chotard et al. (2011), the reddening of supernova flux due to dust in the supernova's host galaxy is consistent with dust with a total-to-selective extinction ratio of $R_V = 2.8 \pm 0.3$. For this analysis, we choose to use a fiducial color law with $R_V = 2.8$ to fit for the contributions from extragalactic dust, and we use the reddening

law from Fitzpatrick (1999). We will examine the validity of this color law, and discuss the value of $R_V$ that best fits our spectra of SNe Ia in Section 3.6.4.

We therefore have functional forms for both extrinsic contributions due to distance uncertainties and host galaxy dust reddening for spectra of SNe Ia. In F15, for each pair of supernovae, the difference in brightness and in color between those two supernovae was calculated by effectively minimizing a chi-square difference between the spectra of the two supernovae while fitting for the coefficients of the difference in brightness and color. If two supernovae are perfect twins, then their spectral features should match perfectly, so only differences due to extrinsic effects like interstellar dust extinction should remain. We found that the estimated differences in brightness and color between two supernovae were consistent even when comparing two supernovae that are not twins, with differences in the estimated magnitudes of less than 0.02 mag for most pairings. This is due to the fact that the spectra of SNe Ia at maximum light are remarkably consistent: the spectral variation of SNe Ia at maximum light is mostly constrained to a handful of spectral lines, and the regions in between those lines have very little spectral variation, as will be shown in Section 3.4.3.

This result motivates a different approach to fitting for the brightness and color of each of the supernovae in our sample. Rather that compute pairwise differences between twins, we determine a "mean spectrum" of a SN Ia at maximum light, and we compare the spectra of each supernova in our sample to this mean spectrum to determine its brightness and color offset. To avoid our estimates of the brightness being biased by spectral features, we simultaneously solve for the amplitude of the intrinsic dispersion of SNe Ia at each wavelength. By weighting by this intrinsic dispersion, we effectively deweight regions of the spectrum with large intrinsic variance, and estimate the brightness and color of the spectrum using the regions where there is low intrinsic variability. We call this procedure "Reading Between the Lines".

One caveat with this model is that any intrinsic variability that modifies the spectrum of an SN Ia in a way that looks like brightness or color will be incorrectly labeled as extrinsic variability at this stage. Assuming that this intrinsic variability also affects the spectrum in some other way, such as modifying the equivalent widths of absorption features, we can recover the intrinsic variability that was confused as extrinsic variability in a later analysis. This procedure is similar to what is done in models like SALT2, where the true B-band maximum brightness is determined by correcting by some function of the SALT2 stretch parameter $x_1$ and color $c$.

### 3.4.2 The Reading Between the Lines Model

Our "Reading Between the Lines" (hereafter: RBTL) model is implemented as follows. For each supernova $i$, we begin with a spectrum at maximum light $f_{max,i}(\lambda)$ with associated uncertainties $\sigma_{f_{max,i}}(\lambda)$. We represent the mean spectrum of an SN Ia at maximum light as $f_{mean}(\lambda)$. Each supernova $i$ then has a parameter $M_i$ representing its difference in brightness compared to the mean spectrum in magnitudes, and a parameter $A_{V,i}$ representing the

coefficient of the color law $C(\lambda)$ that best matches the supernova's spectrum to the mean function. The modeled flux of the spectrum at maximum light of supernova $i$ is then:

$$f_{model,i}(\lambda) = f_{mean}(\lambda) \times 10^{-0.4(M_i + A_{V,i}C(\lambda))} \tag{3.6}$$

We then assume that the intrinsic dispersion of SNe Ia, $\epsilon_{intrinsic}(\lambda)$ can be modeled as an additional normal uncertainty that is uncorrelated in wavelength, with an amplitude that is a fraction of the modeled flux. The total uncertainty of the spectrum at maximum light for a supernova relative to the modeled spectrum is therefore:

$$\sigma^2_{total,i}(\lambda) = \sigma^2_{f_{max,i}}(\lambda) + (\epsilon_{intrinsic}(\lambda) f_{model,i}(\lambda))^2 \tag{3.7}$$

The observed spectrum at maximum light is therefore modeled as:

$$f_{max,i}(\lambda) \sim N(f_{model,i}(\lambda); \sigma^2_{total,i}(\lambda)) \tag{3.8}$$

As in Section 3.3.1, we implement this model using the Stan modeling language (Carpenter et al., 2017), and we fit it to our sample of 207 spectra of SNe Ia at maximum light. To avoid degeneracies in the model, we constrain the sum of each of $M_i$ and $A_{V,i}$ to be zero across the whole sample. As our spectra are binned in 288 wavelength bins, as described in Section 3.2, the model has a total of 288 parameters for the mean spectrum $f_{mean}(\lambda)$, 288 parameters for the intrinsic dispersion $\epsilon_{intrinsic}(\lambda)$, and $207 - 1$ parameters for each of $M_i$ and $A_{V,i}$ (the sum-to-zero constraint removes one parameter).

As in Section 3.3.1, we use Stan to optimize the parameters of this and obtain the MAP estimate of the model parameters. Finally, we apply the inverse of the magnitude and color corrections to obtain "dereddened spectra" $f_{dereddened,i}(\lambda)$ for each of our spectra at maximum light:

$$f_{dereddened,i}(\lambda) = f_{max,i}(\lambda) \times 10^{+0.4(M_i + A_{V,i}C(\lambda))} \tag{3.9}$$

### 3.4.3 Similarity of the Spectra of SNe Ia at Maximum Light

The estimated spectra at maximum light of the 207 supernovae in our sample both before and after dereddening are shown in Figure 3.13, and the modeled intrinsic dispersion of SNe Ia is shown in Figure 3.14. The dereddened spectra show remarkable similarity, especially in wavelength regions away from the main absorption lines, such as between 6600 and 7200 Å where the model finds that an intrinsic dispersion of only roughly 0.02 mag is required to explain the remaining dispersion of SNe Ia after taking the brightness and color into account. The Ca II H&K lines, the Ca II IR triplet and the Si II 6355 Å feature are the locations in the spectra of SNe Ia with the largest intrinsic dispersions. Interestingly, there are several regions of the spectra that are blanketed by lines but that still show relatively low intrinsic dispersion. From 3900 Å to 5900 Å, the spectra of SNe Ia show a variety of absorption lines, but the intrinsic variability is recovered to be ∼0.08 mag at most wavelengths, with the exception of a handful of stronger lines that introduce variability at up to 0.13 mag.

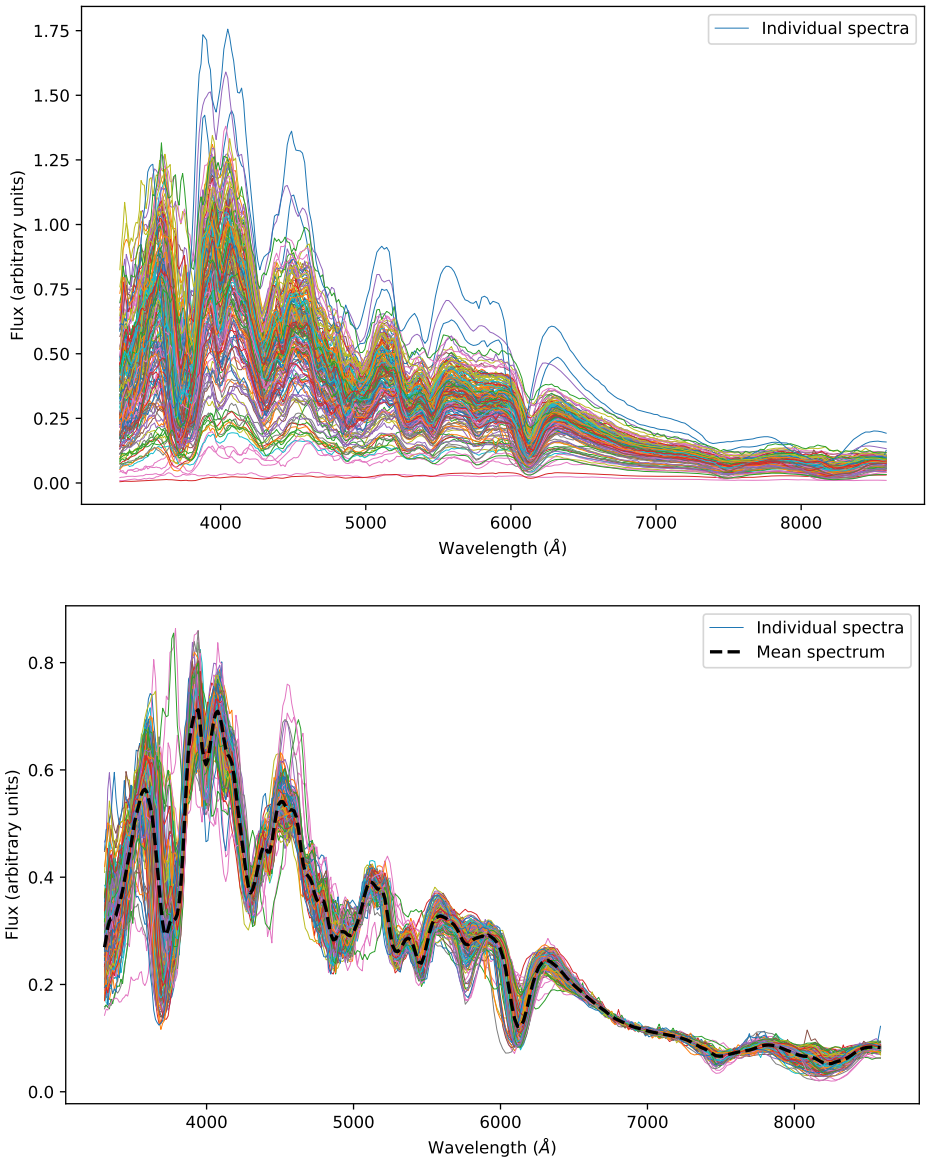



Figure 3.13: Comparison of the diversity of spectra at maximum light before and after dereddening. Top panel: All of the spectra of SNe Ia in our sample estimated at maximum light are shown, after shifting all of these spectra to a common redshift. There is a wide range of diversity in these spectra due to uncertainties in the distance correction and reddening due to host galaxy dust. Bottom panel: The same spectra are shown after estimating and removing the residual brightness and color relative to a mean spectrum following the procedure in Section 3.4.2. The estimated mean spectrum is shown with a dashed black line.

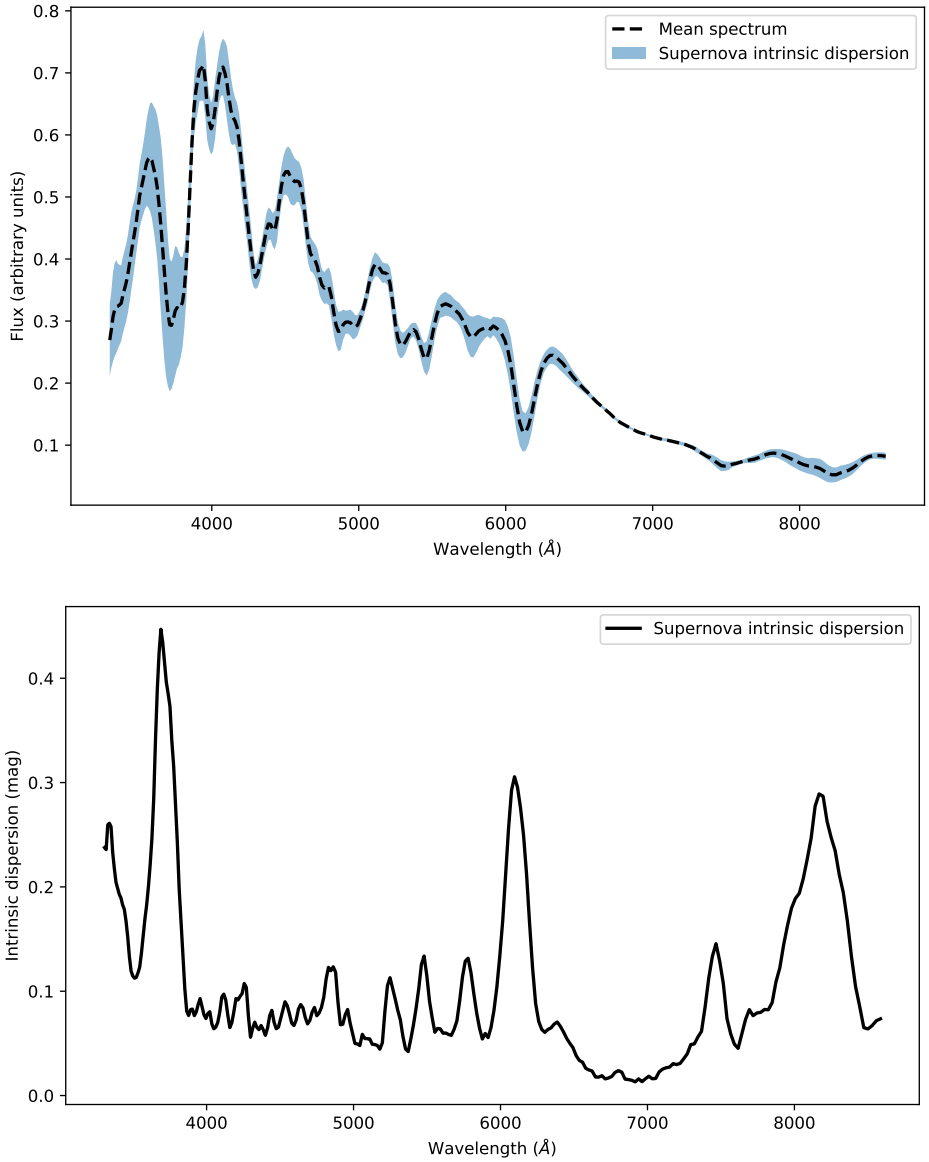



Figure 3.14: Intrinsic dispersion of SNe Ia at maximum light, as modeled by the RBTL modeling procedure. Top panel: The recovered mean supernova spectrum at maximum light is shown along with the one standard-deviation dispersion as a blue contour. Bottom panel: The modeled intrinsic dispersion is shown in magnitudes for each wavelength.

This model is effectively using the intrinsic variance of SNe Ia to weight each wavelength when determining how to fit for the brightness and color of each supernova. Hence, the lines with strong variability are deweighted, and the model effectively "reads between the lines" to estimate the brightness and color.

### 3.4.4 Comparison of the Uncertainties on the Spectra at Maximum Light to the Intrinsic Dispersion of SNe Ia

In Figure 3.15, we compare the uncertainties on the estimated spectra at maximum light to the recovered intrinsic dispersion of SNe Ia. For the vast majority of the SNe Ia in our sample, the uncertainties on the spectrum at maximum light are negligible compared to the recovered intrinsic dispersion of SNe Ia. However, for some of the supernovae in the sample, the uncertainty on the spectrum at maximum light is comparable to the intrinsic dispersion.

In Section 3.5, we will decompose the intrinsic dispersion into its different components. If spectra with large uncertainties are included, this analysis could confuse true variability with the uncertainty in our estimate of the spectrum at maximum light. To mitigate this, we choose to remove any supernovae at this stage of the analysis whose uncertainties on the spectrum at maximum light are large compared to the intrinsic dispersion. We choose to require that the total variance of the estimate of the spectrum at maximum light be less than 10% of the total intrinsic variance of SNe Ia. Choosing this threshold requires somewhat of a trade off: with better measured spectra, we can potentially recover more components of the intrinsic dispersion, but a stricter threshold reduces the number of SNe Ia in the sample which limits the coverage of parameter space. As we will find in Section 3.5, this threshold roughly corresponds to the variance scale of the third non-linear component of the intrinsic dispersion. Out of the original sample of 207 SNe Ia, 175 SNe Ia have an uncertainty on their spectrum at maximum light that passes this threshold.

Note that in the RBTL analysis we included the estimates of the uncertainties for the spectra at maximum light, so the spectra that were cut due to this threshold will not have a major impact on the RBTL analysis. We attempted to refit the RBTL model on only the spectra that pass the uncertainty of the spectrum at maximum light requirement, and found that the changes were negligible (the estimated brightnesses change by $<0.005$ mag). As a result, we choose to use the RBTL model with all of the spectra for further analysis.

## 3.5 Manifold Learning on Spectra of SNe Ia

From the "Reading Between the Lines" analysis in Section 3.4, we have spectra at maximum light for 175 different SNe Ia with extrinsic contributions from the overall brightness/distance to the supernova and the dust in the supernova's host galaxy removed. The resulting "dereddened" spectra, shown in the lower panel of Figure 3.13, exhibit a range of intrinsic diversity. In this section, we will decompose this intrinsic diversity using manifold learning to generate a non-linear parametric model of the intrinsic diversity of SNe Ia.

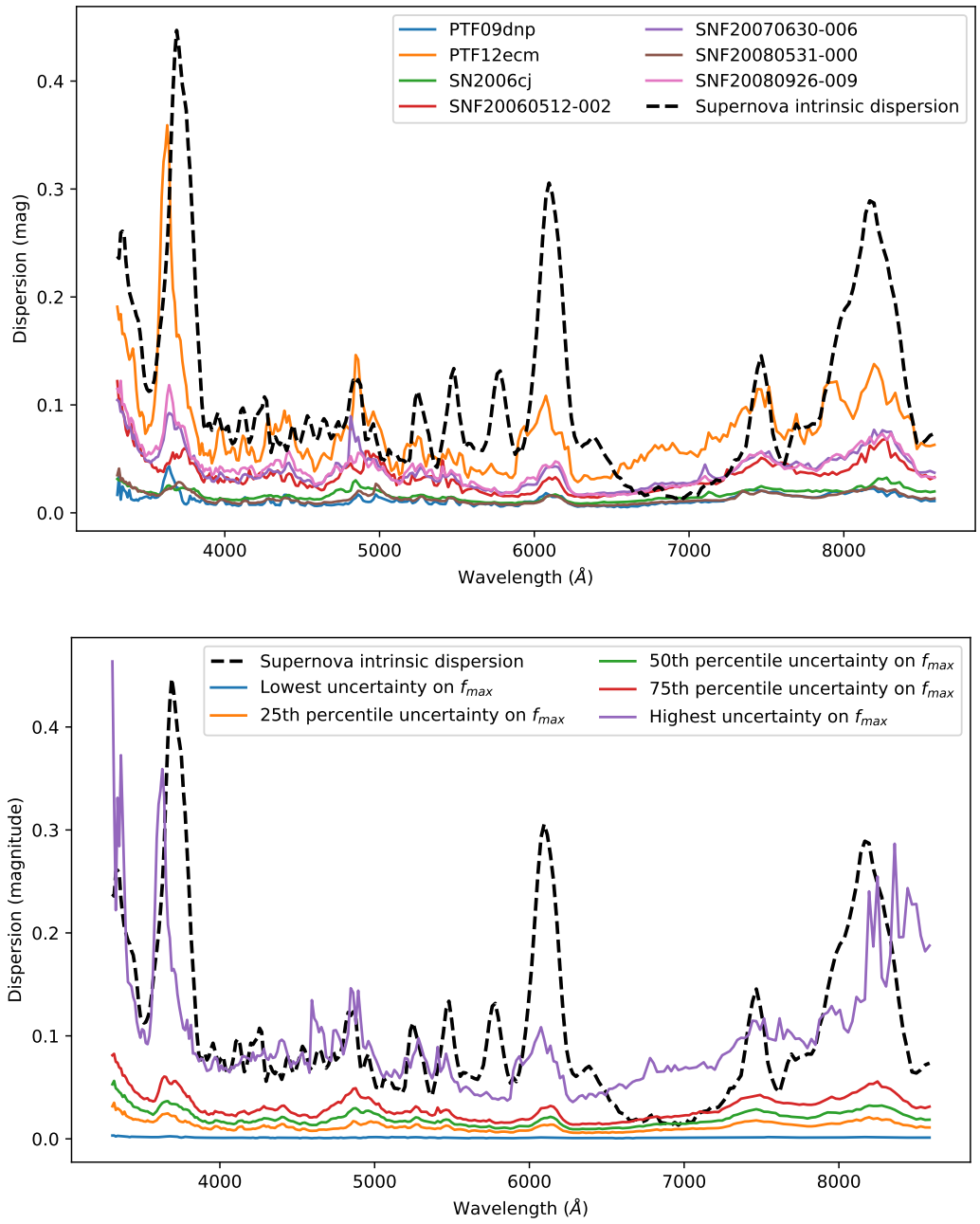



Figure 3.15: Comparison of the uncertainties on the spectra at maximum light to the recovered intrinsic dispersion of SNe Ia. Top panel: The intrinsic dispersion is plotted along with the uncertainties on the spectra at maximum light for the same selection of SNe Ia shown in Figure 3.12. Bottom panel: The intrinsic dispersion is plotted along with the uncertainties in the spectrum at maximum light for the spectrum with the lowest uncertainty, the one with the highest uncertainty, and several percentiles in between.

### 3.5.1 Manifold Learning

Manifold learning has been shown to be very effective for constructing non-linear parametrizations of the spectra and light curves of astronomical objects (Richards et al., 2009, 2012; Daniel et al., 2011; Matijevič et al., 2012; Sasdelli et al., 2016). There are many different manifold learning algorithms, each with different objectives. Fundamentally, most manifold learning algorithms take as input a set of data points represented by high-dimension vectors (here, the dereddened spectra of different SNe Ia), and they embed these data points into a low-dimensional manifold that preserves some measure of the distances between data points in the high-dimensional space. For this analysis, we choose to use the Isomap algorithm (Tenenbaum et al., 2000). The Isomap algorithm produces an embedding using the neighbors of each data point in the high-dimensional space, and effectively aims to preserve the distances between neighboring data points. For neighboring data points, the Isomap algorithm measures a distance between those points that is equivalent to the Euclidean distance in the high-dimensional space. For data points that are far away from each other, however, the Isomap algorithm instead finds a path between these two data points by traversing through pairs of neighboring points and using the total path length as the distance estimate between the far away data points. The Isomap algorithm attempts to preserve these total path lengths, which effectively preserves the local structure and ordering of the data points.

The Isomap algorithm has many similarities to the twinning framework of F15: the neighbors that are found by the Isomap algorithm are the best twins for a supernova. To estimate the distance between two supernovae that are very different, the Isomap algorithm effectively finds the shortest path through pairs of twins. This naturally handles features of the spectra of SNe Ia like different line velocities: a linear model will require multiple linear components to accurately capture the different possible velocities of a line. In contrast, if we have spectra for a large enough sample of SNe Ia, the Isomap algorithm can use pairs of supernovae with increasing line velocities to recover the sequence of line velocities in the sample and use a single component to represent them in the final embedding.

We choose to use as input to the Isomap algorithm the dereddened spectra for each of our SNe Ia divided by the mean spectrum. This means that we are using the relative deviations at each wavelength rather than the absolute deviations. For high S/N spectra, the distances for neighbors that the Isomap algorithm uses are then calculated in same way as the "twinness" in F15. The Isomap algorithm has two parameters that must be set to produce the embedding: the number of neighbors to use, and the number of components (i.e. dimensions) to produce for the low-dimensional embedding. We find that the number of neighbors does not have a major impact on the resulting embedding, and for a range of different values tested between 4 and 20 we obtain similar embeddings and overall performance. We choose to use 10 neighbors for the rest of our analysis.

### 3.5.2 Number of Components of SNe Ia at Maximum Light

Choosing the correct number of components for the Isomap embedding is a more challenging problem. SNe Ia are complex objects, and while most of the variation of SNe Ia might be able to be explained by a small number of components, asymmetries of the explosions, initial conditions and local environment of the SN Ia are likely to introduce countless additional components with small amplitudes that may or may not be relevant for standardization. If we had perfect observations, then we could extract an arbitrarily large number of components. However, for realistic data, at some point measurement uncertainties will overwhelm the small intrinsic variation of the higher order components. In this section, we attempt to determine how many real components of the intrinsic variation of SNe Ia we can extract from our dataset.

Isomap is a deterministic algorithm, and if an embedding with $N > M$ components is produced, then the first $M$ components will be identical to those from an embedding where only $M$ components were produced. Hence, we generate an embedding with an arbitrarily large number of components and use this to determine how many of the components are capturing real intrinsic variation of SNe Ia. We ran several analyses on this embedding. First, we examined the contributed variance from each of the Isomap components. The results of this procedure are shown in Figure 3.16. Isomap will recover the relative variance contributions of different components accurately. However, unlike techniques such as PCA, it does not have a concept of the "total variance" of a set of data points because of how the distances are calculated (described in Section 3.5.1). We can therefore only compare the relative contributions from each component with an arbitrary scale, and we choose to normalize to the variance of the first component. In Figure 3.16, we also show several comparisons with the uncertainties associated with estimating the spectra at maximum light that are described in Section 3.3.5, hereafter referred to simply as the "measurement variances". As described previously, the measurement variances cannot be directly compared to the variance of the Isomap components, so we normalize the overall scale of the measurement variances to the total variance of the first ten Isomap components.

As seen in this plot, the relative contributions of each component to the intrinsic variance of SNe Ia declines rapidly, with the fourth component contributing only 11% as much as the first component. Comparing to our estimates of the measurement variance, where the mean of the measurement variances across all SNe Ia is shown as a dashed green line in Figure 3.16, the first three components are unambiguously detected as being much larger than the typical measurement variance of a SN Ia in the sample, but the fourth and following components are at a comparable scale to the typical measurement variance and could be a result of uncertainties in our estimates of the spectra at maximum light rather than intrinsic variance of SNe Ia. The plot suggests that we limit our analysis to using the first three components. Any additional components are constrained to contribute less than 6% of the total variance of the first three components.

An alternate method of determining how many components are required to explain the intrinsic variation of SNe Ia is to look at how well we preserve the ordering of distances

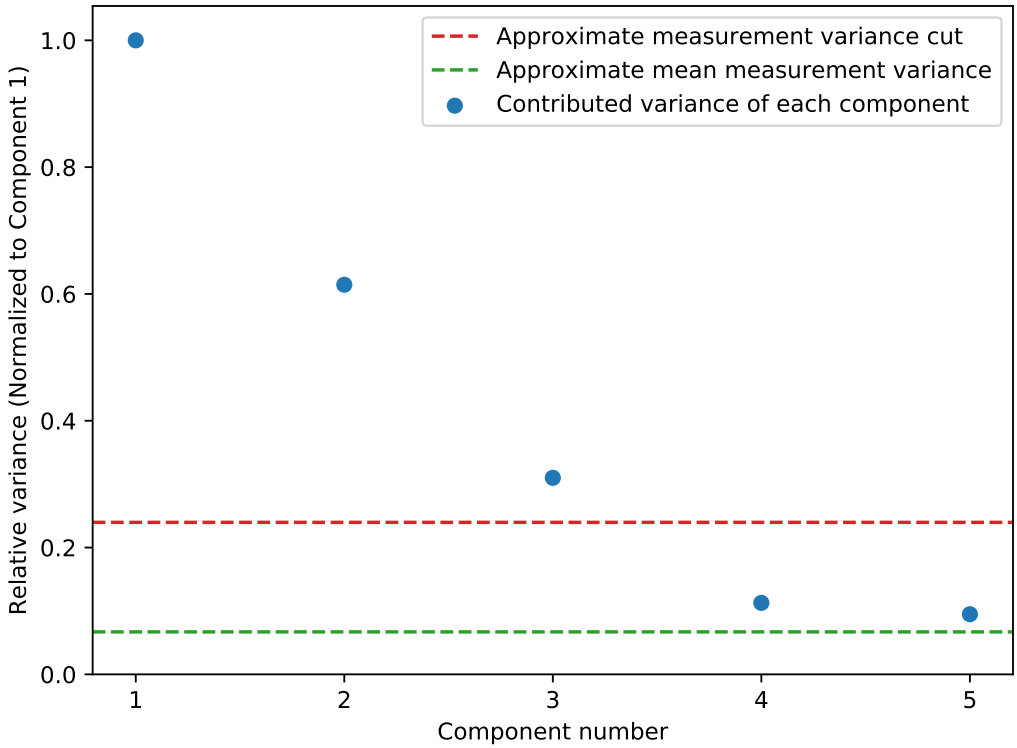



Figure 3.16: Total variance explained by each of the Isomap components. The blue points show the contribution to the total variance from each component. The dashed red line shows an approximation of the cut on the measurement variance. The green dashed line shows an approximation of the mean of the measurement variances for all spectra passing the cut. Note that the relative variances of different components are meaningful for Isomap, but there is no concept of a total variance, so we normalize everything to the variance of the first component. The scale of the measurement variances has been approximated by normalizing to the total variance of the first ten components.

between supernovae (the twinness from F15) in the embedded space. For each pair of SNe Ia, we calculate their spectral twinness as the Euclidean distance between their normalized dereddened spectra at maximum light. For high S/N spectra, this is equivalent to the "twinness" calculated in F15. We then calculate the "Isomap twinness" as the Euclidean distance between the Isomap embedded coordinates of those SNe Ia for a given number of components. We rank the supernovae in terms of their twinnesses. In Figure 3.17, we show how well the spectral twinness rankings are recovered in the embedded space for different numbers of components. We summarize how many "non-twins" are confused as "twins" in the embedded space in Figure 3.18 for different numbers of Isomap components.

We find that three Isomap components are required to correctly preserve the twinness ranking in the embedded space. With only one (two) Isomap components, over 8% (2%) of poorly matches pairs of SNe Ia are incorrectly recovered as "twin supernovae", defined here to be pairs of SNe Ia in the best 20% of pairings in the embedded space. With three components, less than 0.5% of non-twins are incorrectly identified as twins, and adding additional components does make any additional significant improvements. This analysis suggests that three components are sufficient for identifying the main intrinsic differences between spectra of SNe Ia at maximum light, in agreement with our analysis of the variance scales of different components. We will therefore use three components for the rest of this analysis.

### 3.5.3 Understanding the Isomap Components

At this point, we have shown that a three-component non-linear model is sufficient to describe the intrinsic diversity in the spectra of SNe Ia at maximum light (along with the extrinsic contributions from color and brightness removed in Section 3.4). We now examine how the Isomap components affect the different spectral features. For each component, we calculate the median spectrum in ten evenly-spaced bins of that component. We plot these median spectra in Figure 3.19. For comparison purposes, we also show a zoom-in around the Ca II H&K lines in Figure 3.20, a zoom-in around the Si II 5972 Å and 6355 Å features in Figure 3.21, and a zoom-in around the OI triplet and Ca II IR triplet in Figure 3.22.

The first Isomap component primarily affects the pseudo-equivalent widths of the Ca II features. As seen in Figure 3.20, this component maps out a full spectral sequence of the changing pseudo-equivalent width of the Ca II H&K feature, and a similar spectral sequence is seen for the Ca II IR triplet in Figure 3.22. Additionally, as seen in Figure 3.21, this component shows a spectral sequence for the Si II 6355 Å feature. This component is the only one that affects the emission profile of this feature, and the velocity and depth of the absorption profile evolve together. This component has very little effect on the Si II 5972 Å feature.

The second Isomap component primarily affects the pseudo-equivalent widths of the Si II lines. As shown in Figure 3.21, a spectral sequence is identified in both of the Si II 5972 Å and 6355 Å features. Interestingly, this component directly changes the depths of these lines without having a major effect on the line velocities themselves. This component also has

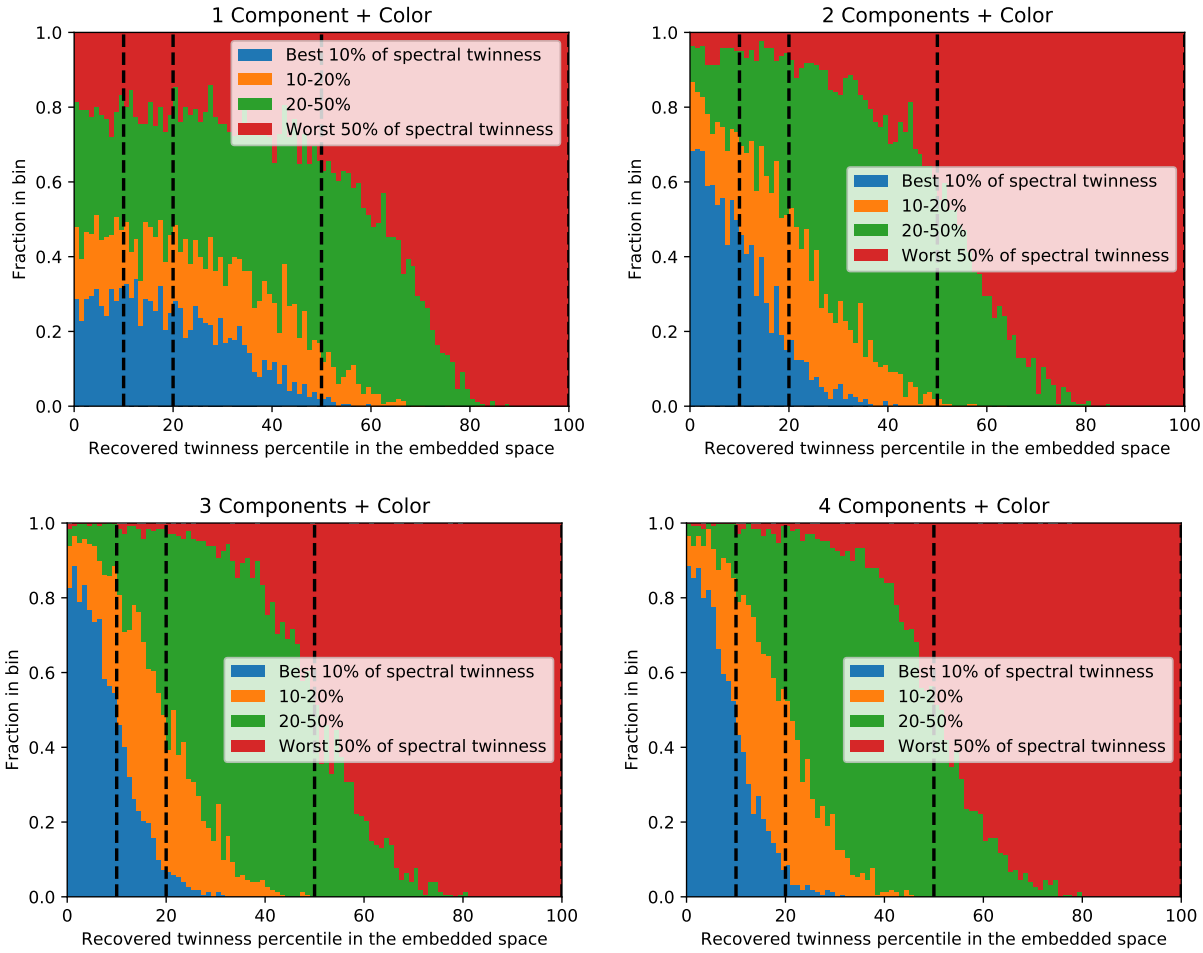



Figure 3.17: Recovery of the F15 spectral "twinness" for different numbers of Isomap components. We rank each pair of SNe Ia by both their spectral twinness and their twinness in the Isomap embedded space (see text for details). We then show the recovered Isomap twinness values for different percentile bins of spectral twinness. Each panel shows the results for a different number of Isomap components. For fewer than three components, a significant number of non-twins are confused as twins, as seen by the number of "non-twins" (shown in red) that leak into the best twinness bins. We measure this leakage as the number of non-twins that end up left of the vertical dashed line at 20%. Note that Isomap does not attempt to preserve the input distances (twinnesses) between far-away points, so even with very large numbers of components there will be differences between the two rankings.

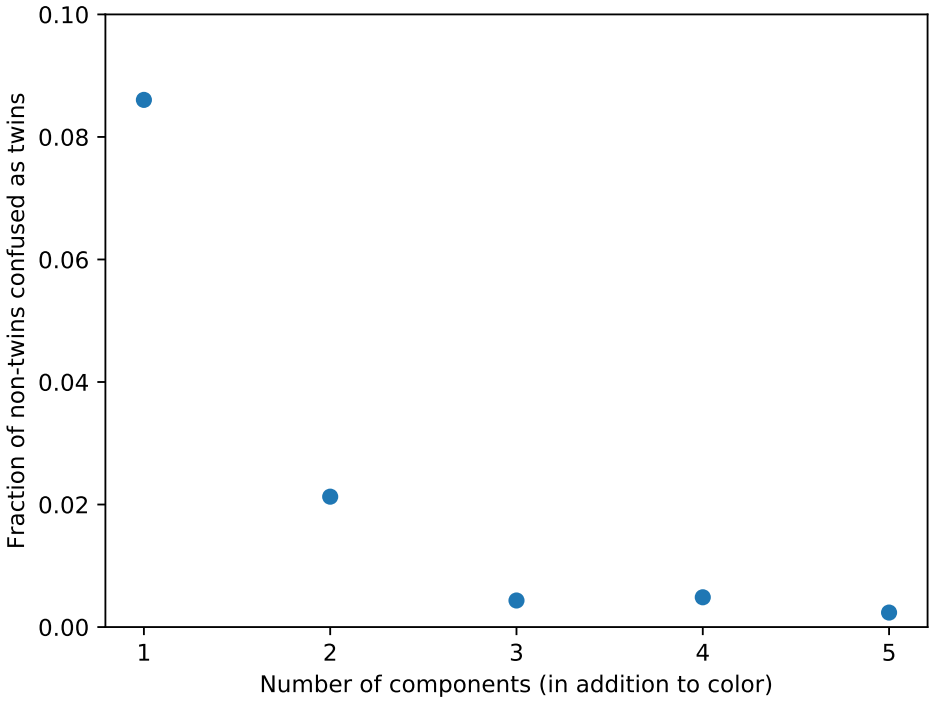



Figure 3.18: Fraction of non-twins confused as twins for different numbers of Isomap components. For a given number of Isomap components, we identify the worst 50% of spectral pairings and determine the fraction of those pairings that are recovered in the best 20% of pairings in the Isomap embedded space. This is equivalent to the fraction of pairings shown in red in Figure 3.17 that are recovered in the best 20% of pairings in the embedded space. When fewer than three components are used, a significant number of non-twins are confused as twins.

a large effect around 3700 Å, as seen in Figure 3.20. This variation appears to be due to absorption by a set of Si II lines at rest frame wavelengths between 3853 and 3863 Å in addition to the Ca II H&K absorption typically associated with these wavelengths. We also see a fairly significant variation in the Si II 4130 Åfeature.

The third Isomap component primarily affects the ejecta velocity of the supernova. As seen in Figure 3.21, this component identifies a spectral sequence in the velocity of the S II doublet feature near 5400 Å and in the velocity of the Si II 6355 Å feature. Similar effects are seen for the velocities of many other lines in the spectra. This component does not appear to have a major effect on any of the line depths in the spectrum. However, we see a sequence in the pseudo-continuum level around the Si II 5972 Å feature that does not appear to be strongly associated with the line depth.

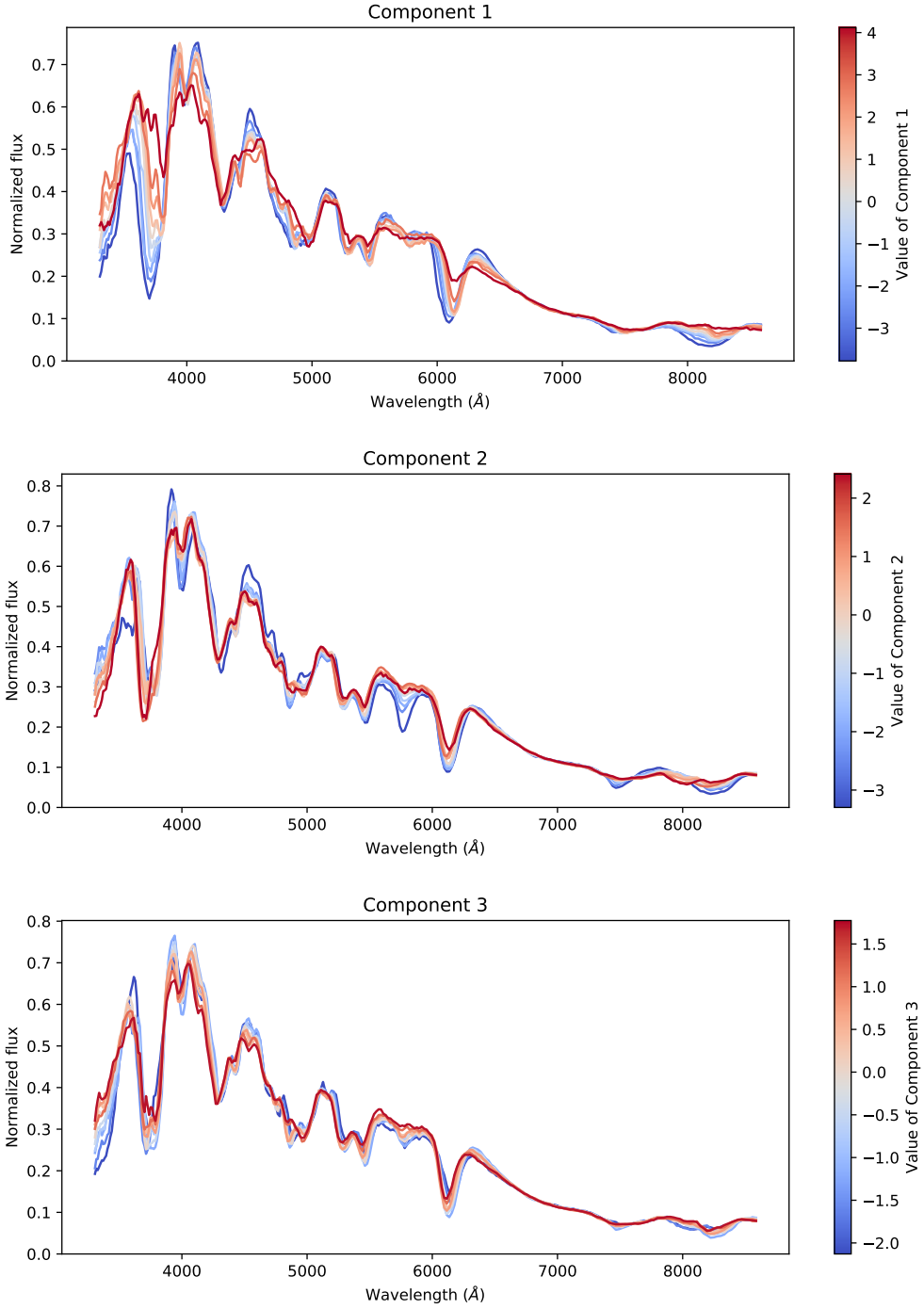



Figure 3.19: Effect of different Isomap components on the spectra of SNe Ia at maximum light. For each component, we show the median spectrum in ten evenly-spaced bins of that component. The spectra are colored according to their component value.

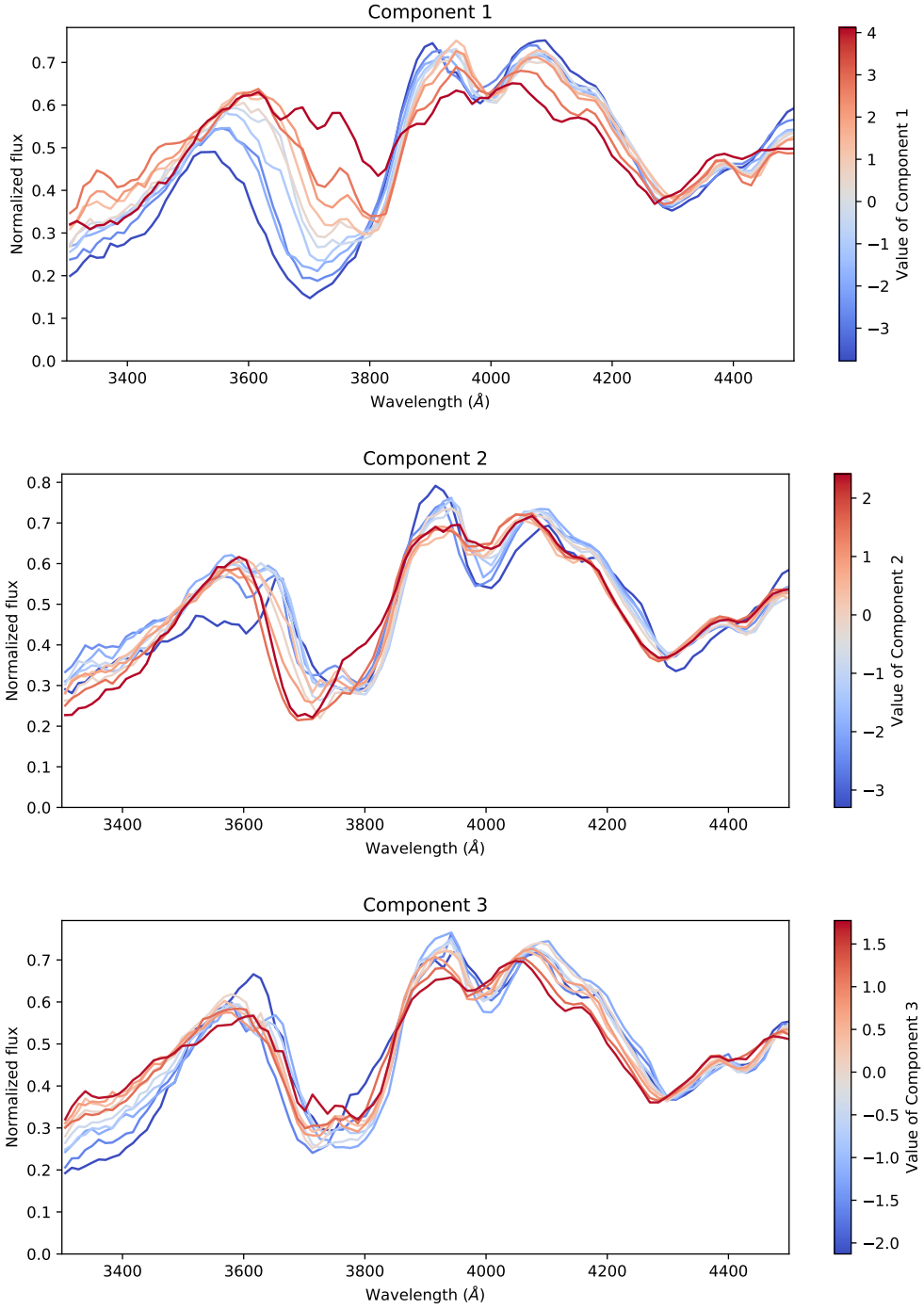



Figure 3.20: Effect of different Isomap components on the spectra of SNe Ia at maximum light between 3300 and 4500 Å, highlighting the Ca II H&K feature. See Figure 3.19 for the details of this figure.

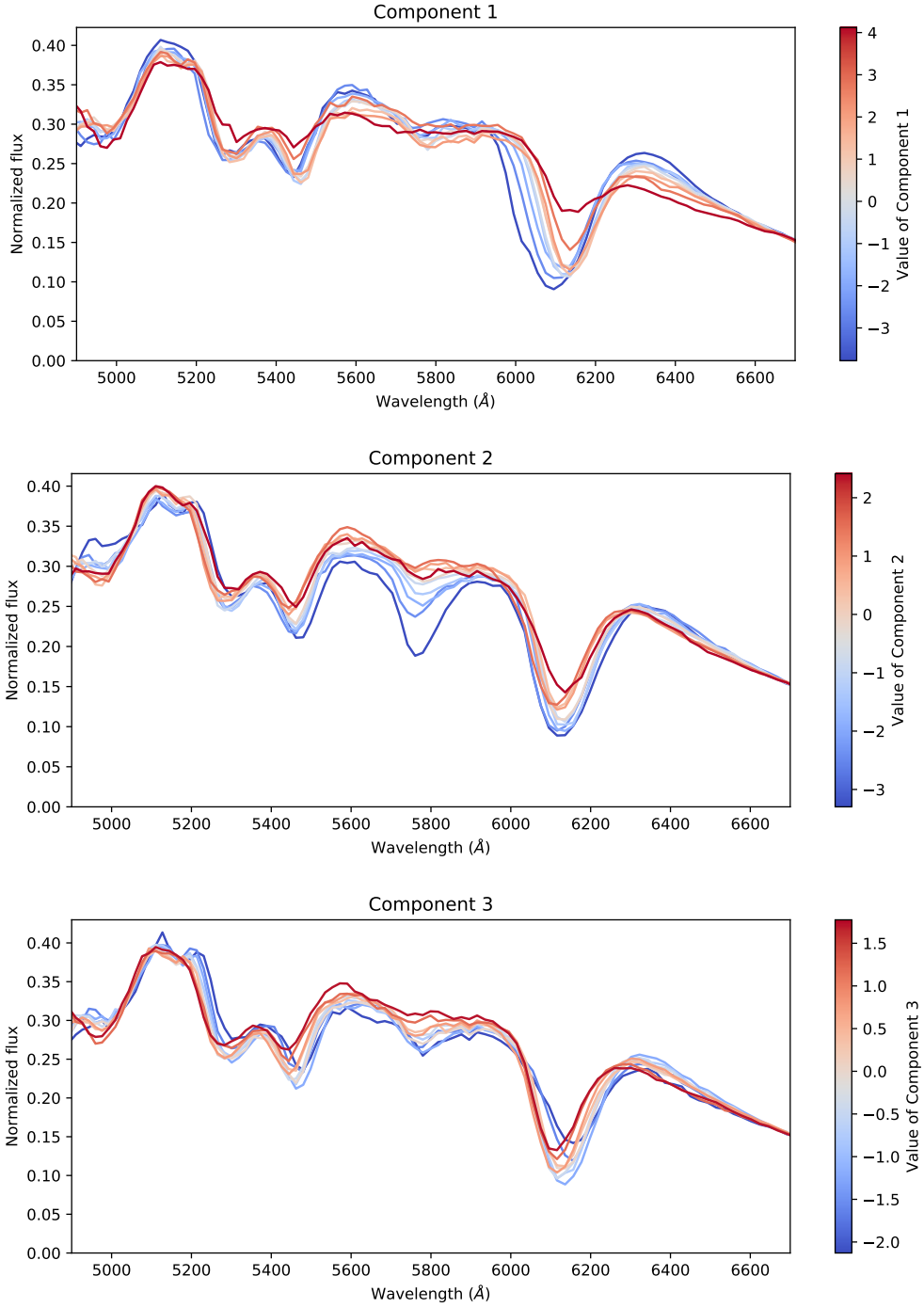



Figure 3.21: Effect of different Isomap components on the spectra of SNe Ia at maximum light between 4900 and 6700 Å, highlighting the Si II 5972 Å and 6355 Å features. See Figure 3.19 for the details of this figure.

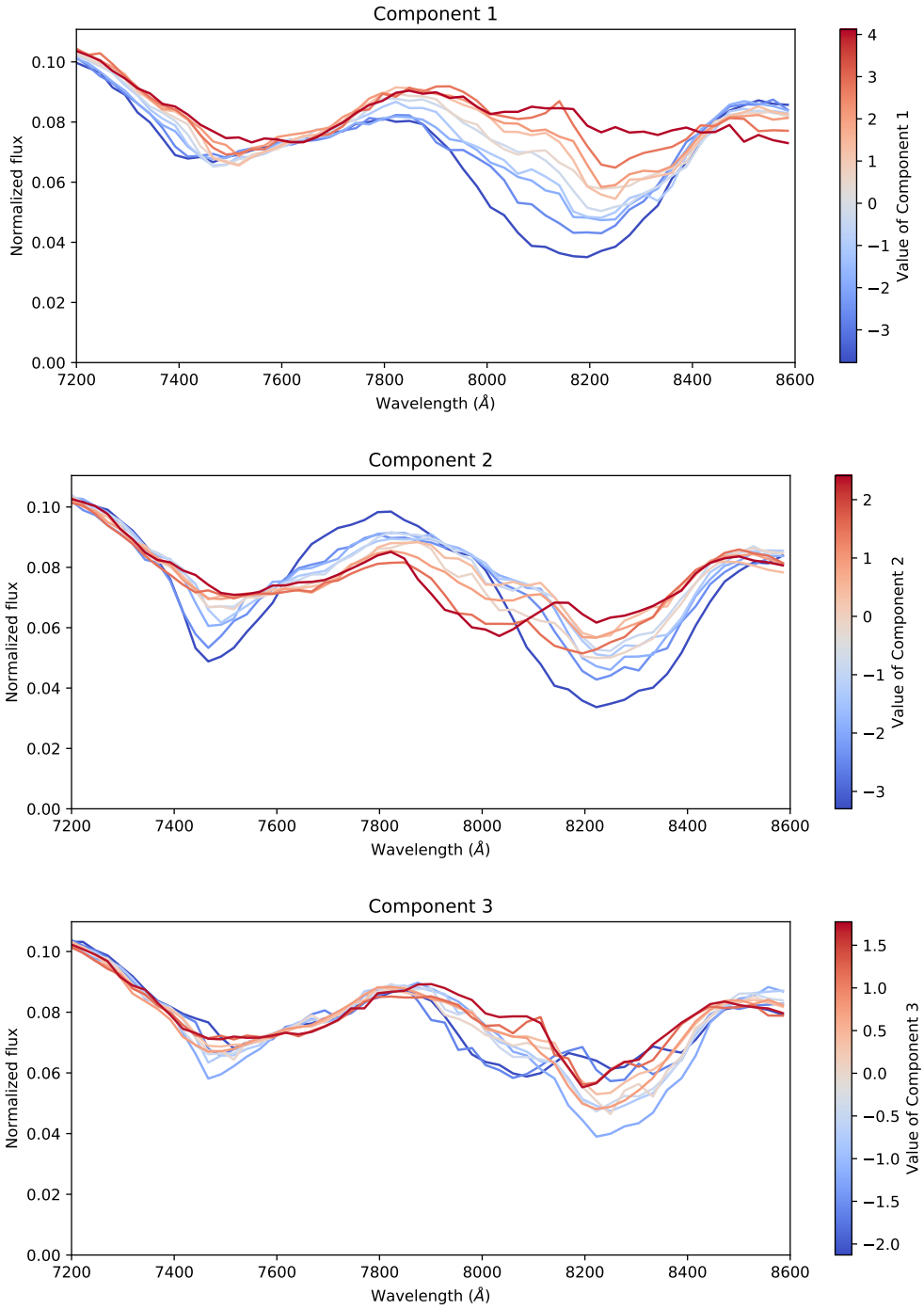



Figure 3.22: Effect of different Isomap components on the spectra of SNe Ia at maximum light between 7200 and 8600 Å, highlighting the Ca II IR triplet features. See Figure 3.19 for the details of this figure.

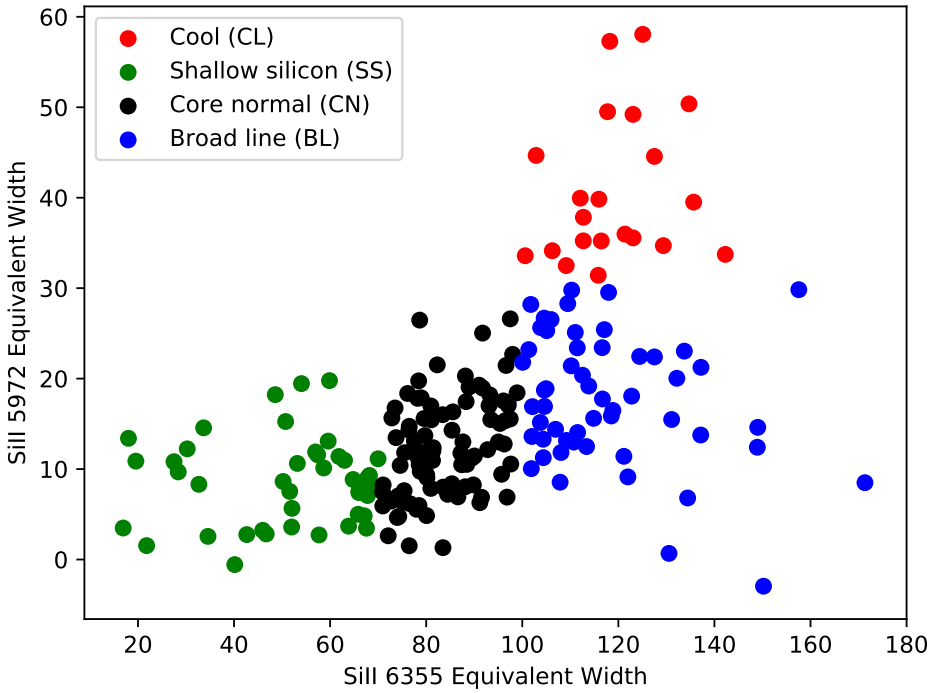



Figure 3.23: Branch classifications for our spectra of SNe Ia at maximum light, using the class definitions from B06.

### 3.5.4 Comparison to the Branch classification scheme

The spectra of SNe Ia near maximum light are often labeled in the literature using the "Branch classification scheme" (Branch et al., 2006, hereafter B06). In this classification scheme, the spectra of SNe Ia are subdivided into four classes based on the pseudo-equivalent widths (pEWs) of the Si II 5972 and 6355 Å absorption features. We measure the pEWs of these lines for our spectra using the methodology described in Chotard et al. (2011). The Branch classifications for our sample are shown in Figure 3.23 using the class definitions from B06. As has been previously shown for large samples of spectra of SNe Ia (e.g. Blondin et al. (2012)), the Branch classifications do not identify distinct subtypes of SNe Ia, and we find a continuous distribution across all of the label boundaries.

We compare the Branch classifications to the result of our Isomap embedding in Figure 3.24. We find that the first two components of our embedding map cleanly separate the different Branch classes from each other: with a set of slices in the Isomap parameter space, we would be able to recreate the Branch classification scheme. There is some confusion between the Shallow Silicon and Core Normal subclasses which appears to be due to uncertainties in the equivalent width measurements used for the Branch classification. In the

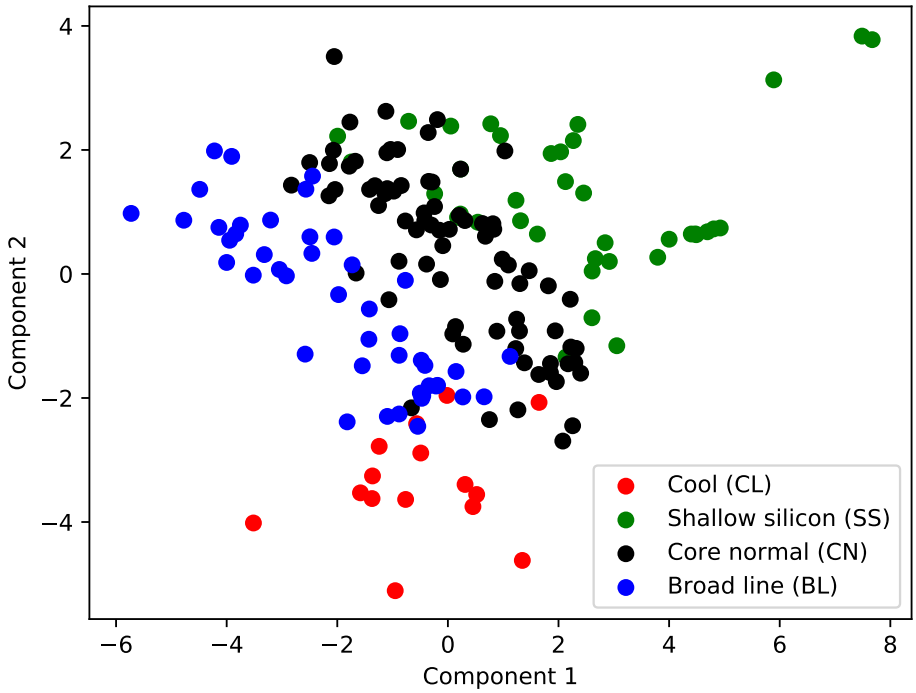



Figure 3.24: Comparison of the Branch classifications to the first two components of the Isomap embedding. We find that with these two components we are able to cleanly recover the Branch classification labels.

Isomap parameter space, the different groups from the Branch classification scheme appear to be "stretched": the Core Normal supernovae lie in a very tight cluster in Figure 3.23, but they are spread over a fairly large region of parameter space in Figure 3.24.

To examine whether or not the Isomap embedding is identifying real differences between different core normal spectra, we calculated the median spectra near different points in the Isomap parameter space where the nearby spectra are identified primarily as "core normal". The results of this procedure are shown in Figure 3.25. All of these median spectra have similar pEWs for the Si II 6355 Å absorption feature, and have a small pEW for the Si II 5972 Å feature. However, the pseudo-continuum level near the Si II 5972 Å feature varies quite dramatically between the different median spectra. As analyses such as B06 focused only on the pEW and ignore the pseudo-continuum level, they were not able to identify this difference. We also see large differences for these spectra near 3800 Å. This is likely due to a set of Si II lines between 3853 and 3863 Å. These differences were not seen in B06 because of a lack of spectral coverage. Finally, we see a spectral sequence for an absorption feature near 8000 Å. This difference was identified in B06 for SN2001el, and was suggested to be due

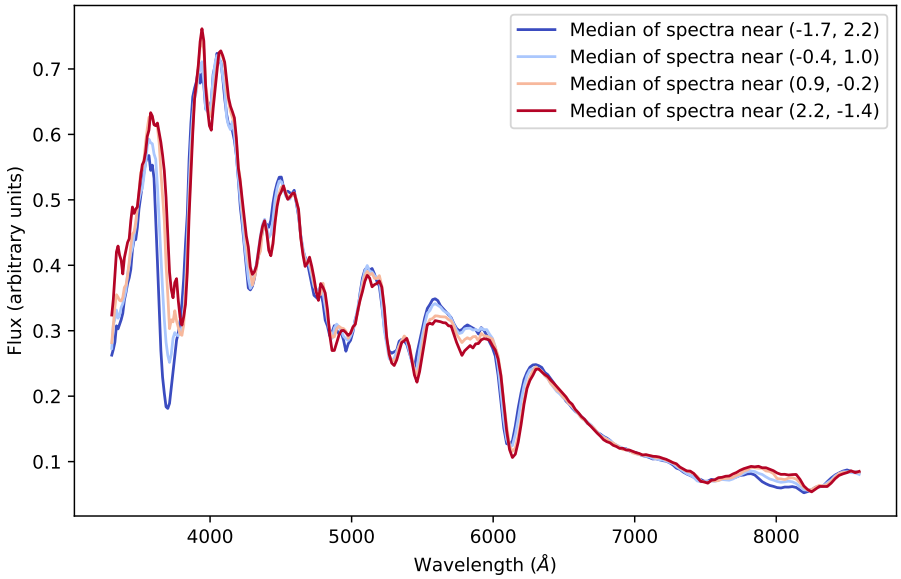



Figure 3.25: Comparison of spectra of Core Normal supernovae that are separated by the Isomap embedding. We calculate the median spectrum for all spectra whose first two Isomap coefficients are within a radius of one unit of a target coordinate. The target coordinates are shown in the legend. All of these spectra are identified as Core Normal by the classification of B06. We find that although these spectra have similar pEWs for the Si II 5972 and 6355 Å absorption features, the Isomap algorithm identifies a spectral sequence in an absorption feature near 3800 Å in the pseudo-continuum level around the Si II 5972 Å feature, and in an absorption feature around 8000 Å.

to high velocity Ca II. Hence, with only the first two components, the Isomap embedding is able to reproduce previous work on the spectral diversity of SNe Ia and identify additional diversity that is not solely limited to the pEWs and velocities of different lines.

### 3.5.5 Recovering the Phillips relation with spectra at maximum light

When looking at broadband light curves of SNe Ia, the diversity of light curves has been primarily explained by the color and stretch of the light curve (Phillips, 1993; Riess et al., 1996; Tripp, 1998). While we can directly measure the color of a supernova's light curve using only a spectrum at maximum light, we can't directly measure the light curve width since we have no baseline in time. One interesting question is whether or not the stretch of the light curve can be recovered directly from a single spectrum at maximum light. It has previously been shown that many of the features of spectra at maximum light correlate with

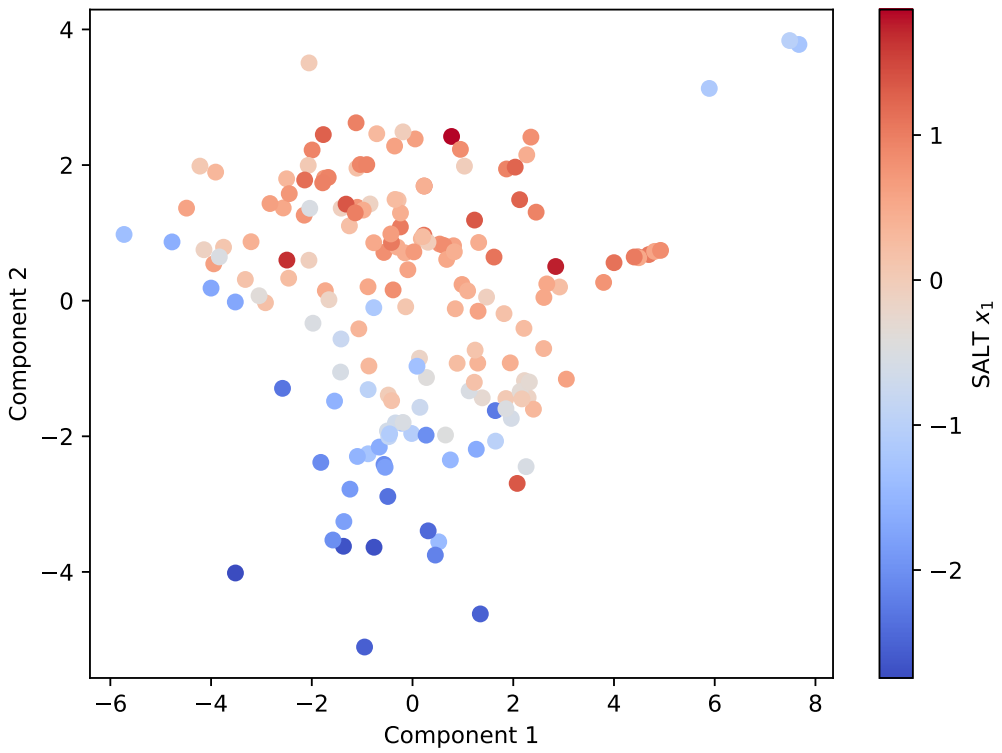



Figure 3.26: Comparison of the SALT2 $x_1$ parameter to the first two components of the Isomap embedding. Each supernova is colored according to its SALT2 $x_1$ value.

stretch, such as the ratio of the pEW of the Si II 5972 Å absorption feature to the pEW of the Si II 6355 Å absorption feature (Nugent et al., 1995).

We calculate the stretch of each of our SNe Ia by fitting the SALT2 model to each of our light curves as described in Section 3.2, and using the SALT2 $x_1$ parameter as our measure of stretch. We show the first two Isomap components, colored as a function of the SALT2 $x_1$ parameter, in Figure 3.26. The sequence of SALT2 $x_1$ values is clearly visible, although it is slightly rotated relative to the individual component axes. To account for this, we fit for the rotation of the components that best predicts SALT2 $x_1$ using a linear combination of the different Isomap components. We find that a combination of the three Isomap components with amplitudes of 0.31, 0.93, and $-0.21$ respectively produces the best prediction of SALT2 $x_1$.

The measured SALT2 $x_1$ value is plotted against this best prediction in Figure 3.27. We also label the SALT2 fits by whether or not they are deemed to be well-fit by SALT2 by the SNfactory pipeline, which is primarily a measure of the normalized median absolute deviation (NMAD) and number of outliers in the SALT2 fit. Note that supernovae can be labeled as having bad SALT2 fits either because they have many observations that are

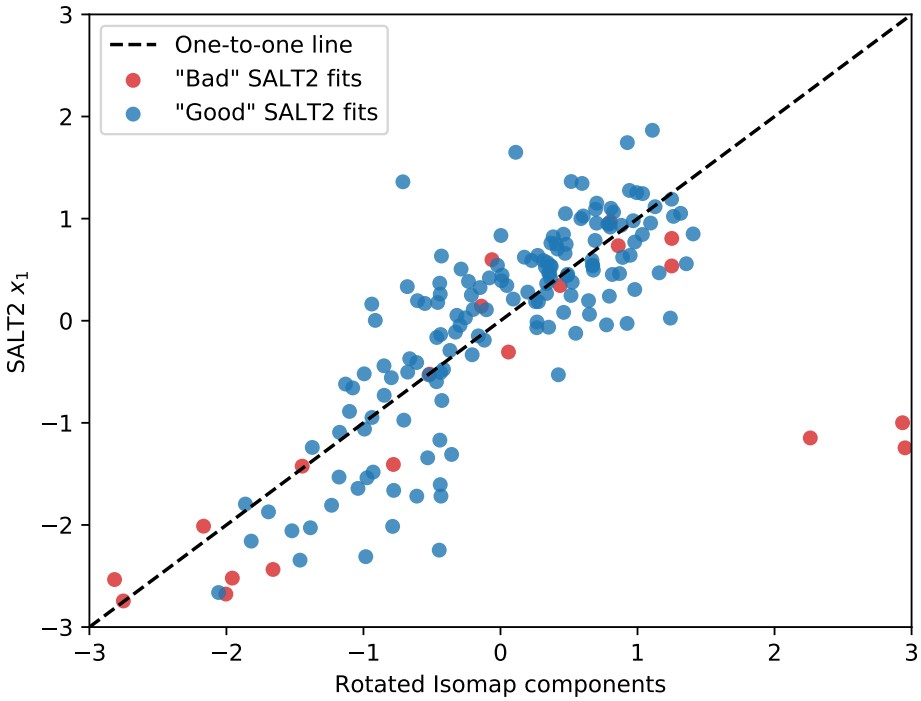



Figure 3.27: Predictions of the SALT2 $x_1$ parameter from the Isomap embedding. We determine the best rotation of the Isomap components to predict the SALT2 $x_1$ parameter, as described in the text. Supernovae with good SALT2 fits are shown in blue, and supernovae whose SALT2 fits have been deemed to be bad are shown in red.

outliers due to data reduction issues or because they are outside of the range of SNe Ia that SALT2 can properly fit. For the SNe Ia with good SALT2 fits, we are able to recover the SALT2 $x_1$ value very well, with a Pearson correlation of 0.79, although the relation appears to be somewhat non-linear.

We find that there are three supernovae for which we are unable to recover their SALT2 $x_1$ values properly. These supernovae can be seen easily in Figure 3.26 as blue outliers in the top right corner of the plot. These three supernovae, LSQ12fhs, SN2005cc and SN2011ay are unusual Type Iax supernovae (Foley et al., 2013). They are not able to be fit very well with SALT2, and are cleanly separated from the rest of the SNe Ia in the Isomap embedding. The spectra of these three supernovae are shown in Figure 3.28. With only three Type Iax supernovae in our sample it is difficult to make definitive statements about this subset, but our results suggest that these supernovae do not belong to the same spectral sequence as the larger sample of SNe Ia.

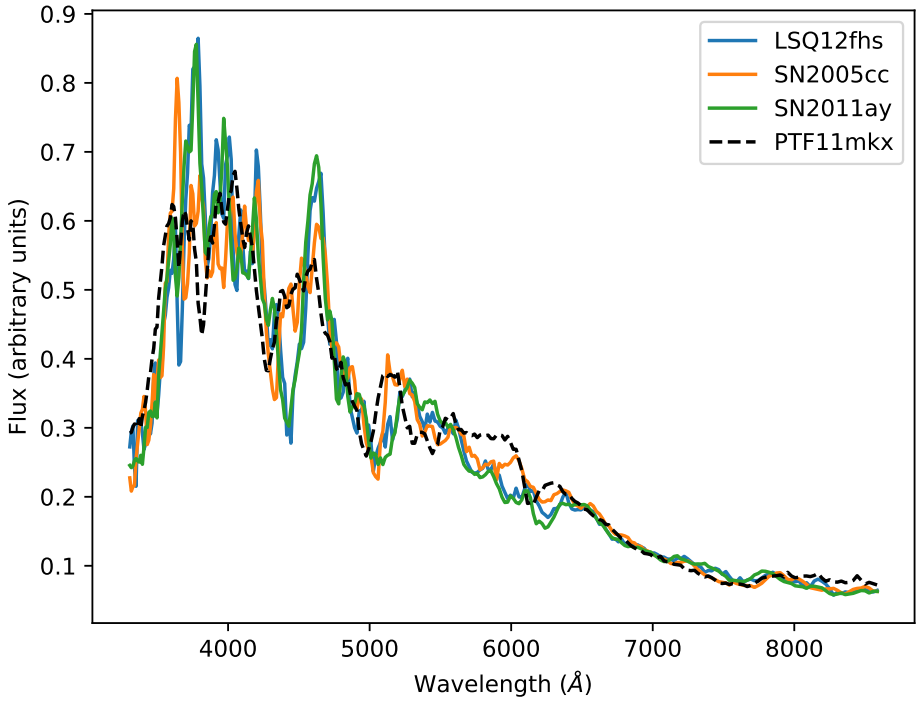



Figure 3.28: Spectra of different Type Iax supernovae that are outliers in the Isomap embedding. For comparison purposes, we also show the spectrum of SNF20070803-005, the next closest supernova to the Type Iax supernovae in the Isomap embedding. While the spectra of these SNe Iax are similar to some of the spectra of SNe Ia at maximum light, they have several unique features, particularly near the Si II 6355Å absorption feature, and are cleanly separated from the remainder of the sample in the Isomap embedding.

## 3.6 Standardization of Type Ia Supernovae With Manifold Learning

### 3.6.1 Sample of Type Ia Supernovae for Standardization Analyses

As part of the Reading Between the Lines (RBTL) algorithm described in Section 3.4, we estimate an overall brightness in magnitudes for each SN Ia in the sample. Because we shifted all of our supernovae to a common redshift, as described in Section 3.2, the output magnitudes can be interpreted as residual magnitudes relative to the mean spectrum of an SN Ia, and ideally would be zero if SNe Ia were perfect standard candles with no observational uncertainties. We look at the properties of these residual magnitudes to evaluate the performance of our techniques for standardizing SNe Ia. To avoid tuning our analysis to op-

timize the distribution of the residual magnitudes, we perform a "blinded" analysis: we split the dataset into a training and a validation set at this stage, and we optimized our analysis while only examining the training set, leaving the validation set to only be examined when the analysis and selection requirements were complete and well-understood for the training set.

The SNfactory pipeline automatically splits the dataset into three categories. For SNe Ia whose light curves are well-fit by the SALT2 model, the pipeline splits them into "training" and "validation" subsets, attempting to keep the observables of SNe Ia such as redshift and color similar between the two subsets. For SNe Ia whose light curves are not fit well by the SALT2 model, using the criteria described in Section 3.2, the pipeline creates a "bad" subset. The "bad" subset contains a combination of unusual SNe Ia and SNe Ia with major issues with their data processing. The "bad" subset is small, containing only 20 of the 175 with spectra at maximum light that pass the criteria for this analysis. Spectra with major issues with data processing could highly influence our analysis, so we chose to use both the SNfactory-labeled "training" and "bad" subsets to develop our model and to decide upon a set of selection requirements that removes spectra with data processing issues. We leave all of the supernovae in the "validation" subset to use as a test once we settle on the final model and selection requirements. A total of 79 of the 175 passing our uncertainty on the spectrum at maximum light requirement are in the SNfactory-labeled "training" subset, 76 are in the "validation" subset, and 20 are in the "bad" subset. The blinding was implemented by immediately deleting all of the magnitudes estimated for SNe Ia in the "validation" subset as soon as the RBTL model had been run, so that it was impossible to accidentally unblind the distributions of the residual magnitudes. At the time of writing, this analysis is still blinded, so all of the following results discussing magnitudes only use the supernovae in our training set comprising the SNfactory-labeled "training" and "bad" subsets.

There are several additional sources of uncertainty that could affect the measured residual magnitudes. As described in Section 3.4.1, because we use the observed redshifts to determine the distances to the supernovae when shifting them to a constant distance away, any additional uncertainties on these redshifts will propagate into our residual magnitudes. Five of the 175 passing our uncertainty on the spectrum at maximum light requirement have redshifts derived from the supernova spectrum rather than ones derived from the host galaxy spectrum. The uncertainties on these redshifts are $\sim$0.005, introducing uncertainties in the measured residual magnitudes of more than 0.10 mag at all redshifts. We therefore reject these supernovae from any analysis of the residual magnitudes.

Similarly, as described in Section 3.4.1, peculiar velocities will introduce an additional contribution to the measured residual magnitudes. Assuming that the peculiar velocities have a dispersion of 300 km/s, the introduced dispersion in brightness becomes comparable to the dispersion in brightness of SNe Ia around $z = 0.02$ where 0.11 mag of dispersion are introduced. We reject 24 SNe Ia that are at redshifts lower than 0.02 out of the 175 passing our uncertainty on the spectrum at maximum light requirement. For redshifts above 0.02, the contribution from peculiar velocities is still significant, so we include these uncertainties in our analyses.

Finally, as discussed in 3.4.1, the dust in the host galaxies of an SN Ia can significantly dim the observed brightness of the SN Ia. While the mean total-to-selective extinction ratio dust was shown to be $R_V = 2.8 \pm 0.3$ in Chotard et al. (2011), there are examples of SNe Ia with very different values of the total-to-selective extinction ratio $R_V$. For example, the extinction of the highly reddened SN2014J has been suggested to be either interstellar dust with $R_V = 1.4 \pm 0.1$ (Amanullah et al., 2014) or circumstellar dust with a similar effect on the observed spectrum (Foley et al., 2014). For this supernova, the difference in assuming a dust law with $R_V = 2.8$ instead of $R_V = 1.4$ would introduce an offset of $\sim$1 mag into the observed magnitude. To avoid having differences in dust properties affect our analysis of residual magnitudes, we choose to reject any supernovae with a measured $A_V > 0.5$. For a supernova with $A_V = 0.5$, a 20% difference in $R_V$ would introduce a $\sim$0.1 mag offset into its residual magnitude. 19 supernovae have measured $A_V$ values larger than this threshold out of the 175 passing our uncertainty on the spectrum at maximum light requirement.

The results of all of these selection requirements are summarized in Table 3.2. We find that a total of 135 SNe Ia pass all of the selection requirements for their magnitudes to be expected to have low contributions from non-intrinsic sources, and a total of 73 of those SNe Ia are included in our training set with 62 held back for the validation set.

### 3.6.2 Residual Magnitudes from Reading Between the Lines

The residual magnitudes measured by the RBTL algorithm are shown in Figure 3.29, both with and without the selection requirements from Section 3.6.1 applied. With the selection criteria applied, we find that the residual magnitudes of the remaining sample of SNe Ia have a dispersion with a standard deviation of $0.123 \pm 0.013$ mag and an NMAD of $0.092 \pm 0.017$ mag. The uncertainties on these measurements, and all other uncertainties on measurements of dispersion in this analysis, are calculated using bootstrapping (Efron, 1979). Note that these residual magnitudes have been corrected using a baseline color law (correcting for both dust and any intrinsic color), but they have not been corrected at all for intrinsic properties of SNe Ia that could affect the overall brightness of the SN Ia, such as the stretch of the supernova. In the rest of this Section, we will use the RBTL-estimated residual magnitudes as a baseline for additional studies on how the intrinsic variation of SNe Ia can be used to improve our estimates of the brightness of SNe Ia.

### 3.6.3 RBTL Residual Magnitudes and Manifold Learning

The RBTL algorithm provides a robust estimate of the brightness and color of the spectrum of an SN Ia. However, as described in Section 3.4.1, if there is intrinsic variability that affects the spectrum in a similar manner to the brightness or the color, then it will be confused as extrinsic variability by the RBTL algorithm. This means that at some level the RBTL residual magnitudes will include contributions from intrinsic variability. Assuming that the intrinsic variability that affects the brightness of the supernova also affects other intrinsic properties of the spectra, such as the absorption features seen in the spectra of

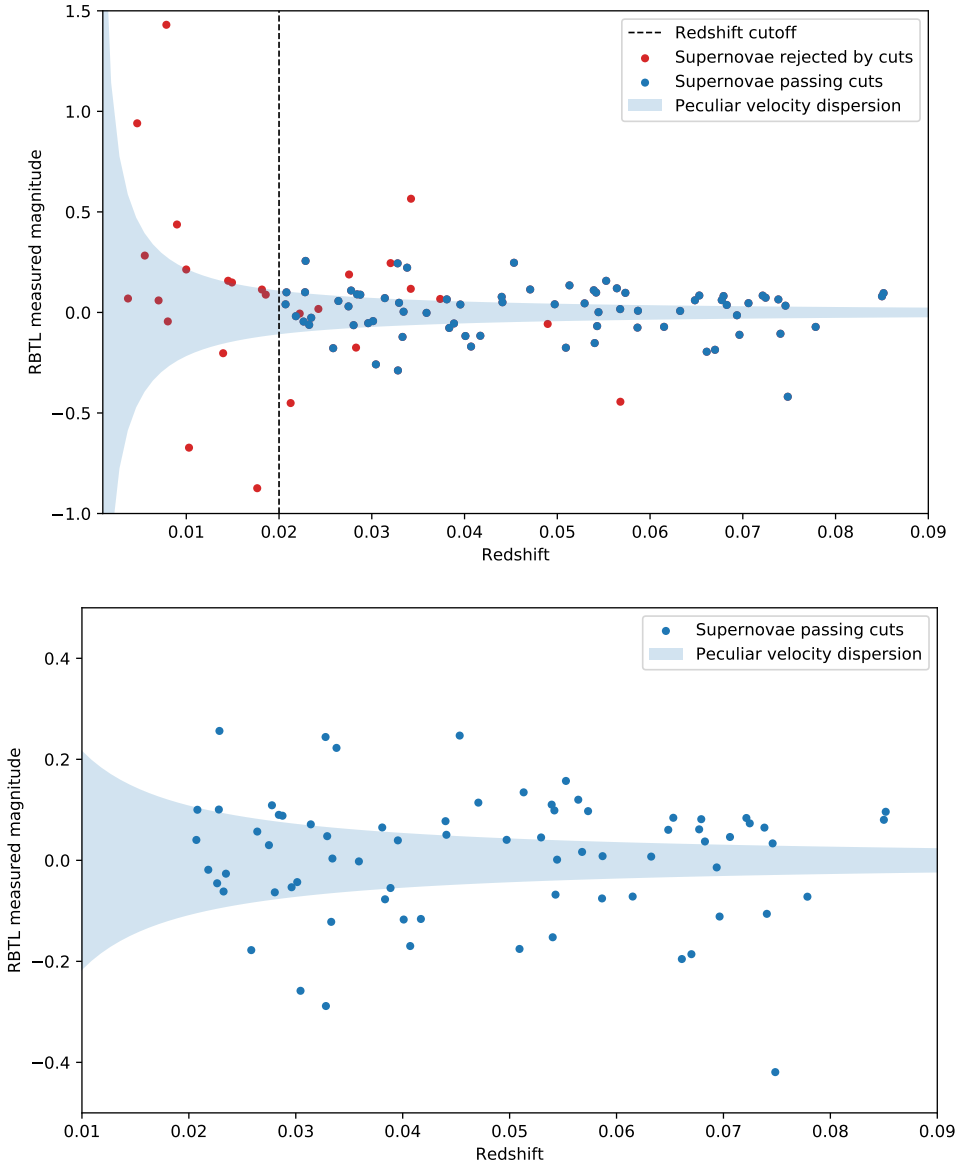



Figure 3.29: Residual magnitudes estimated by the Reading Between the Lines algorithm. Top panel: Supernovae passing our selection requirements of having low contributions to the residual magnitudes from redshift and dust uncertainties are shown in blue, and supernovae failing these selection requirements are shown in red. The shaded blue region indicates one standard deviation of the expected contribution to the residual magnitudes from peculiar velocities assuming a velocity dispersion of 300 km/s. Bottom panel: Zoom in on this figure with only the supernovae passing our selection requirements shown.

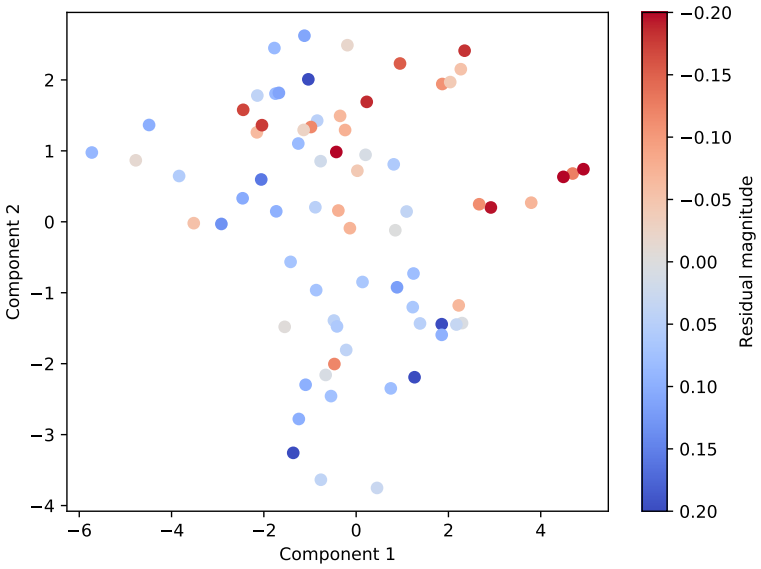



Figure 3.30: RBTL residual magnitudes as a function of the Isomap components 1 and 2. Each data point represents a single supernova. The data points are colored according to their RBTL residual magnitudes, as labeled in the colorbar.

SNe Ia, then we can use the intrinsic diversity measured from the spectral features to remove the intrinsic contributions to the residual magnitudes. For example, light curve fitters such as SALT2 use the measured intrinsic diversity of the light curve stretch to correct for its contributions to the estimated brightnesses of SNe Ia.

The Isomap embedding described in Section 3.5 produces a parametrization of the intrinsic diversity of SNe Ia, and we examine whether this parametrization provides additional information about the peak brightnesses of SNe Ia. In Figures 3.30 through 3.32, we show the RBTL residual magnitudes as a function of the Isomap components. There are visible trends in the residual magnitudes as a function of each of the components, indicating that supernovae with similar intrinsic variation have similar residual magnitudes.

### 3.6.4 Standardization of the RBTL Residual Magnitudes with the Isomap Embedding

As the Isomap embedding is designed to produce a highly non-linear manifold, there is no reason to expect that the resulting embedding will be linearly correlated with the peak brightnesses of SNe Ia, so linear standardization as is traditionally done for classifiers such as SALT2 is unlikely to be sufficient. Instead, we choose to use Gaussian Process (GP) regression to model the residual magnitudes of SNe Ia over the Isomap embedded parameter space. For an introduction to Gaussian process regression, see Section 1.8. We choose

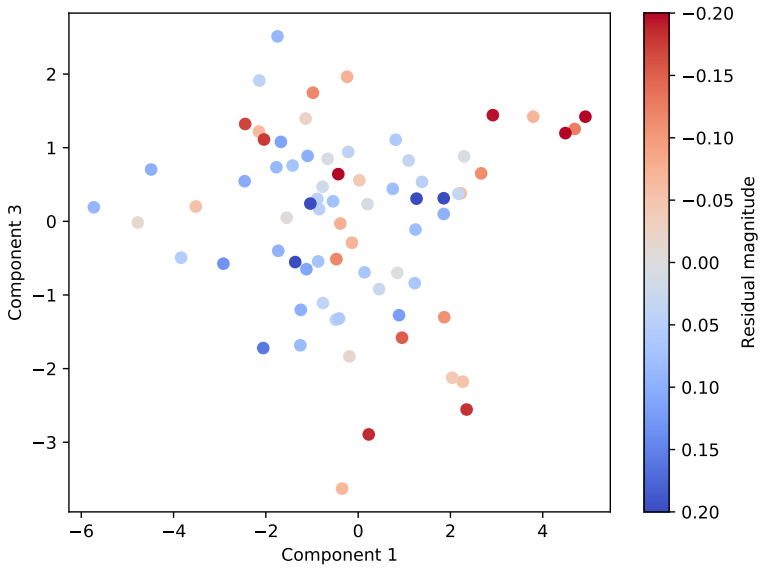



Figure 3.31: RBTL residual magnitudes as a function of the Isomap components 1 and 3. Each data point represents a single supernova. The data points are colored according to their RBTL residual magnitudes, as labeled in the colorbar.

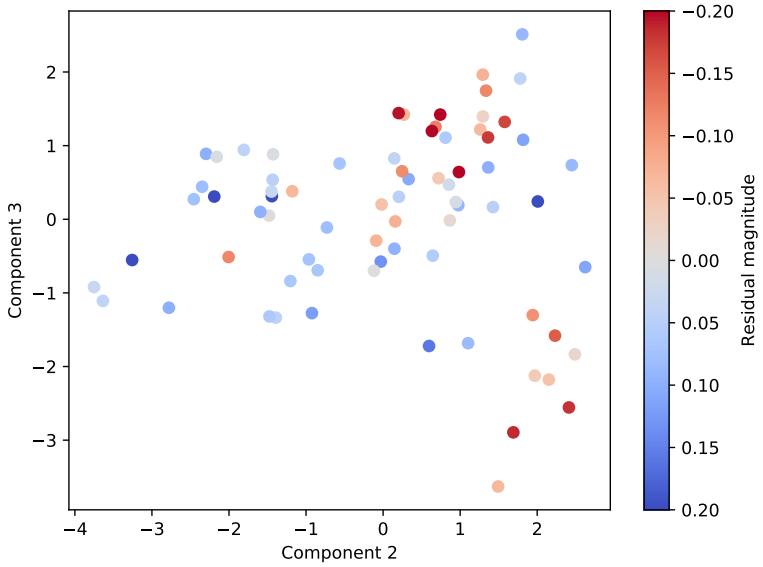



Figure 3.32: RBTL residual magnitudes as a function of the Isomap components 2 and 3. Each data point represents a single supernova. The data points are colored according to their RBTL residual magnitudes, as labeled in the colorbar.

Table 3.4: Parameters for the Gaussian process standardization model described in Section 3.6.4.

| Parameter | Value |
|---|---|
| Color law $R_V$ | $2.77 \pm 0.20$ |
| GP kernel amplitude | $0.098 \pm 0.029$ mag |
| GP kernel length scale | $2.23 \pm 1.40$ |
| Intrinsic dispersion | $0.071 \pm 0.013$ mag |

to use a Gaussian process with a Matérn 3/2 kernel to describe the residual magnitudes of SNe Ia. This kernel has two parameters: an amplitude that sets the amplitude scale of coherent intrinsic brightness variations for different SNe Ia and a length scale that determines the distances in parameter space over which the intrinsic brightness varies. The Isomap embedding is designed to maintain the "spectral distances" between SNe Ia in the embedded space, so we choose to use a three-dimensional Matérn kernel with a single length scale.

When determining the RBTL residual magnitudes, as described in Section 3.4.1, we chose to use a fiducial color law with $R_V = 2.8$. If the wrong mean value of $R_V$ is chosen, then the first order effect is that a correlation is introduced between the measured color and residual magnitude. For this reason, we add an additional linear correction term in color to our model of the residual magnitudes that we interpret as the true $R_V$ value used for the color correction. Finally, we expect there to be some residual dispersion even after these corrections due to additional unmodeled effects. We model this residual dispersion using a Gaussian with an intrinsic dispersion parameter $\sigma_{int}$ that will be determined.

We implement this Gaussian process model using the `George` package (Ambikasaran et al., 2015). The full model has a total of four parameters: the Gaussian process kernel amplitude, the Gaussian process kernel length scale, the color correction and the intrinsic dispersion. We fit this model to the sample of SNe Ia described in Section 3.6.1 optimizing the maximum likelihood. The results of this fit are shown in Table 3.4.

We find that our measurements prefer an $R_V$ value of $2.77 \pm 0.20$, which is consistent with the results of Chotard et al. (2011) and the fiducial color law that we assumed in Section 3.4.1. We verified that different choices of fiducial color laws result in similar recovered $R_V$ values when rerunning the entire analysis. The recovered intrinsic dispersion of $0.071 \pm 0.013$ mag is significantly lower than what is typically found for light curve fitters that only rely on stretch and color, and is consistent with the results of F15 for pairs of twin supernovae, as will be discussed in Section 3.7.1. To understand the behavior of the GP, we show the GP predictions after applying the updated $R_V$ color correction in Figures 3.33 through 3.35. In these figures, we hold one component value fixed to zero, and show the effects of varying the other two components over the Isomap embedded space. Qualitatively, the GP appears to be able to capture the non-linear diversity in RBTL residual magnitudes of SNe Ia.

The GP effectively uses the brightnesses of nearby SNe Ia to generate its predictions. To

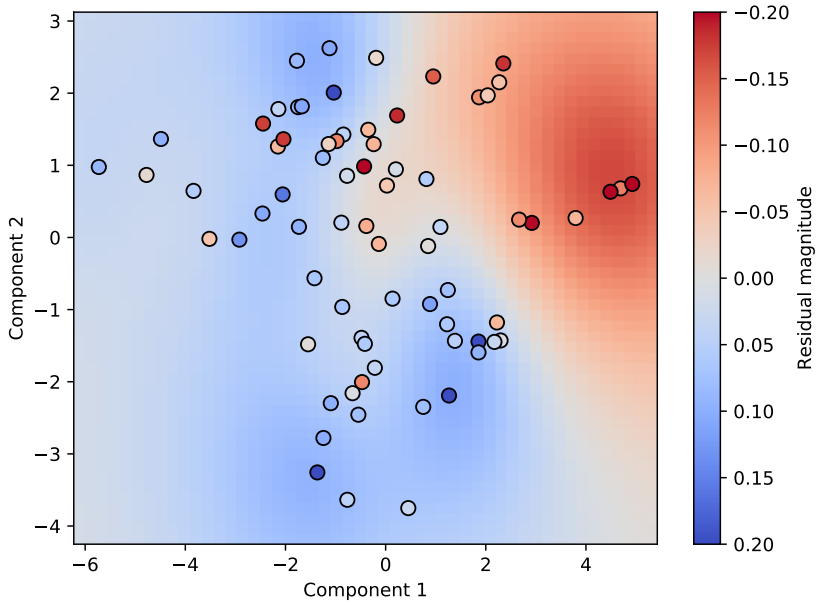



Figure 3.33: GP predictions for the RBTL residual magnitudes as a function of the Isomap components 1 and 2. Each data point represents a single supernova. The data points are colored according to their color-corrected RBTL residual magnitudes, as labeled in the colorbar, and are identical to the data points shown in Figure 3.30. The GP predictions are shown as a smooth color in the background, indicating the amplitude of the GP prediction for each value of component 1 and 2 when component 3 is fixed to zero. Note that this figure only shows a slice through the GP predictions, and the individual predictions will also include information from component 3 that cannot be plotted in only two dimensions.

generate GP predictions of the residual magnitudes for our sample of SNe Ia, we calculate individual "leave-one-out" predictions for each supernova where the GP is conditioned on all SNe Ia except the supernova of interest. This ensures that the measured brightness of the supernova itself cannot contribute to its own predictions. Using these predictions, we calculate GP corrected residual magnitudes for each SN Ia in our sample. We find that the GP corrected residual magnitudes have a dispersion in brightness with a standard deviation of 0.100 ± 0.010 mag and an NMAD of 0.079 mag. These residual magnitudes will be compared with other methods of calculating residual magnitudes in Section 3.7.1.

### 3.6.5 Twins Standardization

As shown in F15, "twin" SNe Ia that have similar spectral features have a much lower dispersion in brightness than the larger sample of SNe Ia. As described in Section 3.5.2, the distances between spectra that we calculate for the Isomap algorithm are equivalent to the "twinness" from F15 for high S/N spectra. The RBTL residual magnitudes are also similar to

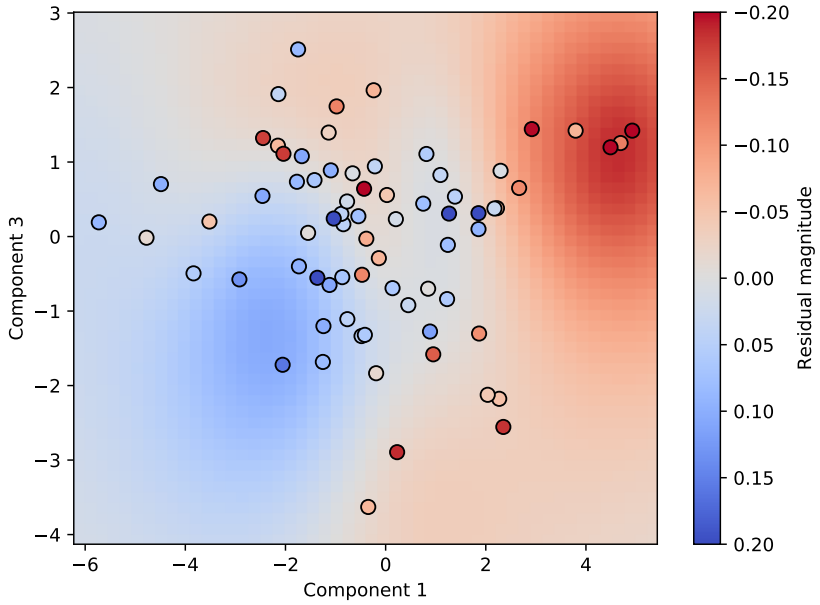



Figure 3.34: GP predictions for the RBTL residual magnitudes as a function of the Isomap components 1 and 3. See Figure 3.33 for details of this figure.

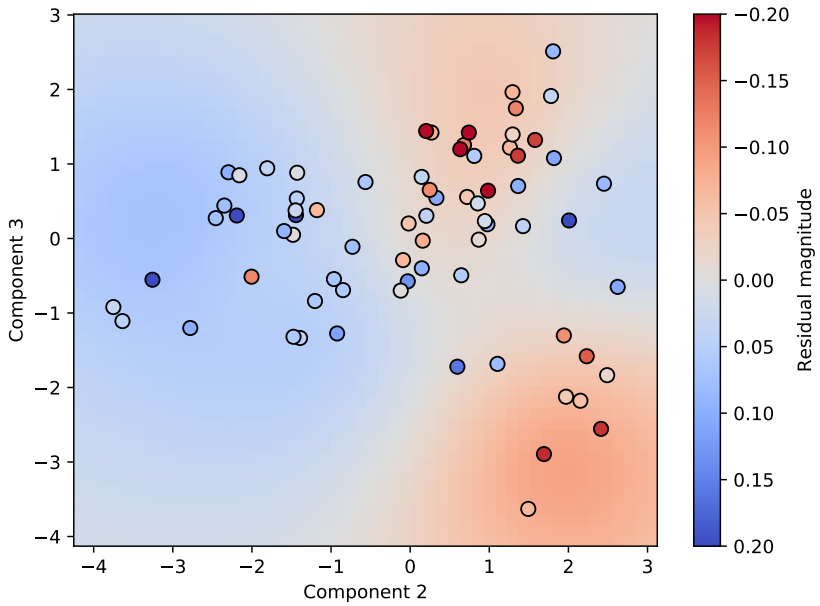



Figure 3.35: GP predictions for the RBTL residual magnitudes as a function of the Isomap components 2 and 3. See Figure 3.33 for details of this figure.

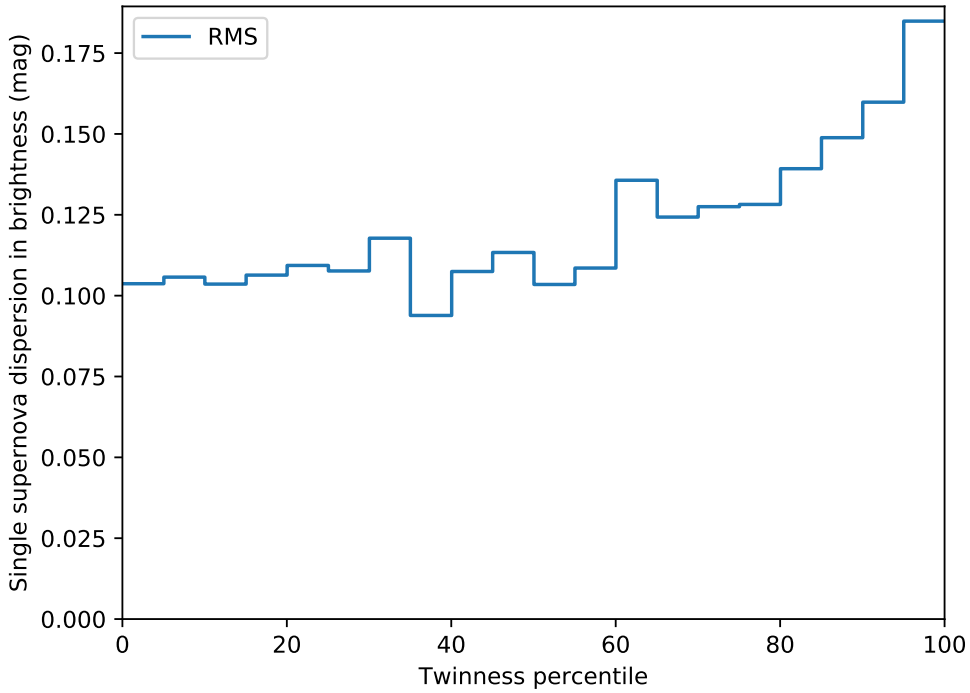



Figure 3.36: Difference in measured RBTL residual magnitudes for pairs of SNe Ia as a function of their twinness, following a similar procedure to F15. We rank each pair of SNe Ia in terms of their twinness, and calculate the RMS of the difference in brightness for all sets of pairs in a bin, corrected to report the implied dispersion in brightness for a single supernova.

the brightness differences calculated in F15, although we weight the spectra by the intrinsic dispersion when estimating the RBTL residual magnitudes.

For direct comparison with the results of F15, for each pair of SNe Ia, we calculate their "twinness percentile", i.e. the percentile of their twinness relative to the twinness values for all sets of pairs. In bins of the twinness percentile, we calculate the RMS of the RBTL magnitude differences for pairs of SNe Ia in that bin. In this analysis, we are measuring the dispersion in brightness of pairs of SNe Ia, so we divide the resulting RMS by $\sqrt{2}$ to report the implied single SN Ia dispersion in brightness that can be directly compared to other standardization techniques. The results of this procedure are shown in Figure 3.36. We find that the best 20% of twins pairs have an implied single SN Ia dispersion with an RMS of 0.105 mag and an NMAD of 0.094 mag.

### 3.6.6 SALT2 Standardization

For comparison purposes, we standardize our SNe Ia using SALT2. From the SALT2 fits described in Section 3.2, we determine the SALT2 parameters $x_{1,i}$ and $c_i$ of each SN Ia along with the observed peak B-band magnitude $m_{B,i}$. Here we restrict our sample to the set of supernovae passing both the SALT2 fit criteria described in Section 3.2 and the redshift and color criteria described in Section 3.6.1. Assuming a fiducial cosmology (described in Section 3.2), we calculate the distance modulus for each supernova $\mu_i$. Given a set of standardization parameters $\alpha$ and $\beta$ along with the absolute magnitude of an SN Ia at maximum light $M_B$, we calculate the SALT2 residual magnitudes $m_{res,i}$ for each supernova as:

$$m_{res,i} = m_{B,i} - M_B - \mu_i + \alpha \times x_{1,i} - \beta \times c_i \tag{3.10}$$

We estimate the uncertainty on $m_{res,i}$ for each supernova by propagating the uncertainties from the SALT2 fits for $x_{1,i}$, $c_i$ and $m_{B,i}$. Additionally, we include the contribution of peculiar velocities $\sigma_{pv,i}$ assuming a dispersion of 300 km/s as described in Section 3.4.1. The SALT2 model is known to not explain all of the dispersion of SNe Ia, so the uncertainty on this residual magnitude must include an intrinsic dispersion term $\sigma_{int}$. The final uncertainty model for $m_{res,i}$ is:

$$\begin{aligned}\sigma^2_{m_{res},i} =& \sigma^2_{m_B,i} + \sigma^2_{pv,i} + \sigma^2_{int} + \alpha^2\sigma^2_{x_1,i} + \beta^2\sigma^2_{c,i} &(3.11)\\ &+ 2\alpha\ \mathrm{Cov}[m_{B,i}, x_1] - 2\beta\ \mathrm{Cov}[m_B, c] - 2\alpha\beta\ \mathrm{Cov}[x1, c] &(3.12)\end{aligned}$$

Given a set of parameters for the above equations, we define the following weighted RMS (WRMS) and chi-square per degree-of-freedom ($\chi^2/DoF$):

$$\mathrm{WRMS} = \sqrt{\frac{\sum_i^N m^2_{res,i}/\sigma^2_{m_{res},i}}{\sum_i^N 1/\sigma^2_{m_{res},i}}} \tag{3.13}$$

$$\chi^2/DoF = \frac{1}{N-4}\sum_i^N \frac{m^2_{res,i}}{\sigma^2_{m_{res},i}} \tag{3.14}$$

where $N$ is the total number of supernovae in the sample. We iteratively minimize these two equations to determine the values of $\alpha$, $\beta$, $M_B$ and $\sigma_{int}$. First, we set $\sigma_{int}$ to a guess of 0.1 mag. We then minimize the WRMS in Equation 3.13 to determine the optimal parameters for $\alpha$, $\beta$, and $M_B$. Given the fitted values of these parameters, we determine the value of $\sigma_{int}$ that sets the $\chi^2/DoF$ in Equation 3.14 to 1. We repeat these two fits until the parameter values converge.

The results of this SALT2 standardization procedure are summarized in Table 3.5. We find a WRMS of 0.149 mag for this sample. Because the light curves of the SNe Ia in our sample are almost all well measured, there is little difference in the uncertainties $\sigma_{m_{res},i}$

Table 3.5: SALT2 standardization parameters fit using the procedure described in Section 3.6.6.

| Parameter | Value |
|---|---|
| $M_B$ | -19.07 mag |
| $\alpha$ | 0.131 |
| $\beta$ | 3.02 |
| $\sigma_{int}$ | 0.126 mag |
| Weighted RMS of $m_{res,i}$ | 0.149 mag |
| Unweighted standard deviation of $m_{res,i}$ | 0.150 ± 0.011 mag |
| NMAD of $m_{res,i}$ | 0.116 ± 0.013 mag |

between SNe Ia, with values ranging between 0.14 mag to 0.18 mag. Hence the unweighted standard deviation of 0.150 ± 0.011 mag is nearly identical to the WRMS. To compare with other standardization techniques and avoid having different weights for different techniques, we therefore use the unweighted standard deviation in further analyses. Interestingly, we also find that this distribution has a tight core with wider wings compared to a Gaussian distribution, as seen in the low NMAD of only 0.116 ± 0.013 mag for this sample. This suggests that while SALT2 standardization is effective for the core of the sample, there are some regions of the parameter space of SNe Ia where there is additional dispersion in the peak brightness that is not captured by SALT2.

We can identify the regions of parameter space where SALT2 standardization is biased using the Isomap embedding. As shown in Figure 3.37, SALT2 corrected residual magnitudes show a strong bias for large values of the first Isomap component. Comparing SALT2 residual magnitudes with Isomap component 1 values of less than 2 to residual magnitudes with component 1 values greater than 2, we find a difference of 0.172 ± 0.038 mag in the mean brightnesses between these two samples. Note that the location of this cut was selected after examining the distribution which reduces the statistical power of this measurement. However, this strongly suggests that the Isomap algorithm is identifying additional intrinsic variation of SNe Ia that affects the peak brightness estimates of SNe Ia and that is not captured by the SALT2 algorithm.

### 3.6.7 Standardization using SALT2 with Isomap components

We investigate a hybrid standardization method where we use the peak brightness and color from a SALT2 light curve fit, and the Isomap components to parametrize the intrinsic variation of SNe Ia instead of the SALT2 $x_1$ parameter. While the RBTL method requires that the input spectra be spectrophotometric to obtain accurate estimates of the brightness and color of each supernova, the Isomap components themselves could potentially be derived from slit-based spectrographs or other forms of spectroscopy that are not flux-calibrated since most of the information is in local spectral feature variation. External estimates of the

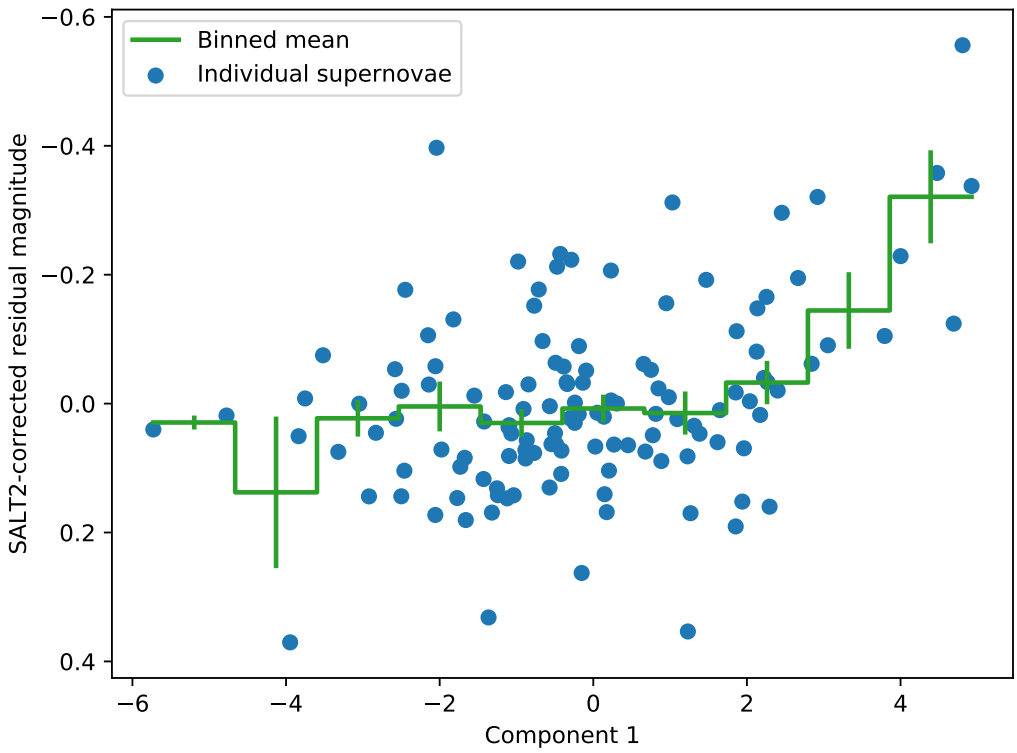



Figure 3.37: Comparison of SALT2 corrected residual magnitudes to the first Isomap component. The SALT2 residual magnitudes are shown for individual supernovae as blue points. The mean residual magnitude in evenly-spaced bins of this component are shown with green lines. We estimate uncertainties on the mean residual magnitude in each bin using the sample standard deviation in each bin. We find that SALT2 Hubble residuals show a strong bias for large values of the first Isomap component.

brightness and color, from SALT2 fits for example, could then be corrected using the Isomap components.

To test whether this method of standardization is effective, we ran the GP standardization procedure described in Section 3.6.4 using the Isomap components, but with the raw SALT2 residual magnitudes and colors (uncorrected for $x_1$) instead of the RBTL residual magnitudes. As for the RBTL analysis, we include a linear term to correct for the SALT2 color which is equivalent to the $\beta$ parameter in traditional SALT2 analyses. The results of the GP fit for this model are summarized in Table 3.6.

As in Section 3.6.4, we calculate the GP predicted brightness for each supernova in our sample using a leave-one-out prediction to ensure that we are not biasing the predictions. When raw SALT2 residual magnitudes are standardized using the Isomap embedding, the standardization performance is significantly better than standardizing only on SALT2 $x_1$.

Table 3.6: Parameters for the hybrid Isomap and SALT2 standardization model described in Section 3.6.7.

| Parameter | Value |
|---|---|
| SALT2 color standardization term ($\beta$) | 3.09 ± 0.16 |
| GP kernel amplitude | 0.423 ± 0.259 mag |
| GP kernel length scale | 9.15 ± 5.50 |
| Intrinsic dispersion | 0.090 ± 0.009 mag |

The resulting standard deviation of the SALT2 + Isomap standardized residual magnitudes is 0.117 ± 0.007 mag compared to 0.146 ± 0.012 mag for SALT2 with traditional $x_1$ standardization for the same set of SNe Ia.

To illustrate the effect of the Isomap components for standardization of raw SALT2 residual magnitudes, we reproduced Figure 3.33 for the SALT2 + Isomap standardization analysis in Figure 3.38. We remove the linear color correction, and show the remaining residual magnitudes for each supernova in the sample along with the GP predicted brightness for each point of the parameter space. First, we note that the amplitude of the differences in the residual magnitudes before GP correction is much larger for the raw SALT2 magnitudes compared to the raw RBTL magnitudes, as reflected in the GP kernel amplitudes of 0.423 ± 0.259 mag and 0.098 ± 0.029 mag respectively. This is expected: in the RBTL analysis we used the full spectrum at maximum light to estimate the peak brightness of SNe Ia, and we explicitly avoided using regions of the spectrum with large variation in this estimate. The SALT2 peak brightness on the other hand is explicitly estimated in the B-band without taking intrinsic variation into account.

We show the difference in the applied standardization between the SALT2 + Isomap standardization and traditional SALT2 standardization in Figure 3.39. We find major differences for these standardization methods, especially for SALT $x_1$ parameters between 0 and 1, that are related to the first Isomap component. The applied standardization varies by as much as 0.5 mag for SNe Ia with the same SALT2 $x_1$ values.

## 3.7 Discussion

### 3.7.1 Comparison of Standardization Techniques

We have defined several different standardization techniques that output residual magnitudes for each supernova in a sample. In this Section, we compare the residual magnitudes from these different techniques. In Section 3.4, we showed that we can obtain a robust estimate of a supernova's overall brightness and color using regions of the spectrum with low intrinsic dispersion in a process that we called "Reading Between the Lines". We refer to the residual magnitudes derived from this technique as the "Raw RBTL" residual magnitudes because they have not been corrected for intrinsic dispersion. In Section 3.6.4, we showed

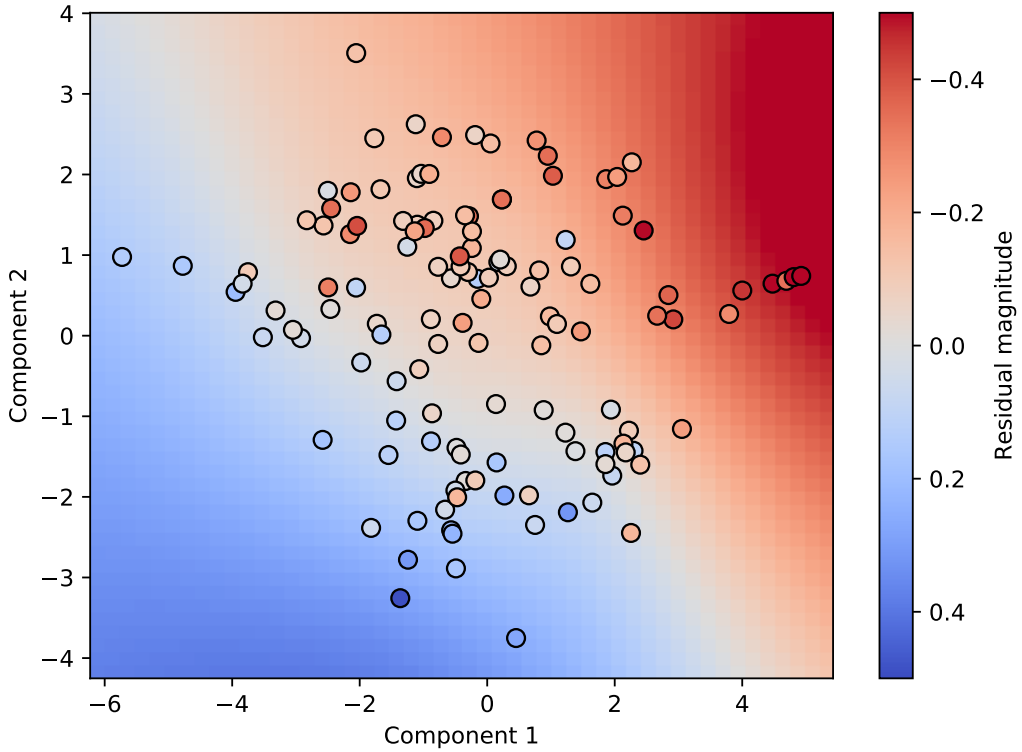



Figure 3.38: GP predictions for the SALT2 residual magnitudes as a function of the Isomap components 1 and 2. Each data point represents a single supernova. The data points are colored according to their color-corrected SALT2 residual magnitudes, as labeled in the colorbar. The GP predictions are shown as a smooth color in the background, indicating the amplitude of the GP prediction for each value of component 1 and 2 when component 3 is fixed to zero. Note that this figure only shows a slice through the GP predictions, and the individual predictions will also include information from component 3 that cannot be plotted in only two dimensions.

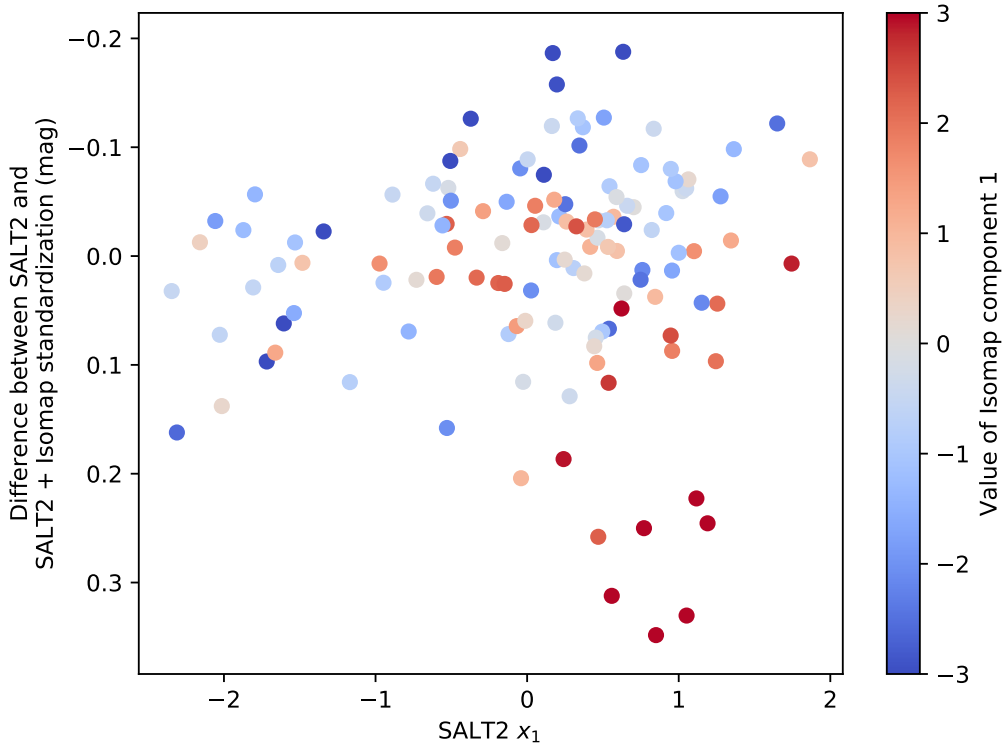



Figure 3.39: Differences between the SALT2 + Isomap standardization and traditional SALT2 standardization. We show the individual differences for all SNe Ia in our sample. The data points are colored by the value of the first Isomap component. We find a large difference in the standardized magnitudes for SNe Ia with SALT2 $x_1$ values between roughly 0 and 1, related to the measured value of the first Isomap component for these SNe Ia.

how a Gaussian process model over the Isomap parameter space can be used to correct the RBTL residual magnitudes, which we will refer to as "RBTL + Isomap standardization". In Section 3.6.6, we applied traditional "SALT2 + $x_1$ standardization" to our sample. In Section 3.6.7, we corrected estimates of the peak brightness and color of an SN Ia from SALT2 using the Isomap embedding, which we call "SALT2 + Isomap standardization".

These standardiztion techniques all required different selection criteria in order to be applied. The raw RBTL, RBTL + Isomap and SALT2 + Isomap standardization techniques both require a spectrum at maximum light with low uncertainty for the Isomap embedding, as described in Section 3.4.4. Both of the SALT2-based standardization techniques require a set of SNe Ia with valid SALT2 fits. Finally, the raw RBTL and RBTL + Isomap analyses are still blinded, so those analyses only include SNe Ia in the training set described in Section 3.6.1. Statistics for the dispersions of residual magnitudes for each of the these techniques are shown for different subsets of SNe Ia in Table 3.7.

Table 3.7: Comparison of standardization results for different methods and samples. For each set of cuts, we show both the NMAD and the standard deviation of the corrected residual magnitudes for each of the different methods of calculating residual magnitudes. We also show an estimate of the the standard deviation with the peculiar velocity removed (see text for details). If a standardization method cannot be applied to some of the supernovae in the chosen sample, (e.g.: SALT2 standardization requires a valid SALT2 fit), then we report "–" instead. Note that with the same sets of cuts, the quoted uncertainties are highly correlated between the different standardization techniques, and therefore do not apply to relative comparisons between different techniques.

| Cuts applied | Number of SNe Ia | Statistic | Raw RBTL dispersion (mag) | RBTL + Isomap dispersion (mag) | SALT2 + $x_1$ dispersion (mag) | SALT2 + Isomap dispersion (mag) |
|---|---|---|---|---|---|---|
| Maximum spectrum + training cuts | 73 | NMAD | $0.092 \pm 0.017$ | $0.079 \pm 0.014$ | – | – |
| | | Standard deviation | $0.123 \pm 0.012$ | $0.100 \pm 0.010$ | – | – |
| | | Pec. vel. removed | $0.109 \pm 0.015$ | $0.082 \pm 0.013$ | – | – |
| SALT2 fit cuts | 152 | NMAD | – | – | $0.116 \pm 0.013$ | – |
| | | Standard deviation | – | – | $0.149 \pm 0.011$ | – |
| | | Pec. vel. removed | – | – | $0.140 \pm 0.012$ | – |
| Maximum spectrum + SALT2 fit cuts | 128 | NMAD | – | – | $0.103 \pm 0.014$ | $0.119 \pm 0.013$ |
| | | Standard deviation | – | – | $0.145 \pm 0.012$ | $0.116 \pm 0.007$ |
| | | Pec. vel. removed | – | – | $0.135 \pm 0.013$ | $0.103 \pm 0.008$ |
| All cuts | 66 | NMAD | $0.106 \pm 0.016$ | $0.074 \pm 0.013$ | $0.101 \pm 0.021$ | $0.103 \pm 0.015$ |
| | | Standard deviation | $0.114 \pm 0.010$ | $0.095 \pm 0.010$ | $0.134 \pm 0.014$ | $0.112 \pm 0.011$ |
| | | Pec. vel. removed | $0.100 \pm 0.012$ | $0.078 \pm 0.013$ | $0.123 \pm 0.015$ | $0.098 \pm 0.013$ |

For the sample of SNe Ia in common to these three standardization methods, we find that the RBTL + Isomap standardization significantly outperforms both of the SALT2 standardization methods. The variance of the residual magnitudes from the SALT2 + $x_1$ standardization are over twice those from the RBTL + Isomap standardization, meaning that a SN Ia corrected with RBTL + Isomap standardization will have over twice the weight in a cosmological analysis compared to a supernova corrected with SALT2 + $x_1$ standardization. Note that the quoted uncertainties on the measured standard deviations do not apply for comparisons between different standardization methods because we are examining the same set of supernovae and most of the residual variation is highly correlated.

One somewhat unexpected result is that we find better dispersions for the Raw RBTL residual magnitudes compared to SALT2 + $x_1$ corrected magnitudes (standard deviations of 0.115 ± 0.010 mag compared to 0.135 ± 0.013 mag for the set of SNe Ia with all cuts applied). This suggests that even without applying any additional corrections using the intrinsic variation of the SNe Ia, the base RBTL method standardizes SNe Ia better than SALT2 does with corrections for intrinsic variation. This was also seen in F15, where even the worst twins had a smaller dispersion in brightness than the SALT2 fits to that same dataset.

Finally, the standard deviations calculated in Table 3.7 have a contribution in their residual magnitudes from the peculiar velocities of the host galaxies of the SNe Ia, as described in Section 3.4.1. Assuming a 300 km/s dispersion in velocity, this contributes an added dispersion of 0.056 mag to the quoted standard deviation for the combination of all cuts and very similar amounts for the other sets of cuts. This dispersion can be removed in quadrature from the quoted values to obtain an estimate of the standard deviations of residual magnitudes that would be obtained for samples of SNe Ia at higher redshifts where peculiar velocity uncertainties have less of an impact on the residual magnitudes. We also show these "peculiar velocity removed" standard deviations in Table 3.7.

### 3.7.2 Correlations with Host Galaxy Properties

As described in Section 1.5, future cosmological analyses of SNe Ia will be limited by their ability to understand the evolution of the peak brightnesses of SNe Ia with redshift. SALT2 standardized residual magnitudes have been shown to have differences of ∼0.1 mag when comparing SNe Ia from host galaxies with different masses, metallicities, local colors, local star-formation rates or global star-formation rates (Kelly et al., 2010; Sullivan et al., 2010; Gupta et al., 2011; Hayden et al., 2013; Rigault et al., 2013, 2015, 2018; Roman et al., 2018). As galaxy properties evolve with redshift, such correlations of the peak brightness of SNe Ia with their host galaxy properties must be well-understood to produce robust cosmological measurements. While it is possible to correct the peak brightnesses of SNe Ia using host galaxy properties, the host galaxy measurements are really indirect measurements of some underlying properties of the SNe Ia. Ideally, it would be better to identify these differences directly from properties of the SNe Ia.

We examined how standardizing using the Isomap embedding affects these correlations with host galaxy properties. We use the measurements of host galaxy properties from Rigault et al. (2018) (hereafter R18) which includes much of the same data used in our analysis. The authors of this analysis found that there is a step of $\Delta_Y = 0.163 \pm 0.029$ mag when comparing SALT2 residuals from younger environments to older ones, and a step of $\Delta_M = 0.119 \pm 0.032$ mag for high-mass hosts compared to low-mass hosts. We reproduced these results for the subset of SNe Ia in our analysis using the published host masses and local specific star formation rates (lsSFR) from R18. The results of this procedure can be found in Figure 3.40.

We estimate the residual steps by estimating the mean of the residual magnitudes on either side of a threshold, using the thresholds defined in R18. For SNe Ia near the threshold we assign a probability of that SN Ia being on either side of the threshold, following the procedure in R18. We then fit a Gaussian mixture model to the data with different means and standard deviations for the residual magnitudes of SNe Ia on either side of the step. In R18, the authors simultaneously fit for the step parameters along with the SALT2 standardization parameters $\alpha$ and $\beta$. Here, we are attempting to determine how large of a step is left over after standardization, so we do not incorporate the step into our standardization model, and instead calculate the size of the step on standardized residual magnitudes. This results in a fundamentally different measurement of the steps than what was done in R18: we are effectively measuring how large the host correlations are in the residual magnitudes of SNe Ia if host property measurements are not available, while R18 instead examined how host property measurements can be incorporated into a standardization procedure.

Using the host data from R18 with the subset of SNe Ia in our dataset and our method of measuring the mass step, we find a step in the residual SALT2 magnitudes in lsSFR of $0.112 \pm 0.024$ mag and a host mass step of $0.093 \pm 0.024$ mag. These numbers are somewhat smaller than what is found in R18 ($0.163 \pm 0.029$ mag for lsSFR and $0.119 \pm 0.032$ mag for host mass) because we are calculating the step after the standardization procedure rather than including it in the standardization procedure. For the full dataset in R18, when the host parameters are not included in the standardization procedure, we find steps of $0.125 \pm 0.023$ mag for lsSFR and $0.102 \pm 0.024$ mag for host mass. These step sizes are consistent with the results seen for our analysis on a subset of the SNe Ia.

We calculate the host property step sizes for our other standardization methods. To compare all of these techniques on the same sample, we unfortunately must apply a large number of different cuts: we require that a supernova have a good interpolation to maximum light for the Isomap analysis, a good SALT2 fit, measured host properties in R18 and because the RBTL + Isomap magnitudes are still blinded, we are limited to the training set. This leaves us with a set of 46 SNe Ia in the blinded sample and 86 SNe Ia in the full sample. We report a preliminary analysis for the host properties for the blinded sample in this Section, although the conclusions will be much stronger with the full sample.

For each of our standardization methods, we plot the residual magnitudes passing these selection cuts as a function of the host properties. The results of this procedure can be seen in Figure 3.41 for the SALT2 + $x_1$ standardization, Figure 3.42 for the SALT2 + Isomap

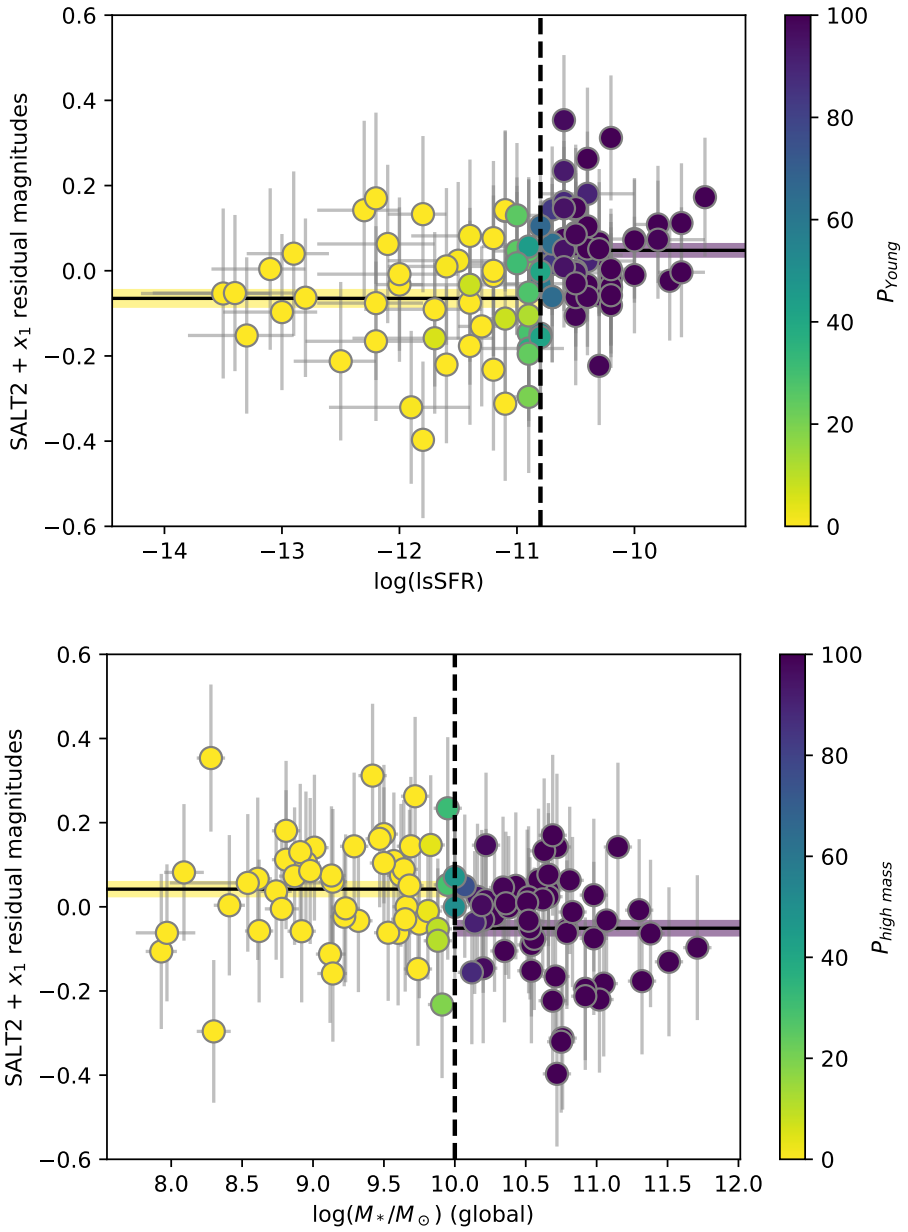



Figure 3.40: Reproduction of the host correlations seen in R18 for the SNe Ia in our analysis. Top panel: Step in SALT2-corrected residual magnitudes as a function of the local specific star formation rate (lsSFR). Bottom panel: Step in SALT2-corrected residual magnitudes as a function of the mass of the host galaxy. We use the thresholds from R18 to define the steps. Each supernova is assigned a probability to be on each side of the threshold according to the measurement uncertainties on its host properties. These probabilities are shown as the color of the data points.

Table 3.8: Measured host-galaxy property step sizes for different standardization methods. For each standardization method and set of cuts, we calculate the step size for various host galaxy properties following the procedure from R18. For all samples, we require that the low-dispersion-magnitude selection requirements described in Section 3.6.1 are satisfied and that the supernova has valid host properties in R18. We apply additional cuts to different samples as listed in the second column, including requiring a valid SALT2 fit (SALT2), requiring a valid spectrum at maximum light (max), and the training sample cut (train). We choose to set the sign of the step to be positive for the SALT2 + $x_1$ analysis: a negative step means that we recover a step in the opposite direction from the one for the SALT2 + $x_1$ analysis.

| Standardization method | Cuts applied | ΔlsSFR (mag) | ΔHost mass (mag) |
|---|---|---|---|
| SALT2 + $x_1$ | SALT2 | 0.112 ± 0.024 | 0.093 ± 0.024 |
| SALT2 + $x_1$ | SALT2 + max + train | 0.111 ± 0.034 | 0.078 ± 0.035 |
| RBTL + Isomap | SALT2 + max + train | 0.060 ± 0.026 | 0.045 ± 0.025 |
| SALT2 + Isomap | SALT2 + max + train | 0.043 ± 0.032 | 0.012 ± 0.034 |

standardization, and Figure 3.43 for the RBTL + Isomap standardization.

We measure the size of the residual steps for each of these standardization methods. The results of this procedure are shown in Table 3.8. We find that the size of the residual magnitude steps in both lsSFR and host mass are decreased for the two Isomap standardization techniques compared to SALT2 + $x_1$ standardization. In both cases, the resulting step sizes are consistent with there being no residual correlation with host properties, although we have a small sample size so it is difficult to make definite conclusions. With the full unblinded dataset, the sample size will roughly double leading to much better constraints on the host steps.

### 3.7.3 Future work

As discussed in Section 3.6.1, the analyses using the magnitudes of SNe Ia are still blinded. We are only using half of the sample to tune the analysis, and have not looked at the residual magnitudes for the other half of the sample. This blinded half of the sample will be used to evaluate the performance of our finalized model. At this stage, we are confident in our analysis techniques, although we identified some issues with the convergence of our host galaxy subtraction models that we are in the process of addressing and that we would like to solve before unblinding the analysis.

Now that we have identified a three-dimensional parameter space that more fully maps out the intrinsic diversity of SNe Ia at maximum light, the next step is to determine what data is necessary to locate a new supernova in this parameter space. As discussed in Section 3.6.7, we are able to reproduce our low dispersions in brightness and reduced correlations with host properties, using a peak brightness and color estimated from a fitter like SALT2. It seems likely that the Isomap coordinates can be identified from a non-spectrophotometric spectrum

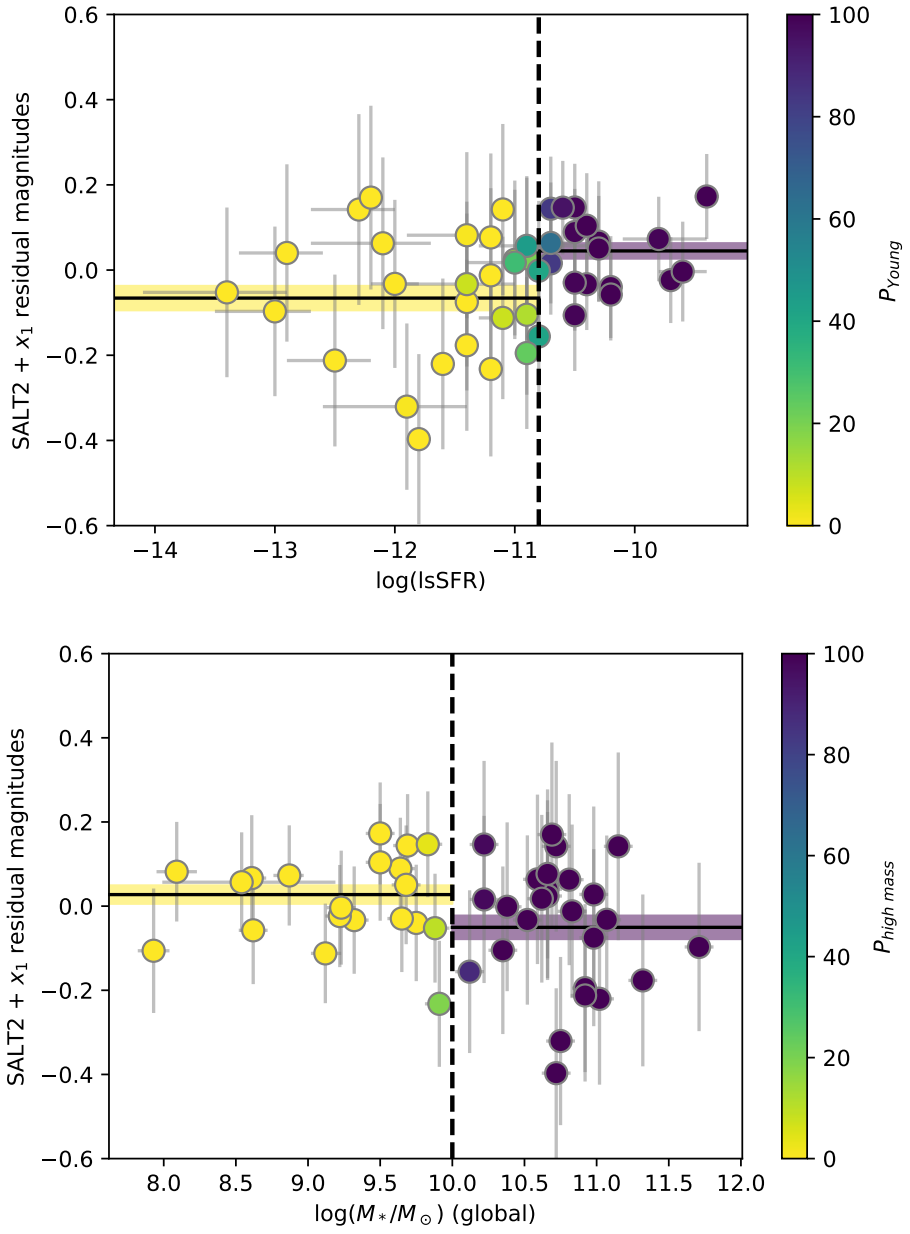



Figure 3.41: Host correlations for SALT2 + $x_1$ standardization for the subset of SNe Ia in common to all of our analyses. See Figure 3.40 for details of this plot.

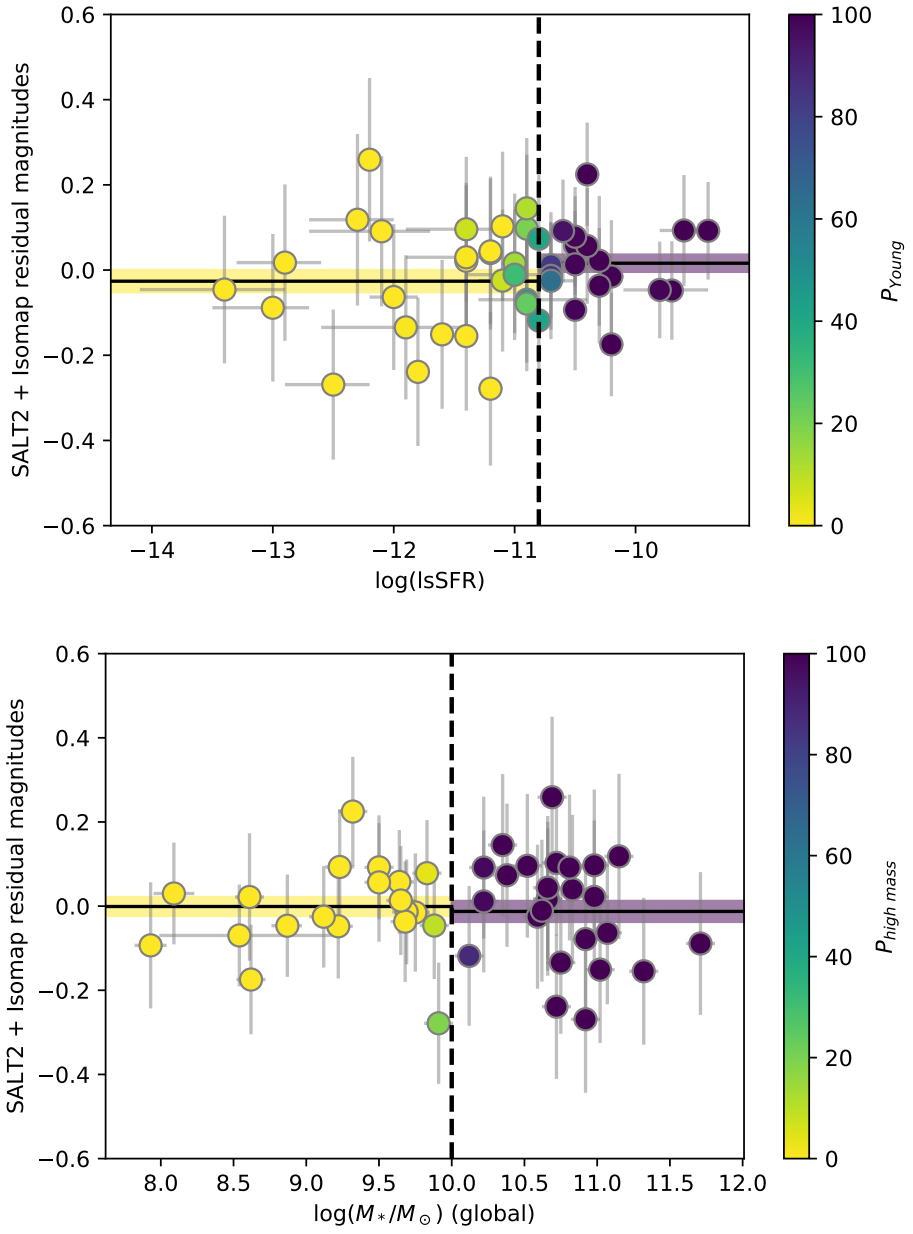



Figure 3.42: Host correlations for SALT2 + Isomap standardization for the subset of SNe Ia in common to all of our analyses. See Figure 3.40 for details of this plot.

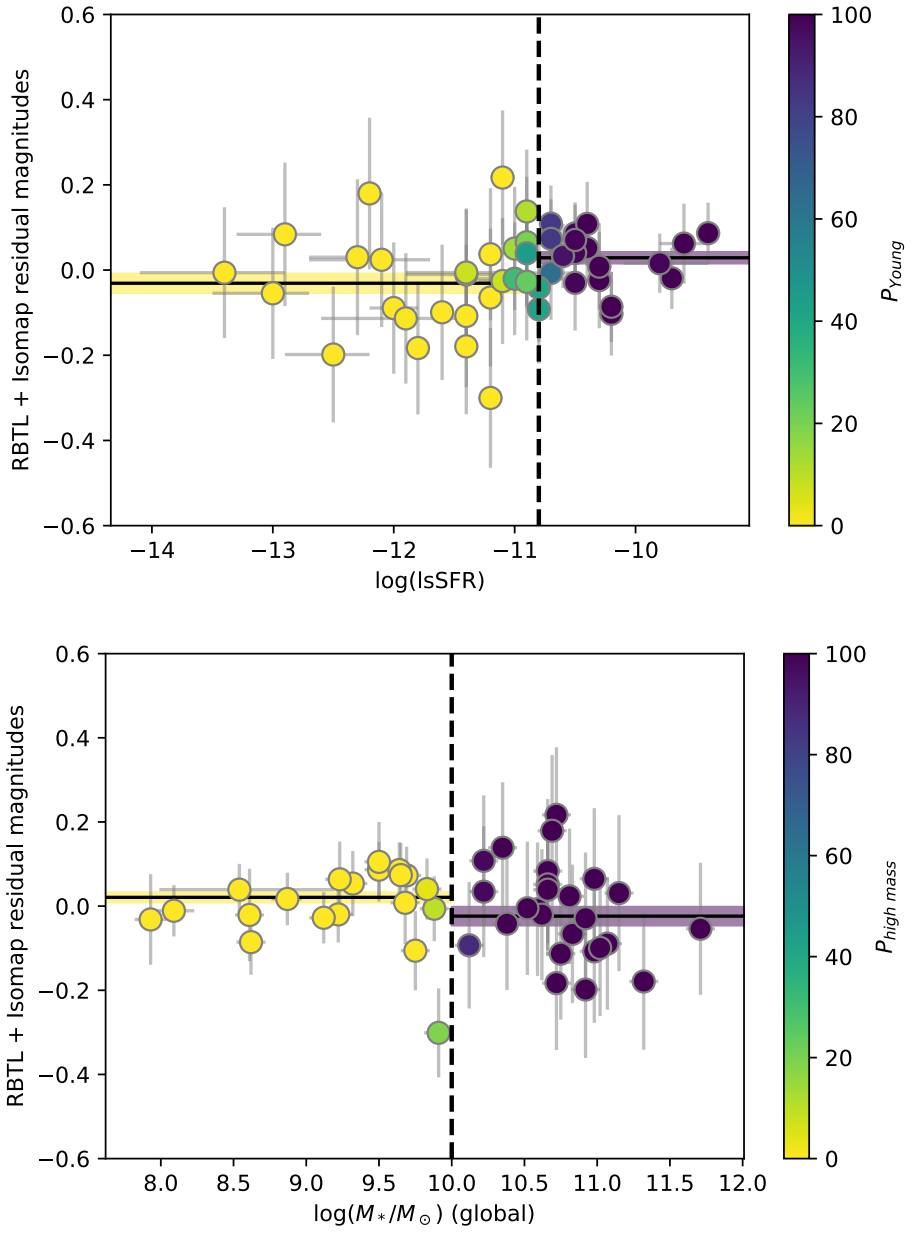



Figure 3.43: Host correlations for RBTL + Isomap standardization for the subset of SNe Ia in common to all of our analyses. See Figure 3.40 for details of this plot.

since most of the information is in variation of different localized features in the spectrum. Further analysis is needed to assess what data quality is necessary from a spectrograph to reproduce these results.

Another option is to build a new spectral timeseries model from the Isomap embedding. Once the Isomap parameter space has been established, a model can be built using the data available for all of our SNe Ia away from maximum light that predicts the spectrum at any phase for a given set of Isomap coordinates, similar to the procedure used for the SNooPy light curve fitter from Burns et al. (2011). This would also potentially allow for the Isomap coordinates for a new SN Ia to be recovered using only photometric observations, perhaps with strategic choices of the photometry bands used and the observation dates.

The Isomap parameter space effectively parametrizes the different subpopulations of SNe Ia. It is possible that some subpopulations of SNe Ia could be more standardizable than others. In Figure 3.30 we see some hints of this with what appears to be lower dispersions for small values of the second Isomap component, but we do not have a large enough sample of SNe Ia to perform a statistically sound study. With a larger sample of SNe Ia, it would be possible to examine the dispersion of the residual magnitudes in different regions of the Isomap parameter space and compare the recovered intrinsic dispersions. This could potentially be used to identify subsets of SNe Ia that are better distance indicators than the larger sample.

Our model was only trained on spectra at maximum light. There could be additional intrinsic diversity in the light curves of SNe Ia that is not captured in the spectra at maximum light, although F15 did show that their twinning procedure performed as well using only spectra near maximum light as it did with the full spectral timeseries. The range of additional intrinsic variability at other wavelengths could be probed by comparing the full timeseries of SNe Ia with similar Isomap coordinates. A similar concern is diversity in the extrinsic variation of SNe Ia. For example, as discussed in Section 3.6.1, there are some SNe Ia with extinction due to dust with very different properties from the color law assumed in our analysis. While we addressed this issue for this analysis by simply removing highly reddened SNe Ia from our sample, the Isomap embedding could actually be used to model the intrinsic diversity of SNe Ia and then probe the variations in extrinsic diversity for these extreme objects.

Finally, as the results of this analysis show strong biases for certain regions of the SN Ia parameter space, this could lead to biases in cosmological analyses if their selection of SNe Ia preferentially select SNe Ia in these regions of the parameter space or not. This is particularly of interest for the measurement of the Hubble constant with SNe Ia as the results rely on the fact that SNe Ia in hosts with Cepheids have similar distributions in brightness to the larger sample of SNe Ia (Riess et al., 2016, 2019; Rigault et al., 2015). By locating these SNe Ia in the Isomap parameter space, we will be able to determine whether or not this assumption affects the analysis, and apply a correction if necessary.

### 3.7.4 Conclusions

In this work, we developed an alternative method of standardizing SNe Ia using only a single spectrum near maximum light. In Section 3.3.1, we showed how we can estimate the spectrum at maximum light for a SN Ia using spectra within five days of maximum light. We use the SNe Ia with multiple spectra to build a model of the spectral evolution of SNe Ia near maximum light, and we can then apply this model to SNe Ia that have only a single spectrum in this phase range. Our model is completely agnostic to any extrinsic effects on the spectrum of an SN Ia like dust extinction or distance uncertainties, and these will be properly propagated into the estimated spectrum at maximum light.

In Section 3.4, we developed a new technique to estimate the brightness and color of SNe Ia at maximum light that we call "Reading Between the Lines" (RBTL). This technique effectively uses the regions of the spectra of SNe Ia where there is little intrinsic variation in between major spectral features to estimate the extrinsic effects on the spectrum such as the overall brightness and reddening due to dust. With this technique, we are able to produce a set of dereddened spectra for all of our SNe Ia at maximum light, along with an estimate of the brightness and color for those SNe Ia. We find that the RBTL method on its own is a very good estimator of the distances to SNe Ia, calibrating the brightness of arbitrary SNe Ia to $0.123 \pm 0.013$ mag. This sample includes many SNe Ia that would normally be labeled as "outliers" by a fitter like SALT2, and the RBTL distance estimate does not take into account any intrinsic variation of SNe Ia beyond that coded in the intrinsic color.

Using the dereddened spectra produced by the RBTL algorithm, in Section 3.5, we showed how manifold learning can be used to parametrize the intrinsic diversity of SNe Ia. Using the Isomap algorithm, we find that a three component non-linear model is able to capture the spectral diversity of SNe Ia. The three Isomap components primarily affect the Ca II features, the Si II features, and the line velocities respectively. We are able to reproduce previously identified subtypes of SNe Ia, such as the subtypes of Branch et al. (2006), directly from the Isomap embedding, and we find a smooth spectral sequence between these subtypes. We also are able to cleanly separate Type Iax supernovae from the rest of the sample, and our results suggest that these supernovae do not belong to the same spectral sequence as the larger sample of SNe Ia.

With the Isomap embedding, we showed that we can improve the standardization of SNe Ia in Section 3.6. We find that SALT2 residual magnitudes corrected for SALT2 $x_1$ and $c$ show a bias of up to 0.3 mag for large values of the first Isomap component. We can use the Isomap embedding to correct either the RBTL residual magnitudes or the SALT2 residual magnitudes directly. With this correction, we find that the dispersion in brightness of the residual magnitudes is greatly reduced compared to SALT2 + $x_1$ standardization. As opposed to other standardization methods such as SALT2, we are able to use the entire sample of SNe Ia, and we do not need to perform any cuts to remove "peculiar" SNe Ia. Standardizing with the Isomap components also significantly reduces the correlations between distance estimates and host galaxy properties seen for traditional SALT2 + $x_1$ standardization. This is important for standardization efforts because it implies that we are identifying

a direct signature of the remaining differences between SNe Ia rather than relying on indirect measurements of the properties of the host galaxies.

A final concern is that the rates of different subpopulations of SNe Ia could evolve with redshift, which could dramatically impact cosmological analyses. As previously described, we identified a region of parameter space where SALT2 residual magnitudes are biased by 0.3 mag. If SNe Ia in this region of parameter space have a different rate at high redshifts compared to low redshifts, then they could significantly bias cosmological measurements. These SNe Ia cannot be identified with a SALT2 light curve fit, but we showed in this analysis that we can distinguish them from the rest of the sample using spectrophotometry. Future work needs to be done to determine whether or not these biased subpopulations can be identified with other techniques such a slit spectroscopy or more advanced light curve fitters.

# Chapter 4

# Photometric Classification of Astronomical Transients with Gaussian Process Augmentation

## 4.1 Overview

In this chapter, we discuss several new techniques that we developed to improve the performance of photometric classifiers when trained on a spectroscopically-confirmed light curve sample that is not representative of the full light curve sample. This chapter was submitted for publication in the Astronomical Journal in July 2019, and was authored by K. Boone. All of the algorithms described in this chapter are implemented in the `avocado` software package[1].

Upcoming astronomical surveys such as the LSST will rely on photometric classification to identify the majority of the transients and variables that they discover. This introduces a potential source of systematic uncertainty: other kinds of supernovae can leak into the sample of SNe Ia and bias the cosmology analysis. In this chapter, we present a set of techniques for photometric classification that can be applied to experiments such as LSST where the training set of spectroscopically-confirmed objects will likely be heavily biased towards bright, low-redshift objects. Using Gaussian process regression to model arbitrary light curves in all bands simultaneously, we "augment" the training set by generating new versions of the original light curves covering a range of redshifts and observing conditions. We train a boosted decision tree classifier on features extracted from the augmented light curves, and we show how such a classifier can be designed to produce classifications that are independent of the redshift distributions of objects in the training sample. Our classification algorithm was the best-performing among the 1,094 models considered in the blinded phase of the Photometric LSST Astronomical Time-Series Classification Challenge (PLAsTiCC), scoring 0.468 on the organizers' logarithmic-loss metric with flat weights for all object classes in the

[1] `https://www.github.com/kboone/avocado`

training set, and achieving an AUC of 0.957 for classification of Type Ia supernovae. Our results suggest that spectroscopic campaigns used for training photometric classifiers should focus on typing large numbers of well-observed, intermediate redshift transients instead of attempting to type a sample of transients that is directly representative of the full dataset being classified.

In Section 4.2, we discuss the PLAsTiCC dataset along with several metrics that can be used to evaluate the performance of photometric classifiers on this dataset. We describe the new techniques that we developed for photometric classification in Section 4.3. In Section 4.4 we discuss the performance and limitations of our classification techniques, and compare them to other techniques. Finally, in Section 4.5, we discuss future work that could be done to improve classifier performance, and how the techniques described in this chapter could be applied to other classifiers.

## 4.2 Dataset

### 4.2.1 The PLAsTiCC dataset

The PLAsTiCC dataset (PLAsTiCC Team & PLAsTiCC Modelers, 2019) is a simulation of transients observed by LSST under realistic observing conditions. The full details of this simulation can be found in Kessler et al. (2019) (hereafter: K19). The PLAsTiCC dataset contains 3,500,734 light curves of 18 different kinds of transient and variable sources. In contrast to the SNPhotCC dataset, which only included different kinds of supernovae, the PLAsTiCC dataset also includes other object types such as variable stars, micro-lensing events and active galactic nuclei. This introduces several challenges, as classifiers must be able to handle more than just supernova-like objects. The details of all of the object types included in the simulations are shown in Table 4.1 along with their counts.

Realistic observing conditions were simulated using the LSST Operations Simulator (Delgado et al., 2014) for a three year period of LSST operations. The `SNANA` package (Kessler et al., 2009) was then used to simulate observations for each of the included models following the generated observing conditions. A simulated trigger model is applied to all of the generated transients following the DES supernova detection model (Kessler et al., 2015), and only objects passing this trigger are kept.

Two distinct LSST survey components were simulated for the PLAsTiCC. The Wide-Fast-Deep (WFD) component consists of observations covering almost half the sky. The Deep-Drilling-Fields (DDF) component consists of 5 different telescope pointings covering $\sim$50 deg$^2$. For the PLAsTiCC simulations, any observations on the same night are co-added, so the DDF observations are effectively $\sim$1.5 mag deeper and $\sim$2.5 times more frequent than the WFD observations. There are significantly more observations in the WFD component, and only 1% of the objects passing the detection trigger are in the DDF sample.

The PLAsTiCC simulations include a model of the photometric and spectroscopic redshifts that will be obtained for LSST. The simulations assume that Galactic objects can

Table 4.1: Summary of the object types included in the PLAsTiCC simulations (Kessler et al., 2019). For each object type, the number of objects both with and without spectroscopic confirmations is listed. The objects with spectroscopic confirmations are a small fraction of the full sample, and they are not representative of the distribution of the full sample.

| ID[a] | Object type | $N_{\text{confirmed}}$ | $N_{\text{unconfirmed}}$ | Galactic | Weight[b] |
|---|---|---|---|---|---|
| 90 | Type Ia SN | 2,313 | 1,659,831 | No | 1 |
| 67 | Peculiar Type Ia SN – 91bg-like | 208 | 40,193 | No | 1 |
| 52 | Peculiar Type Ia SN – SNIax | 183 | 63,664 | No | 1 |
| 42 | Type II SN | 1,193 | 1,000,150 | No | 1 |
| 62 | Type Ibc SN | 484 | 175,094 | No | 1 |
| 95 | Superluminous SN (Magnetar model) | 175 | 35,782 | No | 1 |
| 15 | Tidal disruption event | 495 | 13,555 | No | 2 |
| 64 | Kilonova | 100 | 131 | No | 2 |
| 88 | Active galactic nuclei | 370 | 101,424 | No | 1 |
| 92 | RR Lyrae | 239 | 197,155 | Yes | 1 |
| 65 | M-dwarf stellar flare | 981 | 93,494 | Yes | 1 |
| 16 | Eclipsing binary stars | 924 | 96,472 | Yes | 1 |
| 53 | Mira variables | 30 | 1,453 | Yes | 1 |
| 6 | Microlens from single lens | 151 | 1,303 | Yes | 1 |
| 991[c] | Microlens from binary lens | 0 | 533 | Yes | 2 |
| 992[c] | Intermediate luminous optical transient | 0 | 1,702 | No | 2 |
| 993[c] | Calcium rich transient | 0 | 9,680 | No | 2 |
| 994[c] | Pair instability SN | 0 | 1,172 | No | 2 |
| | Total | 7,846 | 3,492,888 | | |

[a] Each object type was assigned a random ID number to identify it during the blinded phase of the PLAsTiCC.

[b] During the blinded phase of the PLAsTiCC, classifier performance was evaluated using the metric defined Equation 4.1 with the class weights shown in this column.

[c] These object types had no spectroscopically confirmed examples, and were included in the PLAsTiCC to test anomaly detection algorithms. During the blinded phase of this challenge, they were all assigned the same ID of 99 and treated as a single class.

be cleanly separated from extragalactic objects, and the measured photometric and spectroscopic redshifts of the Galactic objects are set to zero. The simulation includes a model of a follow-up survey for extragalactic objects as described in K19. With this follow-up survey, 3.6% of the extragalactic objects have spectroscopic redshifts for their hosts. Extragalactic objects without spectroscopic redshifts are assigned photometric redshifts and uncertainties on those photometric redshifts following a model described in K19. In addition to spectroscopic redshifts, a total of 7,846 of the objects have spectroscopic confirmation of their types, representing only 0.2% of the total dataset. These spectroscopically-classified objects are referred to as a "training" set for the rest of this article as they are used to train the classifiers that will be applied to the "test set" of objects that do not have spectroscopic classifications.

The training set for surveys such as LSST will likely be highly biased since spectra are required to categorize each object in the training set. Typical choices of followup strategies will preferentially select brighter, closer objects. For the PLAsTiCC simulations, this bias can be seen in Figure 4.1. The median redshift of the training set is 0.18, compared to 0.43 for the full dataset. An additional challenge is that the different transients and variables have very different redshift distributions, as illustrated in Figure 4.2, so the biases in the training set will not be the same across different object types.

### 4.2.2 Metrics

#### Flat-weighted metric

As discussed in Malz et al. (2018), the PLAsTiCC team proposed to evaluate the performance of different classifiers on the PLAsTiCC dataset using a weighted multi-class logarithmic loss metric:

$$\text{logarithmic loss} = -\left( \frac{\sum_{i=1}^{M} w_i \sum_{j=1}^{N_i} \frac{y_{ij}}{N_i} \ln p_{ij}}{\sum_{i=1}^{M} w_i} \right) \tag{4.1}$$

where $M$ is the total number of classes and $N_i$ is the number of objects of each class. $y_{ij}$ is 1 if the object $j$ belongs to class $i$ and 0 otherwise. $p_{ij}$ are the predictions of a classifier, and for each object $j$ we should have $\sum_j p_{ij} = 1$. The class weights $w_i$ can be chosen to emphasize the performance of the classifier on specific classes.

An unweighted logarithmic loss is minimized when the classifier outputs predictions for each class that match the conditional probabilities of each class given the observations. By dividing by the total counts for each class, the logarithmic loss is normalized so that each class effectively has the same weight. This is important for the PLAsTiCC simulations because some classes have many more observations than others (e.g., over 10,000 SNe Ia for each kilonova), and an unweighted logarithmic loss would favor a classifier that does not attempt to classify the poorly-represented classes. For an optimal classifier trained on this metric, $p_{ij}$ can be interpreted as a probability. By this, we mean that given a sample with

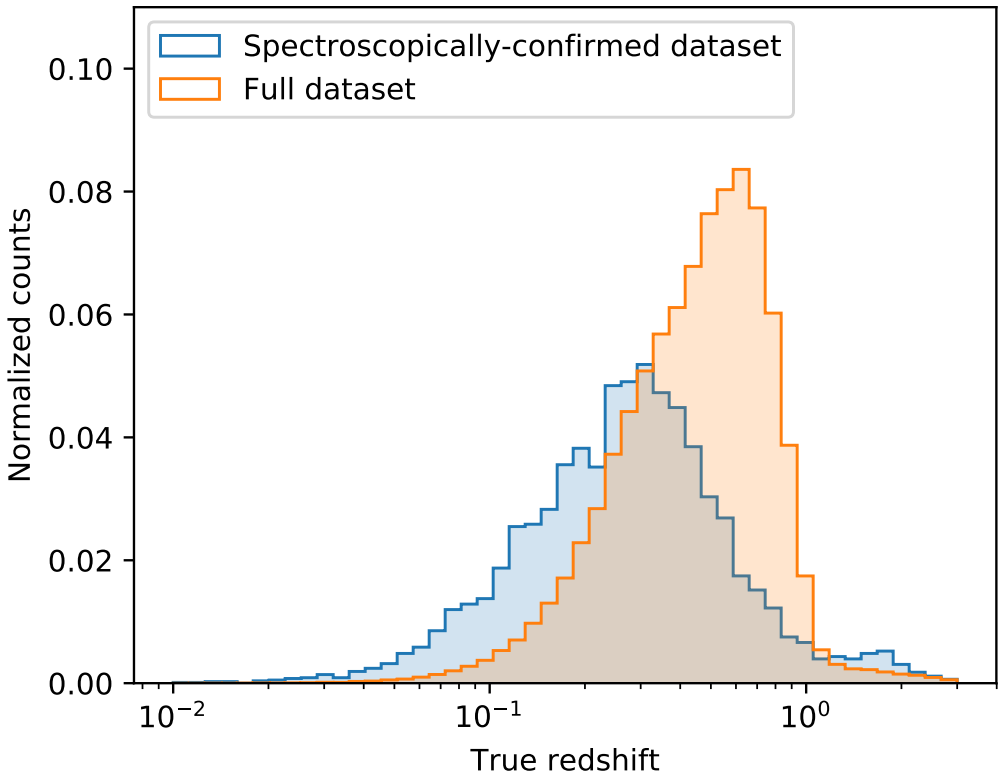



Figure 4.1: True redshift distributions for both the training and test sets in the PLAsTiCC dataset. The simulated followup strategy for the training dataset is strongly biased towards bright low-redshift objects.

equal number of objects from each class, 40% of the objects that are assigned $p_{ij} = 0.4$ will belong to class $j$.

In our main analysis, we choose to use the metric of Equation 4.1 with class weights of $w_i = 1$ for all of the classes present in the training set. We did not attempt to produce a classifier that can identify objects that do not have examples in the training set, so we set $w_i = 0$ for all such classes (the classes with IDs starting with 99 in Table 4.1). We call this metric the "flat-weighted metric" because it gives all of the classes the same weights. The flat-weighted metric can be written as:

$$\text{Flat-weighted metric} = -\left( \frac{\sum_{i=1}^{T} \sum_{j=1}^{N_i} \frac{y_{ij}}{N_i} \ln p_{ij}}{T} \right) \tag{4.2}$$

where the iteration $i$ over classes only considers classes that have examples in the training set, and $T$ is the number of such classes.

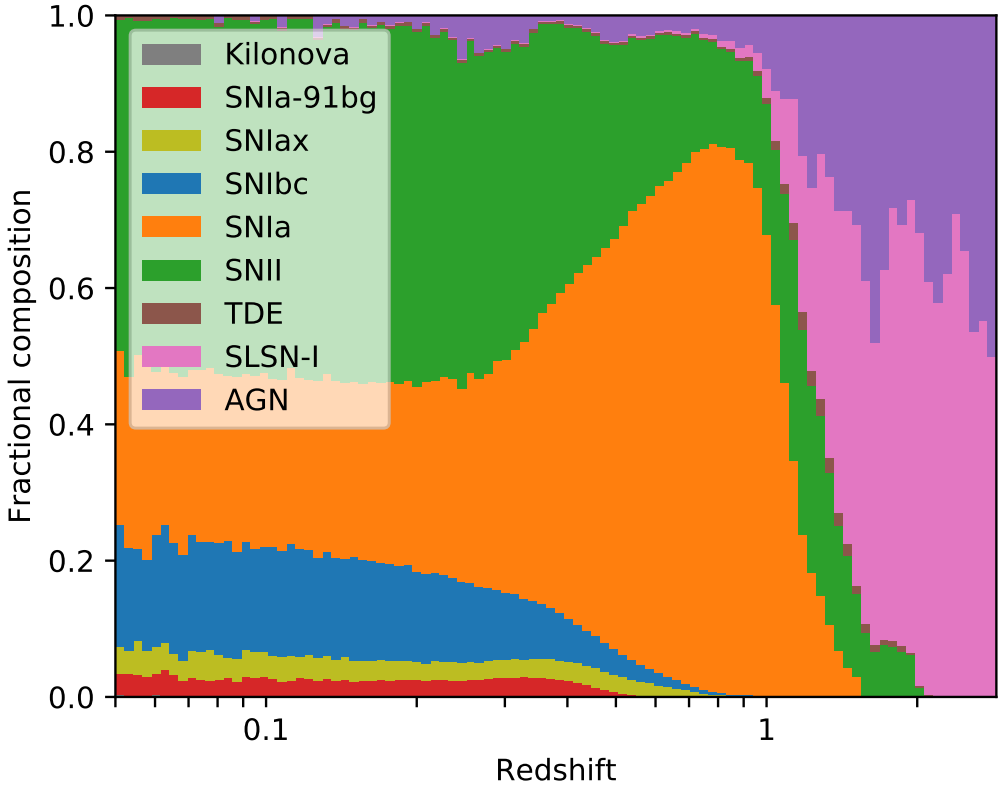



Figure 4.2: Stacked histogram of the fraction of objects belonging to each extragalactic object type as a function of redshift for the PLAsTiCC test dataset. The different object types have very different redshift distributions.

### Redshift-weighted metric

For some science cases, a subtle issue with the flat-weighted metric in Equation 4.2 is that it does not take the redshift of extragalactic transients into account. If a classifier trained on this metric is given information about the redshifts of the different objects, the classifier will learn to use the redshift distributions of different transients in the training set to perform its classification. For example, SNe Ia tend to be discovered at higher redshifts than most other kinds of supernovae, so a classifier trained on the flat-weighted metric will tend to classify ambiguous supernovae at high-redshifts as SNe Ia, and ambiguous supernovae at lower redshifts as other kinds of supernovae. Examples of this effect for the classifiers trained in this analysis will be shown in Section 4.4.2.

For photometric classification, classifiers are typically trained on datasets of spectroscopically-confirmed objects that have biased redshift distributions. It is the redshift distributions of these biased training sets that will be encoded into the predictions, not the redshift dis-

tributions of the test set. For a classifier trained on a metric similar to Equation 4.1, any analysis that depends on understanding the performance of the classifier as a function of redshift (such as cosmology with SNe Ia) requires accurate estimates of the differences between the redshift distributions in the training and full datasets. This is a difficult task if the spectroscopic followup is distributed across many different telescopes with varying observing strategies and objectives. One naive approach to dealing with this issue is to not input measurements of the redshifts of objects into the classifier, in an attempt to prevent it from using the redshifts to make its decisions. However, the redshift affects almost all features of a light curve, so a classifier can obtain a fairly accurate estimate of the redshift of an object through features such as the relative flux levels in different bands or the time dilation of the light curve, and it can use these estimated redshifts to make its predictions.

To mitigate this issue, we introduce the concept of a redshift-weighted metric. We reweight the objects in the training set so that every object class effectively has the same chosen redshift distribution. When a classifier is trained to optimize such a metric, the classifier cannot use the redshift distributions of the objects in the training sample for classification because they are all identical. We implement such a metric by splitting the redshift range into 10 logarithmically-spaced redshift bins between redshifts 0.1 and 3, along with an additional bin for Galactic objects for a total of 11 redshift bins. We then assign weights to each object to normalize the number of objects of each class in each redshift bin. This results in the following redshift-weighted metric:

$$\text{Redshift-weighted metric} = -\left(\frac{\sum_{i=1}^{M} w_i \sum_{k=1}^{K} \sum_{j=1}^{N_i} \frac{y_{ijk}}{C_{ik}} \ln p_{ij}}{\sum_{i=1}^{M} w_i}\right) \tag{4.3}$$

Here, $y_{ijk}$ is 1 if object $i$ belongs to class $j$ and is in redshift bin $k$. $K$ is the total number of redshift bins. $C_{ik}$ is the total number of objects in class $i$ and redshift bin $k$. To avoid extremely large weights for objects in bins that have very few counts, we impose a floor on $C_{ik}$ of 100 objects. For extragalactic objects, we choose to set the weight to $w_i = 1$. A typical extragalactic class is well-represented in roughly half of the different redshift bins while Galactic objects only have a single bin. To roughly maintain a flat class weighting, we set the weight for Galactic objects to the average number of bins that are well-populated for extragalactic classes ($\sim$5). As for the flat-weighted metric, we set $w_i = 0$ for the objects that were not present in the training set since we did not attempt to classify them.

### Kaggle metric

For the blinded phase of the PLAsTiCC hosted on Kaggle, the performance of classifiers was evaluated using the "Kaggle metric" which is of the weighted multiclass logarithmic-loss metric in Equation 4.1 with the class weights shown in Table 4.1. One part of the blinded phase of the PLAsTiCC was identifying new kinds of objects that had no examples in the training set (the "class 99" objects). In this work, we did not attempt to address this part of the challenge. Use of the Kaggle metric requires values for the class 99 objects. Hence,

to evaluate the performance of our classifiers on the Kaggle metric, we generate artificial predictions for the class 99 objects using formulae that were tuned by probing the metric during the blinded phase of the PLAsTiCC. For Galactic objects, we assign a flat predicted probability of 4% to the class 99 objects. For each extragalactic object, we assign a predicted probability according to the following formula:

$$P_{99} = P_{42} + 0.6 \cdot P_{62} + 0.2 \cdot P_{52} + 0.2 \cdot P_{95} \tag{4.4}$$

where $P_X$ is the predicted probability assigned to the class with ID $X$ (see Table 4.1 for the list of IDs). We then rescale all of the predicted probabilities so that they sum to 1 for each object. Note that these formulae are not a proper way of identifying new objects in the data. All of the top 5 performing teams in the blinded phase of the PLAsTiCC used similar formulae, and we are not aware of any successful attempts to identify new objects. For our analyses, we primarily use the flat-weighted metric and redshift-weighted metric, which both ignore the class 99 objects. We do however evaluate our performance on the Kaggle metric using these formulae for the class 99 predictions for comparison purposes.

**Single class metrics**

Finally, we evaluate several standard metrics for the performance of a deterministic classifier when identifying objects of a single specific type. The confusion matrix and corresponding labels for each of the outcomes of classification of one transient type out of a larger sample are shown in Table 4.2. Using the labels from this table, we define the following metrics that will be used in further analysis:

$$\text{true positive rate (TPR)} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{4.5}$$

$$\text{false positive rate (FPR)} = \frac{\text{FP}}{\text{FP} + \text{TN}} \tag{4.6}$$

$$\text{purity} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{4.7}$$

$$\text{completeness} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{4.8}$$

We also calculate the Area under the Receiver Operator Characteristic Curve (AUC) for each of our classes. This metric is defined as the area under the curve of the TPR plotted against the FPR, and ranges between 0.5 for a random classifier to 1 for a perfect classifier. See Lochner et al. (2016) for more complete definitions of all of these metrics.

Table 4.2: Confusion matrix for classification of a single object type (P) out of a larger sample of other object types (N).

| | | True Class | |
|---|---|---|---|
| | | P | N |
| Predicted | P | True positive (TP) | False positive (FP) |
| Class | N | False negative (FN) | True negative (TN) |

## 4.3 Methods

### 4.3.1 Overview

Our approach to photometric classification combines several techniques. We first preprocess the light curves as described in section 4.3.2. In Section 4.3.3, we describe how we use Gaussian process (GP) regression to predict smooth models for each of our sparsely sampled light curves. Using these GP models, as discussed in Section 4.3.4, we augment the spectroscopically-confirmed dataset, generating artificial light curves that are more representative of the full dataset. In Section 4.3.5, we describe how for each light curve in the augmented training set, we calculate a set of features from the GP models. We then train a tree-based classifier on the extracted features to perform the final classification predictions, the details of which can be found in Section 4.3.6.

### 4.3.2 Light curve preprocessing

The fluxes of transients are typically determined by subtracting newly measured fluxes from fluxes measured from a set of reference images. For long-lived transients, these reference images may contain light from the transients themselves. The blinded PLAsTiCC dataset did not provide the reference fluxes of the sources, so for objects such as variable stars, the "background" level of the light curve is simply the flux of the light curve at an arbitrary point in time. To address this issue, we estimate new "background" levels for each light curve using a biweight estimator (Beers et al., 1990). For short-lived transients, this background estimator will return the flux level at times when there is no light from the transient. For sources such as variable stars or active galactic nuclei, this robust estimator will effectively return the mean value of the light curve.

### 4.3.3 Modeling light curves with Gaussian process regression

Gaussian process (GP) regression has been shown to be effective for several applications of astronomical light curve modeling. Kim et al. (2013) used GPs to model the light curves of SNe Ia and predict their peak brightnesses. Fakhouri et al. (2015) and Saunders et al. (2018) modeled the full spectral time-series of SNe Ia with GPs, and used these models to

evaluate the spectra of these objects at arbitrary times. Lochner et al. (2016) (hereafter: L16) introduced GP modeling for astronomical transient classification. Revsbech et al. (2018) (hereafter: RTV18) showed that GP models can be used to augment a biased training set by generating additional training data from the GPs. These works all focused specifically on using GPs to model particular kinds of supernovae. We extend these techniques so that they can be applied to a wider range of transients and variables. An introduction to GP regression can be found in Section 1.8.

Previous works using GPs for photometric classification (e.g., L16 and RTV18) evaluated separate GPs for each band of the light curve, so the GP was not able to take cross-band information into account. However, there are strong correlations in the light curve behavior between different bands. Surveys such as LSST will not observe each band every night, so incorporating cross-band information into the model is essential. As shown in Fakhouri et al. (2015), a GP with a two-dimensional kernel can be used with separate length scales in both time and in wavelength to naturally incorporate this information. For this analysis, we use a GP kernel that is the product of Matérn kernels in both wavelength and time:

$$K_{2D}(t_1, t_2, \lambda_1, \lambda_2; \alpha, l_t, l_\lambda) = \alpha^2 K_{3/2}(t_1, t_2; 1, l_t) K_{3/2}(\lambda_1, \lambda_2; 1, l_\lambda) \tag{4.9}$$

This kernel has three hyperparameters: the amplitude ($\alpha$) and length scales in both time ($l_t$) and wavelength ($l_\lambda$).

We do not attempt to explicitly model the throughputs of the different filters. Instead, we calculate central wavelengths for each of the bands using the estimated LSST throughputs[2] assuming a source with a constant $F_\lambda$ spectrum. We use these central wavelengths as the coordinates for the wavelength dimension of the GP. This effectively means that the GP is producing a model of the spectrum convolved with a broad filter rather than modeling the spectrum directly.

We use the `George` package (Ambikasaran et al., 2015) to implement our GPs. Using maximum likelihood estimation, we fit for both the amplitude ($\alpha$) and time length scale ($l_t$) parameters on a per-object basis. It is difficult to reliably fit the length scale in wavelength on a per-object basis due to the fact that there are only 6 filters used for observations. We choose to fix the length scale in wavelength to 6000Å because we find that this value produces reasonable models for all of the transients and variables in the PLAsTiCC dataset. The results of this analysis are not highly sensitive to the choice of the length scale in wavelength.

Examples of the GP model for a well-sampled SN Ia light curve and a poorly sampled one are shown in Figure 4.3. The GP produces reasonable non-parametric models of the light curves that can be used for further analysis, along with estimates of the model uncertainties. Because we are using a kernel in both wavelength and time, the GP is able to use cross-band information to infer the supernova light curve even in bands where there are few observations. This can be seen in the right plot of Figure 4.3 where the GP produces a reasonable model

[2] `https://github.com/lsst/throughputs`

(with high uncertainty) of the light curve in the LSST $u$ band even though there are no observations in this band.

### 4.3.4 Augmenting the training dataset

The spectroscopically-classified objects that are used as a training set for photometric classifiers tend to be highly non-representative of the full dataset in terms of their redshift and signal-to-noise distributions. These training sets are typically strongly biased towards bright, low-redshift objects. Most previous attempts at producing photometric classifiers have seen strongly degraded performance when trained on non-representative datasets, and have concluded that obtaining representative training sets is essential for photometric classification (e.g., L16). More recently, RTV18 showed that it is possible to apply various transformations to the light curves in the original training set to generate a new training set that is more representative of the test set. We call this process "augmentation" of the training set. Using their `STACCATO` framework for augmentation, RTV18 train a classifier whose performance is significantly better (AUC of 0.96 on the SNPhotCC dataset) than one trained on the original non-representative training set (AUC of 0.93), and approaching the performance of a classifier trained on a representative training set (AUC of 0.977).

In `STACCATO`, GPs are fit to the observations of each object in the training set, with separate GPs for each band. "Synthetic light curves" are then produced for each object by sampling from the GPs. Each sample from a GP produces a different continuous function that can be interpreted as a synthetic light curve that is consistent with the observations of the original object. By repeatedly sampling from the GPs, many synthetic light curves can be produced for each object in the training set. In `STACCATO`, for every object, a "propensity score" is calculated, which is an estimate of how likely the object is to make it into the training set. The propensity score is then used to determine how many different synthetic light curves to make for each object in the training set. By generating different number of synthetic light curves for each object in the training set, an "augmented" training set of synthetic light curves is produced that is more representative of the test set. Finally, a classifier is trained on the set of synthetic light curves.

Our approach to augmentation differs in several ways from the approach of RTV18. In `STACCATO`, different synthetic light curves are generated for each object, but these synthetic light curves all use the same set of observations. Instead, our augmentation procedure involves simulating entirely new sets of observations for each object. When augmenting a light curve from the training set, we throw out large blocks of observations to simulate season boundaries, take originally well-sampled light curves and degrade their sampling, and add noise to the light curve in different bands. We measure the cadence and depth of observations in the test set, and generate augmented light curves that have similar cadences and depths of observations to the test set. This ensures that the light curves in the augmented training set have observations that are representative of the observations of light curves in the test set regardless of the light curve quality in the original training set. We also interpret the GP uncertainty as an uncertainty due to poor measurement rather than intrinsic variation

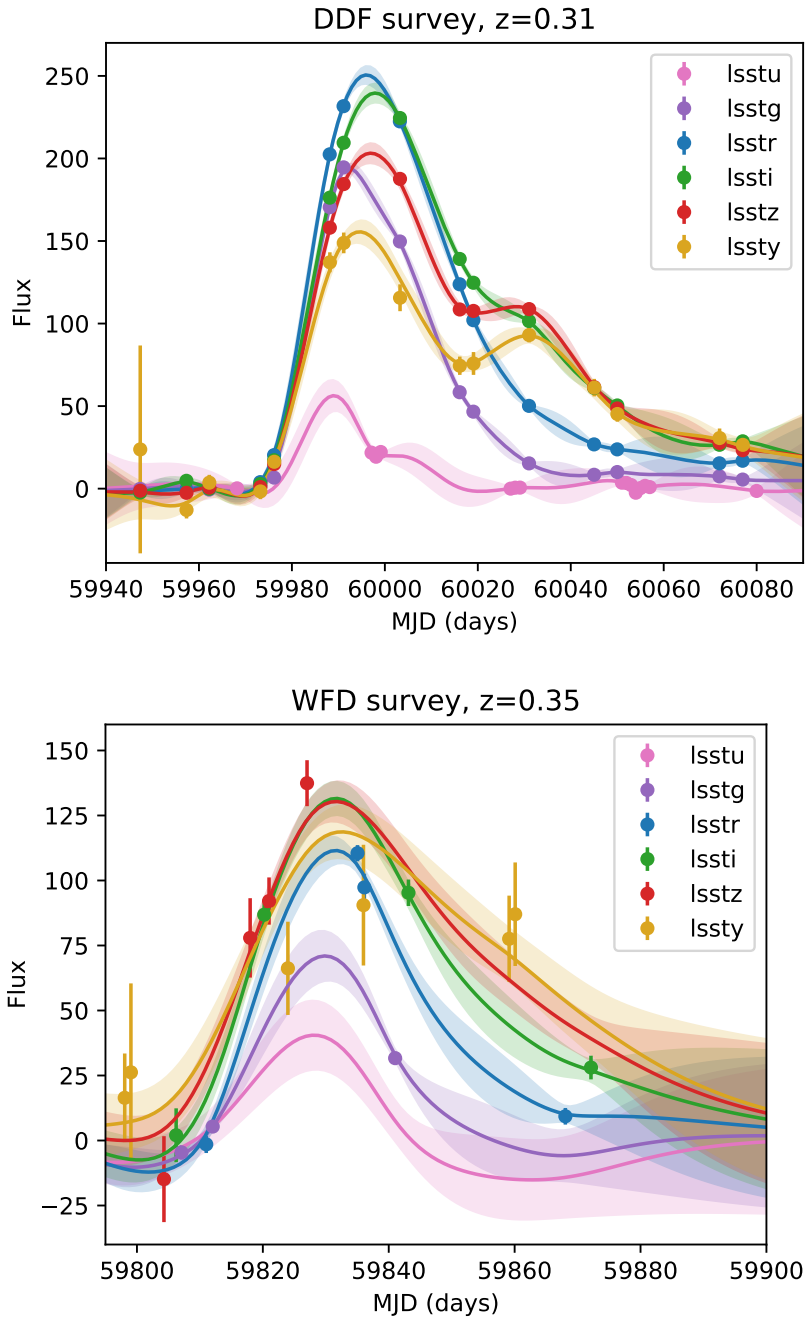



Figure 4.3: Examples of GP models for SN Ia light curves. Top panel: A well-sampled, high signal-to-noise light curve. Bottom panel: a poorly-sampled, lower signal-to-noise light curve. The mean GP flux prediction for each band is shown as a solid line surrounded by a shaded contour indicating the one-standard-deviation uncertainty on the flux prediction.

of the light curve. For this reason, we choose to use the mean prediction of the GP for our augmented light curves, and we propagate the GP prediction uncertainties into the uncertainties of the generated observations.

Additionally, we introduce the concept of "redshift augmentation" where we take an object in the training set and simulate observations of it at different redshifts. Because we are using a Gaussian process in both time and wavelength, we can shift the redshift of an object by evaluating the Gaussian process predictions for the light curve of that object at the redshifted wavelengths. Note that the GP is effectively modeling the spectrum of the object convolved with a filter. Assuming that this convolved spectrum is reasonably smooth and that the different filters have similar profiles, the two-dimensional GP predictions will automatically include k-corrections (Oke & Sandage, 1968) to the observed brightnesses in each filter. There will be higher-order corrections due to sharp structure in the spectrum (e.g., emission lines) and differences in the filter shape, but these are unlikely to significantly affect the classification in most cases. When redshifting the observations of an object, we then update the observed brightnesses by calculating the difference in distance modulus assuming a fiducial cosmology. After these procedures, we have effectively simulated the light curve for an object at a different redshift than it was originally observed at.

With redshift augmentation, if we observe an object at one redshift, then we can effectively use that object for training at all redshifts. Because training samples are typically biased towards bright, low-redshift objects, this means that we can use redshift augmentation to fill in the missing regions of parameter space in the training sample at high redshifts. This differs from the augmentation procedure in `STACCATO`. While `STACCATO` is making additional versions of light curves for objects that were already at high redshifts, we are instead taking low-redshift light curves and simulating what they would look like if they had been observed at high redshifts. A potential caveat with redshift augmentation is that the subpopulations of different object types could evolve with redshift. We discuss how this this can be addressed in Section 4.5.3.

One major challenge with augmentation is determining where in parameter space to generate new objects to match the training set to the test set. `STACCATO` uses a "propensity score" to decide how many new versions should be generated for each object in the training set. Unfortunately, for this technique to be effective, the rates and selection efficiencies in the test set must be known for each object type. The rates of different transients are not currently well known in many regions of parameter space. For SNe Ia and core-collapse supernovae, the current best measurements of the rates above redshift 1 have uncertainties of roughly 50% of the measured rates (e.g., Okumura et al. 2014; Rodney et al. 2014; Strolger et al. 2015). To address this issue, we instead choose to design a classifier whose performance is independent of the rates and selection efficiencies in the training set. This can be done by training a classifier to optimize the metric described in Section 4.2.2; a discussion of the effectiveness of this procedure will be shown in Section 4.4.2. When training a classifier with these properties, for the augmentation procedure, we simply need to ensure that we generate a set of light curves covering the full parameter space for any object type at a given redshift. We therefore simulate each object in our training set the same number of times at a range

of different redshifts.

The full details of our augmentation procedure can be found in Section 4.6. A summary of our approach to augmentation is as follows:

1. Fit a GP to a light curve in the original training sample to use as a template.
2. Choose a new redshift for extragalactic objects, or brightness for Galactic objects.
3. Evaluate the mean GP prediction and uncertainty to obtain a new light curve at the chosen brightness/redshift.
4. Drop observations to simulate poorly-sampled light curves from the well-sampled training light curves.
5. Add measurement uncertainties that are representative of the test dataset.
6. Ensure that the generated light curve would be detected.
7. Simulate a photometric redshift following the mapping between spectroscopic and photometric redshifts in the test dataset.
8. Repeat these steps until a large enough augmented sample has been obtained.

For this analysis, we use the same augmented training set to train all of the different classifiers that will be discussed. We generate up to 100 different versions of each light curve in the training set, which results in a total of 591,410 light curves in the augmented training set.

### 4.3.5 Extracting features from the light curves

To extract features from a light curve, we begin by performing a GP fit as described in Section 4.3.3 and computing the mean GP flux predictions in the LSST $g$, $i$, and $y$ bands. We choose to use the observer-frame LSST $i$ band as the reference for many of our features because this band typically has a reasonable flux level for both the low and high-redshift objects in our sample. We extract a variety of features from the GP flux predictions in each of these three bands, the details of which can be found in Table 4.3.

Table 4.3: Overview of features used for classification. Unless specified otherwise, all features are calculated using the mean flux predictions of a GP fit to the light curve.

| Feature name | Description |
|---|---|
| `host_photoz` | Host-galaxy photometric redshift, taken directly from the PLAsTiCC metadata. |

Table 4.3 continued

| Feature name | Description |
|---|---|
| `host_photoz_err` | Host-galaxy photometric redshift error, taken directly from the PLAsTiCC metadata. |
| `length_scale` | Fitted GP length scale hyperparameter, in days. |
| `max_mag` | Peak magnitude of the GP flux prediction in the LSST $i$ band. |
| `pos_flux_ratio` | Ratio of the maximum positive flux to the maximum-minus-minimum flux in the LSST $i$ band. |
| `[max,min]_flux_ratio _[blue,red]` | Normalized difference of the light curve colors at maximum/minimum light. The blue measurement is the difference between the LSST $i$ and $g$ bands, and the red measurement is the difference between the LSST $y$ and $i$ bands. The normalized difference is calculated by taking the difference of the fluxes in the two bands divided by their sum. |
| `max_dt` | Difference of the time of maximum in the LSST $y$ and $g$ bands in days. |
| `[positive,negative] _width` | An estimate of the light curve "width" that is applicable even for non-supernova-like transients and variables. This is implemented as the integral of the positive/negative parts of the GP flux predictions divided by the positive/negative maximum fluxes. |
| `time_[fwd,bwd]_max _[0.2,0.5]` | Measurements of the rise and decline times of a light curve. This measurement is defined as the time in days for the light curve to rise (bwd) or decline (fwd) to a given fraction (either 20% or 50%) of maximum light in the LSST $i$ band. |
| `time_[fwd,bwd]_max _[0.2,0.5]_ratio _[blue,red]` | Ratio of the rise/decline times calculated as described above in different bands. The blue measurement is the difference between the LSST $i$ and $g$ bands, and the red measurement is the difference between the LSST $y$ and $i$ bands. |
| `frac_s2n_[5,-5]` | Fraction of observations that have a signal greater than 5/less than -5 times the noise level. |
| `frac_background` | Fraction of observations that have an absolute signal-to-noise less than 3. |
| `time_width_s2n_5` | Time difference in days between the first observation with a signal-to-noise greater than 5 and the last such observation (in any band). |

Table 4.3 continued

| Feature name | Description |
|---|---|
| `count_max_center` | Number of observations in any band within 5 days of maximum light. |
| `count_max_rise _[20,50,100]` | Number of observations in any band between 20, 50, or 100 days before maximum light and 5 days after maximum light. |
| `count_max_fall _[20,50,100]` | Number of observations in any band between 5 days before maximum light and 20, 50, or 100 days after maximum light. |
| `peak_frac_2` | Ratio of the maximum flux in the second most prominent peak in the light curve to the maximum flux in the main peak, averaged over all LSST bands. This is intended to identify supernova-like objects that only have a single large peak in most bands. |
| `total_s2n` | Total signal-to-noise of all observations of the object. |
| `percentile_diff _[10,30,70,90]_50` | Measurements of the distributions of the observed fluxes. For each band, we calculate the flux level for a given percentile of observations, and normalize it by the maximum-minus-minimum of the GP flux predictions. We then take differences between this measurement for various percentiles and the $50^{th}$ percentile measurement. The final value is the median of the calculated differences in all bands. |

In general, we find that in the PLAsTiCC dataset, the non-supernova-like variables and transients end up being relatively easy to distinguish, so most our our effort went to identifying features that are effective for distinguishing the different kinds of supernovae. We initially generated hundreds of different features, and we used both the feature importance ranking of our classifier and cross-validation performance (discussed in Section 4.3.6) to select a subset of 41 features that give good performance. Most of the features that we include are standard features that are well-known to distinguish transients, such as the apparent peak brightness of the transient and the photometric redshift. We include measures of the colors of the objects by taking ratios of the peak brightnesses in different bands, and estimates of the rise time and fall times that help distinguish between the different supernova types.

One major difference between this analysis and most previous ones is that our light curve model includes cross-band information. The GP flux predictions in a given band incorporate information from nearby bands because of the GP kernel in the wavelength direction. We only calculate features from the GP predictions in the LSST-$g$, $i$, and $y$ bands, but these features capture the observations in all of the different bands. This can be

seen for the poorly-sampled light curve in Figure 4.3 where we have reasonable models of the light curve even in bands with no observations. We find that calculating features off of GP flux predictions in additional bands beyond the three previously listed does not improve the classification for the PLAsTiCC sample. This kind of analysis is only possible with a light curve model that fits both wavelength and time directions simultaneously as opposed to traditional models where each band is fit independently. Although we chose to perform the GP flux predictions at the wavelengths associated with several LSST bands for simplicity, the GP flux predictions can be performed at arbitrary wavelengths, and observations from different bands or other telescopes can easily be included in the GP model.

There are a handful of notable features that we introduced that have not been included in previous analyses. First, we add several features that are effective at classifying variable object types, most notably the `percentile_diff_X_Y` features that measure the distribution of the photometry values. These features are effective at distinguishing object types with large wings in their photometry distributions (e.g., eclipsing binaries) from object types that have more even distributions (e.g., Mira variables), even with relatively few observations.

Another novel technique in this work is the introduction of features that measure the fit quality. In many cases, the available photometry for an object does not cover its full light curve. An example of such a light curve for a Type Ia supernova can be seen in Figure 4.4. Without any photometry before maximum light, the GP produces a model with large uncertainties, and the measurement of the rise time will be both uncertain and biased. We experimented with several approaches to incorporating this information into the classifier, and found that measuring the number of observations in various bands around maximum light provides an excellent way to evaluate the accuracy of the rise time measurements. The right panel of Figure 4.4 illustrates this for SNe Ia: light curves that have an observation between 20 observer-frame days before maximum light and 5 days after maximum light have a tight distribution of measured rise times, while light curves without this information have a wide, biased distribution of rise times. By adding features measuring the number of observations in various time intervals, the classifier can effectively determine how reliable other features, such as the rise time information, are.

### 4.3.6 Training a LightGBM model

We train a gradient boosted decision tree classifier on features extracted from the augmented training set. Decision trees are a classification technique where objects are filtered through a variety of cuts to attempt to separate them into their different classes. Boosted decision trees combine a large number of these trees to produce a robust classifier, and have proven to be very effective at a variety of different classification tasks in astronomy (e.g: Bailey et al. (2007), L16). Using the LightGBM implementation of gradient boosted decision trees (Ke et al., 2017), we train separate classifiers to optimize objective functions that are direct implementations of the metrics in Equations 4.1 and 4.3.

To evaluate and optimize the performance of the classifier, we use five-fold cross-validation on the augmented training set. We partition the augmented training data into five separate

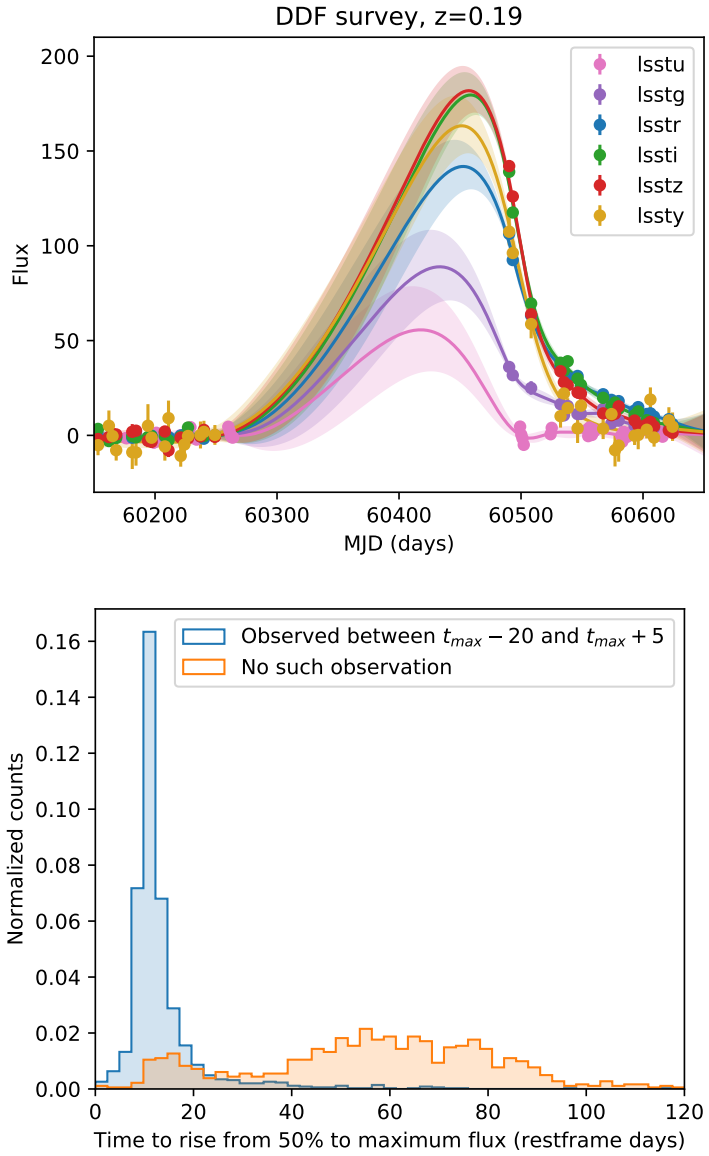



Figure 4.4: Example of a feature that probes the fit quality. Top panel: Example of the GP model for a SN Ia light curve with no data before maximum light. See Figure 4.3 for details of the plot. The GP is unable to constrain the rise time, and produces a model with large uncertainties. Bottom panel: Histogram of the measured rise times for SNe Ia. The rise time is well-constrained (blue histogram) for objects with an observation between 20 observer-frame days before maximum light and 5 days after maximum light. Objects without such an observation (orange histogram) have rise time measurements that are biased to higher values with large dispersions.

Table 4.4: Optimized hyperparameter values used for the LightGBM model.

| Hyperparameter name | Value |
|---|---|
| `boosting_type` | `gbdt` |
| `learning_rate` | `multi_logloss` |
| `colsample_bytree` | 0.05 |
| `reg_alpha` | 0 |
| `reg_lambda` | 0 |
| `min_split_gain` | 10 |
| `min_child_weight` | 2000 |
| `max_depth` | 7 |
| `num_leaves` | 50 |
| `early_stopping_rounds` | 50 |

subsets that each have equal ratios of the different targets. We then train five separate classifiers, each of which is trained on four of the five subsets, and we evaluate its performance on the remaining subset. By repeating this procedure for each of the subsets, we obtain out-of-sample predictions for every object in the augmented training set. We then evaluate the PLAsTiCC metric on these predictions, and use this cross-validation performance to tune our model. For the augmented dataset, we ensure that all light curves generated from the same original light curve are included in the same subset to avoid leaking information across the subset boundaries. When generating predictions for objects that are in the test set, we evaluate the average of the classification probabilities for the new objects from each of the five trained classifiers.

We optimize the hyperparameter values of the LightGBM model by scanning over each hyperparameter individually and then evaluating the cross-validation performance on the flat-weighted metric. The resulting hyperparameter values are shown in Table 4.4. The optimal hyperparameter values are relatively stable across different sets of features and target metrics, so for simplicity we use the same hyperparameter values for all of the analysis variants.

We train two separate versions of our classifier, one of which is optimized for performance on the flat-weighted metric defined in Equation 4.2, and one of which is optimized for performance on the redshift-weighted metric defined in Equation 4.3. Both of these classifiers are trained on the same augmented training set. For the training set, the Kaggle metric is nearly identical to the flat-weighted metric, so we do not train a separate classifier to optimize it.

LightGBM outputs a measure of how much each feature contributed to the classification. We call this measure the "importance" of that feature for classification. The importance of each feature for a classifier trained to optimize the flat-weighted metric are shown in Figure 4.5. The most important features for classification are the photometric redshift of the host galaxy, the peak brightness, and the colors of the light curves at maximum light. The feature importance plot for a classifier trained to optimize the redshift-weighted metric

looks nearly identical.

## 4.4 Results

### 4.4.1 Overall performance

The results of both of our classifiers for many of the metrics defined in 4.2.2 are shown in Table 4.5. In general, we find that both classifiers have similar performance across all of these global metrics. In the blinded phase of the PLAsTiCC, an earlier version of our algorithm won the challenge with a score of 0.680 on the Kaggle metric (lower is better). The updated algorithm presented in this chapter achieves a slightly better score of 0.649 on this metric. This improvement mainly came from restricting the allowable redshift range for the data augmentation procedure and propagating uncertainties. Our original augmentation algorithm was allowed to modify the redshifts of objects arbitrarily. At high redshifts, unreliable extrapolations of the GP models far into the restframe UV were being used to produce the light curves. Additionally, we were not propagating the GP modeling uncertainties into the generated fluxes. These issues were fixed for the version of the algorithm described in this chapter which dramatically improved performance at high redshifts.

Figure 4.6 shows a confusion matrix for the flat-weighted classifier which can be used to evaluate its performance on specific classes. For most classes, the top prediction of this classifier is correct with an accuracy over over 80%. The main challenge for the classifier is distinguishing between the different types of supernovae. For example, Type Iax supernovae are misclassified as Type Ia supernovae 27% of the time, and Type Ibc supernovae are misclassified as Type Ia-91bg supernovae 17% of the time. This misclassification is often highly asymmetrical: only 3% of Type Ia supernovae are misclassified as Type Iax supernovae. This is likely due to the fact that the training set of Type Iax supernovae (183 objects) is small compared to the sample of Type Ia supernovae (2,313 objects), and there is additional diversity in the test set that is not seen in this small training set of Type Iax supernovae. Methods to address these differences will be discussed in Section 4.5.2. The redshift-weighted classifier has a confusion matrix that is nearly identical to the one for the flat-weighted classifier. Note that the confusion matrix only considers the top prediction for each object while our classifier outputs a probability for each object type, meaning that there is additional information available for further analyses that is not captured by the confusion matrix.

### 4.4.2 Redshift-dependent performance

Despite the similarity in performance between the flat-weighted classifier and the redshift-weighted classifier on most global metrics, they exhibit very different performance as a function of redshift. For a fixed overall sample purity of 95%, we calculate the completeness of the SN Ia sample as a function of redshift. The results of this procedure can be seen in Figure 4.7.

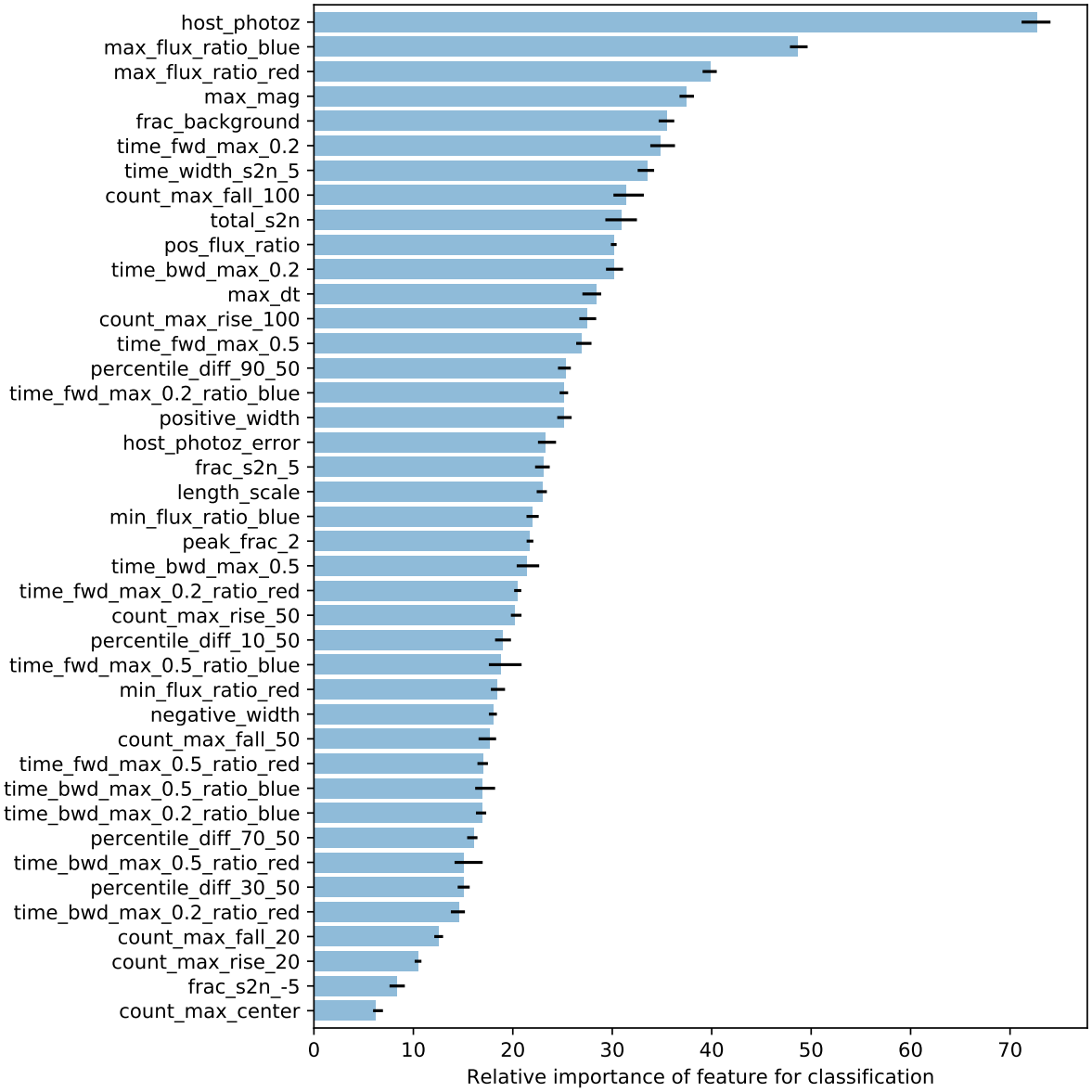



Figure 4.5: Relative importance of each feature for classification for a classifier trained to optimize the flat-weighted metric. Definitions of these features can be found in Table 4.3. Five separate classifiers were trained on different folds of the training set. The blue bars represent the mean importance of each feature across the five classifiers. The thin black bars indicate the range of importance across the five classifiers.

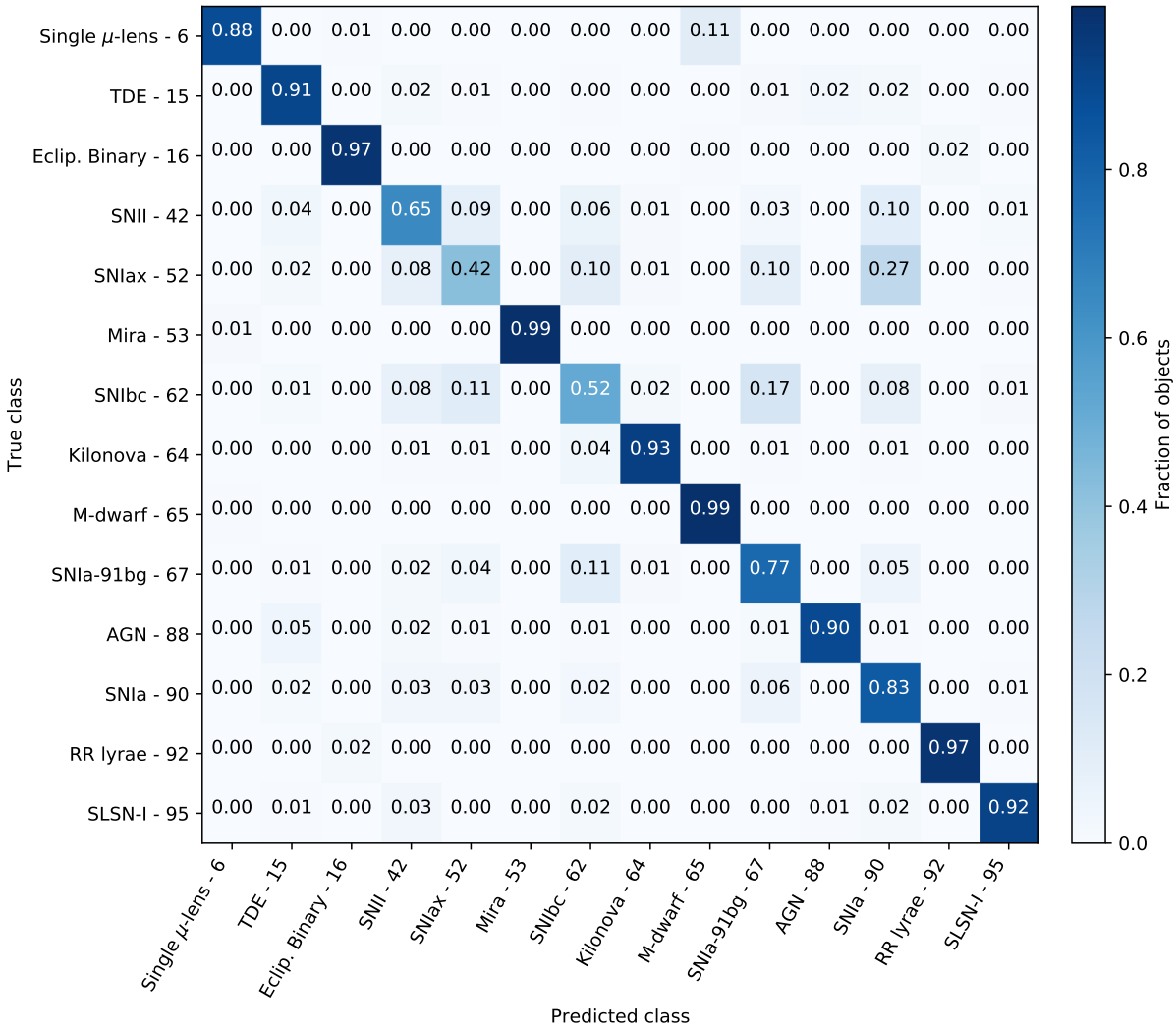



Figure 4.6: Confusion matrix for the flat-weighted classifier. We assign each object a "predicted class" based on whichever class has the highest prediction probability for that object. We then calculate the fraction of objects for each true class that end up being predicted to be a given class. These fractions are shown in the figure for each pairing of true class and predicted class. A perfect classifier would have all ones for the diagonal terms, where the predicted class is the true class, and all zeros for the off-diagonal terms.

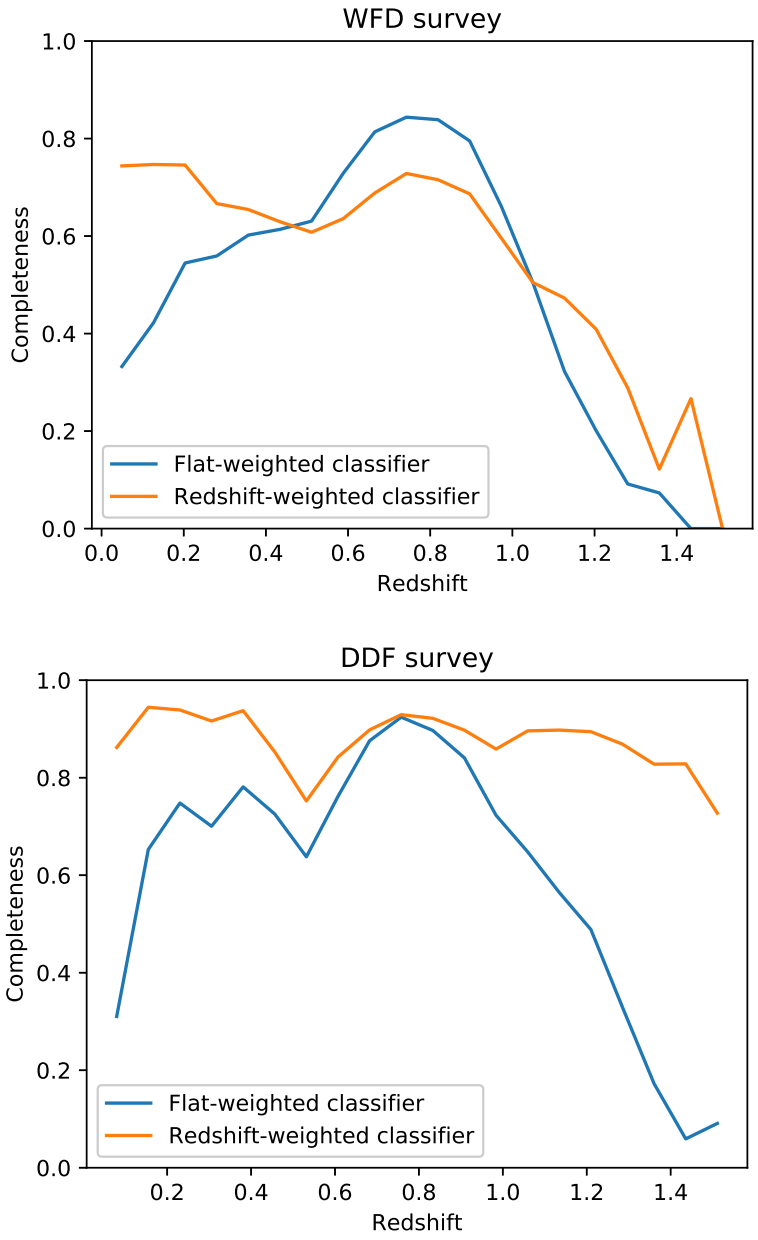



Figure 4.7: Completeness of the sample of SNe Ia in the test set as a function of redshift for a fixed overall sample purity of 95%. Top panel: Results for objects observed in the WFD survey. Bottom panel: Results for objects observed in the DDF survey. The Flat-weighted classifier produces classifications that encode the redshift distributions of different object types in the training set, and shows undesirable performance as a function of redshift. The redshift-weighted classifier produces classifications that are relatively stable with redshift.

Table 4.5: Classifier performance on various metrics. For the flat-weighted, redshift-weighted and Kaggle metrics, lower numbers are better. An optimal classifier will have an AUC of 1, and higher AUCs are better.

| Metric name | Flat-weighted classifier | Redshift-weighted classifier |
|---|---|---|
| Flat-weighted metric | 0.468 | 0.510 |
| Redshift-weighted metric | 0.523 | 0.500 |
| Kaggle metric | 0.649 | 0.709 |
| AUC – 90: Type Ia SN | 0.95721 | 0.95204 |
| AUC – 67: Peculiar Type Ia SN – 91bg-like | 0.96672 | 0.96015 |
| AUC – 52: Peculiar Type Ia SN – SNIax | 0.85988 | 0.84203 |
| AUC – 42: Type II SN | 0.93570 | 0.90826 |
| AUC – 62: Type Ibc SN | 0.92851 | 0.91558 |
| AUC – 95: Superluminous SN (Magnetar model) | 0.99442 | 0.99257 |
| AUC – 15: Tidal disruption event | 0.99254 | 0.99243 |
| AUC – 64: Kilonova | 0.99815 | 0.99579 |
| AUC – 88: Active galactic nuclei | 0.99772 | 0.99706 |
| AUC – 92: RR Lyrae | 0.99987 | 0.99986 |
| AUC – 65: M-dwarf stellar flare | 0.99999 | 0.99999 |
| AUC – 16: Eclipsing binary stars | 0.99983 | 0.99983 |
| AUC – 53: Mira variables | 0.99947 | 0.99937 |
| AUC – 6: Microlens from single lens | 0.99962 | 0.99966 |

We find that the performance of the flat-weighted classifier has undesirable behavior as a function of redshift. For objects observed in the WFD survey, we find that this classifier has its peak completeness at redshift 0.8 where it correctly classifies 80% of the SNe Ia. At nearby redshifts, its completeness drops to below 40%. The reason for this strange performance can be seen in Figure 4.2. Around redshift 0.8, nearly 70% of the observed objects are Type Ia supernovae, so the flat-weighted classifier is using information about the redshift distributions of the different transients at different redshifts as part of its classification. Only 0.6% of the SNe Ia in the sample are at a redshift less than 0.1, so the flat-weighted classifier infers that any object at those redshifts is likely not to be a SN Ia, and therefore has a very poor completeness for low-redshift SNe Ia. We see similar effects for the DDF survey.

Any classifier trained to optimize traditional metrics on a biased training sample will run into these issues if it is allowed to incorporate redshift information into its predictions. Furthermore, the classifier is learning to use the redshift distributions of objects in the training set, not the redshift distributions of ones in the test set. As described in Section 4.3.4, we expect there to be large discrepancies in the redshift distributions of these two sets, even for augmented training sets. A classifier trained on the redshift-weighted metric provides a solution to this problem by outputing a probability for each object type assuming that

each class has the same arbitrarily chosen redshift distribution in the training set. The performance of such a classifier will not depend on the redshift distributions of the different kinds of objects in the training set. This can be seen in Figure 4.7: for the WFD survey, the redshift-weighted classifier shows a stable completeness of ∼70% at low redshifts, and its completeness declines nearly monotonically at higher redshifts as the signal-to-noise decreases. The redshift-weighted classifier is able to classify objects in the DDF survey with a completeness of above 80% at most redshifts, maintaining this performance for even the highest-redshift SNe Ia in the test sample at $z = 1.55$.

To validate the claim that the classifications produced by a classifier trained on the redshift-weighted metric are independent of the redshift distributions of objects in the training set, we simulated modifying the redshift distribution of the SNe Ia in the training set. Starting with the same augmented training set used to train the original classifiers, we create a low-redshift-biased training set by randomly dropping light curves of SNe Ia from the training set with probability $p(z) = \exp(-z)$. Similarly, we create a high-redshift-biased training set by randomly dropping light curves of SNe Ia from the training set with probability $p(z) = \min(\exp(z) - 1, 1)$. We keep all light curves from the transients that are not SNe Ia so that only the redshift distribution of SNe Ia is changing between the original augmented training set and the biased training sets. We then retrain our flat-weighted and redshift-weighted classifiers on these new training sets. The results of this procedure can be seen in Figure 4.8. For the flat-weighted classifier, the classifier performance varies dramatically for these different training sets. At low and high redshifts, the difference in completeness between the different classifiers varies by more than a factor of two. For the redshift-weighted classifier, however, the different classifiers have nearly identical completenesses as a function of redshift. Only small deviations from the original performance are seen at the very edges of the redshift range where we have thrown out almost all of the SNe Ia in the biased training sets.

### 4.4.3 Comparison to other models

A full comparison of our classifier to all of the other classifiers submitted to the blinded phase of the PLAsTiCC will be presented in Hložek et al. (2019, in prep.). Those classifiers used a wide range of techniques that were not explored in this analysis, including template fitting, recurrent neural networks and denoising autoencoders. However, all of those models were trained to optimize the Kaggle metric, and we find that they exhibit the same problematic performance with redshift as discussed for our flat-weighted classifier in Section 4.4.2. The classifiers presented in this chapter are currently the best-performing models on the PLAsTiCC dataset, scoring 0.468 on the flat-weighted metric compared to 0.481 for the next-best model submitted to the blinded phase of the PLAsTiCC challenge. On the redshift-weighted metric, our redshift-weighted classifier scores 0.500, significantly better than the next-best model which scores 0.609. Our classifier can therefore serve as a benchmark for future photometric classifiers.

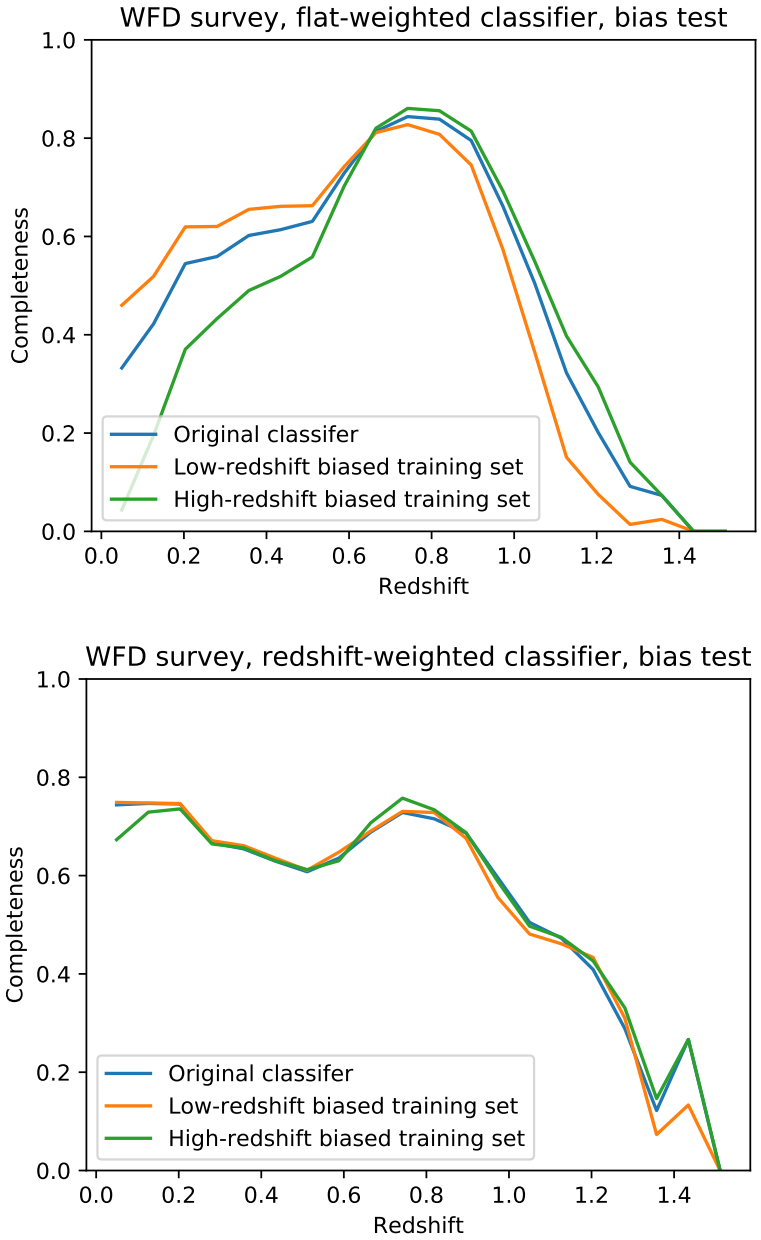



Figure 4.8: Completeness of the sample of SNe Ia in the WFD sample of the test set as a function of redshift for a fixed overall sample purity of 95% when different biases have been introduced in the training set. Top panel: Results for classifiers trained on the flat-weighted metric. Bottom panel: Results for classifiers trained on the redshift-weighted metric. The flat-weighted classifier is highly sensitive to the redshift distribution of objects in the training set while the redshift-weighted classifier shows very little difference in performance.

## 4.5 Discussion

### 4.5.1 Computing resources

The main limitation on computing speed for this model is the GP regression. We perform our computation on an Intel Xeon E3-1270 v3 CPU. A single core on this machine can fit the GP hyperparameters and extract features for ∼10 objects per second. Training the LightGBM classifier takes ∼30 minutes for an augmented training set of 591,410 objects after the features have been computed. Generating predictions with the trained classifier takes a negligible amount of time compared to feature extraction (∼1000 objects per second). Processing the entire PLAsTiCC dataset therefore takes ∼100 core hours of computing time.

Predictions from this classifier can easily be done in real time. A machine with 100 CPU cores similar to the one used for our testing could process ∼1000 objects per second and provide live typing estimates for all of the transients and variables discovered in a survey. As new spectroscopic confirmations of objects are obtained, the classifier could periodically be retrained to incorporate that new data into its training set. The GP fits and feature extractions do not need to be redone for older data so long as the feature extraction algorithm is unchanged, so retraining the classifier on new data can be done in under an hour in most cases.

### 4.5.2 Representativeness of the augmented training set

Ideally, our data augmentation procedure would produce an augmented training set that is fully representative of the test set. In practice, however, there are several reasons why an augmented training set may differ from the test set. First, the training set must cover the full intrinsic diversity of objects of the test set. Note that the augmentation procedure does not attempt to simulate new objects, it simply produces light curves for previously-measured objects under different observing conditions and at different redshifts. If there are rare subtypes of objects that only appear in the test set (e.g., peculiar supernovae), and that have very different light curves from the objects in the training set, then the augmentation procedure will not be able to produce light curves that are similar to the ones observed for these objects. This issue can be addressed by using active learning when obtaining the training set used for classification, as described in Ishida et al. (2019). In this procedure, the output of the classifier is used to determine which objects should be targeted for spectroscopic followup. This helps to ensure that the original training set contains examples of all of the different object types that are present in the full dataset.

The second major challenge for representativeness is in handling the different redshift distributions for object types between the training and test sets. As discussed in Section 4.3.4, the rates of different transients are not currently known well enough as a function of redshift to produce an augmented dataset that is truly representative of the test dataset. Instead, by training a classifier on a redshift-weighted metric, we obtain a classifier whose output is independent of the redshift distributions of the classes in the training sample, as shown

in Section 4.4.2. Such a classifier effectively produces classification probabilities assuming the same arbitrarily chosen redshift distribution for every class. In our implementation, we are using a log-uniform distribution in redshift for our classification, and this assumed distribution can easily be propagated to further analyses. Analyses that depend on photometric classification, such as cosmology with SNe Ia, are already typically required to fit for or model the rates and selection efficiencies of different transient types as a function of redshift to achieve their science goals. A redshift-weighted classifier produces output classification probabilities that depend on known redshift distributions, and the biased redshift distributions in the original training set will have no effect on the classification probabilities.

Our redshift-dependent classifier specifically addressed the issue of having different redshift distributions between the training and test sets. A similar procedure could be applied to other observables of the transients, including but not limited to their peak brightnesses, host properties, or rise and fall times. Assuming that these properties can be measured accurately enough in the training set, we can simply reweight objects in the training set to force the classifier to assume the same arbitrarily chosen distribution over this observable for each object type. This effectively means that the classifier cannot learn anything from the distribution of this observable, or any indirect measurements of the distribution of this observable to classify objects. The observable only needs to be available for the training set, and does not necessarily need to be a feature that is used for classification. For example, our redshift-weighted classifier is reweighted using the spectroscopic host redshifts of the objects in our training sample, but spectroscopic host redshifts are not available for most objects in the full dataset. Nevertheless, the classifier outputs probabilities that are independent of the distributions of spectroscopic host redshifts for different object types in the full dataset.

### 4.5.3 Handling drifting subpopulations with redshift

One potential issue with any kind of photometric classification, including redshift augmentation, is drift in the subpopulations of different types of transients as a function of redshift. For example, the properties of SNe Ia are correlated with the properties of their host galaxies such as host mass (Kelly et al., 2010). As galaxy properties evolve with redshift, the distributions of different subpopulations of SNe Ia also evolve with redshift (Rigault et al., 2013). If the spectroscopically-confirmed training set is biased to low redshifts, then a redshift-augmented training sample will have a different subpopulation distribution than the true one at high redshifts leading to a bias in the classification probabilities. This is a challenge for any photometric classification method, not just ones that use redshift augmentation, because the high-redshift followup strategy itself could be biased towards some subpopulation.

If the subpopulations of an object type can be identified in the training set using some observable, then we can reweight the training set as described previously to produce a classification assuming an arbitrarily chosen distribution over this observable, with the same assumed distribution for all object types in the sample. For instance, a dedicated campaign to measure the host masses of every galaxy in the training set could be used to produce a

photometric classifier that classifies SNe Ia without taking the host mass distributions, or any indirect measurements of the host mass distributions, into account. This classifier will therefore produce classification probabilities that are independent of the change in subpopulations as a function of redshift associated with host mass. Note that this procedure could introduce biases if the observable is not available for all objects in the training sample, e.g., if it were not possible to reliably determine the host masses for higher redshift objects in the training sample.

A final concern about changing subpopulations would be if new subpopulations appear at high redshifts that are not present at low redshifts. While the survey is running, active learning, as described in Ishida et al. (2019) can be used to attempt to identify these subpopulations and trigger spectroscopic followup to add them to the training set. If, however, these new subpopulations are missed entirely in the training set, then the classifier is unlikely to be able to accurately classify objects from them.

Specific science applications may be more or less sensitive to the evolution of subpopulations with redshift. For cosmology with SNe Ia, for example, if a new subtype of SN Ia appeared at high redshift that wasn't identified at low redshift, then it could significantly bias the cosmological analysis. This is an issue regardless of the analysis strategy: at high redshifts, a "representative sample" will likely only be able to spectroscopically type a small number of objects, potentially missing new subpopulations. A lower-redshift followup strategy that intends to use redshift-augmentation may be able to obtain a larger and more complete sample at the lower redshift, but it relies on the assumption that we can produce a reasonable model of the differences between the low and high-redshift samples. In practice, there is a trade-off between all of these concerns, and a variety of different followup strategies should be simulated to determine the optimal strategy for the science objectives.

### 4.5.4 Implications for future surveys

The results of this analysis have several implications for future surveys. For surveys with limited spectroscopic resources devoted to obtaining training samples for photometric classification, previous work (e.g., L16) has suggested that these surveys should attempt to produce a training set of objects with spectroscopically-confirmed types that is as representative of the full dataset as possible. With our augmentation technique, we instead suggest that spectroscopic resources may better be used to obtain spectroscopic classifications of as many intermediate-redshift well-sampled light curves as possible. We can then use these well-sampled light curves to simulate high-redshift light curves rather than having to spend large amounts of spectroscopic followup to type high-redshift objects directly.

With augmentation, a light curve is most valuable if it has good-quality observations with a high cadence. While the labels for all of the objects in the test set that the classifier will be applied to are not known, we do know the cadence, observation depths, whether or not the light curve was impacted by season boundaries, etc. of every object in the test set. A well sampled light curve can always be degraded to simulate all of these effects, as shown in our augmentation procedure. One potential limitation of redshift augmentation is that

the redshift range to which a light curve model can be shifted is limited by the wavelength coverage of its input observations. As discussed in Section 4.6, we find that we must limit the change in wavelength to less than 50% to avoid having large GP extrapolation uncertainties for the bluer bands. This can be somewhat addressed by adding additional followup in bluer bands, even including observations from other telescopes, but it effectively means that there is a maximum change in redshift that can be applied in the augmentation procedure. For this reason, intermediate-redshift light curves that can be augmented to high redshifts are more valuable to have in the training set than very low-redshift light curves.

This suggests that an alternative strategy can be used to generate training sets for photometric classification. At the start of a large survey such as LSST, a small, coordinated campaign can be undertaken to obtain very deep, high-cadence observations in a limited region of the sky. Coincident spectroscopic campaigns can obtain redshifts for many of the transients discovered in this small survey. These light curves can then be used as templates for the augmentation procedure to produce light curves at any redshift, and the well-observed light curves can easily be degraded to match the lower signal-to-noise and cadence light curves from other parts of the survey. At this point, the classifier should be able to accurately classify the majority of "normal" objects in the test sample.

For unusual objects that are not present in this first training set, we can attempt to use active learning following a procedure similar to Ishida et al. (2019) to identify these objects and launch additional followup campaigns to obtain good quality observations for them. We can also develop additional spectroscopic followup strategies tailored to the goals of specific scientific programs where the presence of unusual objects would impact the scientific results, such as searching for new subtypes of SNe Ia at high redshifts that could bias cosmological measurements.

Finally, spectroscopic observations will be obtained for purposes other than simply building training sets for photometric classification or checking for new subpopulations of SNe Ia at high redshifts. For example, there may be spectroscopic campaigns dedicated to specific subtypes or rare subpopulations of other object types. The teams obtaining these observations may have differing goals and selection requirements. It is essential to coordinate spectroscopic efforts with other teams to best utilize the available spectroscopic resources.

### 4.5.5 Improvements to the classifier

There are several improvements that could be made to our classifier. First, for this analysis, we restricted ourselves to a single GP kernel that was required to fit all the different kinds of transients. Different kernels could perform better for different transients, so a natural extension is to investigate the use of different kernels for the GP. As shown in RTV18, a Gibbs kernel provides better fits to supernova-like light curves than the Matérn kernel used for this analysis. Periodic light curves, such as different types of variable stars, could be fit with periodic kernels (Rasmussen & Williams, 2006) to take advantage of the known periodicity. In general, each light curve could be fit with GPs with multiple different kernels, and different features could be computed from each of these fits, at the cost of increased

computing time per object. The features from all of these different fits could be used as input for a single LightGBM classifier. Additional features computed with other means, such as features from a Lomb-Scargle periodogram (VanderPlas, 2018) which have shown to be very useful for classification of periodic light curves, could also easily be added to the classifier, again, at the expense of additional computation time.

An additional approach for further work is to combine the results of multiple independent classifiers. As described in Hložek et al. (2019, in prep.), combining the results of the top classifiers submitted to the blinded phase of the PLAsTiCC resulted in much better performance than any single classifier. These additional classifiers could also be trained on the augmented dataset to improve their performance.

### 4.5.6 Conclusions

We have developed a new framework for photometric classification of transients and variables from surveys such as LSST. Our classifier is designed to be trained on datasets of spectroscopically-confirmed transients and variables and is able to handle training sets that are biased towards bright low-redshift objects. Using GP regression, we augment the set of light curves used for training to generate light curves over a wider range of redshifts and observing conditions than present in the original training set. This procedure is designed so that no specific model or parametrization of the light curve is required: it can be performed even on poorly sampled light curves or ones with large gaps of observations in time.

Our classifier achieves the best performance on the PLAsTiCC dataset of any single algorithm to date, scoring 0.468 on the metric defined by the PLAsTiCC team with flat weights for all objects in the training set. It achieves an AUC of 0.957 for classification of SNe Ia compared to 0.934 for the next best single classifier submitted to the blinded phase of the PLAsTiCC challenge. Our classifier sets a benchmark for the performance of future photometric classifiers on the PLAsTiCC dataset and for the LSST. Additionally, we have shown how a metric can be designed to produce classifiers whose output probabilities are independent of the redshift distributions of the different kinds of transients in the training sample. This leads to a better understanding of the output probabilities of the classifier, which is essential for analyses such as the determination of cosmological parameters using photometrically classified SNe Ia.

All of the results in this chapter can be reproduced with the `avocado` classification package. A Jupyter notebook is provided as part of that package which was used to produce all of the figures shown in this chapter.

## 4.6 Appendix: Implementation of the training set augmentation

An overview of the augmentation procedure can be found in Section 4.3.4. In this section, we provide the details necessary to reproduce our procedure. In general, we attempt to use

as few tuned parameters as possible and to use the available information in the full dataset whenever possible. For each object in the training sample, we generate up to 100 new versions of that object under different observing conditions and at different redshifts.

For each new augmented extragalactic object, we first choose a new redshift for that object. We limit the new redshift so that $0.95\ z_{original} < z_{new} < 5\ z_{original}$. A hard lower bound is used here to avoid making faint objects much brighter and having their new simulated observations be dominated by large modeling uncertainties. As our training set is biased toward bright, low-redshift objects, this does not limit the performance of the classifier at low redshifts. A loose upper bound is used to prevent the augmentation procedure from repeatedly trying to generate objects that are very faint and not able to be detected by the instrument. We impose an additional upper bound on the redshift of $1+z_{new} < 1.5\ (1+z_{old})$. If the wavelength range is shifted too far, the GP is required to extrapolate far from where there is available data, and modeling uncertainties dominate its prediction. For this reason, we limit the possible shift in wavelength to 50% which results in the previous inequality. We choose a new redshift with a log-uniform distribution between the lower and upper redshift bounds described previously. For augmented Galactic objects, we simply modify the brightness of the object by adding an offset in magnitudes drawn from a Gaussian distribution with a mean of 0 and a standard deviation of 0.5 mag.

Once a new redshift or brightness has been chosen, we choose to simulate it either as part of the WFD survey or the DDF survey. Light curves in the DDF survey are much better sampled than light curves in the WFD survey, so for WFD template light curves we only generate WFD light curves. For DDF template light curves, we randomly choose to generate either a WFD or a DDF light curve. We choose a target number of observations for the new light curve using a distribution that roughly matches the distribution of the PLAsTiCC test dataset. For the DDF survey, we choose a target number of observations following a Gaussian distribution with a mean of 330 and a standard deviation of 30. For the WFD survey, we find that the number of observations of each light curve can be described reasonably well with a three component Gaussian mixture model, with probabilities for each component of [0.05, 0.4, 0.55], means of [95, 115, 138] observations and standard deviations of [20, 8, 8] observations. These numbers were all determined empirically by tuning our model to match the observed distribution in the test set.

We then choose the observation times and bands of these new observations. Our goal for this procedure is to generate light curves under a wide range of different observing conditions. To account for a change in redshift, we stretch the light curve in time to account for the time dilation due to the redshift difference. When shifting an object to higher redshifts, we fill in the light curve to account for the fact that there is a lower density of observations after applying the time dilation compared to at lower redshifts. This is done by adding additional observations to the new light curve at the same times as existing observations in the template light curve in randomly selected bands. We do not attempt to sample observations at new times because we find that when doing so the uncertainty in the GP flux predictions leads to unrealistic light curves with large uncertainties. For example, for the light curve shown in Figure 4.4, the GP flux predictions for the rise of the light curve are highly unrealistic

and have large uncertainties. Our method ensures that new light curves generated using this light curve as a template will only include fall-time data where the GP flux predictions are accurate.

We shift the dates of the light curve by up to 50 days either forward or backward. This shifting does not affect our classifier, but will affect classifiers that use absolute date information. To create light curves that are cut off by season boundaries, we randomly drop a block of observations with a width of 250 days. Following this, we randomly drop observations so that we have no more than the target number of observations chosen previously, and we drop at least 10% of observations to ensure that we are introducing some variation in the augmented light curves.

After these procedures, we have chosen a set of observation times and bands for the new light curve. We compute the mean GP flux prediction at those times and at a set of wavelengths corresponding to the bands shifted by the new redshift. We choose to use the mean GP flux predictions rather than drawing flux predictions from the GP because our goal is not to produce all light curves consistent with our data, but instead to provide a reliable interpolation or extrapolation of the template observations. We include the uncertainty of the GP flux predictions in the uncertainties for each new observation.

Using the previously assigned new redshift, we adjust the brightness of the new light curve. This is done assuming a fiducial flat ΛCDM cosmology with $H_0 = 70$ km/s/Mpc and $\Omega_m = 0.3$. We find that the classification results are not sensitive to the choice of fiducial cosmology. We then add simulated observation noise to each of the light curves. To determine the appropriate noise levels for each of the WFD and DDF surveys, we fit lognormal distributions to a random subset of the observations in the test set and draw from those distributions for our new simulated observations. We add this noise in quadrature with the uncertainties from the GP flux predictions.

We naively estimate a photometric redshift for each of our new objects. For each new object, we draw one reference object from the full PLAsTiCC dataset that has a spectroscopic redshift for its host. We then calculate the difference between its photometric and spectroscopic redshifts, and we add this difference to the spectroscopic redshift of the new object to estimate its photometric redshift. This procedure is not a proper model of photometric redshifts, but it does ensure that the augmented training set contains any issues present for the photometric redshifts in the full dataset. Incorporating a proper photometric redshift model into the augmenting procedure would improve performance.

Finally, we apply a selection model to the augmented light curve. The original PLAsTiCC observations have a detection flag that is set in a non-deterministic way. We fit an error function to the observations from the full dataset to predict the probability of detection as a function of signal-to-noise, and use this to predict the probability that an observation is flagged as detected. As was done for the original dataset, we then only keep objects that have at least two detections in their light curves. If the object is rejected, then we repeatedly attempt to generate a new light curve from the same template at the same redshift until the generated object passes the selection criteria. If more than ten attempts at generating a new light curve fail to pass the selection criteria, the template is skipped and we move on to the

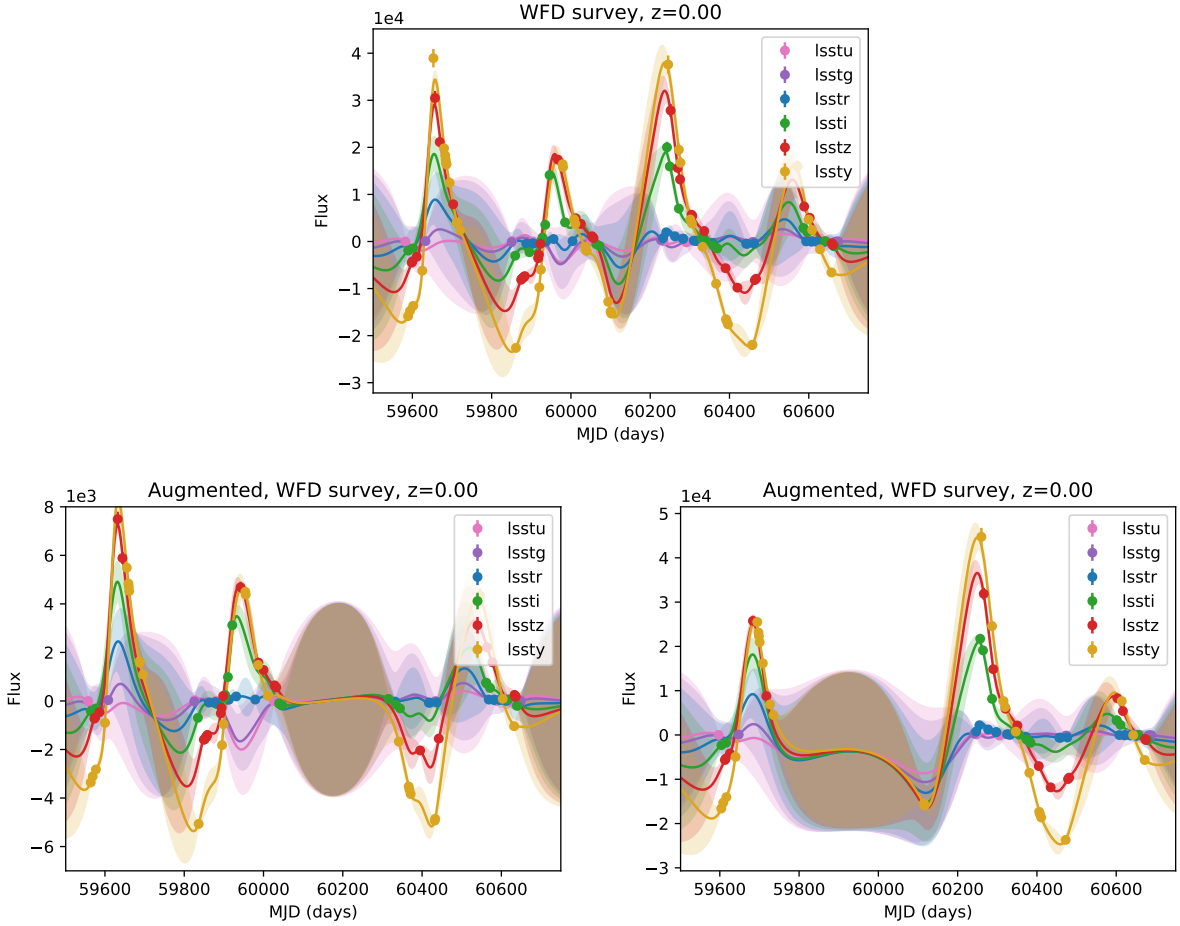



Figure 4.9: Examples of augmented light curves for a Mira variable. Top panel: the original light curve. Bottom two panels: examples of light curves generated using the original light curve as a template. See Figure 4.3 for an explanation of the colors and shading of these plots.

next template. This failure is common, and typically indicates that we chose too high of a redshift for the new light curve, and that the generated fluxes are too faint to be detectable by the telescope. We explicitly try to generate light curves at these high redshifts to ensure that our augmented set covers the full range of light curves that it is possible to detect.

Examples of the final augmented light curves are shown in Figures 4.9 through 4.11.

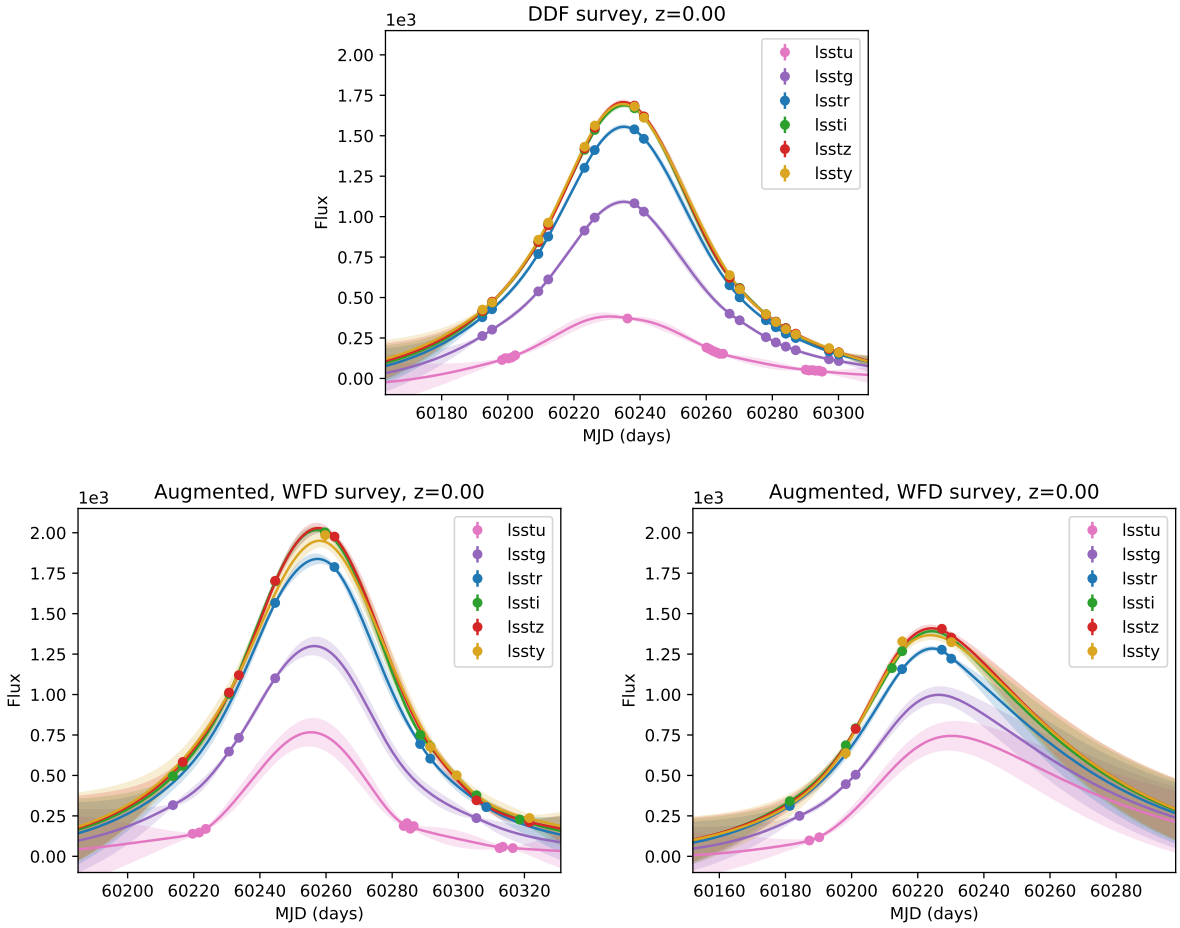



Figure 4.10: Examples of augmented light curves for a single lens microlensing event. Top panel: the original light curve. Bottom two panels: examples of light curves generated using the original light curve as a template. See Figure 4.3 for an explanation of the colors and shading of these plots.

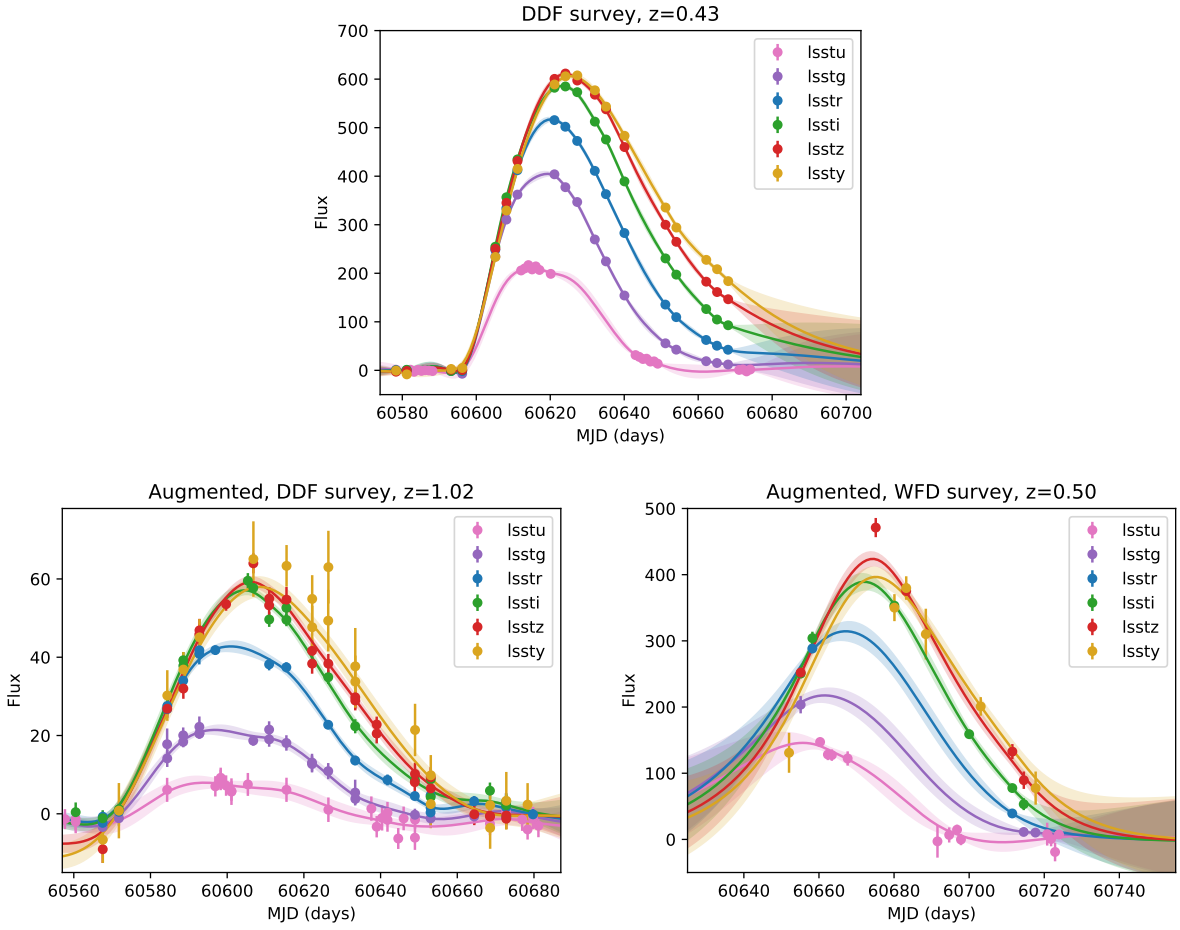



Figure 4.11: Examples of augmented light curves for a superluminous supernova light curve. Top panel: the original light curve. Bottom two panels: examples of light curves generated using the original light curve as a template. See Figure 4.3 for an explanation of the colors and shading of these plots.

# Chapter 5

# Conclusions

In this dissertation, we discussed several different methods of improving the systematic uncertainties in distance measurements to SNe Ia. These methods will be essential for the next generation of cosmological analyses using SNe Ia with experiments such as LSST.

In Chapter 2, we discussed a newly discovered CCD anomaly that affects most astronomical instruments in use, and we showed how it impacts the measured brightnesses of SNe Ia with the SNIFS instrument. We found that there is crosstalk between the binary-coded output of the analog-to-digital converter (ADC) and subsequently read-out pixels. An offset is introduced into the subsequently read-out pixels which is roughly proportional to the number of "1" bits in the binary encoding of a driver pixel. This "binary offset effect" is present in a wide variety of currently-used instruments with different electronics configurations. We characterized the effect and modeled it with high accuracy in the SNIFS instrument used by the Nearby Supernova Factory collaboration on the UH88 telescope. We then showed how the effect can be corrected to prevent it from affecting analyses using data from this instrument.

Another major source of systematic uncertainty in cosmological analyses with SNe Ia is intrinsic variation of SNe Ia. Traditional methods of estimating the distances to SNe Ia, such as SALT2, result in systematic biases that depend on the properties of the host galaxies. In Chapter 3, we discussed a new method of parametrizing SNe Ia using manifold learning to generate a non-linear decomposition of the intrinsic diversity of their spectra near maximum light. We found that a three-component non-linear embedding is able to capture this intrinsic diversity, and we identified regions of the parameter space of SNe Ia where previous standardization methods such as SALT2 have biases of up to 0.3 mag. We showed how our parametrization of SNe Ia can be used to standardize the brightnesses of SNe Ia, resulting in an RMS of $0.100 \pm 0.010$ mag with an intrinsic dispersion of $0.071 \pm 0.013$ mag. We found that correlations between host galaxy properties and distance estimates are greatly reduced using this new parametrization, which is essential for precise cosmological measurements with SNe Ia.

In Chapter 4, we developed a new photometric classifier to address the issue that future surveys such as LSST will only be able to spectroscopically type a small fraction of the

transients that they discover. Our classifier was designed to be able to be applied to training samples of spectroscopically-typed objects that are heavily biased toward bright, low-redshift objects. This classifier was developed as part of the Photometric LSST Astronomical Time-Series Classification Challenge. Out of all of the models submitted during the blinded phase of the PLAsTiCC, our classifier achieved the best performance on the PLAsTiCC test set measured using the weighted log-loss metric proposed by the PLAsTiCC team. We used Gaussian processes in both time and wavelength to build light curve models for each object in the sample that incorporate information from all bands simultaneously. Using these light curve models, we showed how one can take a training set that is heavily biased towards bright, low-redshift objects, and "augment" it by generating new light curves covering a range of redshifts and observing conditions. This augmented dataset contains light curves that are much more representative of the full light curve sample. We then showed how a classifier can be trained whose performance is independent of the redshift distributions of the different object types in the training sample. These techniques described in this chapter will significantly improve the performance of photometric classifiers for surveys such as LSST, and they will help both reduce and characterize the levels of contamination from other transients in cosmological analyses of SNe Ia.

# Chapter 6

# Acknowledgements

This dissertation made use of the following software:

- Astropy (Astropy Collaboration et al., 2013, 2018)
- George (Ambikasaran et al., 2015)
- Jupyter (Kluyver et al., 2016)
- LightGBM (Ke et al., 2017)
- Matplotlib (Hunter, 2007)
- NumPy (van der Walt et al., 2011)
- Pandas (McKinney, 2010)
- scikit-learn (Pedregosa et al., 2011)
- SciPy (Jones et al., 2001)
- sep (Barbary, 2016)

We acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC) who supported the author with a PGS M fellowship for the first year of graduate school.

The results in Chapters 2 and 3 made use of data from the SNfactory taken with the SNIFS instrument. We thank the technical staff of the University of Hawaii 2.2 m telescope, and Dan Birchall for observing assistance. We wish to recognize and acknowledge the very significant cultural role and reverence that the summit of Maunakea has always had within the indigenous Hawaiian community. We are most fortunate to have the opportunity to conduct observations from this mountain. This work was supported in part by the Director, Office of Science, Office of High Energy Physics of the U.S. Department of Energy under

Contract No. DE-AC02-05CH11231. Support in France was provided by CNRS/IN2P3, CNRS/INSU, and PNC; LPNHE acknowledges support from LABEX ILP, supported by French state funds managed by the ANR within the Investissements d'Avenir programme under reference ANR-11-IDEX-0004-02. Some results were obtained using resources and support from the National Energy Research Scientific Computing Center, supported by the Director, Office of Science, Office of Advanced Scientific Computing Research of the U.S. Department of Energy under Contract No. DE-AC02- 05CH11231. We thank the Gordon & Betty Moore Foundation for their continuing support. We also thank the High Performance Research and Education Network (HPWREN), supported by National Science Foundation Grant Nos. 0087344 & 0426879.

The results in Chapter 4 made use of the PLAsTiCC simulations of observations of different transients and variables with LSST. We thank the PLAsTiCC team for producing this dataset and for organizing the PLAsTiCC. We thank the Kaggle platform for hosting the blinded phase of the PLAsTiCC, and all of the teams who participated in this challenge and shared ideas on the Kaggle discussion board. This work was supported by the Director, Office of Science, Office of High Energy Physics of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

In Chapter 2, we used data from a wide range of different telescopes and we are grateful that these data were made available to us.

Funding for SDSS-III has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, and the U.S. Department of Energy Office of Science. The SDSS-III web site is http://www.sdss3.org/. SDSS-III is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS-III Collaboration including the University of Arizona, the Brazilian Participation Group, Brookhaven National Laboratory, Carnegie Mellon University, University of Florida, the French Participation Group, the German Participation Group, Harvard University, the Instituto de Astrofisica de Canarias, the Michigan State/Notre Dame/JINA Participation Group, Johns Hopkins University, Lawrence Berkeley National Laboratory, Max Planck Institute for Astrophysics, Max Planck Institute for Extraterrestrial Physics, New Mexico State University, New York University, Ohio State University, Pennsylvania State University, University of Portsmouth, Princeton University, the Spanish Participation Group, University of Tokyo, University of Utah, Vanderbilt University, University of Virginia, University of Washington, and Yale University.

This project used public archival data from the Dark Energy Survey (DES). Funding for the DES Projects has been provided by the U.S. Department of Energy, the U.S. National Science Foundation, the Ministry of Science and Education of Spain, the Science and Technology Facilities Council of the United Kingdom, the Higher Education Funding Council for England, the National Center for Supercomputing Applications at the University of Illinois at Urbana-Champaign, the Kavli Institute of Cosmological Physics at the University of Chicago, the Center for Cosmology and Astro-Particle Physics at the Ohio State University, the Mitchell Institute for Fundamental Physics and Astronomy at Texas A&M University, Financiadora de Estudos e Projetos, Fundação Carlos Chagas Filho de Amparo à Pesquisa

do Estado do Rio de Janeiro, Conselho Nacional de Desenvolvimento Científico e Tecnológico and the Ministério da Ciência, Tecnologia e Inovação, the Deutsche Forschungsgemeinschaft and the Collaborating Institutions in the Dark Energy Survey. The Collaborating Institutions are Argonne National Laboratory, the University of California at Santa Cruz, the University of Cambridge, Centro de Investigaciones Enérgeticas, Medioambientales y Tecnológicas-Madrid, the University of Chicago, University College London, the DES-Brazil Consortium, the University of Edinburgh, the Eidgenössische Technische Hochschule (ETH) Zürich, Fermi National Accelerator Laboratory, the University of Illinois at Urbana-Champaign, the Institut de Ciències de l'Espai (IEEC/CSIC), the Institut de Física d'Altes Energies, Lawrence Berkeley National Laboratory, the Ludwig-Maximilians Universität München and the associated Excellence Cluster Universe, the University of Michigan, the National Optical Astronomy Observatory, the University of Nottingham, the Ohio State University, the University of Pennsylvania, the University of Portsmouth, SLAC National Accelerator Laboratory, Stanford University, the University of Sussex, and Texas A&M University. Based in part on observations at Cerro Tololo Inter-American Observatory, National Optical Astronomy Observatory (NOAO Prop. I D and PI), which is operated by the Association of Universities for Research in Astronomy (AURA) under a cooperative agreement with the National Science Foundation.

Based in part on observations obtained with MegaPrime/MegaCam, a joint project of CFHT and CEA/DAPNIA, at the Canada-France-Hawaii Telescope (CFHT) which is operated by the National Research Council (NRC) of Canada, the Institut National des Sciences de l'Univers of the Centre National de la Recherche Scientifique (CNRS) of France, and the University of Hawaii.

Based in part on observations obtained at the Gemini Observatory acquired through the Gemini Observatory Archive, which is operated by the Association of Universities for Research in Astronomy, Inc., under a cooperative agreement with the NSF on behalf of the Gemini partnership: the National Science Foundation (United States), the National Research Council (Canada), CONICYT (Chile), Ministerio de Ciencia, Tecnología e Innovación Productiva (Argentina), and Ministério da Ciência, Tecnologia e Inovação (Brazil).

Based in part on observations made with the NASA/ESA Hubble Space Telescope, obtained from the Data Archive at the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5-26555. These observations are associated with programs 9583, 13677, 14327, and 14820.

Some of the data presented herein were obtained at the W. M. Keck Observatory, which is operated as a scientific partnership among the California Institute of Technology, the University of California and the National Aeronautics and Space Administration. The Observatory was made possible by the generous financial support of the W. M. Keck Foundation. This research has made use of the Keck Observatory Archive (KOA), which is operated by the W. M. Keck Observatory and the NASA Exoplanet Science Institute (NExScI), under contract with the National Aeronautics and Space Administration.

Based in part on data collected at Subaru Telescope, which is operated by the National Astronomical Observatory of Japan.

Based in part on data obtained from the ESO Science Archive Facility under requests numbered kboone301279, kboone301093, and kboone301092.

# Bibliography

Aldering, G., Adam, G., Antilogus, P., et al. 2002, in Proc. SPIE, Vol. 4836, Survey and Other Telescope Technologies and Discoveries, ed. J. A. Tyson & S. Wolff, 61–72

Aldering, G., Antilogus, P., Bailey, S., et al. 2006, ApJ, 650, 510, doi: 10.1086/507020

Amanullah, R., Goobar, A., Johansson, J., et al. 2014, ApJL, 788, L21, doi: 10.1088/2041-8205/788/2/L21

Ambikasaran, S., Foreman-Mackey, D., Greengard, L., Hogg, D. W., & O'Neil, M. 2015, IEEE Transactions on Pattern Analysis and Machine Intelligence, 38, 252, doi: 10.1109/TPAMI.2015.2448083

Antilogus, P., Astier, P., Doherty, P., Guyonnet, A., & Regnault, N. 2014, Journal of Instrumentation, 9, C03048, doi: 10.1088/1748-0221/9/03/C03048

Astier, P., Guy, J., Regnault, N., et al. 2006, A&A, 447, 31, doi: 10.1051/0004-6361:20054185

Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33, doi: 10.1051/0004-6361/201322068

Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, AJ, 156, 123, doi: 10.3847/1538-3881/aabc4f

Baade, W. 1938, ApJ, 88, 285, doi: 10.1086/143983

Bacon, R., Adam, G., Baranne, A., et al. 1995, A&AS, 113, 347

Bacon, R., Copin, Y., Monnet, G., et al. 2001, MNRAS, 326, 23, doi: 10.1046/j.1365-8711.2001.04612.x

Baggett, S., Hartig, G., & Cheung, E. 2004, WFC3 UVIS Crosstalk Images, Technical Report, Space Telescope Science Institute, Baltimore, MD

Baggett, S. M., Noeske, K., Anderson, J., MacKenty, J. W., & Petro, L. 2012, in Proc. SPIE, Vol. 8453, High Energy, Optical, and Infrared Detectors for Astronomy V, 845336

Bailey, S., Aragon, C., Romano, R., et al. 2007, ApJ, 665, 1246, doi: 10.1086/519832

Bailey, S., Aldering, G., Antilogus, P., et al. 2009, A&A, 500, L17, doi: 10.1051/0004-6361/200911973

Barbary, K. 2016, The Journal of Open Source Software, 1, doi: 10.21105/joss.00058

Barone-Nugent, R. L., Lidman, C., Wyithe, J. S. B., et al. 2012, MNRAS, 425, 1007, doi: 10.1111/j.1365-2966.2012.21412.x

Beers, T. C., Flynn, K., & Gebhardt, K. 1990, AJ, 100, 32, doi: 10.1086/115487

Beletic, J. W., Gerdes, R., & Duvarney, R. C. 1998, in Astrophysics and Space Science Library, Vol. 228, Optical Detectors for Astronomy, ed. J. Beletic & P. Amico, 103

Bernstein, J. P., Kessler, R., Kuhlmann, S., et al. 2012, ApJ, 753, 152, doi: 10.1088/0004-637X/753/2/152

Bertin, E., & Arnouts, S. 1996, A&AS, 117, 393, doi: 10.1051/aas:1996164

Betoule, M., Kessler, R., Guy, J., et al. 2014, A&A, 568, A22, doi: 10.1051/0004-6361/201423413

Blondin, S., Mandel, K. S., & Kirshner, R. P. 2011, A&A, 526, A81, doi: 10.1051/0004-6361/201015792

Blondin, S., Matheson, T., Kirshner, R. P., et al. 2012, AJ, 143, 126, doi: 10.1088/0004-6256/143/5/126

Bongard, S., Soulez, F., Thiébaut, É., & Pecontal, É. 2011, MNRAS, 418, 258, doi: 10.1111/j.1365-2966.2011.19480.x

Boone, K., Aldering, G., Copin, Y., et al. 2018, PASP, 130, 064504, doi: 10.1088/1538-3873/aab0fe

Branch, D. 1998, ARA&A, 36, 17, doi: 10.1146/annurev.astro.36.1.17

Branch, D., & Tammann, G. A. 1992, ARA&A, 30, 359, doi: 10.1146/annurev.aa.30.090192.002043

Branch, D., Dang, L. C., Hall, N., et al. 2006, PASP, 118, 560, doi: 10.1086/502778

Brown, T. M., & Lupie, O. 2004, Filter Ghosts in the WFC3 UVIS Channel, Technical Report, Space Telescope Science Institute, Baltimore, MD

Burns, C. R., Stritzinger, M., Phillips, M. M., et al. 2011, AJ, 141, 19, doi: 10.1088/0004-6256/141/1/19

Burns, C. R., Parent, E., Phillips, M. M., et al. 2018, ApJ, 869, 56, doi: 10.3847/1538-4357/aae51c

Buton, C., Copin, Y., Aldering, G., et al. 2013, A&A, 549, A8, doi: 10.1051/0004-6361/201219834

Caldwell, D. A., Kolodziejczak, J. J., Van Cleve, J. E., et al. 2010, ApJL, 713, L92, doi: 10.1088/2041-8205/713/2/L92

Capozziello, S., & Fang, L. Z. 2002, International Journal of Modern Physics D, 11, 483, doi: 10.1142/S0218271802002025

Cardelli, J. A., Clayton, G. C., & Mathis, J. S. 1989, ApJ, 345, 245, doi: 10.1086/167900

Carpenter, B., Gelman, A., Hoffman, M., et al. 2017, Journal of Statistical Software, Articles, 76, 1, doi: 10.18637/jss.v076.i01

Carroll, S. M., Duvvuri, V., Trodden, M., & Turner, M. S. 2004, PhRvD, 70, 043528, doi: 10.1103/PhysRevD.70.043528

Castilla, J., Ballester, O., Cardiel, L., et al. 2010, in Proc. SPIE, Vol. 7735, Ground-based and Airborne Instrumentation for Astronomy III, 77352O

Charnock, T., & Moss, A. 2017, ApJ, 837, L28, doi: 10.3847/2041-8213/aa603d

Chevallier, M., & Polarski, D. 2001, International Journal of Modern Physics D, 10, 213, doi: 10.1142/S0218271801000822

Chotard, N., Gangler, E., Aldering, G., et al. 2011, A&A, 529, L4, doi: 10.1051/0004-6361/201116723

Conley, A., Guy, J., Sullivan, M., et al. 2011, ApJS, 192, 1, doi: 10.1088/0067-0049/192/1/1

Daniel, S. F., Connolly, A., Schneider, J., Vanderplas, J., & Xiong, L. 2011, AJ, 142, 203, doi: 10.1088/0004-6256/142/6/203

de Kat, J., Boulade, O., Charlot, Abbon, P. X., et al. 2004, in Astrophysics and Space Science Library, Vol. 300, Scientific Detectors for Astronomy, The Beginning of a New Era, ed. P. Amico, J. W. Beletic, & J. E. Belectic, 517–522

Delgado, F., Saha, A., Chandrasekharan, S., et al. 2014, in Proc. SPIE, Vol. 9150, Modeling, Systems Engineering, and Project Management for Astronomy VI, 915015

Dixon, S., Aldering, G. S., Domagalski, R., et al. 2016, in American Astronomical Society Meeting Abstracts, Vol. 227, American Astronomical Society Meeting Abstracts, 146.15

Dvali, G., Gabadadze, G., & Porrati, M. 2000, Physics Letters B, 485, 208, doi: 10.1016/S0370-2693(00)00669-9

Efron, B. 1979, Ann. Statist., 7, 1, doi: 10.1214/aos/1176344552

Fakhouri, H. K., Boone, K., Aldering, G., et al. 2015, ApJ, 815, 58, doi: 10.1088/0004-637X/815/1/58

Fitzpatrick, E. L. 1999, PASP, 111, 63, doi: 10.1086/316293

Foley, R. J., Hoffmann, S. L., Macri, L. M., et al. 2018, arXiv e-prints, arXiv:1806.08359. https://arxiv.org/abs/1806.08359

Foley, R. J., Challis, P. J., Chornock, R., et al. 2013, ApJ, 767, 57, doi: 10.1088/0004-637X/767/1/57

Foley, R. J., Fox, O. D., McCully, C., et al. 2014, MNRAS, 443, 2887, doi: 10.1093/mnras/stu1378

Freedman, W. L., Madore, B. F., Hatt, D., et al. 2019, arXiv e-prints, arXiv:1907.05922. https://arxiv.org/abs/1907.05922

Gimeno, G., Roth, K., Chiboucas, K., et al. 2016, in Proc. SPIE, Vol. 9908, Ground-based and Airborne Instrumentation for Astronomy VI, 99082S

Gunn, J. E., Carr, M., Rockosi, C., et al. 1998, AJ, 116, 3040, doi: 10.1086/300645

Gupta, R. R., D'Andrea, C. B., Sako, M., et al. 2011, ApJ, 740, 92, doi: 10.1088/0004-637X/740/2/92

Guy, J., Astier, P., Baumont, S., et al. 2007, A&A, 466, 11, doi: 10.1051/0004-6361:20066930

Guy, J., Sullivan, M., Conley, A., et al. 2010, A&A, 523, A7, doi: 10.1051/0004-6361/201014468

Hayden, B. T., Gupta, R. R., Garnavich, P. M., et al. 2013, ApJ, 764, 191, doi: 10.1088/0004-637X/764/2/191

Hlozek, R., Kunz, M., Bassett, B., et al. 2012, ApJ, 752, 79, doi: 10.1088/0004-637X/752/2/79

Hook, I. M., Jørgensen, I., Allington-Smith, J. R., et al. 2004, PASP, 116, 425, doi: 10.1086/383624

Hotokezaka, K., Nakar, E., Gottlieb, O., et al. 2019, Nature Astronomy, doi: 10.1038/s41550-019-0820-1

Humphreys, E. M. L., Reid, M. J., Moran, J. M., Greenhill, L. J., & Argon, A. L. 2013, ApJ, 775, 13, doi: 10.1088/0004-637X/775/1/13

Hunter, J. D. 2007, Computing in Science & Engineering, 9, 90, doi: 10.1109/MCSE.2007.55

Ishida, E. E. O., & de Souza, R. S. 2013, MNRAS, 430, 509, doi: 10.1093/mnras/sts650

Ishida, E. E. O., Beck, R., González-Gaitán, S., et al. 2019, MNRAS, 483, 2, doi: 10.1093/mnras/sty3015

Janesick, J. R. 2001, Scientific charge-coupled devices (SPIE Press)

Jansen, R. A., Collins, N. R., & Windhorst, R. A. 2003, in HST Calibration Workshop : Hubble after the Installation of the ACS and the NICMOS Cooling System, ed. S. Arribas, A. Koekemoer, & B. Whitmore, 193

Jones, D. O., Scolnic, D. M., Riess, A. G., et al. 2017, ApJ, 843, 6, doi: 10.3847/1538-4357/aa767b

Jones, E., Oliphant, T., Peterson, P., et al. 2001, SciPy: Open source scientific tools for Python. http://www.scipy.org/

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

Karpenka, N. V., Feroz, F., & Hobson, M. P. 2013, MNRAS, 429, 1278, doi: 10.1093/mnras/sts412

Kasen, D. 2006, ApJ, 649, 939, doi: 10.1086/506588

Kashikawa, N., Aoki, K., Asai, R., et al. 2002, PASJ, 54, 819, doi: 10.1093/pasj/54.6.819

Ke, G., Meng, Q., Finley, T., et al. 2017, in Advances in Neural Information Processing Systems 30, ed. I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Curran Associates, Inc.), 3146–3154

Kelly, P. L., Hicken, M., Burke, D. L., Mandel, K. S., & Kirshner, R. P. 2010, ApJ, 715, 743, doi: 10.1088/0004-637X/715/2/743

Kessler, R., Bernstein, J. P., Cinabro, D., et al. 2009, PASP, 121, 1028, doi: 10.1086/605984

Kessler, R., Bassett, B., Belov, P., et al. 2010, PASP, 122, 1415, doi: 10.1086/657607

Kessler, R., Marriner, J., Childress, M., et al. 2015, AJ, 150, 172, doi: 10.1088/0004-6256/150/6/172

Kessler, R., Narayan, G., Avelino, A., et al. 2019, PASP, 131, 094501, doi: 10.1088/1538-3873/ab26f1

Kibrick, R. I., Stover, R. J., & Conrad, A. R. 1993, in Astronomical Society of the Pacific Conference Series, Vol. 52, Astronomical Data Analysis Software and Systems II, ed. R. J. Hanisch, R. J. V. Brissenden, & J. Barnes, 277

Kim, A. G., Thomas, R. C., Aldering, G., et al. 2013, ApJ, 766, 84, doi: 10.1088/0004-637X/766/2/84

Kluyver, T., Ragan-Kelley, B., Pérez, F., et al. 2016, in Positioning and Power in Academic Publishing: Players, Agents and Agendas, ed. F. Loizides & B. Schmidt, IOS Press, 87 – 90

Knop, R. A., Aldering, G., Amanullah, R., et al. 2003, ApJ, 598, 102, doi: 10.1086/378560

Komatsu, E., Smith, K. M., Dunkley, J., et al. 2011, ApJS, 192, 18, doi: 10.1088/0067-0049/192/2/18

Kowalski, M., Rubin, D., Aldering, G., et al. 2008, ApJ, 686, 749, doi: 10.1086/589937

Krisciunas, K., Suntzeff, N. B., Phillips, M. M., et al. 2004, AJ, 128, 3034, doi: 10.1086/425629

Kunz, M., Bassett, B. A., & Hlozek, R. A. 2007, PhRvD, 75, 103508, doi: 10.1103/PhysRevD.75.103508

Lantz, B., Aldering, G., Antilogus, P., et al. 2004, in Proc. SPIE, Vol. 5249, Optical Design and Engineering, ed. L. Mazuray, P. J. Rogers, & R. Wartmann, 146–155

Leach, R. W., Beale, F. L., & Eriksen, J. E. 1998, in Proc. SPIE, Vol. 3355, Optical Astronomical Instrumentation, ed. S. D'Odorico, 512–519

Linder, E. V. 2003, Physical Review Letters, 90, 091301, doi: 10.1103/PhysRevLett.90.091301

Lochner, M., McEwen, J. D., Peiris, H. V., Lahav, O., & Winter, M. K. 2016, ApJS, 225, 31, doi: 10.3847/0067-0049/225/2/31

LSST Science Collaboration, Abell, P. A., Allison, J., et al. 2009, arXiv e-prints, arXiv:0912.0201. https://arxiv.org/abs/0912.0201

Malz, A., Hložek, R., Allam, Jr, T., et al. 2018, arXiv e-prints. https://arxiv.org/abs/1809.11145

Mandel, K. S., Narayan, G., & Kirshner, R. P. 2011, ApJ, 731, 120, doi: 10.1088/0004-637X/731/2/120

Matijevič, G., Prša, A., Orosz, J. A., et al. 2012, AJ, 143, 123, doi: 10.1088/0004-6256/143/5/123

McCarthy, J. K., Cohen, J. G., Butcher, B., et al. 1998, in Proc. SPIE, Vol. 3355, Optical Astronomical Instrumentation, ed. S. D'Odorico, 81–92

McKinney, W. 2010, in Proceedings of the 9th Python in Science Conference, ed. S. van der Walt & J. Millman, 51 – 56

Miyazaki, S., Komiyama, Y., Sekiguchi, M., et al. 2002, PASJ, 54, 833, doi: 10.1093/pasj/54.6.833

Nakaya, H., Miyatake, H., Uchida, T., et al. 2012, in Proc. SPIE, Vol. 8453, High Energy, Optical, and Infrared Detectors for Astronomy V, 84532R

Nugent, P., Phillips, M., Baron, E., Branch, D., & Hauschildt, P. 1995, ApJL, 455, L147, doi: 10.1086/309846

Oke, J. B., & Sandage, A. 1968, ApJ, 154, 21, doi: 10.1086/149737

Okumura, J. E., Ihara, Y., Doi, M., et al. 2014, PASJ, 66, 49, doi: 10.1093/pasj/psu024

Pasquet, J., Pasquet, J., Chaumont, M., & Fouchez, D. 2019, A&A, 627, A21, doi: 10.1051/0004-6361/201834473

Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, Journal of Machine Learning Research, 12, 2825

Perlmutter, S., Deusta, S., Gabi, S., et al. 1995, NASA STI/Recon Technical Report N, 96

Perlmutter, S., Aldering, G., Goldhaber, G., et al. 1999, ApJ, 517, 565, doi: 10.1086/307221

Phillips, M. M. 1993, ApJL, 413, L105, doi: 10.1086/186970

Pietrzyński, G., Graczyk, D., Gallenne, A., et al. 2019, Nature, 567, 200, doi: 10.1038/s41586-019-0999-4

Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2016, A&A, 594, A13, doi: 10.1051/0004-6361/201525830

Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2018, arXiv e-prints. https://arxiv.org/abs/1807.06209

PLAsTiCC Team, & PLAsTiCC Modelers. 2019, Unblinded Data for PLAsTiCC Classification Challenge, doi: 10.5281/zenodo.2539456. https://doi.org/10.5281/zenodo.2539456

Poznanski, D., Maoz, D., & Gal-Yam, A. 2007, AJ, 134, 1285, doi: 10.1086/520956

Rasmussen, C. E., & Williams, C. K. I. 2006, Gaussian Processes for Machine Learning (MIT Press)

Ratra, B., & Peebles, P. J. E. 1988, PhRvD, 37, 3406, doi: 10.1103/PhysRevD.37.3406

Reiss, R., Deiries, S., Lizon, J.-L., & Rupprecht, G. 2012, in Proc. SPIE, Vol. 8446, Ground-based and Airborne Instrumentation for Astronomy IV, 84462P

Revsbech, E. A., Trotta, R., & van Dyk, D. A. 2018, MNRAS, 473, 3969, doi: 10.1093/mnras/stx2570

Richards, J. W., Freeman, P. E., Lee, A. B., & Schafer, C. M. 2009, ApJ, 691, 32, doi: 10.1088/0004-637X/691/1/32

Richards, J. W., Homrighausen, D., Freeman, P. E., Schafer, C. M., & Poznanski, D. 2012, MNRAS, 419, 1121, doi: 10.1111/j.1365-2966.2011.19768.x

Riess, A. G., Casertano, S., Yuan, W., Macri, L. M., & Scolnic, D. 2019, ApJ, 876, 85, doi: 10.3847/1538-4357/ab1422

Riess, A. G., Press, W. H., & Kirshner, R. P. 1996, ApJ, 473, 88, doi: 10.1086/178129

Riess, A. G., Filippenko, A. V., Challis, P., et al. 1998, AJ, 116, 1009, doi: 10.1086/300499

Riess, A. G., Strolger, L.-G., Tonry, J., et al. 2004, ApJ, 607, 665, doi: 10.1086/383612

Riess, A. G., Macri, L. M., Hoffmann, S. L., et al. 2016, ApJ, 826, 56, doi: 10.3847/0004-637X/826/1/56

Rigault, M., Copin, Y., Aldering, G., et al. 2013, A&A, 560, A66, doi: 10.1051/0004-6361/201322104

Rigault, M., Aldering, G., Kowalski, M., et al. 2015, ApJ, 802, 20, doi: 10.1088/0004-637X/802/1/20

Rigault, M., Brinnel, V., Aldering, G., et al. 2018, arXiv e-prints. https://arxiv.org/abs/1806.03849

Robberto, M., & Hilbert, B. 2005, The behaviour of the WFC3 UVIS and IR Analog-to-Digital Converters, Technical Report, Space Telescope Science Institute, Baltimore, MD

Rockosi, C., Stover, R., Kibrick, R., et al. 2010, in Proc. SPIE, Vol. 7735, Ground-based and Airborne Instrumentation for Astronomy III, 77350R

Rodney, S. A., Riess, A. G., Strolger, L.-G., et al. 2014, AJ, 148, 13, doi: 10.1088/0004-6256/148/1/13

Roman, M., Hardin, D., Betoule, M., et al. 2018, A&A, 615, A68, doi: 10.1051/0004-6361/201731425

Rubin, D. 2019, arXiv e-prints. https://arxiv.org/abs/1903.10518

Rubin, D., Aldering, G., Barbary, K., et al. 2015, ApJ, 813, 137, doi: 10.1088/0004-637X/813/2/137

Sako, M., Bassett, B., Connolly, B., et al. 2011, ApJ, 738, 162, doi: 10.1088/0004-637X/738/2/162

Sasdelli, M., Ishida, E. E. O., Vilalta, R., et al. 2016, MNRAS, 461, 2044, doi: 10.1093/mnras/stw1228

Saunders, C., Aldering, G., Antilogus, P., et al. 2018, ApJ, 869, 167, doi: 10.3847/1538-4357/aaec7e

Scalzo, R. A., Aldering, G., Antilogus, P., et al. 2010, ApJ, 713, 1073, doi: 10.1088/0004-637X/713/2/1073

Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, ApJ, 500, 525, doi: 10.1086/305772

Scolnic, D. M., Jones, D. O., Rest, A., et al. 2018, ApJ, 859, 101, doi: 10.3847/1538-4357/aab9bb

Siebert, M. R., Foley, R. J., Jones, D. O., et al. 2019, MNRAS, 486, 5785, doi: 10.1093/mnras/stz1209

Silverman, J. M., Ganeshalingam, M., Li, W., & Filippenko, A. V. 2012, MNRAS, 425, 1889, doi: 10.1111/j.1365-2966.2012.21526.x

Stanishev, V., Goobar, A., Amanullah, R., et al. 2018, A&A, 615, A45, doi: 10.1051/0004-6361/201732357

Stein, M. L. 1999, Interpolation of spatial data, Springer Series in Statistics (New York: Springer-Verlag), xviii+247, doi: 10.1007/978-1-4612-1494-6. http://dx.doi.org/10.1007/978-1-4612-1494-6

Strolger, L.-G., Dahlen, T., Rodney, S. A., et al. 2015, ApJ, 813, 93, doi: 10.1088/0004-637X/813/2/93

Stubbs, C. W. 2014, Journal of Instrumentation, 9, C03032, doi: 10.1088/1748-0221/9/03/C03032

Sullivan, M., Conley, A., Howell, D. A., et al. 2010, MNRAS, 406, 782, doi: 10.1111/j.1365-2966.2010.16731.x

Suzuki, N., Rubin, D., Lidman, C., et al. 2012, ApJ, 746, 85, doi: 10.1088/0004-637X/746/1/85

Tenenbaum, J. B., Silva, V. d., & Langford, J. C. 2000, Science, 290, 2319, doi: 10.1126/science.290.5500.2319

The Dark Energy Survey Collaboration. 2005, arXiv e-prints, astro. https://arxiv.org/abs/astro-ph/0510346

The LSST Dark Energy Science Collaboration, Mandelbaum, R., Eifler, T., et al. 2018, arXiv e-prints. https://arxiv.org/abs/1809.01669

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

van der Walt, S., Colbert, S. C., & Varoquaux, G. 2011, Computing in Science and Engineering, 13, 22, doi: 10.1109/MCSE.2011.37

VanderPlas, J. T. 2018, The Astrophysical Journal Supplement Series, 236, 16, doi: 10.3847/1538-4365/aab766

Wilson, E. B. 1927, Journal of the American Statistical Association, 22, 209, doi: 10.1080/01621459.1927.10502953

Wong, K. C., Suyu, S. H., Chen, G. C.-F., et al. 2019, arXiv e-prints. https://arxiv.org/abs/1907.04869

Wood-Vasey, W. M., Friedman, A. S., Bloom, J. S., et al. 2008, ApJ, 689, 377, doi: 10.1086/592374

Wright, C. A., Kibrick, R. I., Alcott, B., et al. 2003, in Proc. SPIE, Vol. 4841, Instrument Design and Performance for Optical/Infrared Ground-based Telescopes, ed. M. Iye & A. F. M. Moorwood, 214–229

Zlatev, I., Wang, L., & Steinhardt, P. J. 1999, Physical Review Letters, 82, 896, doi: 10.1103/PhysRevLett.82.896