Building a Better Candle: The Calibration and Classification of Type Ia Supernovae in the Upcoming Legacy Survey of Space and Time

by

Daniel J. Perrefort

B.S. University of Connecticut, 2015

M.S. University of Pittsburgh, 2017

Submitted to the Graduate Faculty of

The Dietrich School of Arts and Sciences

in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2021

UNIVERSITY OF PITTSBURGH

DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Daniel J. Perrefort

It was defended on

April 27th 2021

and approved by

Michael Wood-Vasey, Department of Physics and Astronomy

Brian Batell, Department of Physics and Astronomy

Rachel Mandelbaum, Department of Physics, Carnegie Mellon University

Jeffrey Newman, Department of Physics and Astronomy

Eric Swanson, Department of Physics and Astronomy

Dissertation Director: Michael Wood-Vasey, Department of Physics and Astronomy

Copyright © by Daniel J. Perrefort
2021

Building a Better Candle:
The Calibration and Classification of Type Ia Supernovae in the Upcoming Legacy Survey of Space and Time

Daniel J. Perrefort, PhD

University of Pittsburgh, 2021

The use of Type Ia Supernovae (SNe Ia) as astronomical distance indicators relies on their intrinsically bright and homogeneous luminosities. By applying empirical relationships to remove any intrinsic, first-order variation in brightness between individual SNe Ia, the apparent brightness of these objects is used to determine a relative measure of distance. Upcoming surveys like the Vera C. Rubin Observatory's Legacy Survey of Space and Time (LSST) will observe on order 100,000 SNe Ia, representing an order of magnitude increase over previous surveys. LSST also promises to provide an impressive sub-percent level of precision between individual measurements. In this work, I present research targeted at two specific challenges faced by SN Ia research in the LSST era.

First, I classify SNe Ia that exhibit non-standard photometric behavior, such as lower luminosities and faster evolution of brightness over time. With LSST promising on order a million new SNe over a 10-year survey, spectroscopic classifications will be possible for only a small subset of observed targets. As such, photometric classification has become increasingly important in preparing for the next generation of astronomical surveys. Using observations from the Sloan Digital Sky Survey II (SDSS-II) SN Survey, I apply an empirically based classification technique targeted at identifying SN 1991bg-like SNe in photometric data sets and classify 16 previously unidentified 91bg-like SNe. Furthermore, I show that these SNe are preferentially found at a further physical distance from the center of their host galaxies and in host environments with an older average stellar age.

Second, I discuss the impact of atmospheric variability on the calibration of LSST observed SNe Ia. LSST will incorporate multiple calibration systems designed to estimate the atmospheric state and isolate systematic errors, including a GPS to quantify the time-dependent column density of precipitable water vapor (PWV) over the observatory. By

combining atmospheric models with near-real-time GPS measurements, I demonstrate that PWV absorption can be removed from observed spectra taken at Kitt Peak National Observatory (KPNO). Using this technique, I use GPS measurements taken at Cerro Tololo Inter-American Observatory (CTIO) to create a model for the PWV absorption over LSST and simulate an LSST-like SN dataset with realistic atmospheric variabilities.

Table of Contents

List of Tables

List of Figures

Preface

Pursuing a Ph.D. is never undertaken alone, and the completion of this thesis would not have been possible without the support of my family, friends, and amazing colleagues. I want to thank my thesis advisor Michael Wood-Vasey, whose continued guidance, patience, and mentorship have been key to my success in graduate school. To my committee members - Rachel Mandelbaum, Jeffery Newman, Brian Batell, and Eric Swanson - thank you for your guiding criticisms and encouragement. To the rest of the amazing astronomy faculty at Pitt, thank you for working so hard to create a supportive atmosphere where graduate students can thrive in their academic pursuits.

The scientific results included in this work, and the personal growth that they represent, would not be possible without the help of so many academic and professional colleagues. To Lluìs Galbany, Brett Andrews, and Yao-Yao Mao, thank you for broadening my scientific horizons and pushing me to achieve new things. To my coauthors on scientific publications, thank you for helping my work take form and improving upon it with your feedback and insights.

I am tremendously grateful to my fellow graduate students who are always willing to help answer questions and tackle challenging issues. To Kara Ponder, thank you for helping me get my graduate school sea-legs during my first few years. To Dritan Kodra and Sumit Sarbadhicary, thank you for the motivation during our many late-night work sessions. To Kevin Wilk and Brian Flores, thank you for helping me understand the complex theory that underlines my work. To Erin Sheridan, Ava Meyers, Paul Justice, and Brian Pardo, thank you for helping share the load of core courses and the graduate school challenges that followed. To my fellow research team members, Jared Hand, Shu Liu, and Troy Raen, thank you for your continuous feedback, support, and insight.

Most importantly, I would like to thank my family for always being incredibly supportive. Thank you to my parents, Patrick and Maria Perrefort, for never letting me accept the way things are over what they could be. To my sister, Alexa Perrefort, and best friend, Steven Vaz, thank you for always believing in my ability to accomplish my goals. To my wonderful

girlfriend, Meghan Cilento, thank you for providing the motivation and focus required to keep me on track. To the rest of my amazing family, thank you for always ensuring I have a solid foundation to stand on as I reach for the future.

Finally, thank you to the outstanding teachers who helped set me on this path and the friends who helped keep me there. In particular, thank you to my high school teachers Reine Issa for inspiring my love of mathematics and Silvia Nedelcov for my love of the physical sciences. Also, thank you to Anthony Pensiero, Liana Hotte, Filip Dul, and Maxwell Hotkowski for helping me decide at the very beginning of this journey that graduate school was, in fact, the right path for me to take.

1.0 Introduction

Estimating distances on scales beyond our galaxy is fundamental to understanding the evolution of our Universe. Among the various physical mechanisms used to measure intergalactic distances, the application of exploding stars known as Supernovae (SNe) has grown significantly over the past few decades. The underlying physics driving these explosions is still not well understood. However, observational study of Type Ia Supernovae (SNe Ia) – a subgroup of the SN population – has shown that this subgroup is incredibly uniform in peak luminosity. Coupled with the fact that they are also extremely bright, this allows SNe Ia to be used as well calibrated distance indicators known as *standard candles*.

The current generation of large-scale astronomical surveys have collectively observed ∼5,000 SNe Ia which have been used to measure the Universe with unprecedented precision. Upcoming surveys like the Vera C. Rubin Observatory's Legacy Survey of Space and Time (LSST) plan to take this even further by observing on order 100,000 SNe Ia (LSST Science Collaboration et al. 2009). Achieving the precision level promised by LSST requires a significant reduction in the systematic and statistical uncertainties commonly found in SN studies. In general, variation in calibrated SN luminosities is caused by a broad and diverse collection of well known systematics. In this work, we address a subset of the specific challenges that will be faced by the LSST SN science effort.

1.1 What Are Supernovae

The name "Supernova" is derived from their appearance as new ("nova") unusually bright ("super") stars in the night sky. The oldest of these events known to be observed by mankind were recorded nearly two thousand years ago during the first and second centuries. By the early 1990's, enough SNe had been observed that astronomers began to identify them as a formally distinct class of astronomical objects (Osterbrock 2001). Of these historical transients, seven have been categorized by modern astronomers (Clark et al. 1976) as being

SNe: SN 185, SN 393, SN 1006, SN 1054 (also known as the "Crab Nebula"), SN 1181, SN 1572 ("Tycho's SN"), and SN 1604 ("Kepler's SN")[1]. With the advent of charge coupled devices (CCD) in the 1980's, large scale photometric surveys grew in popularity, and the number of observed SNe increased significantly. Modern SN data sets now include thousands of observed SNe with upcoming astronomical surveys expected to provide another order of magnitude increase. By analyzing spectroscopic observations of SNe, two physically distinct kinds of SNe have been identified.

Core-Collapse SNe (ccSNe) occur when the gravitational pressure at the center of a massive star is no longer sufficient to sustain nuclear fusion. The high pressure found in a star's core drives the nuclear fusion of lighter elements such as Hydrogen into heavier elements such as Helium. As a star increases in age, its source of lighter elements is exhausted and the star is forced to fuse heavier and heavier elements together. For stars that are at least 8 times the mass of our sun ($M_\odot$), this process continues until the formation of Iron. Since Iron requires more energy to fuse than it generates, the star is no longer able to continue nuclear burning. Without the force of nuclear fusion to counteract the opposing force of gravity, the stellar atmosphere begins to collapse rapidly. Depending on the mass of the star, this collapse will either continue into a black hole ($M > 25M_\odot$), or the outer layers of the star will rebound off the Iron core and cause an explosion ($9M_\odot < M < 25M_\odot$) (Heger et al. 2003). It is the latter of these two scenarios that constitutes a ccSN.

Thermonuclear SNe (also referred to as SNe Ia) occur on significantly longer time scales. In this case, the progenitor star is much lighter, and the gravitational pressure is only sufficient to support nuclear fusion up to the creation of Carbon and Oxygen. Once nuclear fusion stops, the dead stellar core (known as a Carbon and Oxygen White Dwarf; CO WD) compresses under gravity into a degenerate plasma of unbound nuclei and electrons. After the CO WD stabilizes, further gravitational compression is opposed by the fundamental forces of quantum mechanics. According to the Pauli exclusion principle, no two half-spin particles (e.g., electrons) can occupy an identical quantum state. On macroscopic scales,

[1]SNe are named using the year of their discovery followed by a unique alphabetical identifier indicating the order of discovery. For example, SN 1991B is the second SN discovered in 2020 where as SN 1991bg is the 59^{th}. Standard convention is to drop the use of the trailing letters when only one SN was discovered in the given year and to use capital letters only when a single letter is required.

this manifests as an electron degeneracy pressure that resists the compression of degenerate matter in the core of the CO WD.

If left undisturbed, the resulting CO WD will continue the remainder of its existence in an inert state. However, if the mass of the WD increases beyond 1.4 $M_\odot$ (also known as the Chandrasekhar mass limit), the increased gravitational pressure at the center of the WD will reignite the fusion of carbon. The energy released by this fusion creates a feedback loop, and the resulting thermonuclear runaway causes an explosion. Since the mass limit of a CO WD is imposed by fundamental quantum mechanics, it is approximately constant for all SNe Ia. As a result, the brightnesses of SNe Ia explosions are intrinsically very uniform.

The mechanism by which a CO WD increases in mass is still not clear, but it is theorized to involve the participation of a second, gravitationally bound companion. One possible explanation is that the CO WD is in an orbit with a second, still burning star (referred to as the *single degenerate* scenario; Whelan et al. 1973). In this case, the CO WD and its companion star are physically close enough to permit the gravitational accretion of matter from the companion star onto the surface of the WD. This gradual buildup continues until the 1.4 $M_\odot$ limit is reached, and the WD explodes.

While the single degenerate theory does provide an explanation for the detonation of a CO WD progenitor, it is not without limitations. For example, thermal excitation of the gravitationally accreted matter from the companion star would eventually cause the newly added matter to combust on the WD's surface. The net result of this detonation is the removal of matter from the WD, causing an overall decrease in the WD's mass. This decrease in mass would prevent the WD from ever reaching the Chandrasekhar mass limit. Modifications to the theory have suggested that seismic pressure waves caused by the surface explosion are necessary for instigating the WD explosion (e.g., by Arnett 1969; Nomoto et al. 1976). However, observations of SNe Ia have yet to discover a companion star in the vicinity of the explosion, and the single degenerate scenario remains theoretical.

An alternative theory is that the companion object is another CO WD (the *double degenerate* scenario; Webbink 1984). In the double degenerate scenario, the orbiting pair of WD's gradually lose angular momentum through the emission of gravitational waves. As the two CO WD's spiral inwards, gravitational perturbations induced by the collapsing orbit break

apart the less massive of the two WD objects. Similar to the single degenerate scenario, the remnants of the less massive WD are gravitationally drawn to the surface of the larger WD, and the increased mass causes an explosion.

The double degenerate scenario is also not without its issues. Typical CO WD's have a mass between 0.4 and 0.8 $M_\odot$. If the combined mass of the two objects is too large, the gravity of the combined system will cause the colliding matter to collapse and form a neutron star. Alternatively, if the combined mass is too small, thermal excitation of the material will cause the degenerate matter to expand and eliminates any possibility for thermonuclear runaway (Pakmor et al. 2010).

Although the progenitor system for a SN Ia has yet to be identified, spectroscopic observations demonstrate a chemical makeup strongly indicative of a CO WD. Photometric observations also demonstrate a strong uniformity in the observed brightness of the explosions. This finding further supports the theory that SNe Ia are caused by a well-constrained, universal process and is key in their use as standard candles.

1.2 Using Supernovae as Distance Indicators

The use of SNe Ia as distance indicators relies on the fact that the observed brightness of an object is a function of its distance from the observer. As an example, consider two identical light sources placed at different distances from an observer in a vacuum (e.g., two light-bulbs). Since the intrinsic luminosity of these objects is the same, the light source that appears dimmer is known to be further away from the observer. Furthermore, if the apparent brightness of each object is known at some predefined distance, the difference between the known and observed brightness provides a relative distance measurement for each object. In astronomy, objects used to measure distances in this way are referred to as *standard candles*.

The observed brightness of astronomical objects is typically quoted in *magnitudes* (m) using the logarithm of the observed flux (f) relative to a known reference flux (f_0):

$$m = -2.5 \log_{10} \left(\frac{f}{f_0} \right). \tag{1.1}$$

The *absolute magnitude* (M) is defined as the brightness an object would have if it were observed from a fixed distance of 10 parsecs (32.6 light-years). Using the *distance modulus* (μ), the distance of an object can be expressed using the difference between the apparent and absolute bolometric magnitudes:

$$\mu = m - M. \tag{1.2}$$

Using SNe Ia as standard candles relies on them having intrinsically uniform absolute magnitudes at the time of peak brightness. SNe Ia are brightest between 15 and 20 days after the initial explosion, during which point they peak at M≈ -19.5 mag in the rest-frame B band. Instead of fading immediately, SNe Ia fade by approximately 3 mag in the month following peak brightness and by about 1 mag each month thereafter (Taubenberger 2017). This sustained, post-explosion luminosity is caused by the decay of ^{56}Ni synthesized during the explosion. Since ^{56}Ni is unstable, it decays into ^{56}Co and then later into ^{56}Fe. The gamma rays emitted by these transitions deposit energy into material ejected during the explosion, predominantly through Compton scattering. This scattering excites electrons in the ejecta, resulting in the emission of photons (Bersten et al. 2017).

SNe Ia appear in the night sky without warning, making it difficult to predict when the explosion will reach peak brightness. Instead, multiple observations of the SN are taken in several broad-band filters to measure the per-band brightness of the object over time. The apparent magnitude at peak brightness is then estimated by fitting the observed light-curve with templates derived from other well-known SNe Ia. The parameters fitted in this process depend on the chosen template. However, most templates include one or more parameters for the light-curve width (also called *stretch*; x_1) and the ratio of light observed between different bandpasses (also called *color*; c).

The distance (and therefore brightness) of astronomical objects is dependent on the makeup and evolutionary history of the Universe. Assuming the absolute luminosity of an object (L) is known, its *luminosity distance* is defined in terms of the observed flux (F) as:

$$d_{\mathrm{L}}^2 \equiv \frac{L}{4\pi F} \rightarrow \mu = 5\log_{10}\left(\frac{d_{\mathrm{L}}}{10\mathrm{pc}}\right), \tag{1.3}$$

where the right hand side of the implication follows from Equations 1.1 and 1.2. The luminosity distance can also be expressed in terms of the *comoving distance*, which represents the constant distance between two objects as the Universe expands

$$d_{\mathrm{L}}(z) = (1+z)\,\chi(z). \tag{1.4}$$

Assuming the Universe is both homogeneous and isotropic, the co-moving distance is related to the energy density of matter (Ω_{M}), the density of dark energy (Ω_{DE}), and the equation of state parameter (w) as

$$\chi(z) = \frac{c}{H_0}\int_0^z \frac{dz}{\left[\Omega_{\mathrm{M}}\,(1+z)^3 + \Omega_{\mathrm{DE}}\,(1+z)^{3(1+w)}\right]^{\frac{1}{2}}}. \tag{1.5}$$

Here we have assumed the Universe is flat (i.e., $\Omega_{\mathrm{M}} + \Omega_{\mathrm{DE}} = 1$). The leading prefix term describes the current expansion rate of the Universe (H_0) relative to the speed of light (c).

Note that for $w = -1$ the dark energy term in Equation 1.5 becomes

$$\Omega_{\mathrm{DE}}\,(1+z)^{3(1-1)} = \Omega_{\mathrm{DE}}\,(1+z)^0 = \Omega_{\mathrm{DE}}, \tag{1.6}$$

which is equivalent to a vacuum energy density. Assuming the dark energy equation of state is constant with time, estimates for the makeup of the Universe provide best fit parameters of $\Omega_{\mathrm{M}} = 0.321 \pm 0.018$, $\Omega_{\mathrm{DE}} = 0.679 \pm 0.018$, and $w = -0.978 \pm 0.059$ (DES Collaboration, Abbott, Alarcon, et al. 2018). Alternatively, w can be allowed to vary over time as

$$w = w_0 + w_a\left(1 - \frac{1}{1+z}\right). \tag{1.7}$$

In this case, current estimates for $w_0 = -0.885 \pm 0.114$ and $w_a = -0.387 \pm 0.43$ are consistent with a cosmological constant (i.e., w_0, w_a = -1, 0; ibid.). By comparison, the upcoming LSST will provide constraints on w_0 and w_a of ± 0.02 and ± 0.14 respectively (The LSST Dark Energy Science Collaboration et al. 2018).

In general, the specific value of the absolute magnitude of SNe Ia is unimportant since their distances are determined as a relative measurement (it is absorbed by the above-mentioned fit as a fitted nuisance parameter). However, minimizing any variation in the calibrated residual brightness (either for individual objects or for subgroups of a larger sample) is key for maximizing the precision of SN Ia distances. Typical variation in SN Ia

distances can be as high as 0.6 mag and correlate strongly with the parameters of stretch and color. SNe Ia with larger fitted stretch and bluer colors tend to be brighter, and the value of these fitted parameters can be included in the calculation of the distance modulus to reduce the intrinsic scatter found in observed SN Ia magnitudes (M. M. Phillips 1993; M. M. Phillips, Lira, et al. 1999; Riess, Press, et al. 1996; Tripp et al. 1999):

$$\mu = m - M + \alpha x_1 - \beta c. \tag{1.8}$$

where α and β are nuisance parameters absorbed as part of the overall cosmological fit. By calibrating for differences in stretch and color between SNe, existing surveys have reduced this scatter to as low 0.1 mag (for example, see Abbott et al. 2019; Betoule et al. 2014; D. M. Scolnic et al. 2018).

Following the discovery of the accelerating expansion of the Universe (Perlmutter et al. 1999; Riess, A. V. Filippenko, et al. 1998), several SN surveys were undertaken to increase the precision of cosmological measurements and to further constrain the evolutionary history of the Universe. A key LSST science goal is increasing the precision of SN Ia cosmology over previous surveys. LSST will be capable of performing observations with an impressive 5 mmag level of precision and a zero-point stability across the sky of 10 mmag (Željko Ivezić et al. 2019). However, even with these tight photometric constraints, accurately calibrating the peak brightnesses of SNe Ia is essential for maximizing the precision of cosmological fits and for identifying systematic differences in SN populations that may present as cosmological effects. In addition to precise photometry, comparing the brightnesses of SNe Ia at different redshifts requires an accurate representation for the wavelength-dependent transmission of light through each filter. This includes contributions from the telescope and camera hardware in addition to the absorption of light by the atmosphere.

1.3 The Diversity of Supernovae

Driven mostly by the advent of large scale astronomical surveys, the number of SNe Ia recorded in recent years has grown exponentially. Modern supernova studies include thou-

sands of individual SNe Ia; a significant increase over the few hundred being used thirty or so years ago (Perlmutter et al. 1999; Riess, A. V. Filippenko, et al. 1998). Among the many discoveries this new data has brought to light is the realization that SNe Ia are not as uniform as once believed. In fact, SNe Ia represent a diverse collection of objects with a wide range of intrinsic behaviors.

Over the course of several decades, a classification scheme was put in place for categorizing SNe based on their observed spectro-photometric properties. Beginning in 1941, SNe were first classified as Type I or Type II depending on the absence or presence of a Hydrogen absorption feature (Minkowski 1941). As shown in Figure 1, the category of Type I SNe was further refined based on the presence of strong Silicon (Type Ia) or strong Helium features (Ib; Elias et al. 1985; Wheeler and Levreault 1985). Any remaining Type I SNe were classified as Type Ib (Wheeler and Harkness 1990). By the 1990's, the number of observed targets had grown to the point that even SNe Ia, the most uniform type of SNe, began to show significant intrinsic diversity. At this point, several uncoordinated efforts were made to extend the existing classification structure.

Most definitions for SNe Ia subtypes rely on the Si II (λ6355) absorption feature uniquely found in SNe Ia. In Benetti et al. 2005, SNe were classified using the velocity of the λ6355 feature and how quickly the observed luminosity changed in the B-band (Δm_{15}(B)). This system introduced three Ia subtypes: the Faint group, the High Velocity Gradient group, and the Low Velocity Gradient group. A similar approach was later taken by Wang et al. 2009 where SNe were categorized as either High Velocity or Low Velocity events depending only on the velocity of their λ6355 feature. Using the relative strengths of the SiII λ5972 and λ6355 absorption features, Branch et al. 2006 classified SNe Ia into four groups: Core Normal, Cool, Broad Line, and Shallow Silicon. In addition to attempts at developing a quantitative classification system, there is also a strong historical precedent for identifying subsets of peculiar SNe based on their spectro-photometric similarity to a single representative object. Despite efforts to unify the multitude of available classification schemes into a single system (for example, Gal-Yam 2017), SN Ia classifications remain a confusing collection of overlapping definitions.

It is still unclear whether these subgroups are driven by underlying physical differences

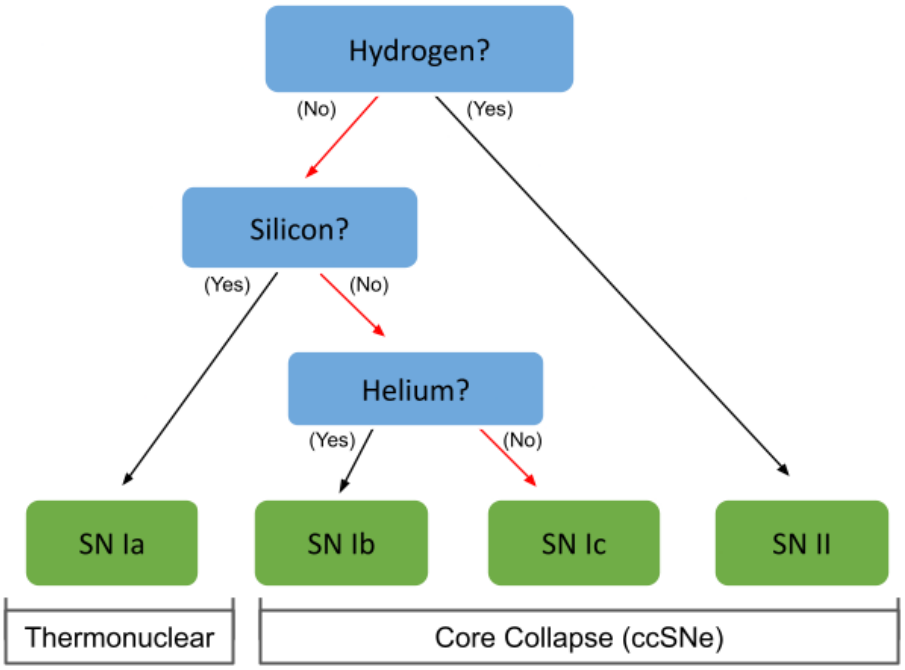


Figure 1: The classification of SNe begins with the identification of key absorption features in their spectra. The strength of these features indicates the relative abundance of different elements, which is used to subdivide SNe into four categories. SNe falling into the Ia category are entirely comprised of thermonuclear SN events. Similarly, the remaining three subtypes belong to core-collapse SNe (ccSNe).

in the explosion mechanism or if SNe Ia are drawn from a continuous distribution of properties. Leading theories indicate that SNe Ia tend to follow a continuous distribution primarily driven by variations in the explosion temperature (Nugent, M. Phillips, et al. 1995). Regardless, the application of existing classification schemes to increasingly large SN data sets remains the leading tool by which astronomers explore the SN diversity. For the remainder of this work, we restrict our discussion to only the most commonly identified Ia subtypes shown in Figure 1 of Taubenberger 2017 (herein Figure 2). Among these classifications, SNe Ia similar to SN 1991T and 1991bg are the most commonly identified peculiars in modern SN datasets.

The definitive classification of a SN is usually determined via spectroscopic observation, ideally performed at or near the time of peak brightness. Unfortunately, not all observatories are equipped for performing such observations. For example, the Rubin Observatory is equipped with an auxiliary spectroscopic telescope, but its use is primarily reserved for the quantification of atmospheric effects. The availability of spectroscopic classifications for LSST SNe will therefore depend on observations performed by coordinated follow-up programs performed at other observatories. Given the large number of transient objects LSST will observe, the availability of these observations will be extremely limited, and an alternative classification technique is needed to fully exploit the 100,000 SNe Ia that will be discovered.

Some subtypes can be identified photometrically due to their distinct differences from the SN Ia population. For example, SNe Ia similar to SN 1991bg are both redder and fainter than normal SNe Ia, most likely due to a lower production of ^{56}Ni in the explosion (Nugent, M. Phillips, et al. 1995). They also decline more rapidly after peak brightness, and their peak brightness in redder bands is delayed compared with the normal Ia population. Most notably, the secondary light-curve maximum normally found in redder bands is missing for 91bg-like events (see Figures 4 and 5). By identifying the unique photometric behavior of these and other peculiar SNe, they can be accurately identified in photometrically dominated astronomical surveys like LSST. The net result is an order of magnitude increase in observe peculiar SNe and a significant improvement in statistical power over existing data sets.

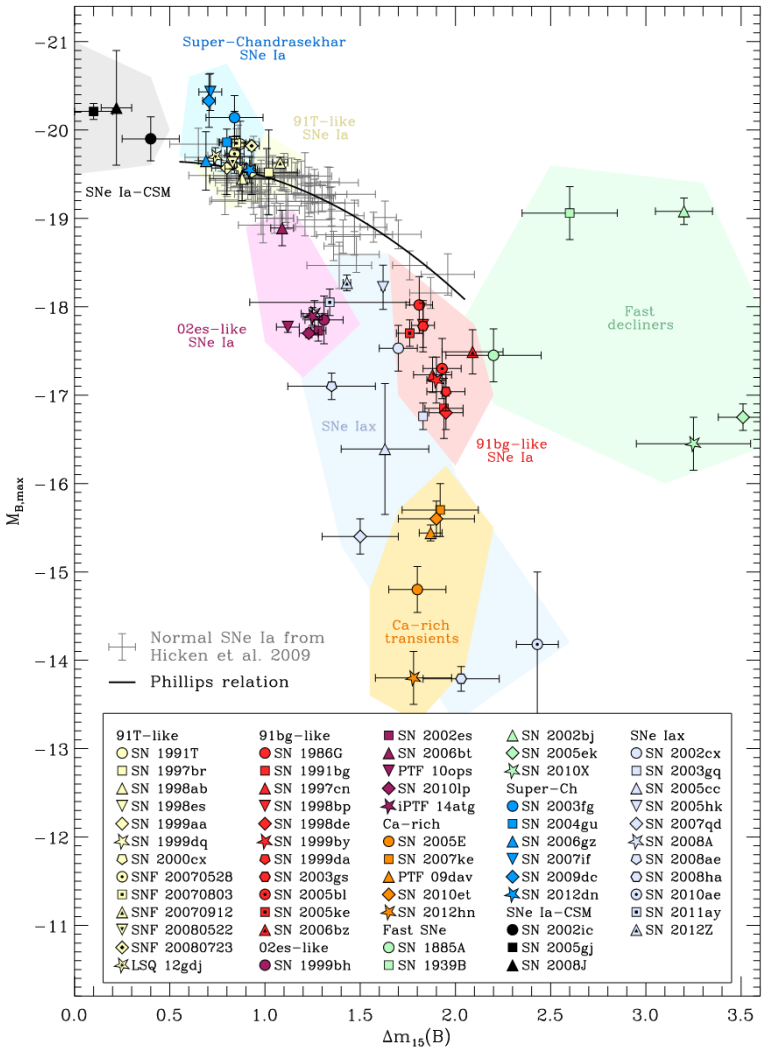


Figure 2: The absolute B-band magnitude of select SNe is shown against how quickly the luminosity of those objects changes over time. Absolute magnitudes are determined at the time of peak brightness and the change in observed magnitude is measured over the following 15 days. Commonly identified subgroups of SNe Ia are color coded based on existing classifications found in the literature (image credit: Taubenberger 2017).

1.4 The Calibration of Future Large Scale Surveys

The current generation of ground-based astronomical surveys already require tight constraints on photometric accuracy and precision. In addition to accounting for the intrinsic variability of SNe, these surveys require accurate calibration for external variabilities such as the atmosphere. Surveys like the Dark Energy Survey (DES; Dark Energy Survey Collaboration et al. 2016) achieve this through a Forward Global Calibration Method (FGCM; D. L. Burke et al. 2018). The FGCM approach combines data taken with auxiliary instrumentation with data from the primary imaging survey to evaluate models for the hardware and atmospheric effects. This process is typically performed for observations taken during ideal *photometric* observing conditions. The same stellar catalog is then used to adjust the calibration of exposures taken during poorer observing conditions (i.e., non-photometric conditions).

Approaches like FGCM use auxiliary telescopes to perform dedicated observations of younger, well behaved stars. A-type stars (which have surface temperatures of about 4,000 K) are well suited for this purpose since they have few intrinsic absorption features and are well described by existing models. By spectroscopically observing these targets, models for the atmospheric absorption can be used to fit the wavelength dependent atmospheric transmission. Alternatively, this approach can also be performed using narrow-band filters centered on key atmospheric features (Baker et al. 2017; T. Li et al. 2014).

Unfortunately, not all observatories are equipped with dedicated auxiliary hardware for determining the atmospheric transmission function. For surveys like the Panoramic Survey Telescope and Rapid Response System (Pan-STARRS; Magnier et al. 2020), broad-band observations of reference stars are used instead. By comparing the observed brightnesses with values from a pre-tabulated catalog, a polynomial is fit to determine the necessary photometric correction (ibid.). In general, the fitted terms implicitly correct for first-order effects introduced by hardware and atmospheric opacity. However, this process does not fully account for second-order effects induced by differences between the spectral energy distributions (SED) of the calibrated target and the reference catalog.

The absorption of light due to precipitable water vapor (PWV) dominates the atmo-

spheric transmission in the red optical and near-infrared (NIR) wavelengths ($\lambda > 5,500$ Å). Since redder stars emit much more of their light in these wavelengths than bluer stars, their brightness varies differently with changes in atmospheric conditions. This effect is even more significant when using a stellar catalog to calibrate observations of objects with significantly different SEDs, such as SNe. When not fully accounted for, this effect can introduce significant photometric errors of over 1% (Ž. Ivezić et al. 2007; T. S. Li et al. 2016). Fortunately, the complex absorption profile of PWV is well known and can be accurately approximated using atmospheric modeling software.

The standard convention is to report PWV measurements as the number of mm of liquid water that would be collected if it were all condensed along zenith into a tube of unit radius (although some works in the literature alternatively quote this value along the line of sight; $\mathrm{PWV_{los}}$). The localized PWV concentration can change significantly across the sky and vary by up to 10% per hour (D. Li et al. 2017). Correcting for atmospheric variability via atmospheric modeling therefore requires multiple measurements taken frequently over a given observing night. By measuring the time dependent PWV concentration, atmospheric models can be used to improve the calibration of ground based surveys (Wood-Vasey et al. 2020).

A growing technology in astronomy is the use of dual-band GPS receivers to measure the time dependent PWV concentration. The speed of light through a medium is dependent on both the properties of that medium and the wavelength of light. By measuring the relative delay of signals traveling through the atmosphere at two different frequencies, atmospheric models are used to estimate the corresponding PWV (Blake et al. 2011; Manandhar et al. 2018; Nahmias et al. 2004). In general, GPS measurements cannot constrain the transmission due to non-PWV components of the atmosphere. However, GPS measurements can be performed in near real-time and correlate strongly with spectroscopic results (D. Li et al. 2017).

The Rubin Observatory will be equipped with a dedicated auxiliary telescope in addition to a GPS measurement system. This allows for the improved calibration of observed targets while providing a means for cross-validating atmospheric measurements taken by each system. Frequent GPS measurements will also be used for assisting in the scheduling of LSST

observations. By monitoring and correcting for atmospheric conditions, the overall survey calibration can be improved.

1.5 Dissertation Overview

In Chapter 2, we explore the classification of *peculiar* SNe Ia that exhibit non-standard spectro-photometric behavior (this work is also available as a peer reviewed publication in Perrefort, Zhang, et al. 2020). Recent studies have identified an increasing number of peculiar Ia subtypes that are not suitable for use in cosmological studies. The identification of these SNe is an important step in removing them as outliers from cosmological datasets. Furthermore, the ability to accurately classify peculiar SNe is key to developing a physical explanation for the source of observed diversity. To address this, we develop an improved technique for identifying peculiar SNe similar to SN 1991bg. Using data taken by the Sloan Digital Sky Survey (SDSS), we demonstrate this approach on real world observations and discuss the properties of SNe identified as 91bg-like.

In Chapter 3, we discuss the calibration of time variable atmospheric effects (the first section of this chapter was originally published in Perrefort, Wood-Vasey, et al. 2019). The transmission of light through the atmosphere varies with both wavelength and time. Accurately estimating the atmospheric transmission is therefore crucial when calibrating ground based astronomical observations. The absorption of light by atmospheric water vapor is particularly problematic due to its complex absorption profile and highly variable nature. We begin by demonstrating the use of dual-band GPS technology to accurately estimate the atmospheric opacity due to water vapor. Using real-world GPS measurements, we generate a simulated collection of SNe as observed by LSST. From these simulations, we quantify the impact of PWV on cosmological deliverables and discuss methods for mitigating those effects.

In Chapter 4, we present a selection of customized software developed to support the above mentioned work. Documentation for these software packages is provided to ensure the reproduciblity of scientific results and to support future research efforts. All software

documented in this work is open-source and free for public use. Finally, we present our concluding remarks in Chapter 5.

2.0 Supernova Classification

Following their fundamental role in the discovery of the accelerating expansion of the universe (Perlmutter et al. 1999; Riess, A. V. Filippenko, et al. 1998), SNe Ia have been used to determine cosmological parameters with an increasing level of accuracy and precision (Betoule et al. 2014; DES Collaboration, Abbott, Allam, et al. 2018; Jones et al. 2018; Rest et al. 2014; D. M. Scolnic et al. 2018; D. Scolnic et al. 2014). The use of SNe Ia as cosmological probes relies on the fact that SN Ia luminosities at the time of maximum brightness are not only bright but also have low intrinsic scatter. This scatter can be reduced even further by calibrating their intrinsic peak luminosity with light-curve width (M. M. Phillips 1993; M. M. Phillips, Lira, et al. 1999) and optical color (Riess, Press, et al. 1996; Tripp et al. 1999). However, even after applying these corrections SNe Ia remain a heterogeneous collection of objects spanning a diverse collection of subtypes (Taubenberger 2017).

Early attempts at classifying peculiar SNe Ia quickly identified categories of overluminous, SN 1991T-like objects (Alexei V. Filippenko, Richmond, Matheson, et al. 1992; M. M. Phillips, Wells, et al. 1992) and subluminous, fast-declining objects like SN 1991bg (Alexei V. Filippenko, Richmond, Branch, et al. 1992; Leibundgut et al. 1993; Turatto et al. 1996). More recent works have introduced additional classifications based on spectroscopic properties such as SN 2002es-like SNe (Ganeshalingam, W. Li, Alexei V. Filippenko, Silverman, et al. 2012), which are subluminous but slow declining, super-Chandrasekhar-mass candidates (Howell et al. 2006), 2002cx-like SNe, also known as SNe Iax (Jha et al. 2006; W. Li, Alexei V. Filippenko, et al. 2003; Meng et al. 2018), and others with fewer observed targets. However, the presence of SNe like SN 1991T and SN 1991bg make up the predominant population of observed peculiar SNe.

There is some disagreement in the literature when it comes to the rate of 91bg-like events. Recent rate estimates using SNe observed by the Lick Observatory Supernova Search (LOSS) range from 11 to 15% of the SN Ia population (Ganeshalingam, W. Li, Alexei V. Filippenko, C. Anderson, et al. 2010; W. Li, Leaman, et al. 2011). Alternatively, González-Gaitán et al. 2011 and Silverman et al. 2012 estimate that 91bg-like SNe make up a more modest 6 to 9%

using various, low-redshift data sets. However, González-Gaitán et al. 2011 notes that their estimates increase dramatically with the inclusion of transitional SN 1986G-like SNe, which have luminosities that lie in the intermediate range between normal and SN 1991bg-like.

The upcoming Vera C. Rubin Observatory will conduct the Legacy Survey of Space and Time[1] (LSST; LSST Science Collaboration et al. 2009) and observe hundreds of thousands of new SNe over a ten-year survey, promising a dramatic increase in the number of observed peculiar SNe. However, the availability of spectroscopic follow-up observations and, as a result, spectroscopically determined classifications, will be heavily limited. The ability to provide accurate, photometric classifications will thus be increasingly important in the coming years for maximizing the science that can be done with the Rubin Observatory.

One approach to this challenge is the development of machine learning classifiers designed to reproduce existing classification schemes (e.g., Dai et al. 2018; Karpenka et al. 2013; Lochner, McEwen, et al. 2016; Möller et al. 2016; Muthukrishna et al. 2019; Pasquet, Johanna et al. 2019; Richards et al. 2012; Sasdelli et al. 2016; Varughese et al. 2015). Although machine learning classifiers benefit from the ability to scale to large data sets, they don't reveal the underlying physics that lead to a classification. The ability of a machine learning classifier to identify unexpected, peculiar objects is also extremely sensitive to the quality and diversity of the initial training sample.

An alternative is to classify SNe based on their light-curve properties. Empirically based classification schemes are not only transparent in how they work, but can simultaneously provide physically motivated values such as light-curve color, standardized peak luminosities, and decline rates. Furthermore, many SN analyses already employ the use of a light-curve fitter, making it easy to incorporate classification into existing analysis procedures.

In González-Gaitán et al. 2014, G14 hereafter a photometric identification technique was introduced for discriminating SN 1991bg-like objects in photometric samples. Using several low-redshift samples from the literature, G14 demonstrated that this method is not only capable of identifying dim, fast-declining SNe, but can potentially identify other peculiar transients such as SNe Iax-like, SN 2006bt-like, and super-Chandrasekhar SNe Ia. We apply

[1]Previously known as the Large Synoptic Survey Telescope.
`https://www.lsst.org/news/vro-press-release`

here the same classification technique to a larger target sample and compare results against spectroscopically determined subtypes.

We make two significant changes to the original approach of G14. The first is the use of a newer model for 91bg-like SNe that has been extended further into the near infra-red (NIR) and ultraviolet (UV). By using this extended model, we are able to apply the classification to a larger, higher-redshift sample of SNe Ia. Secondly, we consider multiple implementations of the technique and discuss potential biases that may arise.

The layout of this chapter is as follows: In Section 2.1 we discuss the supernova sample considered by this paper. In Section 2.2 we present our chosen classification method, including a detailed outline of our procedure in Section 2.2.1 and the models employed by our analysis in Section 2.2.3. Our photometric classifications are then presented in Section 2.3.

2.1 Classification Data

We focus on the photometric classification of SNe observed by the Sloan Digital Sky Survey II (SDSS II) Supernova Survey. Since spectroscopic followup is limited for the SDSS SN sample, we additionally consider spectrophotometric observations taken by the Carnegie Supernova Project I (CSP-I), allowing us to better estimate the performance of our classification technique.

2.1.1 SDSS-II

The SDSS II SN survey (Sako, Bassett, A. C. Becker, et al. 2018, S18 hereafter) was performed using the 2.5-m Sloan Foundation Telescope (J. E. Gunn et al. 2006; York et al. 2000) at Apache Point Observatory (APO). The SDSS II SN survey ran over three observing seasons from 2005 to 2007, covering a 300 square-degree stripe of sky along the celestial equator in the Southern Galactic hemisphere in the *ugriz* bands (Doi et al. 2010; Fukugita et al. 1996). All SDSS SNe are referred to in this work using their associated Candidate Identifier (CID) published in S18.

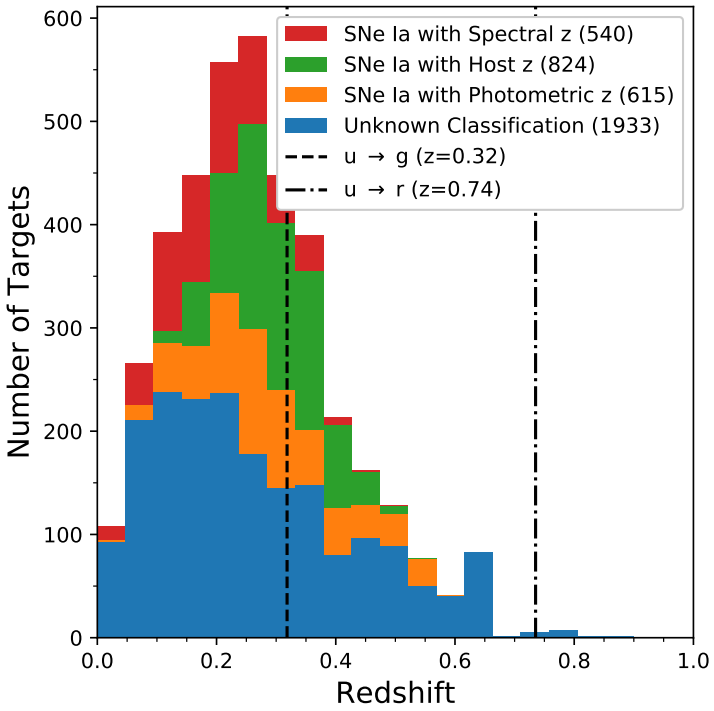


Figure 3: Stacked redshift distribution for SDSS objects considered in this work, grouped by their classification in S18. Spectroscopic classifications are available predominately for lower redshift targets, while higher redshift targets have fewer photometric or spectroscopic classifications. Reference lines mark redshift values at which rest-frame u-band observations are equivalent to the observer frame g and r band as determined by effective wavelength. Not included are 76 objects with unknown classifications for which a redshift value could not be determined.

The SDSS SN data release provides light-curve data for 10,258 variable and transient sources. This includes 540 objects spectroscopically classified as SN Ia, 64 SN II, 3 superluminous SN (SLSN), 22 objects classified as either SN Ib or Ic, and 4,131 sources that are either variable or AGNs. Additionally, there are 2,009 targets with light curves that were deemed too sparse or noisy to provide a classification. Shown in Figure 3, this combined sample spans redshifts out to $z \leq 0.9$.

Initial classification of the SDSS spectra were performed in Zheng et al. 2008 using the `rvsao.xcsao` cross-correlation package of IRAF (Tody 1993). Additional photometric classifications using the SDSS photometry were performed by Sako, Bassett, Connolly, et al. 2011 using an extension of the Photometric SN IDentification (PSNID) software (Sako, Bassett, A. Becker, et al. 2008). Further delimitations to each classification were added manually by Sako, Bassett, Connolly, et al. 2011 depending on whether the classification was made photometrically (denoted with a prefix p) spectroscopically (no prefix) or using a host galaxy redshift (prefix z). These classifications are referred to throughout this work as a reference for those familiar with the SDSS data set or other uses of PSNID.

No systematic search or subtyping effort for peculiar SNe was performed beyond the assignment of basic SN types (Sako 2020). However, a small selection of targets was manually flagged based on their spectroscopic or photometric properties. These SNe are listed in Table 1 and include 4 SNe possibly similar to SN 1991bg, 1 SN possibly similar to SN 2000cx, 2 SNe possibly similar to SN 2002ci, and 3 SNe possibly similar to SN 2002cx.

Photometric zero points for SDSS-II were determined using stars from the Ivezić catalog (Ž. Ivezić et al. 2007). The position, band-specific flux, and host-galaxy intensity for each target were then fitted for using Scene Model Photometry (SMP; Holtzman et al. 2008). Betoule, M. et al. 2013 provides the most updated calibration of these data using a position-dependent correction, which were applied in S18 and are also used in this work.

2.1.2 CSP-I

CSP-I was a five-year survey in the optical and NIR run at Las Campanas Observatory (LCO) from 2004 through 2009. Optical observations were taken in the $ugriBV$ bandpasses

(Stritzinger et al. 2011) using the SITe3 and Tek5 CCD cameras on the Swope 1 m and du Pont 2.5 m telescopes respectively. NIR imaging was performed in the *YJH* bands (Contreras et al. 2010) using the Wide-Field IR Camera (WIRC) (Persson et al. 2002) on the du Pont 2.5 m telescope and later RetroCam on the Swope 1 m telescope.

CSP-I includes spectrophotometric observations of 134 SNe Ia spanning $z < 0.085$ (M. Hamuy et al. 2006; Krisciunas et al. 2017). Out of these targets, 96 were classified as normal SNe Ia, 13 as being like SN 1991bg, 5 like SN 1991T, 5 like SN 2002cx, and 4 were unclassified.

In Mosher et al. 2012 overlapping observations between CSP-I and SDSS-II were compared for 9 spectroscopically confirmed SNe Ia, five of which were classified as peculiar. Photometric observations in the *gri* bands were found to agree within a 1% level in flux with a typical epoch-to-epoch scatter no greater than 0.05 magnitudes. The u band scatter was slightly higher at 0.077 magnitudes, but flux values were still consistent within 1%. Taking into account the small sample size along with systematic uncertainties in the analysis, it was estimated that offsets in observer frame u were conservatively within 0.04 magnitudes.

2.2 Classification Method

The classification of 91bg-like SNe relies on them having distinct spectrophotometric differences from the normal SN Ia population. Most notably, the presence of strong Ti lines in their spectra indicates that 91bg-like SNe tend to be cooler than normal SNe Ia, which can generally be attributed to a lower yield of ^{56}Ni synthesized in the explosion (Nugent, M. Phillips, et al. 1995). In photometric terms, this means 91bg-like events have light-curves that are both fainter at maximum than normal SNe Ia ($M_B(91bg) \sim -18$) and decline more rapidly after peak brightness. The latter of these effects is typically well demonstrated by the light-curve parameters stretch or Δm_{15}, where typical values are found to be $\Delta m_{15}(B) \gtrsim 1.7$ (L. Galbany et al. 2019).

The cooler explosion temperature of 91bg-like SNe also has an impact on the evolution of light-curve color. The lower temperatures allow for the recombination of Fe III to Fe II to happen sooner, which results in redder colors at maximum. This causes the epoch

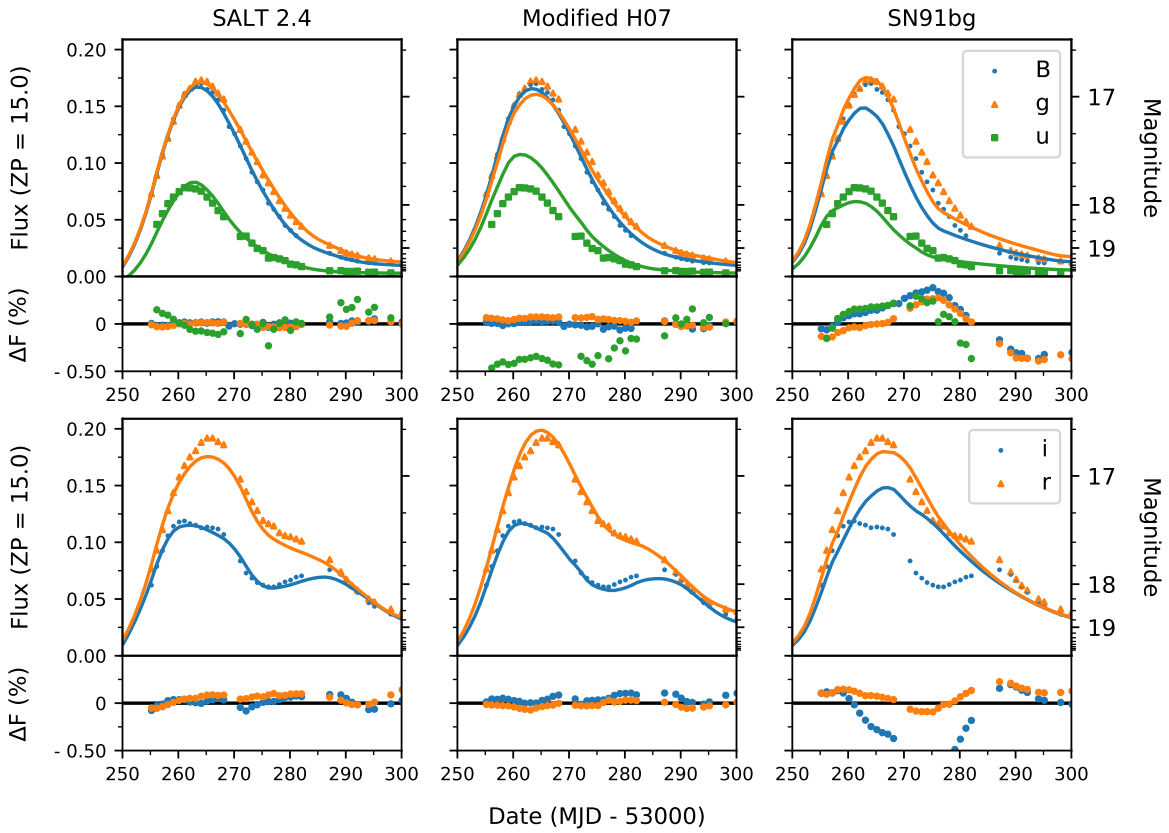


Figure 4: CSP observations of the spectroscopically normal SN 2004ef are fit separately in the rest-frame blue (top) and red (bottom) bands using the SALT 2.4 (left), H07 (middle), and SN 1991bg-like (right) models. We note in the blue bands that the fitted 91bg model is narrower and fainter at peak than the observations. We also note that the H07 model, which is trained using a more heterogeneous set of template spectra than SALT 2.4, overestimates the u-band. In the red bands we see that the morphology of each model plays a greater role. In particular, the lack of a secondary maximum in the 91bg-like model lends greater importance to late time observations (phase $\gtrsim 30$ days) in constraining the width of the modeled light-curve.

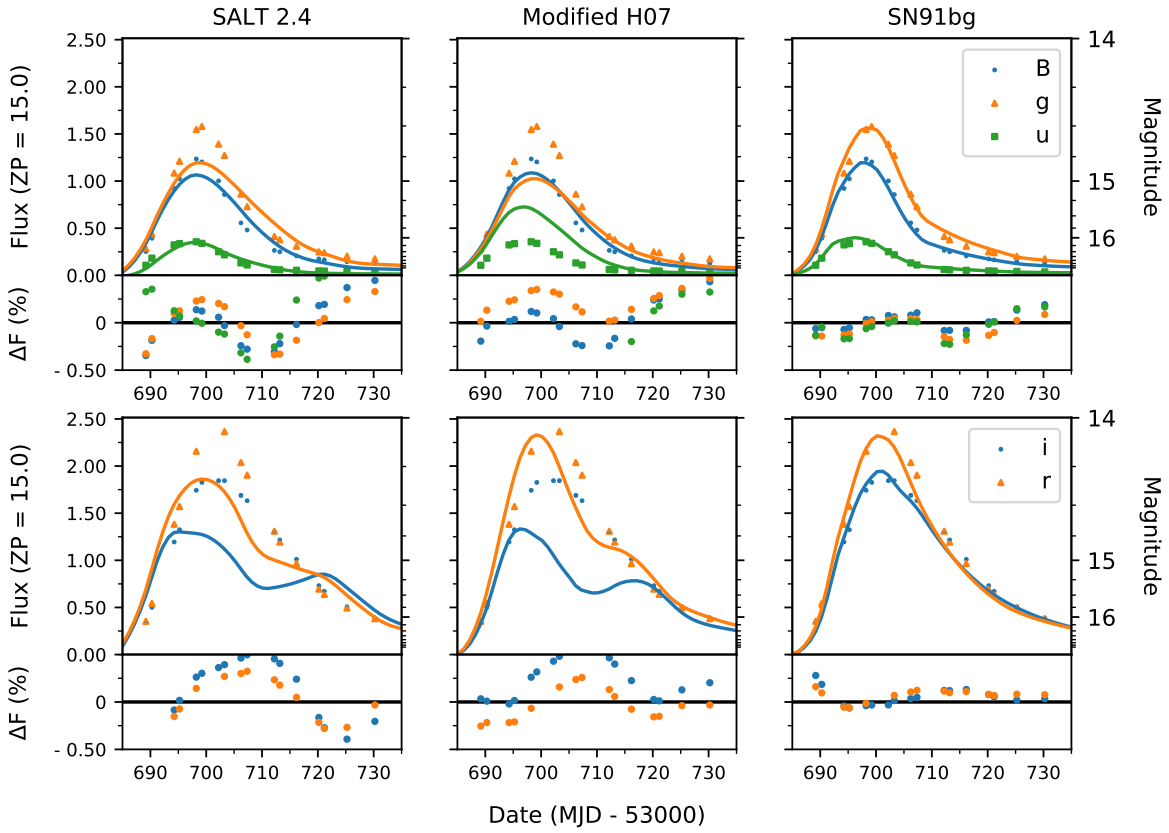


Figure 5: CSP observations of the 91bg-like SN 2005ke are fit separately in the rest-frame blue (top) and red (bottom) bands using the SALT 2.4 (left), H07 (middle), and 91bg-like (right) models. We note that the SALT 2.4 and H07 fits — which represent normal SNe Ia — are significantly bluer when fit in the *ugB* bands than the SN 1991bg model which is a significantly better fit to the data.

of peak brightness for redder bands to be delayed when compared with the normal SNe Ia population. SN 1991bg-like explosions also lack the secondary maximum seen in the redder bands in normal SNe Ia. The delay in the peak brightnesses combined with the lack of a shoulder or secondary maxima in redder bands make 91bg-like SNe identifiable by their photometric properties.

In principle, one might attempt to distinguish 91bg-like events by directly looking for these photometric properties — specifically the lack of a secondary maximum. However, this requires an observational cadence with even and complete sampling from maximum light through the secondary maximum. These problems can be alleviated by performing an overall fit to the data and selecting targets based on a χ^2 value or a set of model parameters, but this raises additional challenges since there may be some normal SNe with intrinsically lower stretch or redder colors. For example, highly reddened, normal SNe with a low stretch would look potentially similar to 91bg-like SNe in terms of χ^2. We here instead analyze SDSS-II SN data using the classification method of G14.

2.2.1 Classification Procedure

We employ our classification method following an adaptation of G14. To start, photometric observations for each target are split into collections of rest-frame red and blue bandpasses as defined by the rest-frame effective wavelength of each band $\lambda_{\rm z,eff}$. When doing so, we define *blue* bandpasses as having $\lambda_{\rm z,eff} < 5,500$ Å and *red* bandpasses as having $\lambda_{\rm z,eff} > 5,500$ Å. The use of 5,500 Å is chosen to separate the rest-frame *ug* bands from the rest-frame *riz* and thus puts the secondary maximum in the red bandpasses.

We separately fit each of the blue and red data using two light-curve models: one representing normal SNe Ia and one representing 91bg-like objects. We calculate the χ^2 for each combination based on the modeled flux F for a set of parameters $\bar{p}$, the observed flux f, and the of degrees of freedom, $d = N - \text{len}(\bar{p})$.

$$\chi^2(\bar{p}) = \frac{1}{d}\sum_{i=1}^{N}\frac{(F(\bar{p}) - f_i)^2}{\sigma_i^2}\,. \tag{2.1}$$

Using the resulting χ^2 values, we classify targets based on their position in the following phase space:

$$x \equiv \chi^2_{\rm blue}({\rm Ia}) - \chi^2_{\rm blue}({\rm 91bg})\,, \tag{2.2}$$

$$y \equiv \chi^2_{\rm red}({\rm Ia}) \ - \chi^2_{\rm red}({\rm 91bg}). \tag{2.3}$$

By construction of the above coordinates, we expect 91bg-like SNe to fall in the upper right (first) quadrant while normal SNe Ia should fall in the lower left (third) quadrant of the x, y plane.

Our normal SNe Ia and 91bg-like models are imperfect, particularly in the model variances. Thus the classification quadrants may not be best separated by intersecting lines at $(0, 0)$. An alternative origin may instead provide a higher level of purity when classifying peculiar SNe Ia. Following G14, the quadrant boundaries are determined by using spectroscopically classified targets to maximize the figure of merit (FOM) parameter

$$\mathrm{FOM} = \frac{N_{\rm true}}{N_{\rm total}} \frac{N_{\rm true}}{N_{\rm true} + N_{\rm false}}\,, \tag{2.4}$$

where $N_{\rm total}$ represents the total number of objects with a given type (e.g., 91bg-like objects), $N_{\rm true}$ is the number of objects correctly classified as a given type, and $N_{\rm false}$ is the number of objects falsely classified as a given type.

2.2.2 Fitting Procedure

We choose to use the SNCosmo Python package to handle light-curve fits since it allows us to easily implement, modify, and apply a variety of template-based models (Barbary et al. 2016). Unless otherwise stated, we use the `iminuit` minimization routine (`sncosmo.fit_lc`) to determine best-fit parameters. By default, the SNCosmo package fits each model using a single set of global, model-dependent parameters. However, we note that this behavior is significantly different from the original implementation of the classification technique in G14.

In G14, SNe were classified using the SiFTO light-curve fitter, which is an empirical fitter that uses magnitude, light curve shape (i.e., stretch), and color to fit a light-curve (Conley

et al. 2008). It is important to note that SiFTO uses band-specific flux normalizations as opposed to a set of global, light-curve specific parameters. Instead of relying on a dedicated color term, SiFTO allows the scale factor of the template to vary independently for each band. Similarly, the shape of the SiFTO model is described by a single stretch parameter that is applied differently in each observed filter as a function of effective wavelength.

The choice of how parameters are varied across bands can potentially have a major impact on the resulting classification. In principle, we expect fits performed to red and blue data as collective sets to be more constrained by intrinsic color. However, by allowing parameters to vary across bands, the overall morphology of the light-curve is allowed to have a more significant impact. The drawback to this approach is a potentially higher sensitivity to the cadence of the observations. To understand the impact of this choice, we implement fitting routines for both approaches and compare the results.

Our resulting fitting procedure is as follows:

1. For each target, the Milky Way extinction is determined using the Schlegel et al. 1998 dust map and the Fitzpatrick 1999 extinction law. This value is never varied in any fit.
2. To determine a fiducial set of fit parameters, each light-curve is fit using both the H07 and SN91bg models and all available data points. At this step, all model parameters are varied except the redshift, which is only varied if it has not been determined with a spectroscopic observation.
3. Using the redshift value determined in the previous step, the observed bandpasses are separated into the rest-frame blue ($\lambda_{\rm z,eff} < 5,500$ Å) and red ($\lambda_{\rm z,eff} > 5,500$ Å) bandpasses.
4. The red and blue data sets are fit using both models. So that we can investigate the resulting effect, we perform fits twice: once allowing fit parameters to vary independently across bands, and again using a single set of parameters for each red / blue data sets. At this stage, the redshift and time of B-band maximum remain fixed to the value determined in step 2.

2.2.3 Supernova Models

Although the classification scheme described above only requires two SN models, we consider three models in our analysis: two used for target classification and a third as a baseline reference for comparison with existing results in the literature. When discussing the general properties of an observed light-curve, we default to the fitted parameters of the SALT 2.4 model (J. Guy et al. 2007). The remaining two models are chosen to closely mimic those used in G14.

For normal SNe, we use the spectral time series template from E. Y. Hsiao et al. 2007, H07 hereafter. This template was intentionally constructed to incorporate a large and heterogeneous sample of observed spectra and is the same model used by G14. Although the H07 model is already built into SNCosmo, the default model does not include a stretch-like parameter. This is problematic for two reasons. Firstly, it lends the other models a potentially unfair advantage in their flexibility to fit a given light-curve. Secondly, it limits our ability to investigate the impacts of fitting each bandpass independently versus as red / blue sets. To address these issues we add a stretch parameter $-0.5 < x_1 < 0.5$ to the preexisting parameters of amplitude A, redshift z, and time of B-band maximum t_0.

Here and throughout this paper, we have chosen the variable x_1 to represent a stretch-like parameter similar in significance to that of SALT 2.4. However, we note that the full meaning of this parameter is uniquely dependent on the model being discussed. For the H07 model, we implement the x_1 parameter such that the flux F is determined from the template $f_{\mathrm{H}}(t, \lambda)$ as:

$$F_{\mathrm{H}}(t, \lambda, z, A, t_0, x_1) = A f_{\mathrm{H}} \left(\frac{t - t_0}{x_1 + 1}, \frac{\lambda}{(1 + z)} \right) . \tag{2.5}$$

As a model for 91bg-like SNe, we employ the same spectroscopic template used in the PLAsTiCC challenge (The PLAsTiCC team et al. 2018), which includes the same parameters as the H07 model in addition to a color parameter c. This template is based on the 91bg model from Nugent, Kim, et al. 2002 but is extended further into the NIR and ultraviolet (UV) using synthetic spectra from Hachinger et al. 2008 and light-curves observed by Swift (Brown et al. 2009). This allows the model to cover a broader wavelength range from 1,000

to 12,000 Å and permits the fitting of targets at higher redshift values. The parameters of the 91bg model span stretch values from $0.65 \leq x_1 \leq 1.25$ and color values from $0 \leq c \leq 1.0$.

Using observations from CSP (Krisciunas et al. 2017), we compare fits of all three models to a spectroscopically normal and 91bg-like SN in Figures 4 and 5 respectively. We note in the blue bands that the most significant difference between the fitted models is their color, although the decline rate does play a secondary effect. Conversely, the biggest difference between the normal and 91bg-like models in the red bands is the overall morphology (i.e., the existence or lack of a secondary maximum). In particular, we see that late time observations ($\gtrsim$ 30 days past maximum) play an important role in constraining the stretch of the SN 1991bg model when fitting normal SNe Ia light-curves. We also note that there is a key difference in the epoch of the first maximum, particularly in the blue bands, which plays an important role in determining the resulting chi-square.

2.3 Photometric Classification Results

Table 2 presents the fitted parameters for each SN using both the SN 1991bg and modified H07 models. Fits are performed for all targets not classified by S18 as being non-transients (i.e., as either Variable or AGN like objects). To ensure fits are well constrained by the data, we disregard any targets not having observations with a signal to noise ratio SNR ≥ 5 in two or more bandpasses. Furthermore, we require at least one of these observations to fall between -15 and 0 days of the fitted SALT 2.4 B-band maximum, and the other between 0 and 25 days. Targets not matching these criteria are dropped from our sample, leaving a total of 3,882 remaining targets.

2.3.1 FOM Optimization

In order to optimize Equation 2.4, we require a set of spectroscopically classified SNe Ia. Spectroscopic classification of SDSS targets was attempted for this work following the pre-

scription of Silverman et al. 2012 using the SN IDentification software[2] (SNID; Stéphane Blondin et al. 2007) (see Appendix A for details). Although excellent agreement was found with S18 when assigning SN types, subtyping results proved to be unreliable. Overall, this was attributed to either poor wavelength coverage, strong host galaxy contamination, and / or a low SNR in the observed spectra.

As an alternative, we supplement our data set with spectroscopically classified objects from CSP (Folatelli et al. 2013). By virtue of spanning a lower redshift range, photometric observations taken by CSP have, on average, a higher SNR than targets observed by SDSS. This is problematic since our classification scheme relies on a coordinate system that is based on chi-squared values and thus scales inversely with the average SNR (see Equation 2.2). Our solution is to rescale the classification coordinates of CSP targets to more closely resemble SDSS using the median SNR in the rest-frame blue ($\mathrm{SNR_B}$) and red bands ($\mathrm{SNR_R}$) as follows:

$$x'_{\mathrm{CSP}} \equiv \frac{\mathrm{SNR}_{\mathrm{B},SDSS}}{\mathrm{SNR}_{\mathrm{B},CSP}}\, x_{\mathrm{CSP}}, \tag{2.6}$$

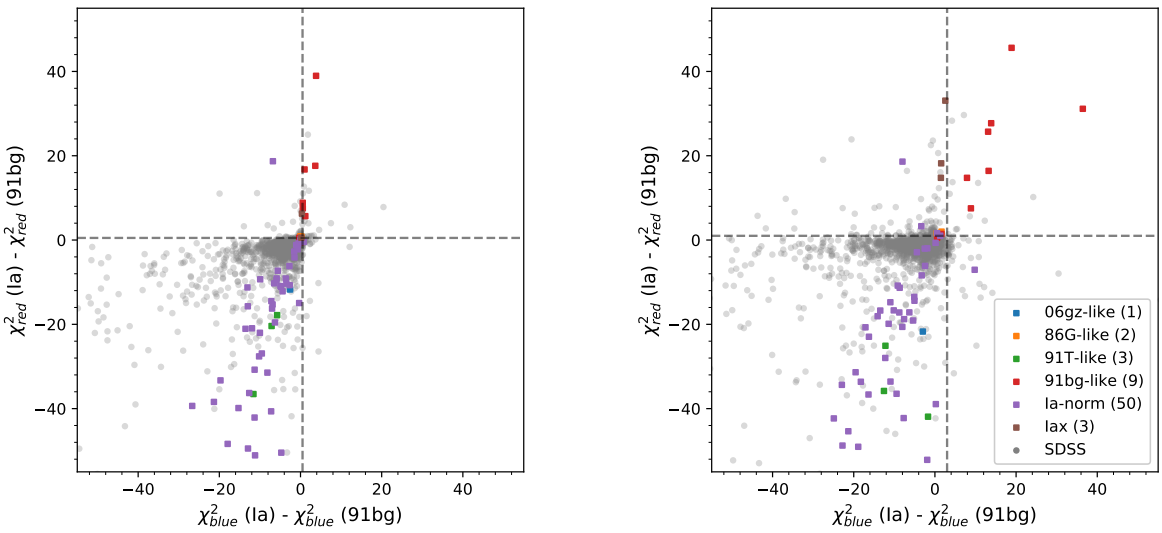
$$y'_{\mathrm{CSP}} \equiv \frac{\mathrm{SNR}_{\mathrm{R},SDSS}}{\mathrm{SNR}_{\mathrm{R},CSP}}\, y_{\mathrm{CSP}}. \tag{2.7}$$

Using the above definitions, we determine scale factors of 0.023 and 0.030 for the x and y coordinates respectively.

Following the prescription of G14, the rescaled CSP coordinates are used to optimize the value of Equation 2.4 using a bootstrap technique (Efron 1979; Felsenstein 1985). In total, we draw 100 random samples, each containing 75% of the available classification coordinates, and recalculate the FOM each time. So that the FOM can be evaluated, each sample is guaranteed to contain at least one 91bg-like SN. Shown in Figure 6, the final boundaries are chosen using the average over all randomly realized samples. This results in a peak FOM value of 0.78 when fitting observations as red / blue sets and 0.75 when fitting bandpasses independently.

Shown in Figure 7, we note there exists a significant degeneracy in the maximized FOM value. This makes it possible to shift the classification boundaries in such a way that the classification of some targets changes despite the FOM remaining constant. To address this,

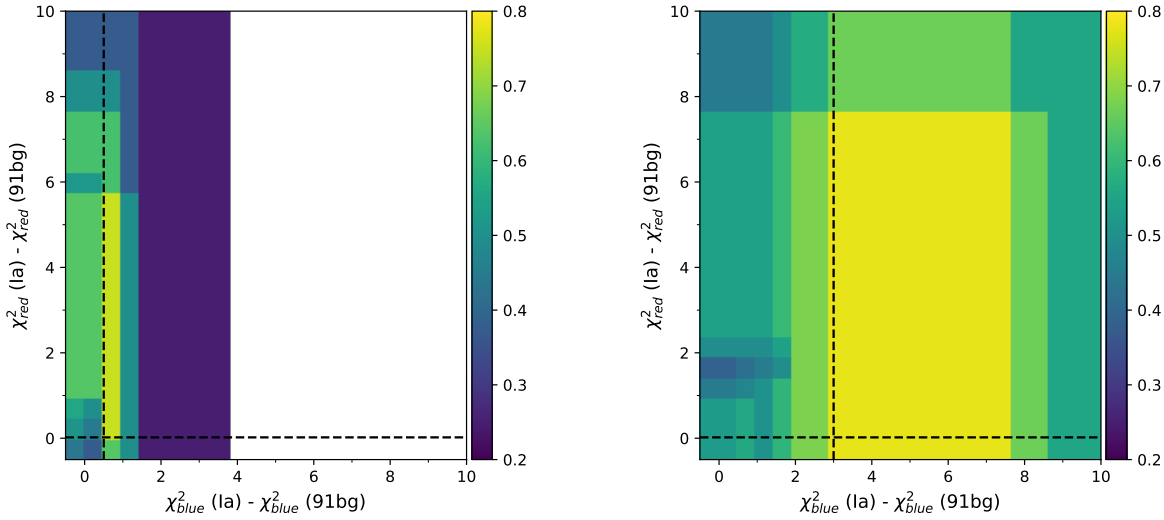
[2]Version 5.0: `https://people.lam.fr/blondin.stephane/software/snid/index.html`



(a) Classification coordinates determined by fitting observed bandpasses independently. Dashed lines are used to indicate classification boundaries $x = 0.5$, $y = 0$.

(b) Classification coordinates determined by fitting rest-frame red / blue bandpasses as collective sets. Dashed lines are used to indicate classification boundaries of $x = 3$, $y = 0$.

Figure 6: Classification coordinates are shown for objects observed by CSP (colored squares) and SDSS (grey circles) as determined by fitting rest-frame red / blue bandpasses independently (top) and as collective sets (bottom). CSP coordinates in both panels have been scaled by a factor of 0.023 along the x-axis and 0.030 along the y-axis to match the median SNR of SNe Ia observed by SDSS. The increased dispersion of points in the bottom panel indicates a lower sensitivity of the classification on the chosen classification boundaries.



(a) FOM values determined by fitting observed bandpasses independently. Dashed lines indicate classification boundaries of $x = 0.5$, $y = 0$ which align with the peak FOM value of 0.75. FOM values are not available beyond $x \approx 4$ as no CSP SNe have classification coordinates in that region.

(b) FOM values determined by fitting rest-frame red / blue bandpasses as collective sets. Dashed lines indicate classification boundaries of $x = 3$, $y = 0$ which align with the peak FOM value of 0.78.

Figure 7: FOM values are shown as a function of classification boundaries. The higher dispersion of points when fitting red / blue observations as collective sets results in a significant area of degenerate FOM values that can only be broken by the inclusion of additional spectroscopically classified SNe Ia.

we choose to use classification boundaries having the largest FOM while also independently minimizing the x and y cutoffs to be as close to the origin (0, 0) as possible. This results in classification boundaries of (3, 0) when fitting bandpasses as sets and (0.5, 0) when fitting bandpasses independently.

2.3.2 Classification Results

Figure 6 shows the classification coordinates resulting from fitting bands independently and as collective red / blue sets. We see that when fitting the red and blue bands as collective sets, fits to red bandpasses have a stronger impact on the classification. This is demonstrated by the increased vertical dispersion of points where $y > 0$. Similarly, when fitting bandpasses independently the classification of SN 1991bg-like objects is primarily driven by poor template fits in the red bands. We attribute this to differences in the light-curve morphology.

We note that fits to the data as collective red / blue sets require fewer overall light-curve fits than when fitting bandpasses independently. Fitting the data as two sets only requires fitting each model twice, whereas fitting individual bands requires a number of fits equivalent to twice the number of observed bands. As a result, fitting the observations as collective sets introduces fewer opportunities for a fit to diverge and the resulting classification coordinates are more stable. Collective fitting results also have a higher optimized FOM value. For these reasons, we choose to use coordinates determined from the collective fitting process to classify 91bg-like SNe. This leaves us with 16 remaining objects: CID 2778, 11570, 12689, 15204, 16215, 16309, 16692, 17094, 17468, 17886, 18218, 18751, 18890, 19065, 21678, 21898. Classification coordinates for these objects are listed in Table 3.

Sown in Figure 8, we note that some targets are classified as normal SNe Ia in the red vs. blue χ^2_ν comparison despite the SN 1991bg model having a lower χ^2_ν when fit to all available data. An example case is shown in Figure 9, where we examine fits to observations of SDSS object CID 15749. In this case, we see that significant influence is exerted by poor fits to observations in *ug* bands. However, visual inspection of the light-curve shows that the morphology of these bands is unusual despite normal behavior in the other bands. By

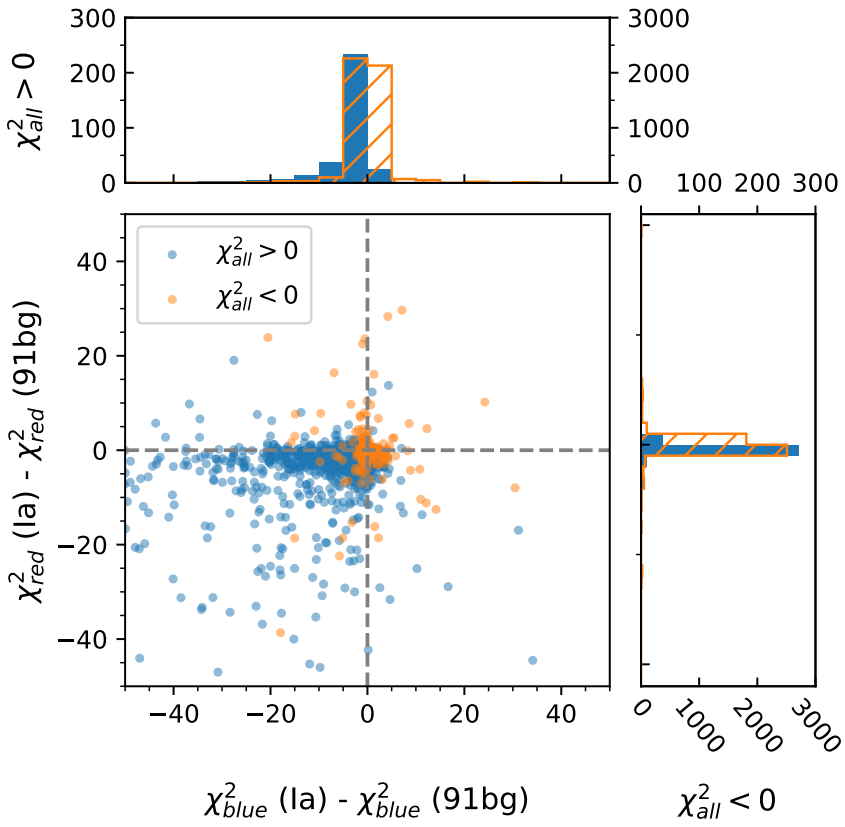


Figure 8: Differences between the reduced χ^2 of models for normal and SN 1991bg-like SNe. Fits are performed separately in the rest-frame blue ($\lambda_{\rm z,eff} < 5,500$ Å) and red ($\lambda_{\rm z,eff} > 5,500$ Å) bandpasses using a collective set of parameters for the redder and bluer bandpasses. We expect SN 1991bg-like objects to fall in the upper right quadrant (Q1) and normal SNe Ia in the lower left (Q3). Light-curves with better overall fits (smaller χ^2) to all the data with the H07 (SN 1991bg) model are shown in blue (orange).

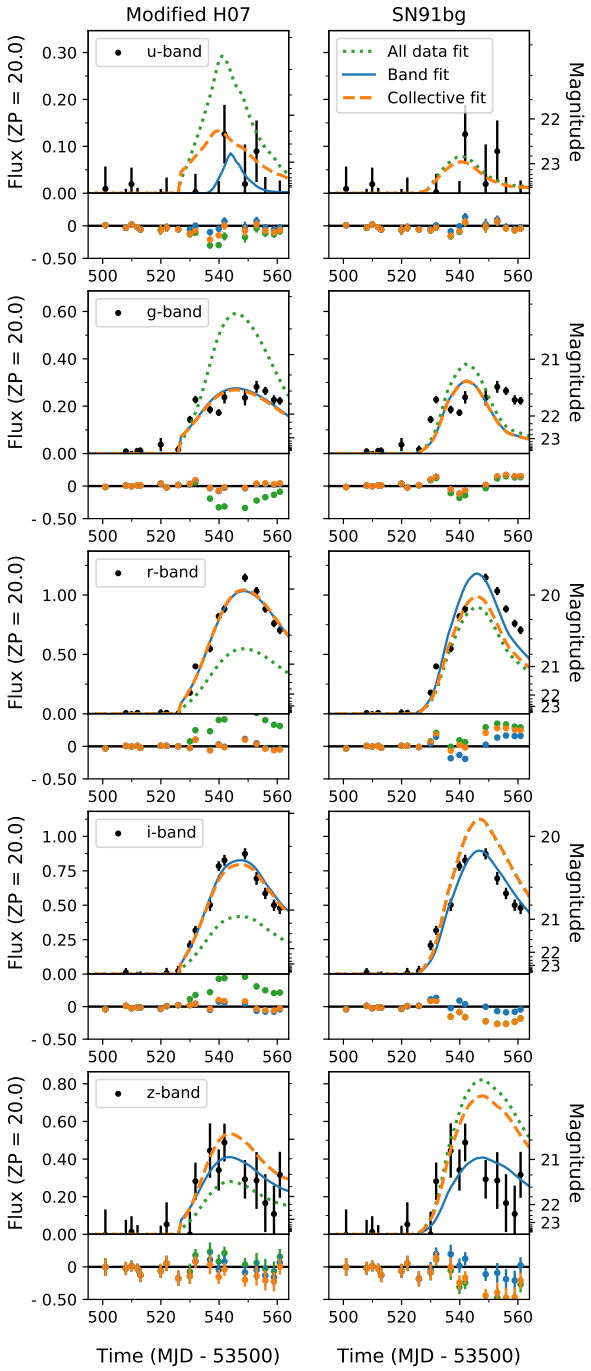


Figure 9: A comparison of fits to SDSS object CID 15749 using models for normal (left) and SN 1991bg-like (right) supernovae. Fits are performed to the entire data set (dotted green), the blue and red bandpasses as separate sets (dashed orange), and each bandpass independently (solid blue). By fitting each model to subsets of the data, the impact of the observed color in constraining the fit is lessened and the morphology each band allowed to play a more influential role.

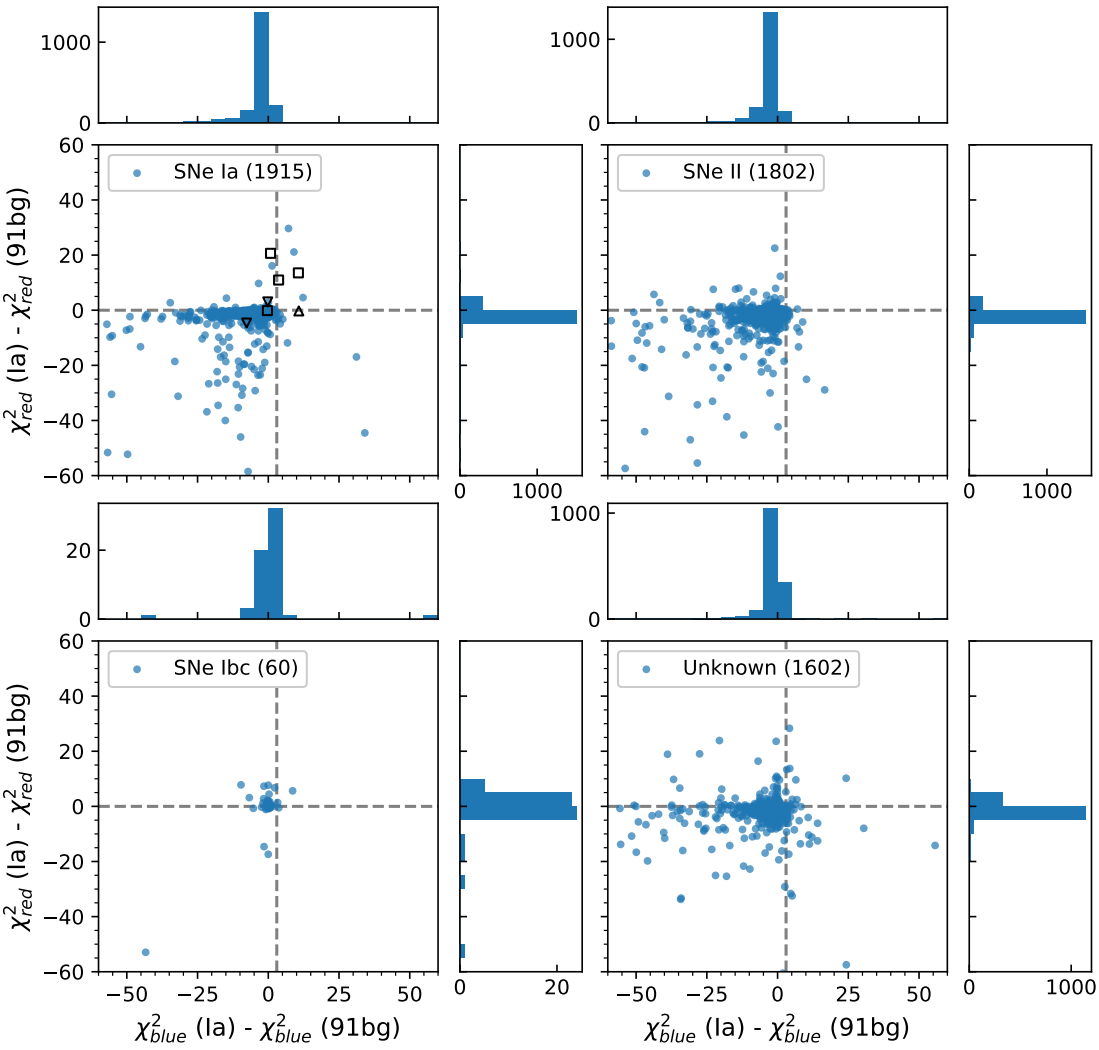


Figure 10: SNe observed by SDSS are broken down into four categories determined by their spectroscopic classification. If a spectroscopic classification is not available, the photometric classification determined by the PSNID software is used instead. The difference in reduced χ^2 for models of normal and SN 1991bg-like SNe are shown for normal SNe Ia (top left), SNe II (top right), SNe Ib/Ic (bottom left), and targets with light-curves too noisy to determine a classification (bottom right). We note that SN II events are constrained to quadrants 2 through 4 while SN Ib and Ic-like objects are clustered near the center of the phase space and SNe Ia are primarily scattered across quadrants 1 and 3. Dashed lines are used to indicate $x = 3$, $y = 0$.

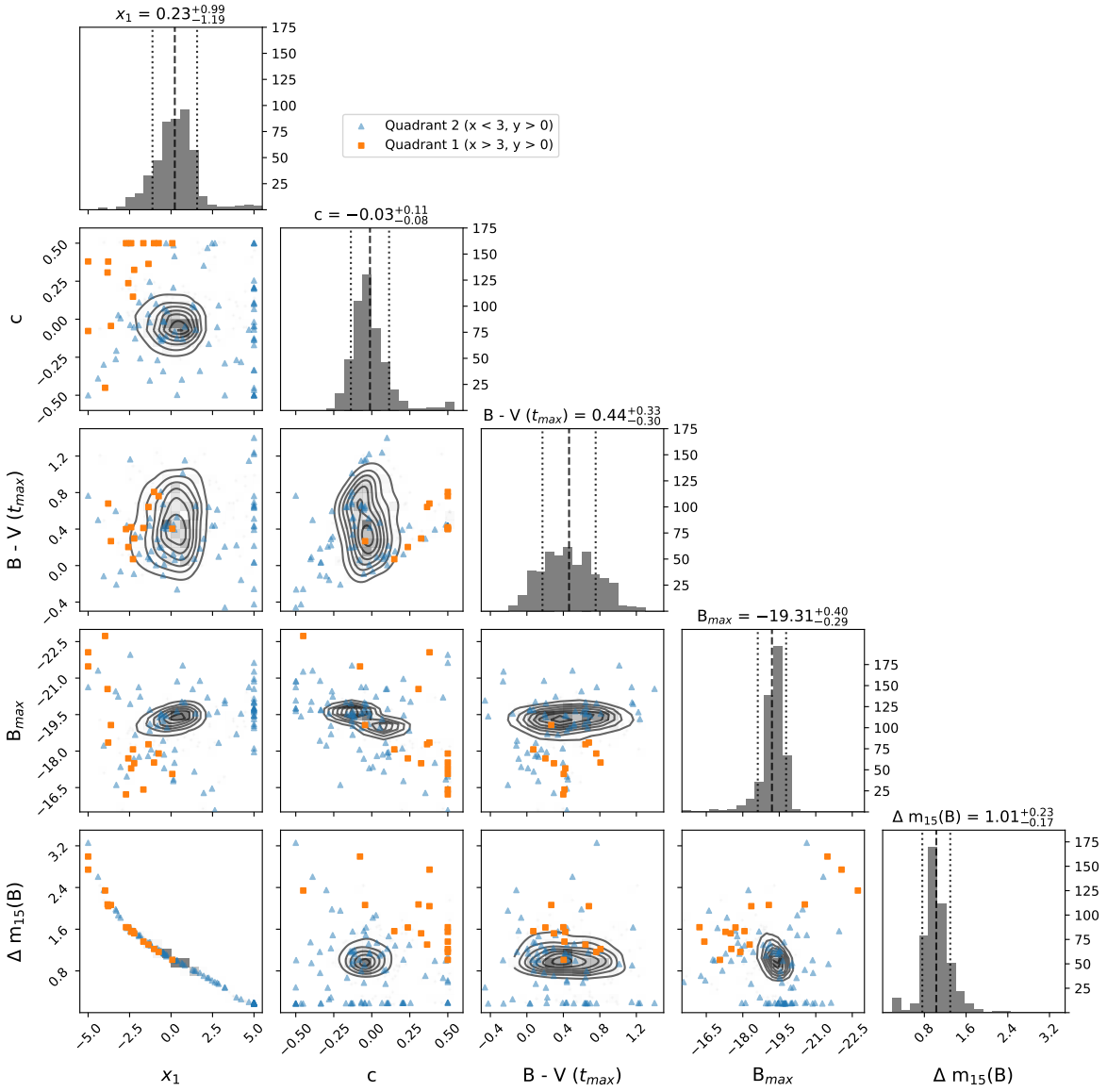


Figure 11: Distributions of the SALT 2.4 light-curve model fit to SDSS photometric observations. Points are color coded according to their position in the phase space $x = \chi^2_{\rm blue}({\rm Ia}) - \chi^2_{\rm blue}(91{\rm bg})$, $y = \chi^2_{\rm red}({\rm Ia}) - \chi^2_{\rm red}(91{\rm bg})$. We note that that SN 1991bg-like objects in quadrant 1 of the x, y phase space (Q1; orange squares) are fainter and redder than the normal SNe Ia population in quadrant 3 (Q3; grey density plot). Objects in quadrant 2 (Q2; blue triangles) are expected to be non-1991bg-like peculiar SNe.

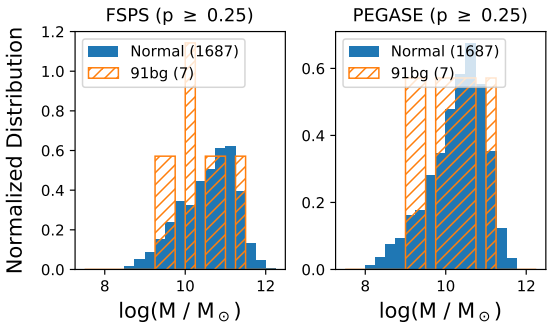
separating the fits to the blue and red data, the impact of the *ug* bands is mitigated, and the morphology of the other bands is allowed to play a more influential role in the classification.

To further understand the behavior of the employed classification scheme, in Figure 10 we compare our classification results against spectroscopic and photometric classifications from S18. We note that two of the four SNe with light-curves that were visually flagged by the SDSS SN team as potential 91bg-like objects are also labeled as 91bg-like SNe by our classifier. We also note that targets classified in the original data release as Type II SNe (SNe II), either spectroscopically or photometrically, are primarily constrained to quadrants two and three. The same can not be said for other types of core-collapse (CC) events, however both Type Ib and Ic SNe tend to be clustered in the center of the phase space. By asserting a cut at $x \geq 3$, $y \geq 0$, we find that all but one SNe II event can be excluded from the first quadrant (CID 5052). This indicates that, given the models we have chosen and their implementation, the classification scheme is robust against contamination by CC events.

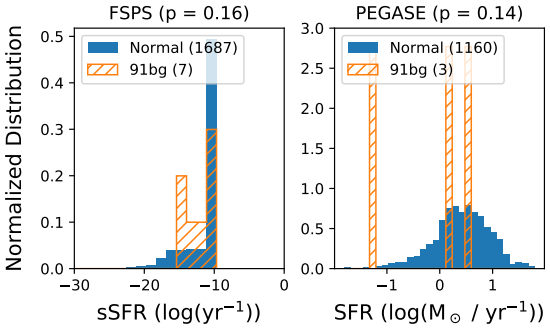
Given that SN 1991bg-like SNe are redder at peak than their normal SNe Ia counterparts (de Jaeger et al. 2018; Lluís Galbany, Mario Hamuy, et al. 2016), it is expected for them to be more easily mistaken as CC explosions. It is thus surprising that so few CC SNe are classified as SN 1991bg-like SNe. In G14, contamination by CC SNe was explored using observations of 64 CC SNe from J. P. Anderson et al. 2014 and other sources in the literature. Using a combination of cuts on magnitude, color, and the quality of fit in various bands, all but one SN were successfully removed.

Our implementation of the classification procedure displays a similar level of resistance to CC contamination as the original implementation of G14. However, in this work we are able to exclude CC events without the need for additional cuts on individual SN properties. One possible explanation is a significant intrinsic bias in the types of targets observed by the SDSS survey. Another possibility are differences in the models chosen for normal/SN 1991bg-like events, but this is minimized since we have specifically chosen models similar to those used in G14.

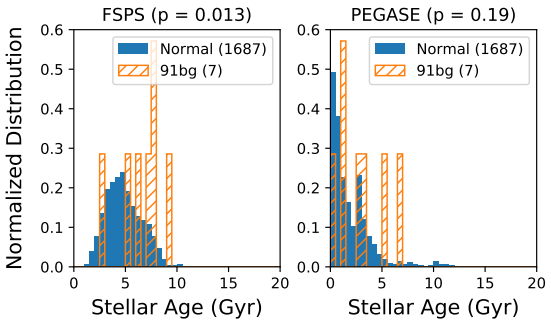
We thus conclude that this change in behavior is primarily driven by changes we have imposed in the way parameters are varied to fit each model. By choosing to vary parameters either independently across bands or as collective red / blue sets we have chosen to favor



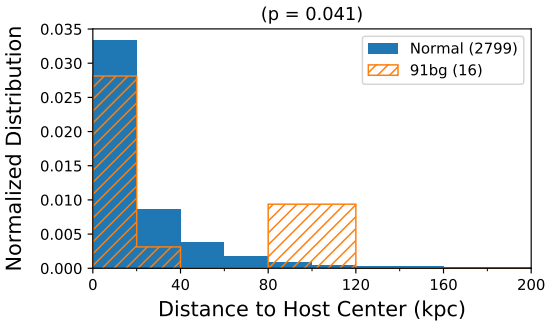
(a) Mass estimates for SNe host galaxies as calculated by FSPS (left) and PEGASE (right).



(b) Star formation rates for SNe host galaxies as calculated by FSPS (left) and PEGASE (right). We note the number of targets with available sSFR estimates is different for each distribution.



(c) Average stellar age for SNe host galaxies as calculated by FSPS (left) and PEGASE (right).



(d) Physical distance of SNe from the photometric center of their host galaxies.

Figure 12: Area normalized distributions for host galaxy properties of objects identified as normal (solid blue) and SN 1991bg-like (dashed orange) SNe. The Aderson-Darling test is used to determine whether the two distributions are drawn from different underlying distributions and the resulting p-value is displayed for each host property. For p-values below the 5% threshold, we conclude SN 1991bg-like events are drawn from a different underlying distribution. We find that 91bg-like events prefer further distances from the centers of their host galaxies and are more likely to occur in hosts with an older average stellar age than normal SNe Ia.

either light-curve morphology or color in the fitting process. In comparison, the way in which SiFTO varies parameters across bands while enforcing inter-bandpass relationships provides a middle-ground between these approaches. SiFTO also relies on significantly more free parameters, thus improving the quality of the overall fit to non-SNe Ia.

Although the SDSS data release did not include a dedicated SN subtyping effort, objects classified by SDSS were submitted to the Open Supernova Catalog (Guillochon et al. 2017), and that collection was collectively analyzed by Pruzhinskaya et al. 2019 for anomalous light-curves using a random-forest machine learning classifier. Listed in Table 4, a total of 37 SN from the SDSS sample were identified as peculiar objects. Out of these, none of these targets are included in our selection of 91bg-like SNe.

We note that the number of selected targets is significantly less than what is expected according to the rate of 6 – 15 % claimed in the literature (Ganeshalingam, W. Li, Alexei V. Filippenko, C. Anderson, et al. 2010; González-Gaitán et al. 2011; W. Li, Leaman, et al. 2011; Silverman et al. 2012). However, we expect to observe fewer SNe than the predicted rates due to intrinsic survey bias towards the identification of normal (brighter) SNe Ia. Additionally, the faintness and narrowness of SN 1991bg-like light-curves mean they spend less time over the SNR = 5 limit. This makes them less likely to be selected for followup and also more likely to be removed by quality cuts.

2.3.3 Properties of Selected SNe

To understand the intrinsic behavior of our selected SNe, we fit each target with SALT 2.4 and list the results in Table 5. Figure 11 shows that objects identified by our classifier follow many of the trends we expect of 91bg-like objects. In terms of color, selected objects tend to be redder than normal SNe with an average B - V color of 0.88 mag. They also fall on the dimmer and faster-declining extremes with an average $\Delta m(B)_{15}$ of 1.56. We note that SNe selected in quadrant 2 are much more diverse in their parameter distribution, but still have a notable subset of SNe lying on the faint and fast extremes of the SNe population.

Figure 12 shows the relationship between SNe and their host galaxies using properties determined in S18 with the FSPS (Conroy and James E. Gunn 2010; Conroy, James E.

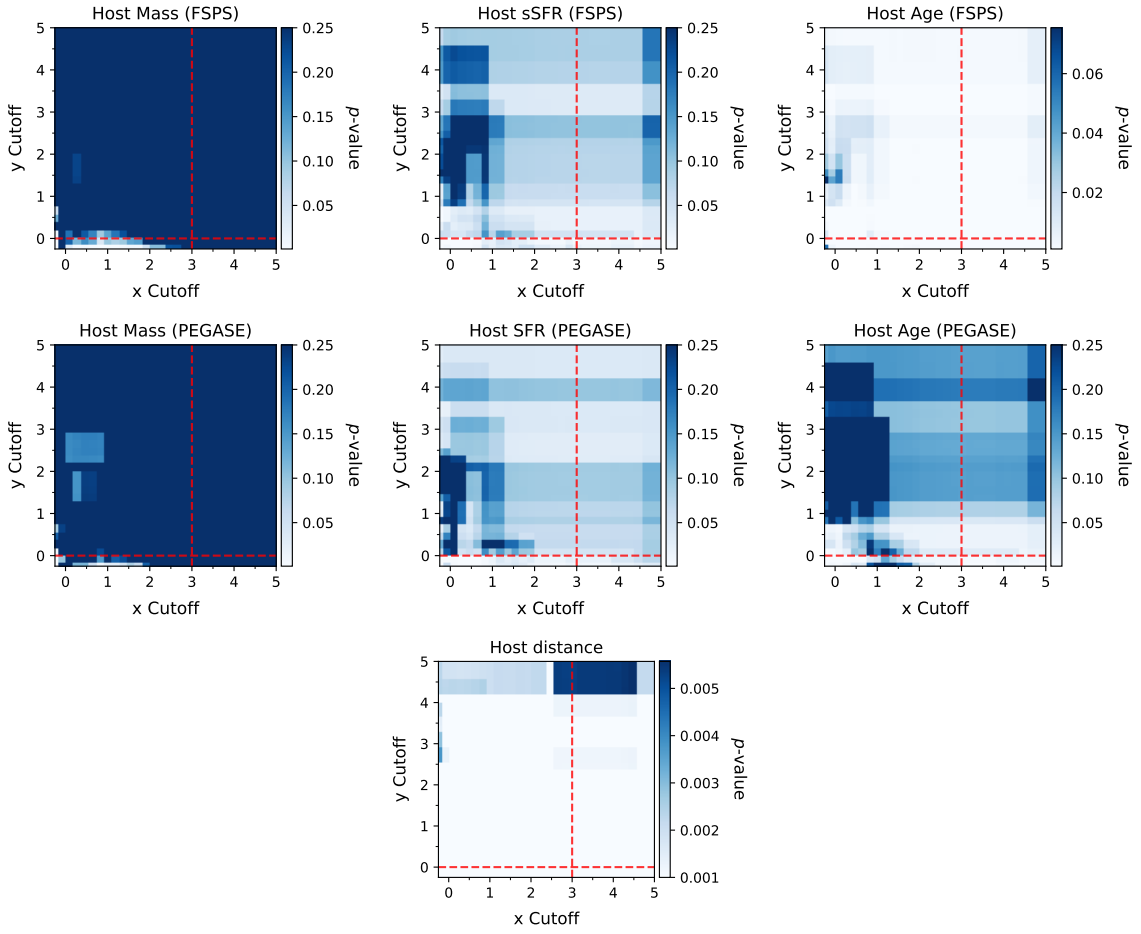


Figure 13: The Anderson-Darling test is used to determine whether normal and 91bg-like SNe are drawn from the same underlying distributions of host galaxy properties. This test is performed as a function of quadrant boundaries for the classification coordinates. The resulting p-values are shown for a collection of host galaxy properties as calculated by FSPS (top row) and PEGASE (middle row). Considered properties in these rows include host galaxy mass (left), star formation rate (center) and average stellar age (right). The calculation is also repeated for the physical distance of SNe from their host galaxies (bottom row). The actual quadrant boundaries used in this work are shown in red for reference. The p values are not qualitatively sensitive to the choice of quadrant boundary — for host mass, average stellar age, and distance the p values consistently fall either above or below 0.05. The normal SNe Ia and 91bg-like SNe Ia identified in this work are clearly different in their host galaxy stellar age and physical distance from the host galaxy center.

Gunn, and White 2009) and PEGASE (Fioc et al. 1997) software routines. To understand whether the selected SNe are drawn from the same underlying host galaxy distribution as normal SNe, we perform an Anderson-Darling test (T. W. Anderson et al. 1952). We take as a null-hypothesis that the two populations are drawn from the same underlying distribution. For a p-value $< 5\%$ we reject the null hypothesis and assert that the underlying distributions are different. In all cases presented by this work, p-values are bound to the range $0.001 \leq p \leq 0.25$.

We find no statistical evidence to indicate the selected SNe are drawn from a different underlying distribution of galaxy mass or Star Formation Rate (SFR). Although the p-value determined for SFR using FSPS and PEGASE differ significantly (24% and 16% respectively), this can be attributed to the fact that the FSPS routine was able to determine SFR values for more 91bg-like targets than the PEGASE routine. In practice, the number of available points is considered in the calculation of the p-value (see Scholz et al. 1987). However, we also cannot rule out the possibility that the subsample enforced by PEGASE is somehow biased towards a particular distribution of SFR.

We find the only considered properties to indicate a different underlying distribution is the average stellar age and distance from the center of the host galaxy. Visual inspection of Figure 12 shows that objects identified by our classifier as being 91bg-like prefer galaxies with older stellar populations and have a higher probability of occurring further away from the center of the galaxy. This confirms a previous result found using SDSS data in Lluís Galbany, Miquel, et al. 2012, hereafter G12. Using a subset of 200 spectroscopically or photometrically confirmed SNe Ia at redshifts $z \leq 0.25$, G12 found that the average fitted color term (c) from SALT decreased with the projected distance for SNe Ia in spiral galaxies. It was also determined that SNe in elliptical galaxies tend to have narrower light-curves if they explode at larger distances, although the impact of selection effects was unclear.

In principle, the results of the Anderson-Darling test are dependent on the quadrant boundaries used to classify targets. Figure 13 shows the recalculated p-values for a range of classification boundaries. We see that for targets that have passed our quality cuts, there is minimal variation in p-values surrounding our chosen quadrant boundaries for host galaxy mass, age, SFR, and distance. For an x and y cutoff large enough, we do see a slight increase

in the p-value for some properties. However, as shown in Figure 14, there are only a small number of points selected at those extremes and the existence of additional selection effects becomes unclear.

2.4 Discussion

Using SN observations from SDSS, we explore the implementation of an empirically based classification technique targeted at the identification of SN 1991bg-like SNe. In the presented approach dedicated light-curve fits are performed for observational data in rest-frame blue ($\lambda_{\rm z,eff} < 5,500$ Å) and red ($\lambda_{\rm z,eff} > 5,500$ Å) bandpasses. Using models for both a normal and 91bg-like SN, targets are classified based on the difference in reduced χ^2 values for each model in blue and red wavelengths.

We consider two distinct implementations of this technique. In the first implementation each observed bandpass is fitted independently and the χ^2 values from each fit are summed to determine the overall χ^2 for the blue and red bandpasses respectively. The second implementation is performed in the reverse order: Observed bandpasses are split into sets of bluer and redder data and then fit as two collective sets. We find no significant differences in the classifications generated from either approach. However, we note that the latter approach requires a larger number of spectroscopically classified SNe Ia to fully train the classification procedure.

To understand the potential for contamination by non-SNe Ia, we compare our classification results with spectroscopic classifications for a limited subset of SDSS targets. We find that our classification procedure is robust against contamination from core collapse events with only one SNe II being classified as SN 1991bg-like. When performing the same comparison against photometric classifications from PSNID, we reach the same conclusion.

In total our classifier identifies 16 SNe from the SDSS-II SN sample: CID 2778, 11570, 12689, 15204, 16215, 16309, 16692, 17094, 17468, 17886, 18218, 18751, 18890, 19065, 21678, 21898. Existing subtypes for SDSS SNe in the literature is limited, restricting our ability to compare results. A total of 37 targets from the SDSS sample were identified as anomalous

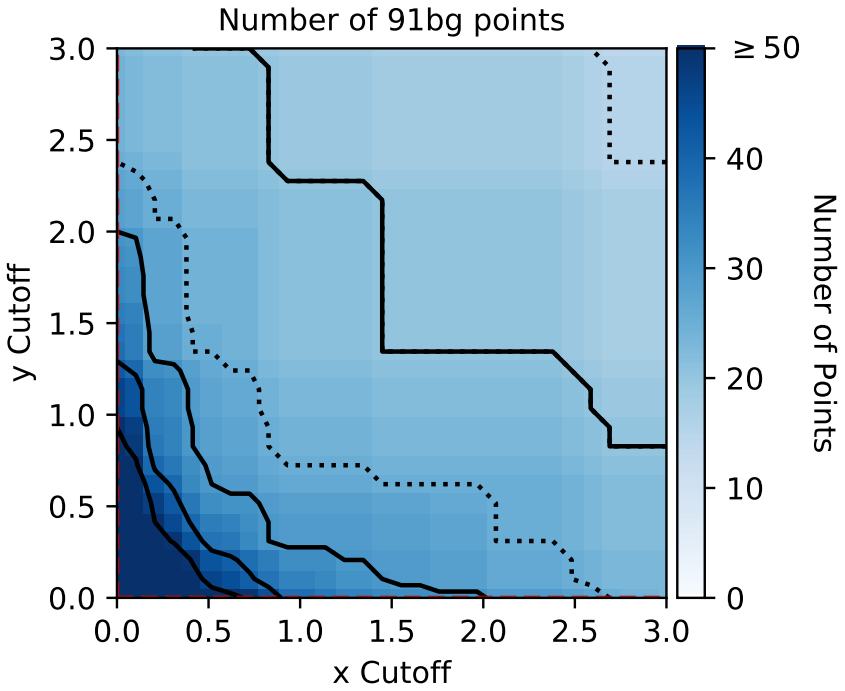


Figure 14: The number of objects classified as SN 1991bg-like as a function of quadrant boundaries for the classification coordinates. Contours are shown in steps of 10 (solid) and 5 (dashed) SNe.

objects in an external analysis using a random-forest classifier on photometric data from the OSC. Out of these objects, we classify none of them as SN 1991bg-like events.

Using host galaxy properties from SDSS, we investigate potential differences in the distribution of host galaxy properties for normal SNe and those selected by our classifier. An inspection of host mass as measured by the FSPS and PEGASE routine reveals no statistically significant bias in host galaxy mass. We also find no significant trend in host galaxy SFR. However, selected objects are seen to prefer galaxies with older stellar populations and have a higher probability of occurring further away from the center of the galaxy.

Future work is currently planned to extend the presented classification technique to classify other kinds of peculiar SNe. A more detailed investigation is also planned to explore potential biases that may be introduced in the way spectroscopic templates are varied to fit photometric observations.

3.0 Impact and Calibration of Atmospheric Water Vapor

A key requirement for LSST is achieving strict constraints on photometric repeatability. This includes the ability to perform observations with a 5 mmag level of precision and a zero-point stability across the sky of 10 mmag (Željko Ivezić et al. 2019). Understanding and calibrating for the effects of atmospheric absorption is an important part of achieving this precision level (see T. S. Li et al. 2016, David L. Burke, Saha, et al. 2014, and David L. Burke, Axelrod, et al. 2010). In principle, we expect grey variations of the atmospheric transmission to be implicitly removed by the calibration pipeline that will produce calibrated magnitudes in the *natural* LSST system. However, atmospheric variations are not just grey, and a second-order effect is to distort the effective shape of the survey passbands. This introduces undesired, color dependant effects, herein referred to as *atmospheric variability*.

Fully correcting for atmoshperic variability requires the determination of an effective atmospheric transmission function which is both time and wavelength dependent. However, most photometric surveys take a simpler approach using supplemental broad-band observations of well understood astronomical targets. Astronomers traditionally calibrate broad-band imaging using a reference catalog of well observed stars to compute correction terms for color, airmass, and perhaps a higher-order color-airmass term. In general, the color term accounts for the difference in filter and detector sensitivity with wavelength, but also includes some average contribution of the atmosphere above the telescope being used.

Although this approach implicitly accounts for the effects of atmospheric opacity on observed images, the applied correction is an incomplete approximation. Differences between the underlying spectral energy distribution (SED) of observed targets and those included in the reference catalogue lead to errors in the corrected brightnesses. This is particularly problematic when calibrating observations of targets like SNe, whose SEDs often look significantly different than those of well behaved stars (see Section 3.2.2).

A more detailed description of the atmosphere can be obtained by spectroscopically observing a telluric standard star. These bright stars of known SED are well suited for determining the absorption and scattering of the atmosphere. In order to describe atmospheric

effects, spectroscopy should be performed on a telluric standard at the same airmass as a desired target. This is also ideally performed at the same time and on-sky position as the target being calibrated. The total atmospheric absorption per wavelength can then be found by dividing the observed spectrum by the tabulated SED already corrected for absorption.

While this method is effective, the majority of telescopes are not configured to have an auxiliary spectrograph for observing telluric stars. Because atmospheric absorption is variable over time, observations of a standard star must be performed repeatedly and within a short time interval of other targets. Even in setups with the capability to easily switch back and forth between mosaic imaging and single-object spectroscopy, such observations require diverting valuable observation time away from other targets.

As an alternative, astronomers commonly express the atmospheric absorption as a linear function of airmass. Using photometric observations taken over a range of airmass values, corrections are performed by fitting a linear function in each band. This approach assumes that the absorption scales linearly with airmass. However, the absorption spectrum of atmospheric water vapor is a complex series of very narrow absorption lines. These individual lines can saturate, and thus the absorption does not scale linearly with airmass. This non-linearity introduces errors due to higher order effects when calibrating photometric images (Blake et al. 2011).

In the redder range of CCD sensitivity ($5,500 < \lambda < 12,000$ Å), the atmospheric transmission function is dominated by absorption due to precipitable water vapor (PWV)[1]. The use of GPS to measure the localized PWV column density is a recently emerged technology in astronomy and an accurate alternative to traditional methods (Dumont et al. 2001). Through the use of atmospheric modeling, these PWV measurements can be used to simulate the atmospheric transmission. The resulting transmission function can then be used to correct photometric observations of sources with known spectral energy distributions for atmospheric absorption. For systems with existing auxiliary equipment for measuring the atmosphere, GPS systems also provide a means for cross validating the output of those systmes.

[1]While ozone and aerosol scattering also play a role, their opacity is relatively smooth with wavelength and varies over much longer time scales.

In this chapter we explore the calibration of PWV effects using GPS measurements in a real world setting (Section 3.1) and use that data to simulate how PWV absorption will impact SNe Ia observed by the upcoming LSST (Section 3.2). Section 3.1.1 continues discussing standard calibration techniques and provides a general overview of the calibration process. In Section 3.1.2, we demonstrate the GPS based measurement of PWV at Kitt Peak National Observatory (KPNO) and how the PWV at KPNO can be predicted from nearby GPS stations on the surrounding desert floor. Section 3.1.3 provides further validation of this approach by combining GPS measurements with existing atmospheric models to correctly model PWV absorption features present in spectra taken at KPNO.

Using real world PWV measurements, Section 3.2.1 outlines a procedure for simulating SNe observed by LSST. From these simulations, we demonstrate the fundamental impact of PWV on LSST SNe in Section 3.2.2.

3.1 Calibrating For PWV Effects Using GPS Measurements

Shown in Figure 15, the atmospheric transmission between $5,500$ and $12,000$ Å is dominated by absorption due to PWV. The strength of PWV absorption lines in observed spectra correlates strongly with measurements of localized PWV column density (Blake et al. 2011). This indicates that PWV measurements can be combined with atmospheric models to provide estimates of the atmospheric transmission at a given date and time. However, accomplishing this requires a source of accurate and readily accessible PWV measurements. Furthermore, since PWV levels can change by over 10% per hour, measurements must be available in close to real time.

By measuring the delay of dual-band GPS signals traveling through the atmosphere, it is possible to determine the PWV column density along the line of sight (see Braun et al. 2001, Dumont et al. 2001, and Nahmias et al. 2004). This approach is made even more appealing by the existence of several established GPS networks dedicated to the measurement of geological and meteorological data on the international scale. The SuomiNet project[2] (Ware et al. 2000)

[2]For more information see `https://www.suominet.ucar.edu`

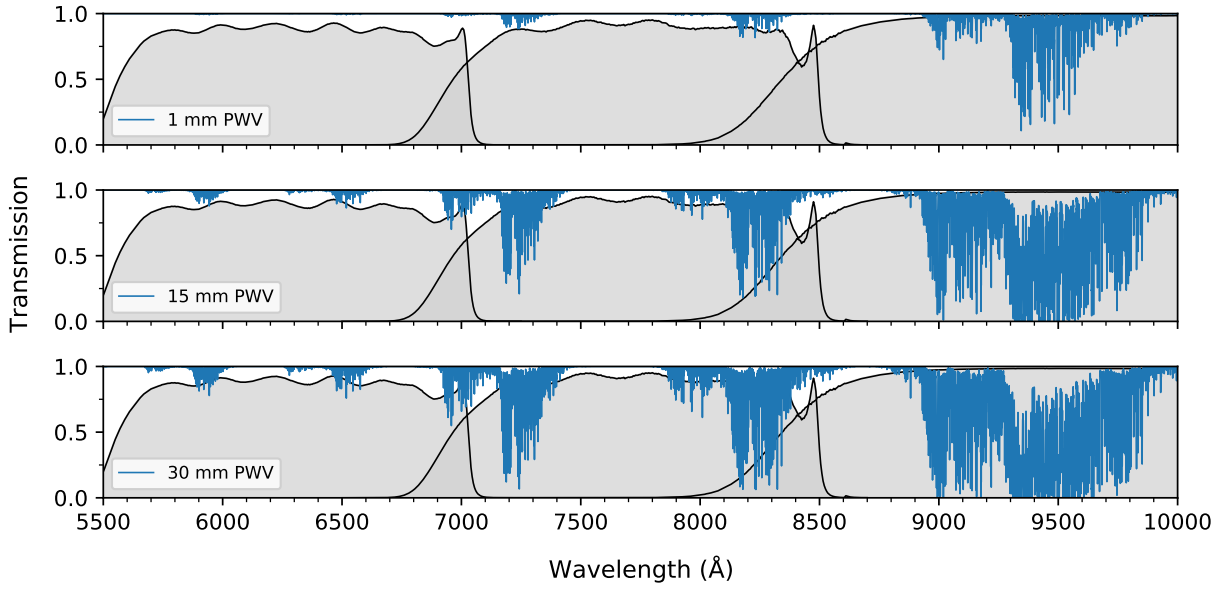


Figure 15: The r, i and z band mosaic filters of Kitt Peak National Observatory (grey) compared against the MODTRAN modeled atmospheric transmission function due to precipitable water vapor (blue). Atmospheric transmission functions are shown for an airmass of one and a precipitable water vapor (PWV) column density of 1 mm (top), 15 mm (middle), and 30 mm (bottom). Note that absorption features do not scale linearly with PWV, and some saturate at relatively low column densities.

is a meteorological initiative that uses data from multiple GPS networks to provide semi-hourly PWV measurements. It currently publishes meteorological data from hundreds of receivers throughout the United States and Central America. For the remainder of this work, we rely on data published by SuomiNet when real-world PWV measurements are needed.

3.1.1 Effects of PWV on Photometric Calibration

When correcting photometric observations for atmospheric effects, astronomers commonly express atmospheric absorption as a linear function of airmass. In this approach photometric observations are corrected by fitting for a set of extinction coefficients k' and k'' in each band. For example, given an airmass X, the observed i and z band magnitudes of a standard star are related to the tabulated, intrinsic magnitudes z_0 and i_0 by a set of linear equations

$$\mathrm{z} = \mathrm{z}_0 + k'_\mathrm{z} \cdot X + k''_\mathrm{z}(\mathrm{b}-\mathrm{v}) \cdot X \tag{3.1}$$

$$\mathrm{i} = \mathrm{i}_0 + k'_\mathrm{i} \cdot X + k''_\mathrm{i}(\mathrm{b}-\mathrm{v}) \cdot X \tag{3.2}$$

The first order extinction term k' accounts for the decrease in a star's observed flux with airmass. The inclusion of a second order coefficient k'' accounts for the fact that the observed flux of blue stars decreases faster than red stars as they approach the horizon.

To measure the second-order extinction, observations are taken of a red and blue star over a wide airmass range. The second-order extinction in each band can then be found by fitting for the difference in magnitude between the two stars.

$$\Delta\mathrm{z} = k''_\mathrm{z}\Delta(\mathrm{b}-\mathrm{v}) \cdot X + \Delta\mathrm{z}_0 \tag{3.3}$$

$$\Delta\mathrm{i} = k''_\mathrm{i}\Delta(\mathrm{b}-\mathrm{v}) \cdot X + \Delta\mathrm{i}_0 \tag{3.4}$$

Using the resulting value for k'', the first order extinction coefficient is then found by fitting Equations 3.1 and 3.2. Although this method does account for a first order airmass dependence, it does not directly account for any nonlinear effects. This is to say it does not account for parts of the atmospheric transmission having a nonlinear airmass dependence.

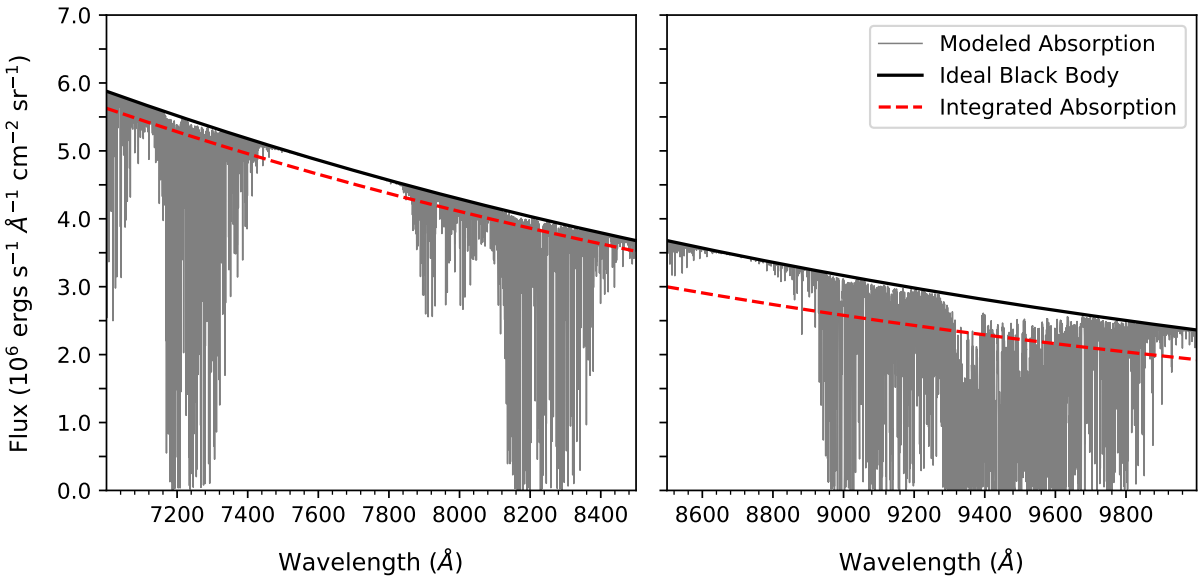


Figure 16: The SED of a blackbody at 8,000 K (black) across the i-band (left) and z-band (right) ranges. Shown in grey, the modeled atmospheric absorption for a PWV column density of 15 mm is applied to the SED. This is compared to the black body SED scaled using the integrated absorption in each band in red.

For a PWV column density at zenith PWV_z, the column density along the line of sight is given by

$$\mathrm{PWV}_{\mathrm{los}} = \mathrm{PWV}_z \cdot X \tag{3.5}$$

However, due to saturation, not all absorption features scale linearly with PWV concentration – some features saturate at relatively low concentrations (< 10 mm). Thus a linear function of airmass and color is not sufficient to describe the atmospheric transmission from PWV.

Figure 16 details the error introduced in a black-body by considering PWV absorption averaged over a bandpass versus the actual absorption spectrum. Because atmospheric absorption varies with wavelength, it affects stars differently depending on their spectral type. This means that variations in the spectral types of photometric standards used to correct an image introduce errors in the magnitudes of observed targets. This effect is more pronounced for higher airmass due to the increased PWV along the line of sight, and is an important consideration for KPNO where PWV_z exceeds 20 mm over 13% of the time.

Demonstrated in Figure 17, when using a type A star to correct cooler G or M type stars,

spectral variations between stars used in the atmospheric correction can introduce errors as large as -0.02 mag. This error is particularly important when performing high accuracy photometry to 1% or better. An alternative is to correct photometric observations using atmospheric models. For an atmospheric transmission $T(\lambda)$, the photometric correction for an object with a spectral energy distribution $S(\lambda)$ is given by

$$C = \frac{\int_{\lambda_i}^{\lambda_j} S(\lambda) \cdot T(\lambda)\, d\lambda}{\int_{\lambda_i}^{\lambda_j} S(\lambda)\, d\lambda} \tag{3.6}$$

where the integration bounds are defined by the wavelength range of the photometric bandpass. Using atmospheric models, measurements of the PWV column density are used to determine $T(\lambda)$ at a given date, time, and airmass. If tabulated values for $S(\lambda)$ are not available, spectral templates are used instead. For example, the SED of a star is well estimated by its color, due to the strong relationship between stellar spectral type and intrinsic color.

3.1.2 Use of GPS at Kitt Peak

In March of 2015, we installed SuomiNet connected weather station on top of the WIYN 3.5 meter telescope building at Kitt Peak National Observatory. In addition to a GPS receiver, the station includes barometric, temperature, and wind speed sensors. SuomiNet compiles measurements from its affiliated weather stations at thirty minute intervals. These semi-hourly measurements, in addition to the local PWV column density along zenith, are then released publicly on an hourly basis.

In order to prevent equipment damage, the weather station at Kitt Peak is powered down during lightning storms. This creates gaps in the available SuomiNet data for Kitt Peak. Additionally, the barometric sensor was malfunctioning in 2016 from January through March, so we ignore any SuomiNet data published for Kitt Peak during this time period. The sensor has since been repaired, but occasionally records a non-physical drop in pressure. We disregard these measurements by ignoring any meteorological measurements taken for Kitt Peak with a pressure below 775 mbar.

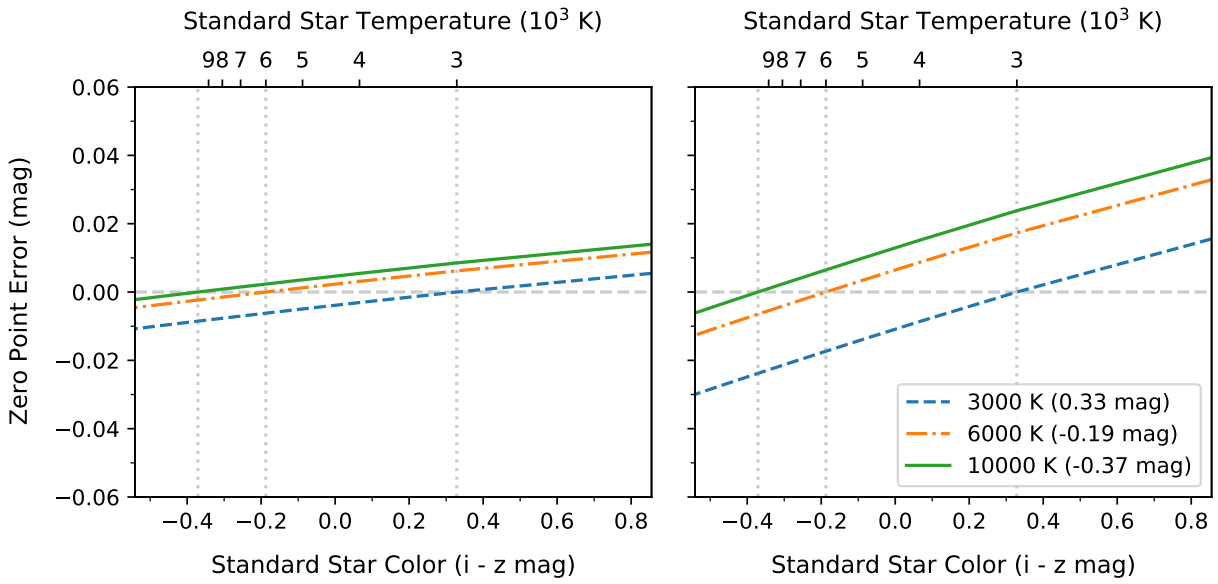


Figure 17: Correcting photometric observations using tabulated values of a standard star introduces residual error in the magnitudes of other stars with different spectral types. The residual error in z band photometric zero point due to absorption by precipitable water vapor is shown for three black bodies at $3,000$ (M type), $6,000$ (G type), and $10,000$ K (A type). Results are shown as a function of the color of the reference star used to calculate the zero point. Error values are shown for a PWV column density of 5 (left) and 30 mm (right).

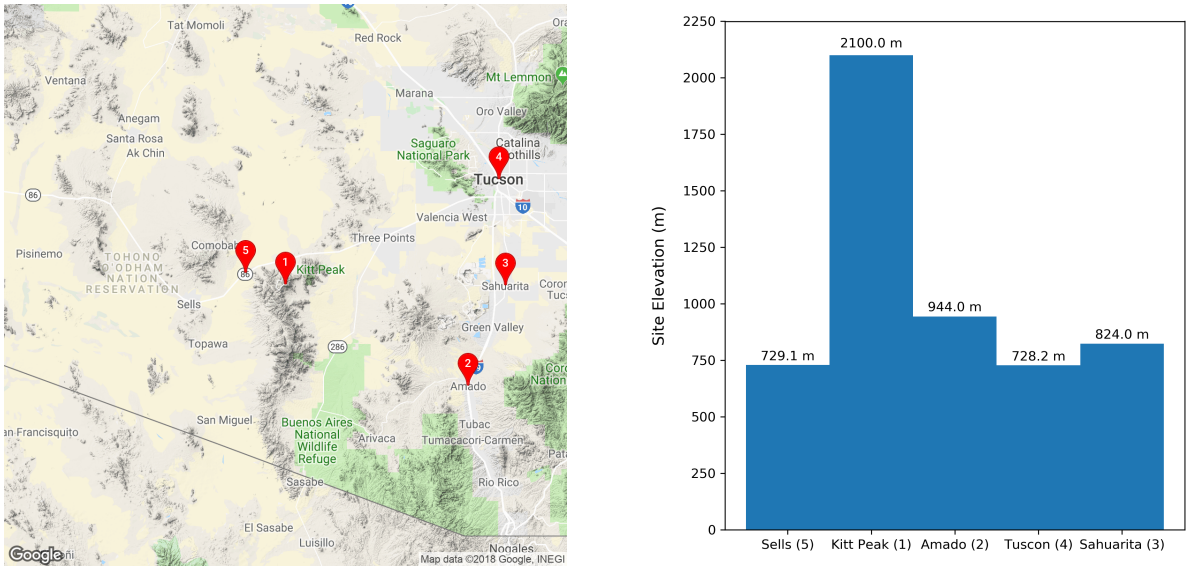
In order to determine the PWV level during periods without SuomiNet data, measurements from other nearby receivers can be used to model the PWV level at Kitt Peak. This model can also be used for times before the Kitt Peak receiver was installed. In addition to data taken at Kitt Peak, we use measurements from four other receivers within a 45 mile radius at varying levels of altitude. This includes receivers located at Amado (SuomiNet receiver Id *AMAZ*), Sahuarita (*P014*), Tucson (*SA46*), and Sells (*SA48*) Arizona. The location of these receivers is shown in Figure 18a, with SuomiNet measurements for Kitt Peak, Amado, and Sells shown in Figure 19.

Note that the PWV level at each location follows the same seasonal trend, but the PWV concentration at Kitt Peak tends to be lower. Since each of the chosen receivers are geographically close together, variations in PWV between Kitt Peak and the four supplementary locations are predominantly caused by differences in altitude. Shown in Figure 20, the PWV level at each location can thus be related to the PWV level at Kitt Peak by applying a linear fit. Each fit is able to predict the PWV column density at Kitt Peak to a precision of 1 mm plus 10% of the predicted value.

For times when SuomiNet data is unavailable for Kitt Peak, each of the linear fits are used to estimate the PWV column density at Kitt Peak. The resulting estimations are then averaged and used to supplement data taken by the Kitt Peak weather station. This full data set provides a model for the PWV column density at zenith over time. To determine the PWV column density for a specific date and time, we first determine the concentration along zenith by interpolating from the supplemented PWV data. The PWV column density along the line of site is then calculated using Equation 3.5. Using this value, the corresponding atmospheric transmission dues to PWV is determined using the MODTRAN modeling software (Berk et al. 2014)

3.1.3 Validation of GPS Based Atmospheric Models

From 2010 September 16th through September 20th, an observation run was performed on 18 standard stars using the R.C. spectrograph on the Mayall 4m telescope. To reduce flux loss due to atmospheric dispersion, the spectrograph was configured to use a wide 7”



(a) (b)

Figure 18: The `pwv_kpno` package uses PWV measurements for five locations within a 45 mile radius of Kitt Peak. Shown on the left, these locations include Kitt Peak (1), Amado (2), Sahuarita (3), Tucson (4), and Sells (5) Arizona. The elevation of each location is shown on the right.

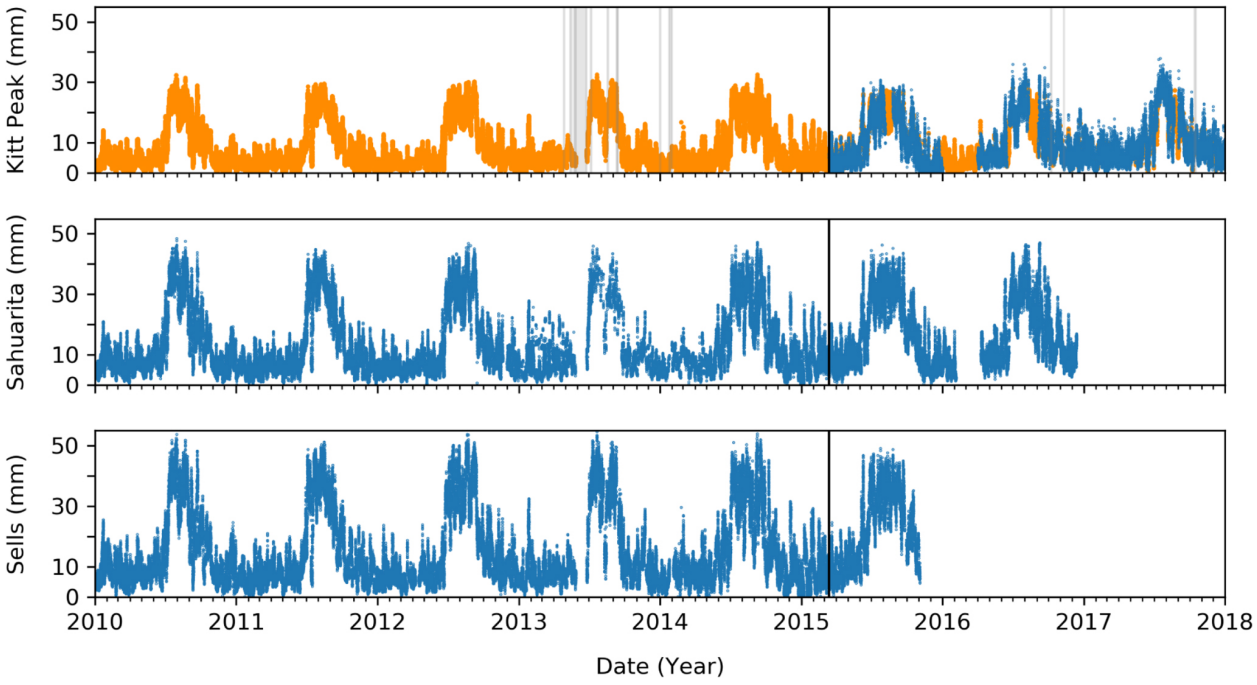


Figure 19: Measurements of precipitable water vapor (PWV) from the SuomiNet project from 2010 onward. From top to bottom, SuomiNet measurements for Kitt Peak National Observatory, Sahuarita AZ, and Sells AZ (Blue). The modeled PWV level at Kitt peak is shown in Orange. Periods of one day or longer where there are no modeled PWV values are shown in the top panel in grey. The geographic proximity of these locations means that the primary difference in PWV between locations is due to differences in altitude. Measurements taken at Kitt Peak National Observatory begin in March of 2015.

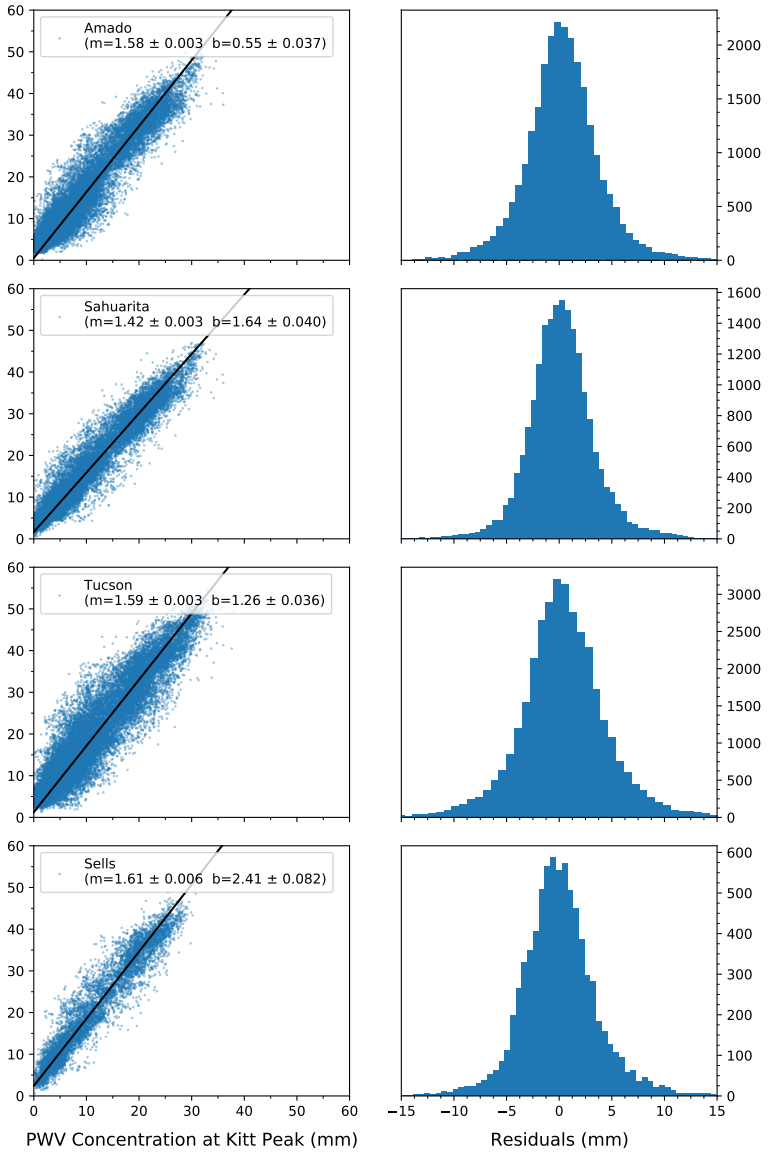


Figure 20: Linear fits to measurements of precipitable water vapor (PWV) column density taken at four different locations versus simultaneous measurements taken at Kitt Peak. Each row corresponds to a different location being compared against Kitt Peak, with measurements shown on the left and binned residuals shown on the right. The slope (m) and y-intercept (b) is shown for each fit. The correlation in PWV column density between different sites allows the PWV column density at Kitt Peak to be modeled using measurements from other locations.

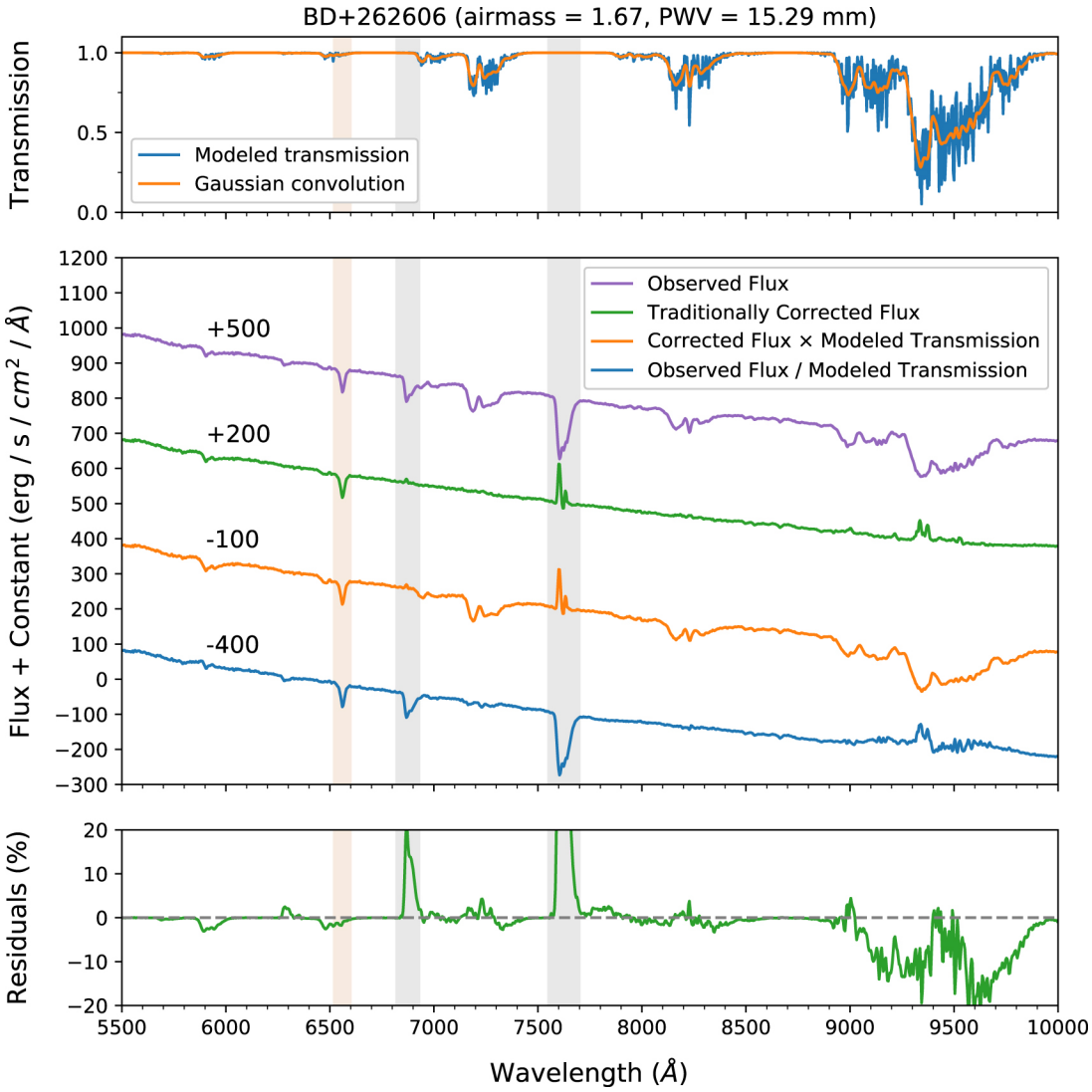


Figure 21: Observations of BD+262606 were taken using the R. C. Spectrograph at Kitt Peak National Observatory. The top panel demonstrates the modeled PWV transmission function at the time of observation (blue) smoothed by a Gaussian kernel (orange). The middle panel shows the observed spectrum (purple) and the spectrum corrected using catalog values (green). These are compared against the catalog corrected spectrum multiplied by the modeled transmission (orange) and the observed spectrum divided by the smoothed transmission (blue). Residuals between the catalog corrected and model corrected spectrum are shown in the bottom panel. H_α lines are highlighted in red and O_2 lines in grey.

slit. Observations were recorded between 5,500 and 10,200 Å with an average dispersion of 3.4 Å per pixel. Seeing for all observations varied between 1 and 2".

As an example, Figure 7 shows the SED of BD+262606 observed at an airmass of 1.67. To flux-calibrate the observed spectrum, low-airmass observations were taken of BD+17 4708 each night. This minimized the introduction of additional telluric effects in the calibrated spectrum. To correct the observed spectrum for atmospheric effects, the absorption in the standard star was scaled to match the airmass of the other observations following the prescription of Wade et al. 1988.

The atmospheric model does not directly account for the smoothing that occurs in observed spectra due to a spectrograph's spectral resolution function. As a result, directly dividing the observed spectra and modeled transmission will produce a very high, unphysical flux for wavelengths where the transmission function is saturated. To account for any saturated features, the modeled transmission is first binned to approximately match the observed spectrum's resolution. The transmission is then smoothed further using a Gaussian kernel.

To correct for atmospheric effects, the observed spectrum is divided by the smoothed PWV transmission function. We note that the observed spectrum was taken before a GPS receiver was installed at Kitt Peak. This means that no direct PWV measurements are available for the time of observation, and we instead determine the modeled PWV transmission using measurements from GPS receivers on the surrounding desert floor.

In the model-corrected spectrum, the absorption feature at 6,550 Å is an Hα line intrinsic to the observed spectrum. Furthermore, the absorption features at 6,875 and 7,650 Å are caused by O_2 absorption. Since the atmospheric model used in this work only provides models for the PWV absorption[3], these two features remain uncorrected. Given that there are no emission lines relative to the continuum, the feature at 9350 Å is categorized as an unidentified artifact from the reduction process.

Corrections for the PWV absorption features agree reasonably well between the catalog and model corrected spectrum. The largest deviations between the corrected spectrum oc-

[3]In general MODTRAN does support contributions to the atmospheric transmission other than PWV. However, the inclusion and discussion of these effects is outside our current scope.

cur redward of 9,000 Å. Some of these deviations can be attributed to cloudy observation conditions, creating large spatial and time variations in the PWV concentration along the line of sight (Querel et al. 2014). However, correcting this feature is also difficult since it is in fact a number of thin, saturated lines that have been blended together. Overall we find that the model struggles to correct the observed spectrum past 9,000 Å, but performed well enough overall to be used to satisfactorily correct photometric observations.

3.2 Impact of PWV on LSST

Having validated our ability to measure PWV and model the corresponding atmospheric transmission, we now focus on the impact of PWV on cosmological measurements performed by the upcoming LSST. The use of SNe Ia as cosmological probes relies on them having homogeneous light-curves with standardizable luminosities at the time of peak brightness. By calibrating for intrinsic variations between the absolute magnitudes of individual SNe Ia, cosmological models are used to fit SN Ia distance moduli (μ) as a function of redshift (for example, see; Abbott et al. 2019; Betoule et al. 2014; Perlmutter et al. 1999; Riess, A. V. Filippenko, et al. 1998; D. M. Scolnic et al. 2018). However, as SN Ia samples grow larger, the variation in brightness between individual SNe no longer becomes the limiting issue. Instead, common systematics that change the effective average difference for large subsets of the sample become important. Photometric calibration, and a proper treatment of PWV, are thus key in achieving the goals of large surveys such as LSST.

The LSST will incorporate multiple calibration systems designed to estimate the atmospheric state and isolate systematic errors (Ingraham et al. 2016; Sebag et al. 2014). Among these systems is the Rubin Auxiliary Telescope (AuxTel; see LSE-379[4]), which will use state-of-the-art atmospheric models in conjunction with spectroscopic observations of bright, isolated stars to provide dedicated measurements of the atmospheric transmittance. By leveraging the positions of several stars across the sky, observations will be performed simultaneously with LSST providing a description of the atmosphere on a per-pointing ba-

[4]https://ls.st/LSE-379

sis. This will provide a complete description of the atmospheric behavior summed over its collective components and effects (e.g., Rayleigh scattering plus aerosol, O_2, O_3, and H_2O absorption).

Current plans for the LSST calibration system also include the use of a dual-band GPS receiver to determine the localized density of Precipitable Water Vapor (PWV) in the atmosphere (see Braun et al. 2001; Dumont et al. 2001; Nahmias et al. 2004). In addition to its integration with other systems (e.g. the LSST scheduler), the use of a dedicated GPS system allows for the real-time monitoring of PWV separate from the other atmospheric components. This is beneficial since the PWV absorption can vary significantly with time (quote number; see source) has a distinct and complex transmission function with lines that saturate even during relatively dry, *photometric* conditions (PWV ≈ 5 mm).

Existing works have already explored techniques for implementing the survey level calibration of LSST (David L. Burke, Axelrod, et al. 2010; David L. Burke, Saha, et al. 2014) and of LSST predecessor surveys (e.g., Pan-STARRS; Magnier et al. 2020, DES; D. L. Burke et al. 2018). We here instead focus on a much lower level of the LSST pipeline by creating a simulated LSST data set of SNe Ia with residual atmospheric effects that are left uncorrected by the global calibration strategy. These simulations will incorporate real-world atmospheric measurements to create realistic atmospheric variability in the simulated observations. The ultimate goal of this work is to establish a standing set of simulations that that will support future research into the incorporation of atmospheric calibration measurements into LSST subsystems.

3.2.1 Simulating PWV Effects in LSST SNe

To better understand the expected impact of atmospheric variability on LSST SNe, we create a simulated photometric data-set of SNe Ia as observed by LSST. In general, this involves describing both the intrinsic behavior of each observed target in addition to the observational conditions through which each observation is performed (e.g., the atmospheric transmission function). For our purposes here, we require:

1. A template for the spectro-photometric behavior of the observed targets and how the

SED of each target varies over time.

2. A distribution of parameters used to evaluate the given model, including the location, time, and filter through which each astronomical object is observed (i.e., the *cadence* of the simulated survey).
3. The wavelength dependent response curves for each photometric filter used by the given telescope, including contributions from the time-variable atmospheric transmission due to PWV.
4. A simulated reference star catalog used for calibrating the simulated flux values in each band.

Fortunately, many of the requirements listed above can be met with models already available in the literature. For convenience, and to facilitate easier comparison with existing works, we rely on existing models and datasets whenever possible.

3.2.1.1 LSST Filter Response Curves and Atmospheric Transmission

The filter response curves used in our simulations are shown in Figure 22. For the optical throughput of each filter, we use baseline estimates published by the LSST systems engineering team[5]. The left side of Figure 22 details the contribution of the LSST hardware to the overall filter response. Although the effective throughput of these components may change slightly over the course of LSST (e.g., through the accumulation and eventual cleaning of dust on reflective surfaces), hear we treat the throughput of each hardware component as being constant with time. However, the complete shape of the filter response curve includes the absorption of light due to the atmosphere, and therefore remains time dependent.

We note that the overall calibration of the response curves is irrelevant when fitting SN Ia, as any uniform offset drops out when discussing changes in relative brightness or color. This means that the contribution of time independent components does not have a significant effect so long as they are the same for all observations. Since our primary concern is changes in the shapes of the filter curves that vary with time, we ignore atmospheric effects induced by non-PWV components of the atmosphere. Not only are these effects relatively smooth with wavelength, but they vary on significantly longer timescales than PWV and

[5]Version 1.5; See https://github.com/lsst/throughputs for more details.

play a subdominant effect. We instead narrow our focus exclusively to the time-variability of PWV.

As a model for the PWV transmission function, we used the same model discussed in Section 3.1 and shown in Figure 15. Evaluating this model requires knowledge of how the PWV concentration above the Rubin Observatory changes over time. To quantify this temporal evolution, we use GPS PWV measurements taken at the nearby Cerro-Tellolo Inter-American Observatory (CTIO). GPS measurements for CTIO are available intermittently from 2013 through 2018. Shown in Figure 23, measurements taken over 2016 and 2017 provide the most complete temporal coverage.

The CTIO GPS measurements do not span a long enough period of time to directly evaluate an atmospheric model for the full, 10 year LSST. Instead, we use the CTIO measurements to create a model for how the PWV concentration varies across seasons. To accomplish this, we used PWV measurements taken during 2016 and supplement any missing data with measurements taken during 2017. This supplemented data set is herein referred to as the *PWV Model*. Shown in Figure 24, the PWV concentration at zenith is determined for a given observation by interpolating from the PWV model for the corresponding time of year. The PWV concentration along the line of sight is then calculated using Equation 3.5.

3.2.1.2 SN Ia Spectral Template As a model for the spectroscopic behavior of SNe Ia, we use a version of the Salt2 SN model (J. Guy et al. 2007) which has been extended by Pierel et al. 2018 further into the UV and near-IR[6]. This choice is primarily motivated by the already wide-spread use of Salt2 in the literature. Furthermore, the broader wavelength coverage of the extended model allows SNe Ia to be simulated to a higher redshift ($z \approx 0.82$) than the original Salt2 model ($z \approx 0.55$), beyond which the underlying SED template no longer spans bluer filters such as the LSST u band.

The Salt2 template is evaluated using the wavelength (λ) plus five parameters:

$$F_{spec}(\lambda, t) = x_0[M_0(\lambda, t - t_0) + x_1\, M_1(\lambda, t - t_0)]e^{-c\,CL(\lambda)} \tag{3.7}$$

[6]The spectral template corresponding to this model in available in the `sncosmo` package under the alias `salt2-extended`. See `https://sncosmo.readthedocs.io/` for the official documentation.

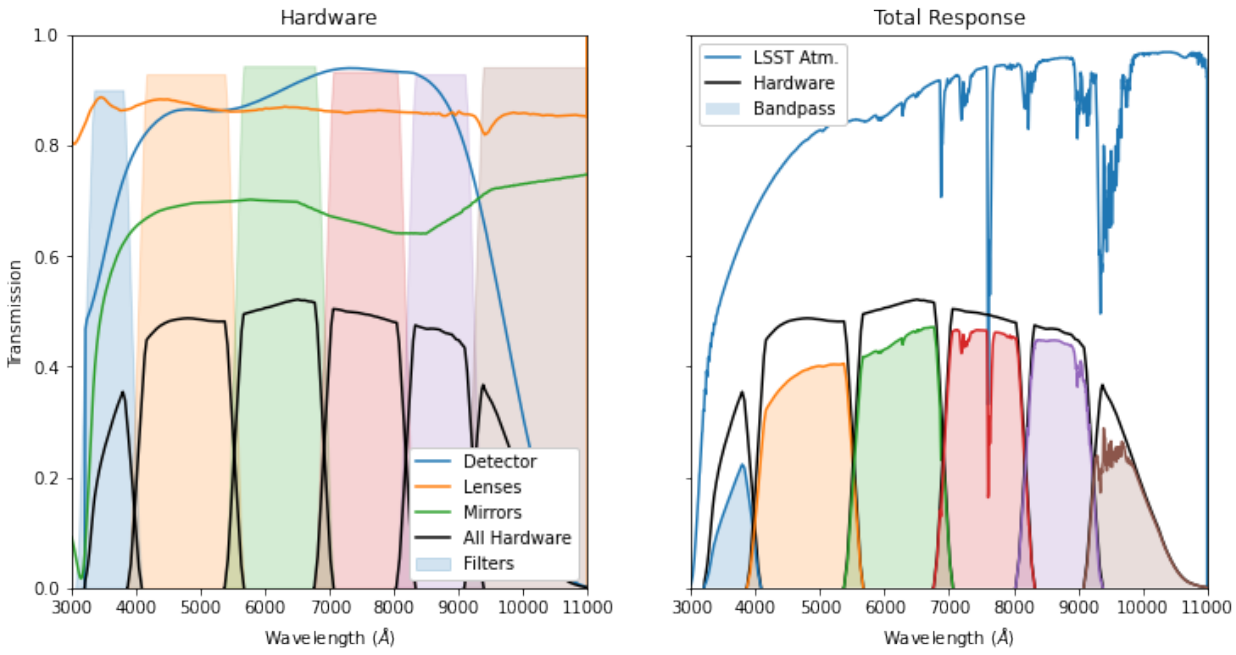
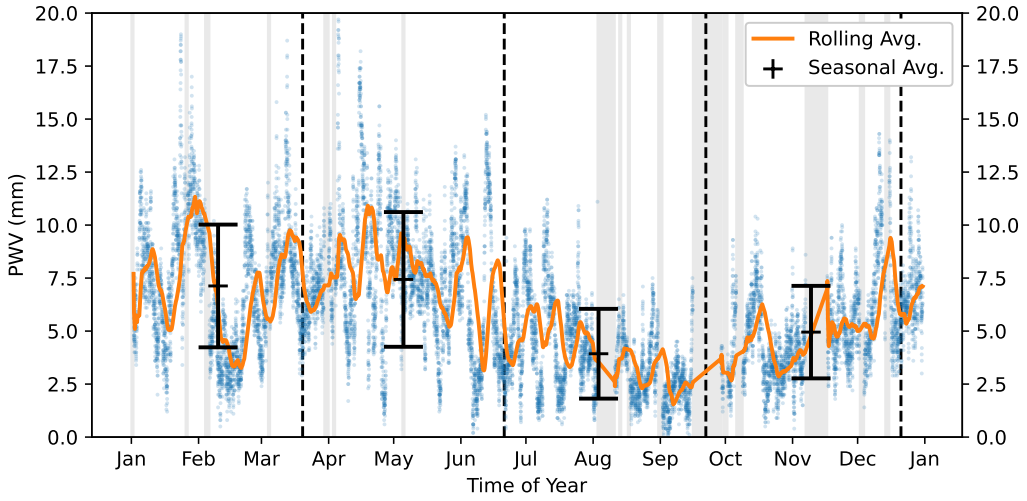
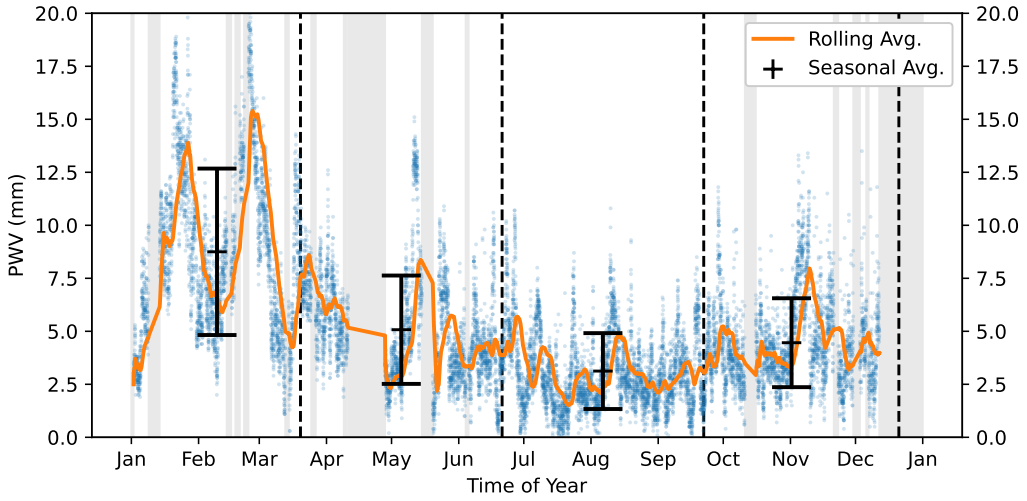


Figure 22: The hardware contribution to the LSST filter response curves (left) is compared against the total filter response when including the combined atmospheric transmission (right). The complete fiducial atmosphere estimated by the LSST systems engineering team is shown in the right hand figure as a solid blue line. This includes the absorption of light due to PWV, with a PWV concentration of ∼4 mm (plus absorption due all other non-PWV components).



(a) CTIO PWV measurements taken over 2016.



(b) CTIO PWV measurements taken over 2017.

Figure 23: GPS based PWV measurements are publicly available for the Cerro-Tellolo Inter-American Observatory from 2013 through 2018. Out of these years, 2016 (top) and 2017 (bottom) provide the most complete coverage. Gaps of missing data (highlighted in grey) are caused by maintenance and poor weather conditions. Each of the four seasons are separated by a vertical dashed line. Note the average PWV changes significantly between seasons, and that the overall standard deviation of PWV within each season (represented by black error bars) is significant.

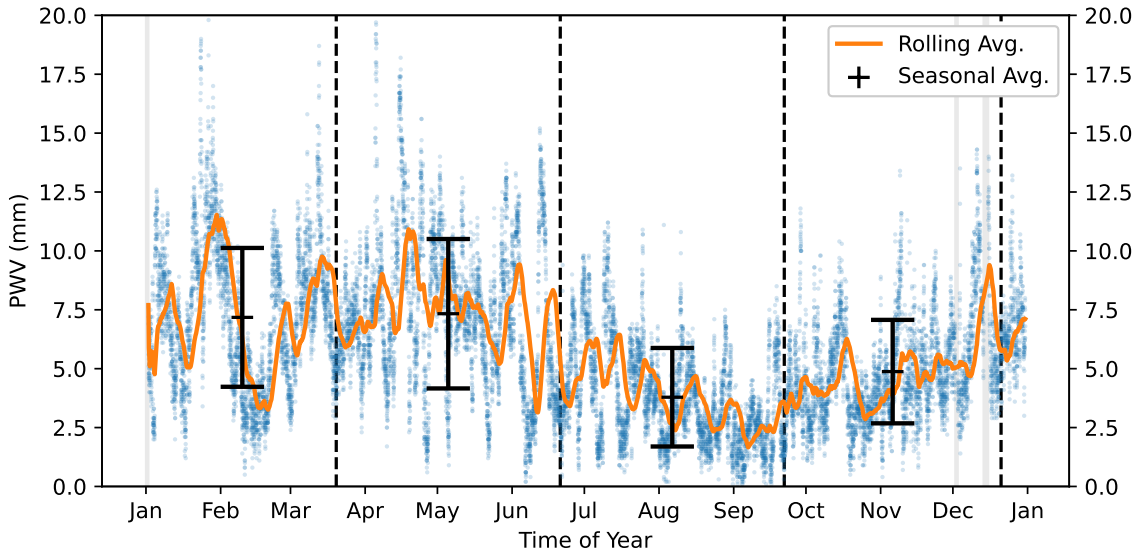


Figure 24: Our model for the PWV concentration at Zenith above CTIO is created by supplementing data from 2016 with data from 2017. To evaluate the model, the PWV is interpolated for using a given time of year. This PWV model is used as a direct estimate for the PWV above the Rubin Observatory due to the close proximity of the two surveys.

Here M_0 and M_1 each provide a two dimensional surface describing how the spectral flux varies with wavelength and time. The M_0 term is used to represent the fiducial SED of a SN Ia while M_1 provides a first order correction that provides for some intrinsic variation in the SN Ia population. The x_0 parameter acts as an overall scale factor, whereas x_1 only scales the contribution of M_1. The parameter c varies the ratio of flux in each band according to an assumed color law (CL), while t_0 represents the time of peak brightness of the explosion in the B band.

To simulate broad-band photometric observations, the spectral template is evaluated for a set of given parameters and multiplied by the response curve of each filter. This product is then integrated across all wavelengths to determine the corresponding broad-band flux. The extended Salt2 template is shown for three different redshifts in Figure 25. As the spectral template is redshifted to further distances, the broad absorption features of the template begin to overlap with the complex absorption profile of PWV, changing the overall magnitude of absorbed flux. The impact of PWV on simulated flux values is thus dependent on not just the PWV concentration itself, but also on the parameters used to evaluate the spectral template.

3.2.1.3 Input Cadence Data To evaluate our chosen model, we rely on the parameter distributions used in the Photometric LSST Astronomical Time-series Classification Challenge (PLAsTICC; The PLAsTiCC team et al. 2018). Simulations performed by the PLAsTICC team include the evaluation of multiple SNe models (including Salt2) using multiple proposed cadences for LSST (for more information on proposed cadences, see Lochner, Dan Scolnic, et al. 2021). Although the PLAsTICC data already includes pre-constructed SNe Ia light-curves, these light-curves include a predefined atmospheric model and are thus not suitable for our current purpose. We instead rely only on data from the PLAsTICC simulations related to the overall cadence. This includes the time and on-sky position (Ra, Dec) of each photometric observation and Salt2 parameters for approximately 4 million SNe Ia. Our implementation of the overall simulation procedure is flexible and can be run for any of the cadences simulated by the PLAsTICC team. We include with this work simulations run for all of the included cadences.

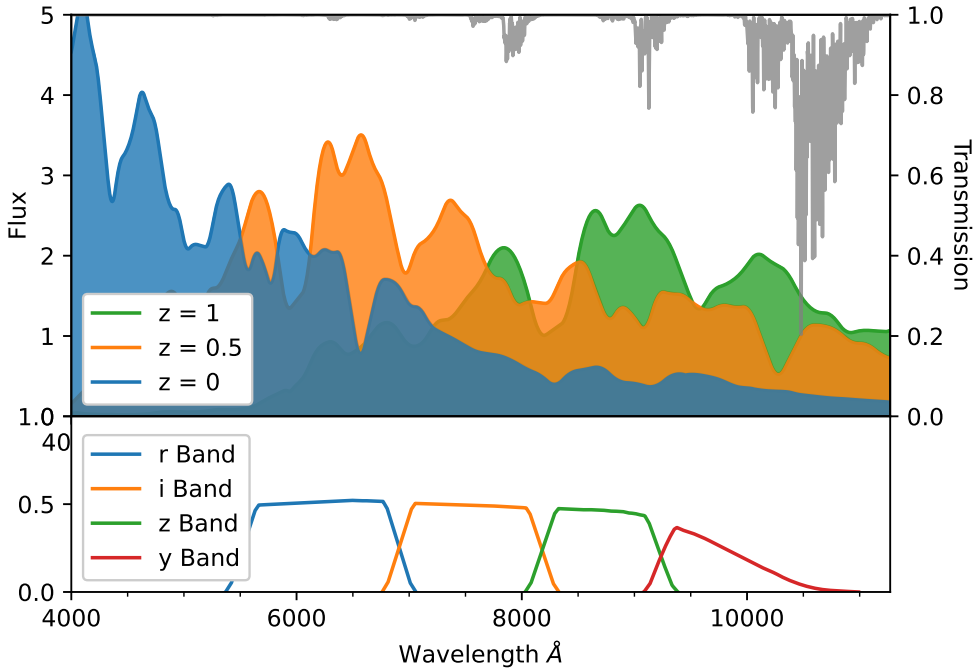


Figure 25: The spectral template of the Salt2 supernova model for a redshift of 0.1 (blue), 0.5 (orange), and 1 (green) against the PWV transmission (grey) for a PWV concentration of 4 mm. SN Ia spectra are dominated by several broad absorption features, many of which have equivalent-widths similar to the widths of PWV absorption features in the zy bands. As the template spectrum is shifted to higher redshifts, different spectral features begin to overlap with the strong PWV absorption feature at $\lambda \approx 9500$ Å.

3.2.1.4 Reference Star Catalog

The calibration of photometric observations using a reference catalog is discussed in Section 3.1.1. To provide for this calibration procedure in our simulations, we use stellar spectral templates from the Goettingen Spectral Library (Husser et al. 2013). For each simulated SN observation, we assume there exists at the same airmass three stars with spectral types G2, M5, and K2. An example stellar spectrum is shown for a G2 star in Figure 26.

The broadband flux of these stars is simulated using the same approach taken in our simulation of SN fluxes. The spectral template of each star is multiplied by the filter response curve in each band (which includes the the time dependent absorption of light scaled by the airmass of the observation) and integrated over all wavelengths. This simulation process is then repeated without the presence of atmospheric effects to determine the ratio in flux decrease for each star. These ratios are averaged across all three stars and used to correct for the simulated atmospheric effects in each SN observation.

3.2.2 Validation and Discussion of Simulation Results

It is important for both the validation of our pipeline and the interpretation of its outputs to understand how the effects of PWV absorption propagate through each step in a SN analysis. Shown in Figure 27, we demonstrate a light-curve simulated using the proposed **alt_sched** cadence for LSST[7]. When determining the peak brightness of a SNe Ia, the apparent peak magnitude is estimated by fitting the observed flux with a spectroscopic template (such as Salt2). The quality and resulting parameters from this fit are sensitive to the temporal sampling of the observed light curve. For example, the availability of observations within $\sim$10 days of peak brightness is crucial to estimating the time of maximum brightness (t_0).

For our present purpose, it is important to isolate the impact of PWV variability and study it separately from any effects introduced by the overall cadence. To accomplish this, we run a second set of light-curve simulations demonstrated in Figure 27b. The dense temporal

[7]At present, the official cadence that LSST will use is still being determined. The cadence chosen here is one of many that have been submitted for consideration, and was selected arbitrarily for the purpose of demonstrating simulation results. An overview of other proposed cadences for LSST is provided by Lochner, Dan Scolnic, et al. 2021.

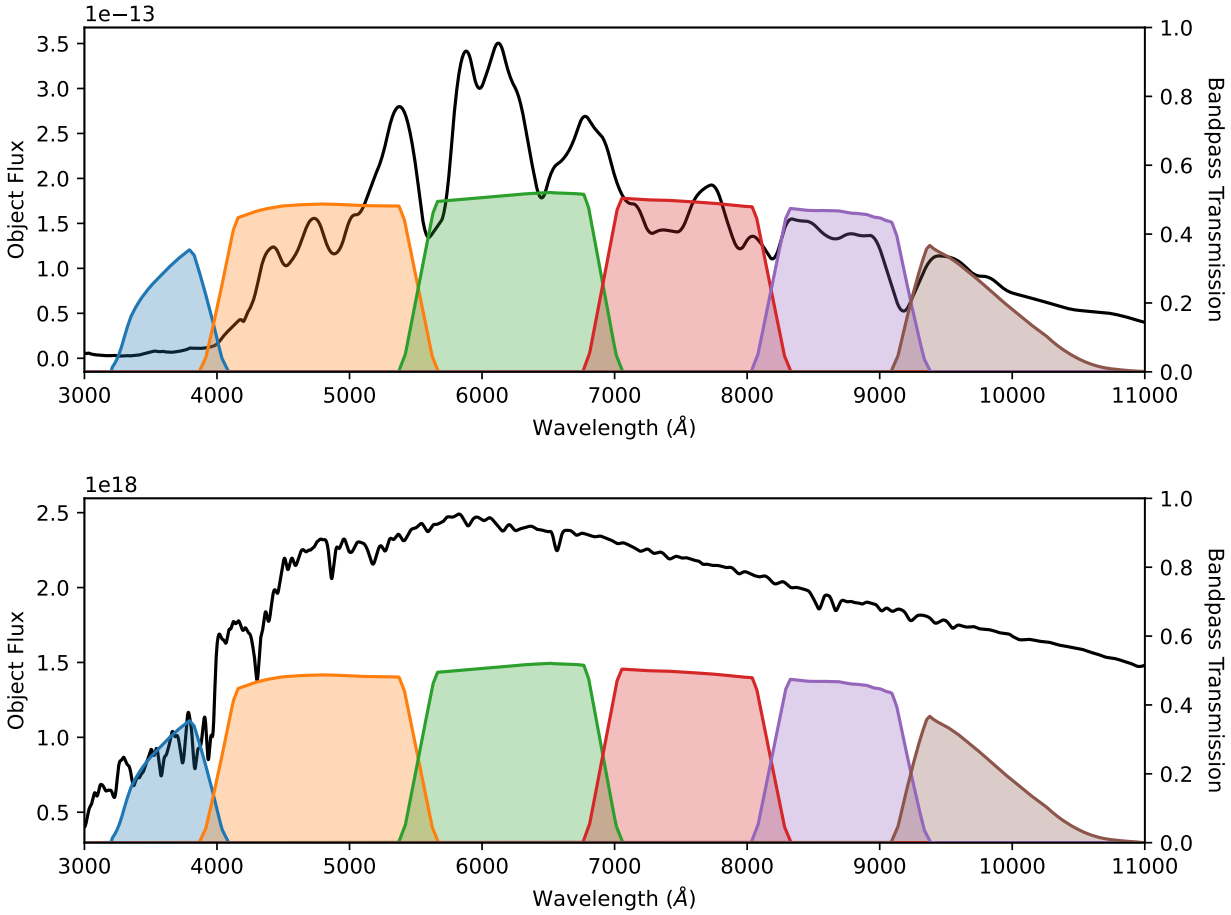


Figure 26: The SED of a SN Ia (top) at redshift $z = 0.5$ is plotted over the LSST hardware throughput in each filter and is compared against the SED of a G2 type star (Bottom) that has been binned to a similar resolution. Note that the overall shape of both SEDs are significantly different, which causes PWV to impact the observed broad-band flux in each filter (plus the corresponding colors) differently.

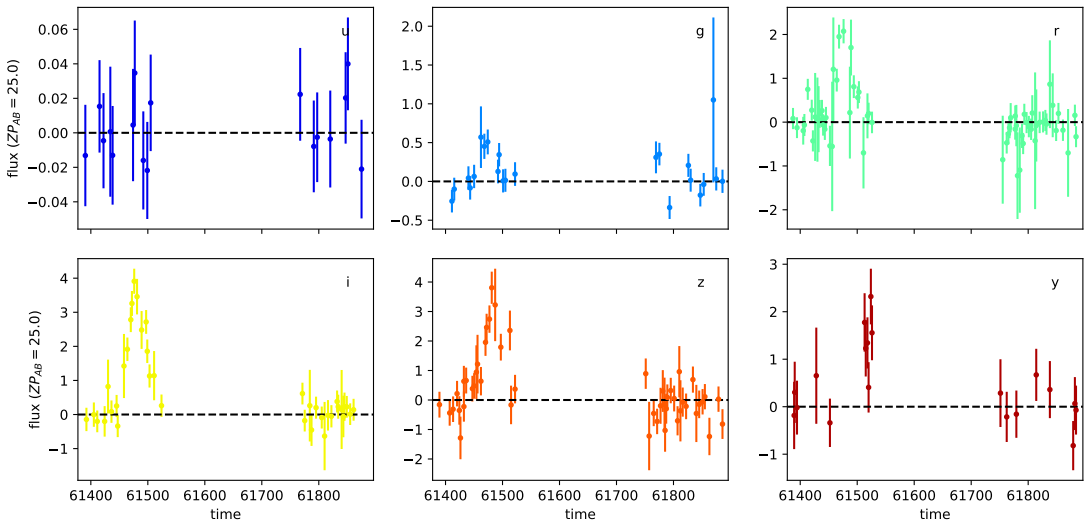
sampling across multiple bands tightly constrains the chi-squared minimization process when fitting a spectral-template. To further simplify our analysis, we restrict our simulations to include no variation in light-curve stretch (i.e., we fix $x_1 = 0$) and color ($c = 0$), set the airmass of each observation to 1 (zenith), and set the PWV concentration to a constant value for each light-curve.

Traditionally, SNe are fit using an estimation of the filter response curves that do not vary with time. Instead of relying on estimates for the time variability of the atmosphere (which isn't always available), the average atmospheric state is used instead. The atmospheric transmission function used in estimating the fiducial LSST filters assumes a PWV concentration of $\sim$4 mm. For consistency with this choice, we discuss the impact of atmospheric effects relative to a 4 mm concentration of PWV along the line of sight.

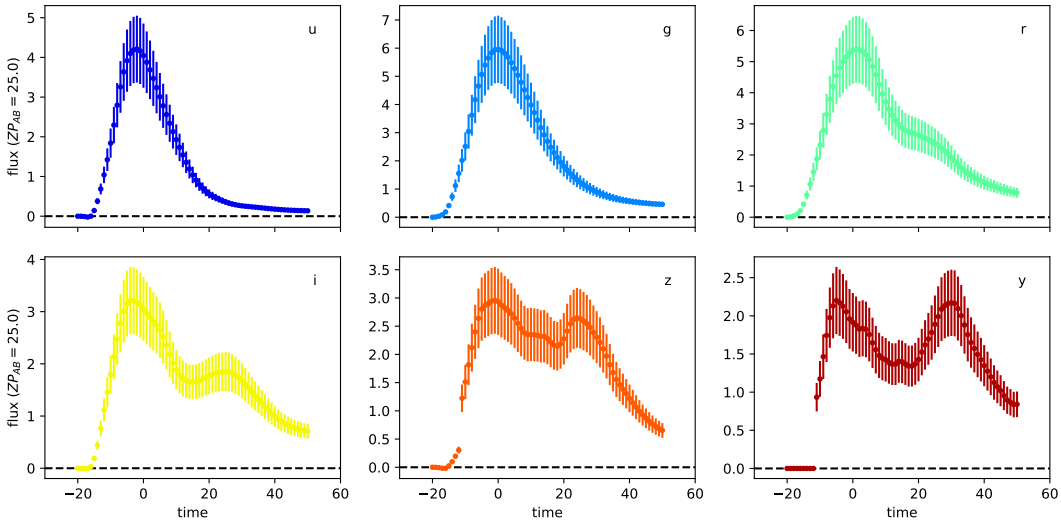
In Figure 28 we demonstrate the change in fitted magnitude for each band as a function of redshift and PWV. We note that despite PWV absorption predominately affecting redder wavelengths, there is a significant change in the fitted magnitudes for all bandpasses. This can be attributed to the assumption of a fiducial atmosphere with a fixed PWV concentration in the fitting process. The PWV absorption profile assumed by the spectral template is different from the actual PWV transmission used in the simulated light-curve. The χ^2 minimizer treats this difference as being intrinsic to the SN itself and modifies the color of the template to minimize the χ^2.

We also note that the change in fitted magnitude for all bands share a complex structure as a function of redshift. We attribute this complex structure with redshift in part to the broad features of SN Ia spectra moving across the water vapor absorption features. However, another driving factor is the differences between how the SED's of a SN and a reference star are impacted by PWV absorption atmosphere. As shown in Figure 26, the SED of a star is significantly different from that of a SN Ia. These differences result in an incomplete correction of atmospheric effects when calibrating a SN relative to the reference star.

Figure 29 shows the residuals in the distance moduli as determined from the simulated SNe. Note that this makes the unphysical assumption that PWV is constant for all points in the SN Ia light-curve. In realistic conditions, the PWV would vary, and these biases would be averaged down. However, a season dependence of average PWV would translate into a



(a) A SN Ia light-curve simulated using the alt_sched cadence.



(b) A SN Ia light-curve simulated using an over-idealized, highly sampled cadence.

Figure 27: An example light-curve simulation is shown for the alt_sched cadence (left) and for an over-idealized, highly sampled cadence (right). Depending on the final cadence strategy adopted by LSST, the left light-curve represents a realistic expectation for the performance of LSST. However, oversimplified light curves, like the one on the right, are a valuable tool for validating analysis and simulation procedures. Their dense sampling eliminates any effects a more sparse sampling may have on the χ^2 minimizer when fitting a spectroscopic template.

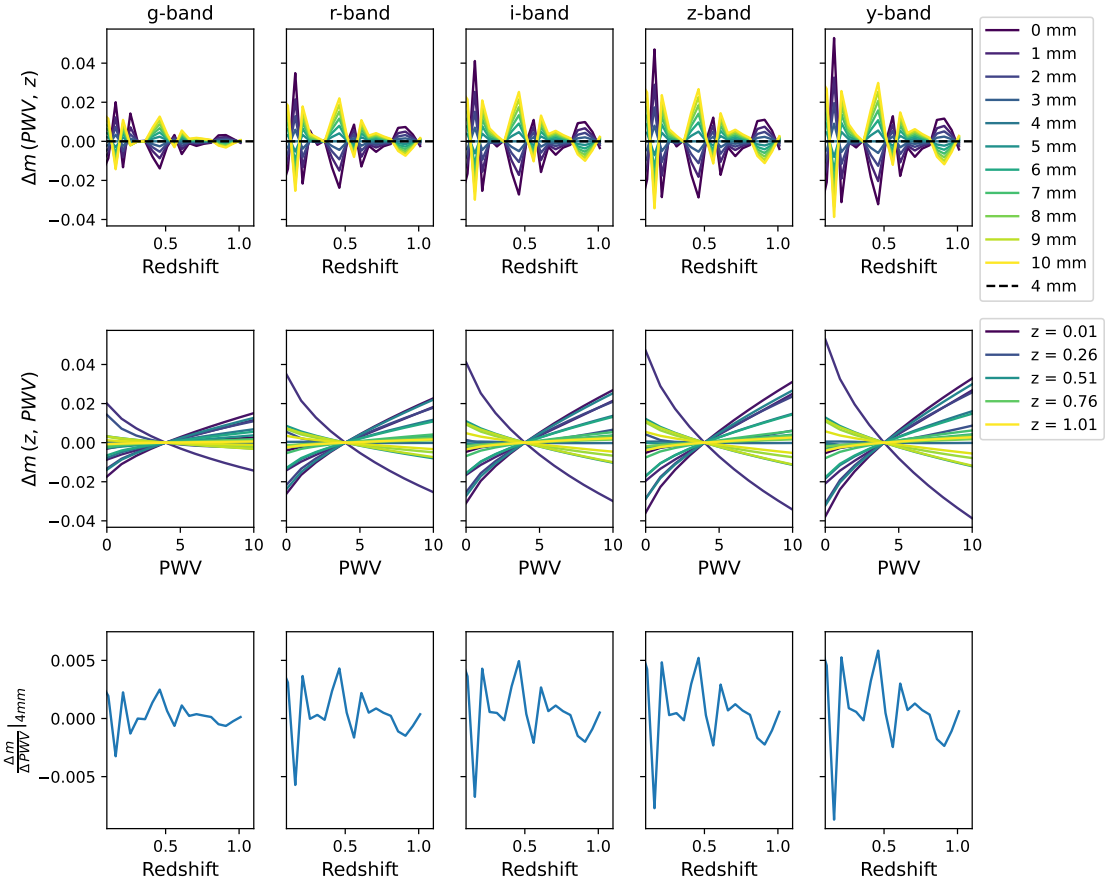


Figure 28: The change in fitted magnitude is shown for a series of simulated SNe, each simulated with a different redshift and PWV concentration. Results are plotted for each band against the simulated redshift (top) and simulated PWV (bottom). The slope of the change in magnitude with PWV is also shown in the bottom row.

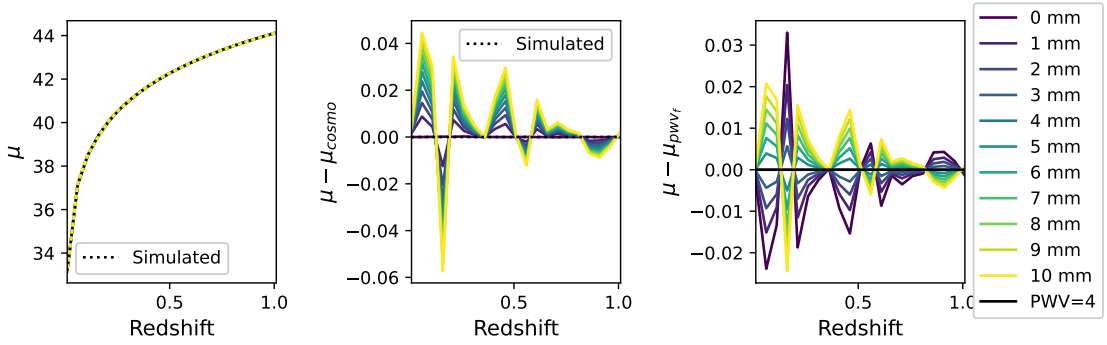


Figure 29: Distance modulus residuals ($\Delta\mu$) determined from a set of simulated SNe Ia, each suffering from a different level of PWV absorption (assumed constant for the full light curve). Residuals are shown both relative to the underlying cosmology of the simulation (center) and to a fiducial atmosphere of 4 mm PWV (right). The pattern with redshift comes from the broad features of SN Ia spectra moving across the water vapor absorption features.

bias as a function of RA, as there is a correlation between average PWV and observed RA. Even for relatively dry, photometric conditions (PWV $\lesssim 4$), uncorrected PWV absorption introduces a bias of up to 0.02 mag in the estimated μ. A similarly sized shift is also found for low redshift targets suffering from PWV absorption that varies by more than a few mm away from the assumed fiducial atmospheric value. This is an indicator that the assumption of a fiducial transmission model that is constant from image to image, such as when constructing the effective bandpass throughputs for a given survey, is not sufficient for achieving the millimag level of precision anticipated from future large scale surveys.

We note that our results are similar in size to Hubble residuals found in existing cosmological analyses. For example, after assuming a mean atmospheric absorption model to calibrate instrumental magnitudes, the analysis of Betoule et al. 2014 found average residuals in the fitted μ of approximately 0.01 and 0.06 mag for their low and high redshift SN samples, respectively. Calibration uncertainties were the largest contributing factor in measurement uncertainty and made up over 36% of their overall uncertainty.

4.0 Supplemental Software

As the size and diversity of astronomical data sets continue to grow, the use of custom built software has become increasingly commonplace. LSST is scheduled to run for a ten year period, and the analysis of its data products will likely continue for many years afterwords. The documentation and maintenance of software built for science efforts like LSST is crucial for ensuring published results can be reliably reproduced. It is also an important step in supporting future research efforts by reducing the need to rebuild previous functionality from scratch.

The following sections document custom software built to support research presented in this work. All presented software is open source and free for public use with attribution. A copy of the most up to date documentation is also available online (see the documentation of each software package in the following sections for more details).

4.1 SNData: Data Access for Supernova Data Releases

In recent years the astronomical community has made several attempts to build and maintain a centralized system for storing and accessing astronomical data products (for example, see Guillochon et al. 2017; Ochsenbein et al. 2000). Although some of these efforts have been met with relative success, there does not presently exist a standardized data-access software used by the supernova research community at large. Custom data access solutions are typically implemented on an as-needed basis by individual researchers, often resulting in multiple (sometimes incompatible) solutions being used within a single research team or project. This lack of a unified system results in added overhead and longer on-boarding times for new team members that are required to learn the functionality of multiple APIs.

The `sndata` package is built to provide a unified data-access API for data releases published by a variety of supernova (SN) surveys. Unlike existing data access solutions, the

package is designed to prioritize usability over scope or completeness[1]. A summary of data sets accessible through `sndata` is provided in Table 6. At the time of writing, `sndata` provides access to ten different data sets published by eight different supernova surveys. These surveys were chosen either by user request or as needed by the development team to supporting ongoing research projects.

No attempt was made during development to migrate third party data into a single, proprietary storage system. Instead, the package automates the processes of downloading data from the original publication into the working environment and provides for the parsing of that data. The details of this process are abstracted away from the user using a modular, object oriented design, allowing developers to instead focus on building scalable analysis pipelines that translate with minimal effort between and across data sets.

For the most up-to-date documentation of the `sndata` package, see https://sndata.readthedocs.io/en/latest/.

4.1.1 Installation and Setup

The `sndata` package is publicly available via PYPI, and can be installed using the `pip` command line tool:

```
pip install sndata
```

The `pip` package manager will automatically install any missing dependencies in the host Python environment. If `pip` is unavailable on the target system, the package source code is publicly available online and can be download directly from the GitHub software repository service:

```
# Clone the source code from GitHub
git clone https://github.com/sncosmo/sndata/ sndata

# Install the package from source
cd sndata
python setup.py install --user
```

By default, data downloaded by `sndata` is stored in the installation directory. This allows

[1] Although in general, these two concepts are often highly correlated when discussing data access solutions.

multiple `sndata` installations to independently manage their downloaded data at the cost of redundant data being potentially written to different locations on disk. Alternatively, the package can be configured to store data in a customized location by specifying the `SNDATA_DIR` variable within the working environment (e.g., in a `.bash_profile` or `.bashrc` file). For example:

```
export SNDATA_DIR="/your/data/directory/path"
```

4.1.2 Basic Data Access

`sndata` is designed such that individual data releases are represented as dedicated classes, each of which is subclassed from the abstract `SpectroscipicRelease` or `PhotometricRelease` classes depending on the type of published data. Data releases published by the same research team / astronomical survey are grouped together into package modules named using the abbreviation of that survey. To access data from a specific release, import and initialize the corresponding data access class. For demonstration purposes we will be using the third data release from the Carnegie Supernova Project (CSP):

```
from sndata.csp import DR3

# The DR3 class locates data on your machine at instantiation
dr3 = DR3()

# Information about the parent survey
print(dr3.survey_name)
print(dr3.survey_abbrev)

# A summary of the DR3 data set
help(dr3)

# Where to go for more information
print(dr3.survey_url)

# The type of data in this release
print(dr3.data_type)

# The primary publication(s) and NASA ADS link(s) describing the data
print(dr3.publications)
print(dr3.ads_url)

# Photometric data releases include filters
print(dr3.band_names)
```

All classes adhere to the same user interface design whenever possible (i.e., each class provides the same core methods, attributes, and functionality). A summary of each data release, including any (minor) deviations from the standard UI, can be accessed by calling the builtin help function, or by visiting the online package documentation[2].

4.1.2.1 Downloading Data

To minimize disk space usage, `sndata` does not come pre-installed with any survey data. Instead, users must manually tell `sndata` to download (or delete) data from their local machine:

```
# Download data for the given survey / data release
dr3.download_module_data()

# Delete any downloaded data for the given survey / data release
dr3.delete_module_data()
```

It is useful to note that any data already downloaded to the local environment is skipped when calling `download_module_data`, making the function safe to use in a automated/repetitive environment (such as at the beginning of a frequently run script). This behavior can be disabled by specifying the `force=True` argument, in which case all data remote data will be downloaded and any local data will be overwritten.

`sndata` downloads data directly from the source of publication. Since survey data is often hosted across multiple servers, it is possible the server responsible for hosting a subset of a data release (e.g., the filter transmission curves) is temporarily offline. In this case, `sndata` will raise a warning and continue downloading any data that is still accessible. The `download_module_data` function can then be re-run once the server is back online.

4.1.2.2 Parsing Data

Observational data can be retrieved for individual astronomical objects using a unique identifier value specified as a string. By default, `sndata` uses the same object Id values as specified in the corresponding data set. Returned data is provided as an `astropy.Table` object; a choice ultimately motivated by the wide spread use of the `astropy` package (Price-Whelan et al. 2018) by the astronomical community. For convenience, the right ascension, declination, redshift, and redshift error are included in the table's meta data

[2]See https://sndata.readthedocs.io/en/latest/.

when available.

```
# Get a list of available objects
list_of_ids = dr3.get_available_ids()

# Get data for a given object
demo_id = list_of_ids[0]
data_table = dr3.get_data_for_id(demo_id)
print(data_table)

# Don't forget to check the meta data!
print(data_table.meta)
```

Data tables returned by `sndata` are automatically formatted for use with the `sncosmo` python package (Barbary et al. 2016); another package commonly used by the supernova cosmology community (see Section 4.1.2.3 for more details). In doing so, the values of the table may be manipulated from the original file data into different units, column names, etc.. To disable this feature, specify the `format_table=False` argument.

It is not uncommon for an analysis to iterate over all astronomical objects included in a given data release. When faced with this requirement, the `iter_data` function is provided for convenience:

```
for data in dr3.iter_data():
    ...  # Custom analysis logic here
```

The `iter_data` function also allows users to optionally select a subset of the yielded data using a user defined filter function. This function should accept a data table yielded by `iter_data` and return a boolean: `True` if the data table should be returned by the iterator and `False` if the data should be skipped. For example, to only select observed targets with a redshift less than .1:

```
def filter_func(data_table):
    return data_table.meta["z"] < .1

for data in dr3.iter_data(filter_func=filter_func):
    ...  # Custom analysis logic here
```

We note that the `iter_data` function reads reads in data from disk before checking the filter function. For this reason, filter functions should not be used in an attempt improve I/O related bottlenecks.

Some surveys include supplementary tables in their data releases. These tables often include additional information such as object meta-data or summary statistics. The inclusion of tables from published papers in a data release is also common. Any supplementary tables included in a data release are listed by their unique identifier using the `get_available_tables` function and can be retrieved by calling the `load_table` function as follows:

```
# Retrieve a list of unique IDs for any available tables
published_tables = dr3.get_available_tables()

# Read one of those tables by referencing the table name or number
demo_table_name = published_tables[0]
demo_table = dr3.load_table(demo_table_name)
```

Note that the `load_table` function caches any returned tables in memory. This improves the speed of successive calls and limits overhead from repeated I/O operations.

4.1.2.3 SNCosmo Compatibility

The `sncosmo` package (Barbary et al. 2016) is a widely used Python library for supernova cosmology analysis. Although a full overview of `sncosmo` is outside the scope of this work, the `sndata` package is built to be compatible with `sncosmo` and has been adopted as an official affiliate package. For this reason, we provide a brief overview of the cross package compatibility here.

`sndata` automatically formats data into a data model recognized by `sncosmo`. This includes renaming table columns and converting data into common units. Furthermore, `sndata` is able to register the filter transmission curves for a given survey into the `sncosmo` registry[3] (the system by which `sncosmo` keeps track of what each filter, model, etc., are called). This allows `sncosmo` to automatically load filter profiles when fitting observational data with a given supernova model. Filter profiles can be registered with `sncosmo` using the `register_filters` function:

```
# The names of the bands that will be registered
print(dr3.band_names)

# Register the band-passes of the survey with SNCosmo
# You can optionally specify 'force=True' to re-register band-passes
dr3.register_filters()
```

[3]Filter profiles are not available for data releases that do not include photometric data. For example, spectroscopic data releases will raise an error when trying to register filter profiles.

Once a survey's filters are registered, the observed flux values can be fit with a supernova model as follows:

```
import sncosmo

# Get data for supernova 2004dt
data_table = dr3.get_data_for_id("2004dt")
print(data_table)

# Fit the data
model = sncosmo.Model("salt2")
model.set(z=data_table.meta["z"])
fit_result, fitted_model = sncosmo.fit_lc(
    data=data_table,
    model=model,
    vparam_names=["t0", "x0", "x1", "c"])
```

4.1.3 Combining Data Sets

`sndata` allows users to combine individual data releases into a single `CombinedDataset` object. The resulting object provides the same general user interface as a single data access module, but provides access to data from multiple surveys / data releases. To create a combined data set, import the data access classes for each of the data releases you want to join and pass them to the `CombinedDataset` object at instantiation. For demonstration purposes, we combine data from the third CSP data release (`csp.DR3`) and the three year cosmology release of the Dark Energy Survey (`des.SN3YR`):

```
from sndata import CombinedDataset, csp, des

combined_data = CombinedDataset(csp.DR3(), des.SN3YR())
```

The resulting object provides the same user interface as the rest of the `sndata` package, including having the same method for accessing tables:

```
 # Download all data for the combined data releases
 combined_data.download_module_data()

 # Load a supplementary table
 list_of_table_ids = combined_data.get_available_tables()
 demo_table_id = list_of_table_ids[0]
 demo_sup_table = combined_data.load_table(demo_table_id)
```

The access pattern for individual object data is also the same:

```
# Get a list of available objects
list_of_ids = combined_data.get_available_ids()

# Get data for a single object
demo_id = list_of_ids[0]
data_table = combined_data.get_data_for_id(demo_id)
print(data_table)

# Iterate over data for all objects in the combined data set
for data in combined_data.iter_data():
    ...  # Custom analysis logic here
```

It is important to note that the format of object and table Id's for `CombinedDataset` objects is slightly different than for a single data release. Unlike the object and table Id's for a single data release, the default Id's for a `CombinedDataset` are tuples instead of strings. Each tuple contains three elements including (in order) the individual object identifier, data release name, and survey name. For example, the ID value for supernova `2007S` from CSP DR3 would be `("2007S", "DR3", "CSP")`. By specifying object Id's in this way, it is ensured that objects in combined data releases always have unique identifiers. However, in the case where the object Id's from two data releases are already unique (as is the case when combining `csp.DR3` and `des.SN3YR`), `CombinedDataset` objects are smart enough to mimic the behavior of a normal / single data release and can take object Id's as strings. For example:

```
# You can specify object ID's as tuples
combined_data.get_data_for_id(("2007S", "DR3", "CSP"))

# or if the object names across the joined surveys are unique, as a string
combined_data.get_data_for_id("2007S")
```

It is possible for two different data releases to include the same astronomical object. In this case, object Id's from different surveys can be "joined" together so that when requesting data for a given object Id, data is returned for all Id's that have been joined together. Accomplishing this is as simple as:

```
# Note that you can join an arbitrary number of object Id's
combined_data.join_ids(obj_id_1, obj_id_2, obj_id_3, ...)
print(combined_data.get_joined_ids())

# To undo the above joining action
combined_data.separate_ids(obj_id_1, obj_id_2, obj_id_3, ...)
```

When retrieving data for a joined ID, the returned data table is simply the collective data tables for each joined ID stacked vertically. It is also worth noting that CombinedDataset objects are aware of successive join actions. This means that the following two examples are functionally equivalent.

```
# You can join multiple Id's at once
combined_data.join_ids(obj_id_1, obj_id_2, obj_id_3)

# Or join them successively
combined_data.join_ids(obj_id_1, obj_id_2)
combined_data.join_ids(obj_id_2, obj_id_3)
```

4.1.4 Creating Custom Data Classes

`sndata` uses an object oriented design to automate basic data management tasks and ensure a consistent user interface. Prebuilt template classes are provided to represent spectroscopic or photometric data, allowing users to create customized data access objects. In general, the steps for creating a new data access class include:

1. Inheriting from one of the abstract `SpectroscopicRelease` or `PhotometricRelease` template classes
2. Defining some meta data describing the data release
3. Defining a few private functions for parsing the new data set

In the following sections we work through examples of building data access classes for photometric and spectroscopic data.

4.1.4.1 Spectroscopic Data To create a data access class for spectroscopic data, we start by inheriting from the `SpectroscopicRelease` class. We then add some meta data about the survey and data release that is being represented.

```python
from astropy.table import Table
from sndata.base_classes import SpectroscopicRelease
from sndata import utils


# Class should be named using the official acronym of the data release
# If an acronym is not available, the publication source can be used
class Smith20(SpectroscopicRelease):
    """Describe the contents of the data release here
    (Source: Smith, Jhonson et al. 2020)

    Deviations from the standard UI:
        - None

    Cuts on returned data:
        - None
    """

    # General metadata
    survey_name = "Demo Supernova Survey"  # Full survey name
    survey_abbrev = "DSS"  # Official survey abbreviation
    release = "Smith20"  # Name of the data release
    survey_url = "www.url.com"  # URL of the official data release
    publications = ("Smith et al. 2020",)  # Any relevant publications
    ads_url = "https://ui.adsabs.harvard.edu/"  # Link to publication
```

Next we use the `__init__` method to define the local paths of the data being represented and, if desired, the remote URL's the data should be downloaded from. All data should be stored in a sub-directory of the `self._data_dir`, which is determined automatically by `sndata`.

```python
    def __init__(self):
        """Define local and remote paths of data"""

        # Call to parent class defines the self._data_dir attribute
        # All data should be downloaded to / read from that directory
        super().__init__()

        # Local data paths. You will likely have multiple directories here
        self._spectra_dir = self._data_dir / "spectra_dir_name"

        # Define urls for remote data.
        self._spectra_url = "www.my_supernova_spectra.com"
```

Finally, the logic for downloading and parsing individual data files is added as private methods. These private functions are ultimately called by their public counterparts defined by the parent class, ensuring a consistent, public facing user interface.

```python
    def _get_available_tables(self) -> List[int]:
        """Get Ids for any supplemental tables data tables"""

        # Find available tables - for example:
        # Find available tables - assume standard Vizier naming scheme
        table_nums = []
        for f in self._table_dir.rglob('table*.dat'):
            table_number = f.stem.lstrip('table')
            table_nums.append(int(table_number))

        return sorted(table_nums)

    def _load_table(self, table_id: int) -> Table:
        """Return a Vizier table published by this data release

        Args:
            table_id: The published table number or table name
        """

        readme_path = str(self._table_dir / "ReadMe")
        table_path = str(self._table_dir / f"table{table_id}.dat")

        # Read data from file and add meta data from the readme
        data = ascii.read(table_path, format="cds", readme=readme_path)
        description = utils.parse_vizier_table_descriptions(readme_path)
        data.meta["description"] = description[table_id]
        return data

    def _get_available_ids(self) -> List[str]:
        """Return a list of target object IDs for the current survey"""

        # Returned object Ids should be sorted and unique.
        # For example:
        files = self._spectra_dir.glob('*.dat')
        return sorted(Path(f).name for f in files)

    def _get_data_for_id(
        self, obj_id: str, format_table: bool = True
    ) -> Table:
        """Returns data for a given object ID

        Args:
            obj_id: The ID of the desired object
            format_table: Format for use with 'sncosmo' (Default: True)

        Returns:
            An astropy table of data for the given ID
        """

        # Read in data for the object ID and return an astropy table
        pass

    def _download_module_data(
```

```
        self, force: bool = False, timeout: float = 15
    ):
        """Download data for the current survey / data release

        Args:
            force: Re-Download locally available data
            timeout: Seconds before timeout for individual files/archives
        """

        # If you do not wish to include download functionality,
        # do not include this method

        # The 'utils' module includes functions for downloading files
        # See the 'utils.download_tar' and 'download_tar.download_file'
        # functions. Here is an example:
        utils.download_tar(
            url=self._spectra_url,
            out_dir=self._data_dir,
            skip_exists=self._spectra_dir,
            mode='r:gz',
            force=force,
            timeout=timeout
        )
```

Notice that there is no need to explicitly raise errors for invalid object Id's (exception class `sndata.exceptions.InvalidObjId`) or handle errors where there is no downloaded data (`sndata.exceptions.NoDownloadedData`). This is handled automatically by the public functions of the parent class.

It is common for astronomical papers to publish supplementary tables using the Vizier catalog service (Ochsenbein et al. 2000). The formatting of Vizier tables is fairly standard and in many cases the `_get_available_tables` and `_load_table` methods can be copied exactly from the above example. Instead of copy and pasting these two methods, users can alternatively inherit the `DefaultDataParser` class.

4.1.4.2 Photometric Data

Photometric data is represented the same way as spectroscopic data, but with a few differences. The first is to inherit from the `PhotometricRelease` class and include some extra meta data.

```python
from sndata.base_classes import PhotometricRelease

class Smith20(PhotometricRelease):

    # Include all of the meta data for a spectroscopic data release
    # plus the name and zero point of each photometric filter
    band_names = tuple("u", "g", "r", "i", "z")
    zero_point = tuple(25, 25, 25, 25, 25)
```

If not already using transmission filters built into `sncosmo`, you will also need to define the directory where the filter transmission curves are stored and add some small logic for registering those filters with `sncosmo`. The location of the filter data on should be defined at instantiation:

```python
    def __init__(self):
        """Define local and remote paths of data"""

        # Call to parent class defines the self._data_dir attribute
        # All data should be downloaded to / read from that directory
        super().__init__()

        # Local paths of filter transmission curves
        self._filter_dir = self.data_dir / 'filters'
```

The registration of new filters with `sncosmo` is then handled by the `_register_filters` method as follows:

```python
    def _register_filters(self, force: bool = False):
        """Register filters for this survey / data release with SNCosmo

        Args:
            force: Re-register a band if already registered
        """

        bandpass_data = zip(self._filter_file_names, self.band_names)
        for _file_name, _band_name in bandpass_data:
            filter_path = self._filter_dir / _file_name
            wave, transmission = load_filter_data(filter_path)
            band = sncosmo.Bandpass(wave, transmission)
            band.name = filter_name
            sncosmo.register(band, force=force)
```

The process of constructing the remainder of the class follows exactly from Section 4.1.4.1.

4.2 pwv_kpno: PWV Data Access and Atmospheric Modeling

We here introduce `pwv_kpno`[4]: a Python package that provides models for the atmospheric transmission due to H_2O at user-specified sites. By using MODTRAN models (Berk et al. 2014) in conjunction with publicly available PWV measurements, `pwv_kpno` is able to return models for the atmospheric transmission between $3,000$ and $12,000$ Å. The package was beta tested at Kitt Peak National Observatory (KPNO) using a dual-band GPS system that was installed at the WIYN 3.5-m telescope in 2015. Thus, direct PWV measurements at Kitt Peak are available starting in early 2015, but by using measurements from stations on the surrounding desert floor, the package is capable of modeling the atmospheric transmission for years 2010 onward. The package also provides access to tabulated PWV measurements, along with easy to use utility functions for retrieving and processing newly published PWV data.

For the most up-to-date documentation of the `pwv_kpno` package, see https://mwvgroup.github.io/pwv_kpno/.

4.2.1 Features and Use of pwv_kpno

The `pwv_kpno` package provides access to models for the atmospheric transmission due to PWV at any location within the SuomiNet GPS network. However, the package is configured by default to return models for Kitt Peak National Observatory. We here demonstrate the features of `pwv_kpno` using the default model for Kitt Peak and further discuss modeling custom sites in Section 4.2.2

`pwv_kpno` is registered with the Python Package Index and is compatible with both Python 2.7 and 3.5 through 3.7. Using PWV measurements published by the SuomiNet project, the package is able to determine the atmospheric transmission between $3,000$ and $12,000$ Å. The package also provides methods for the automated retrieval and processing of published SuomiNet data.

[4] `pwv_kpno` can be downloaded using the pip package manager or at `https://mwvgroup.github.io/pwv_kpno/`

4.2.1.1 Accessing PWV Data

In order to model the atmospheric transmission for a given date and time, `pwv_kpno` requires there to be SuomiNet data stored on the user's local machine. Each package release contains the necessary data to return models for Kitt Peak from 2010 through the end of the previous year. This data is automatically included when installing the package.

Access to tabulated PWV data and modeling of the PWV transmission function is provided by the `pwv_atm` module. A list of years that have been downloaded from SuomiNet to the user's local machine can be retrieved using the `downloaded_years` method.

```
from pwv_kpno import pwv_atm
pwv_atm.downloaded_years()

> [2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017]
```

The returned list includes all years for which any amount of data has been downloaded.

In order to update the locally stored data, `pwv_kpno` can be used to automatically retrieve and processes new data from SuomiNet. This is achieved using the `update_models` method.

```
pwv_atm.update_models()

> [2017, 2018]
```

Here the returned list includes any years for which new data was downloaded. By default, the function will download all published data for any years not currently present on the local machine. In addition, it will also download data for the most recent year that is locally available. This method ensures there are no years with incomplete measurements in the locally available data. If desired, the user can alternatively specify a specific year to download from 2010 onward.

In addition to downloading data for Kitt Peak, the `update_models` method also downloads measurements taken at the four supplementary locations shown in Figure 18a. Each time the method is run, a new set of linear fits is created to describe the PWV concentration at Kitt Peak as a function of the PWV concentration at each supplementary location. These new fits are then used to recreate the entire supplemented PWV model for Kitt Peak. The error in PWV modeled using each of these fits is taken as the standard deviation of that fit's residuals.

Users can access the locally available SuomiNet data using the `measured_pwv` method. Results are returned as an Astropy table (Price-Whelan et al. 2018) and can be independently filtered by year, month, day, and hour.

```
pwv_atm.measured_pwv(year=2016, month=11, day=14)

>         date       KITT KITT_err P014 ...
>         UTC         mm      mm     mm  ...
> ------------------ ---- -------- ---- ...
> 2016-11-14 00:15  4.7     1.025  6.9 ...
> 2016-11-14 00:45  4.3     1.025  6.7 ...
>               ...  ...       ...  ... ...
```

Excluding the date column, each column is labeled using the SuomiNet identification codes for the GPS receivers.

`pwv_kpno` also provides access to the modeled PWV column density at Kitt peak via the `modeled_pwv` method. As in the previous example, these results can also be filtered independently by year, month, day, and hour.

```
pwv_atm.modeled_pwv(year=2016, month=11, day=14)

>     date       pwv pwv_err
>     UTC         mm    mm
> ------------------ --- -------
> 2016-11-14 00:15 4.7   1.025
> 2016-11-14 00:45 4.3   1.025
>               ... ...     ...
```

4.2.1.2 Modeling the Atmosphere For a known PWV column density, the package provides access to the modeled atmospheric transmission via the `trans_for_pwv` function. This method returns the modeled transmission function as an Astropy table with wavelengths ranging from $3,000$ to $12,000$ Å. For example, given a PWV column density of 13.5 mm:

```
pwv_atm.trans_for_pwv(13.5)

> wavelength transmission
>  Angstrom
> ---------- -------------
>    3000.00 0.9999999916
>    3000.05 0.9999999916
>        ...           ...
```

Atmospheric models can also be accessed for a given datetime and airmass using the function `trans_for_date`.

```
from datetime import datetime
import pytz

obsv_date = datetime(
    year=2013,
    month=12,
    day=15,
    hour=5,
    minute=35,
    tzinfo=pytz.utc)

pwv_atm.trans_for_date(date=obsv_date, airmass=1.2)

> wavelength transmission transmission_err
>  Angstrom
> ---------- ------------- -----------------
>    3000.00 0.9999999916 1.7305648359e-08
>    3000.05 0.9999999916 1.7305648359e-08
>        ...           ...               ...
```

If `pwv_kpno` does not have any supplemented SuomiNet data within a day of the requested datetime, an exception is raised. Both the `trans_for_pwv` and `trans_for_date` functions determine the atmospheric transmission by returning a set of MODTRAN transmission models.

4.2.1.3 Modeling a Black Body

The `blackbody_with_atm` module provides functions for modeling the effects of PWV absorption on a black body SED. For example, consider a black body at 8,000 K under the effects of atmospheric absorption due to 15 mm of PWV. For a given array of wavelengths in angstroms, the `sed` method returns the corresponding spectral energy distribution.

```
from pwv_kpno import blackbody_with_atm as bb_atm

temp = 8000
wavelength = np.arange(7000, 10000, 100)
pwv = 15

sed = bb_atm.sed(temp, wavelength, pwv)
```

The SED from the above example can be seen in Figure 16. If desired, the SED of a black

body without atmospheric effects can also be achieved by specifying a PWV column density of zero.

Using the `magnitude` function, users can determine the magnitude of a black body in a given band. For example, in the i band, which ranges from $7,000$ to $8,500$ Å, the AB magnitude of a black body is found by running

```
band = (7000, 8500)
mag = bb_atm.magnitude(temp, band, pwv)
```

Here the i band is treated as a top-hat function, however, the `magnitude` function also accepts `band` as a two dimensional array specifying the wavelength and response function of a real-world band. As in the previous example, the magnitude of a black body without the effects of atmospheric absorption can be found by specifying a PWV level of zero.

4.2.2 Modeling Other Locations

By default, `pwv_kpno` provides models for the PWV transmission function at Kitt Peak National Observatory. However, `pwv_kpno` also provides atmospheric modeling for user customized locations. Modeling multiple locations is handled by the `package_settings` module, and allows modeling at any location with a SuomiNet connected GPS receiver.

Each site modeled by `pwv_kpno` is represented by a unique configuration file. Using the `ConfigBuilder` class, users can create customized configuration files for any SuomiNet site. As an example, we create a new model for the Cerro Tololo Inter-American Observatory (CTIO) near La Serena, Chile.

```
from pwv_kpno.package_settings import ConfigBuilder

new_config = ConfigBuilder(
    site_name="cerro_tololo",
    primary_rec="CTIO",
    sup_rec=[],
    wavelength=custom_wavelengths,
    cross_section=custom_cross_sections
)

new_config.save_to_ecsv("./cerro_tololo.ecsv")
```

Here `site_name` specifies a unique identifier for the site being modeled, `primary_rec` is

the SuomiNet ID code for the GPS receiver located at the modeled site, and `sup_rec` is a list of SuomiNet ID codes for nearby receivers used to supplement measurements taken by the primary receiver. Unlike the default model for KPNO, there are no additional receivers near the CTIO and so `sup_rec` in this example is left empty (the default value). By default, `pwv_kpno` models use MODTRAN estimates for the wavelength dependent cross section of H_2O from 3,000 to 12,000 Å. The optional `wavelength` and `cross_section` arguments allow a user to customize these cross sections in units of Angstroms and cm^2 respectively.

If desired, users can specify custom data cuts on SuomiNet data used by the package. Data cuts are defined using a 2d dictionary of boundary values. The first key specifies which receiver the data cuts apply to. The second key specifies what values to cut. Following SuomiNet's naming convention, values that can be cut include PWV (`"PWV"`), the PWV error (`"PWVerr"`), surface pressure (`"SrfcPress"`), surface temperature (`"SrfcTemp"`), and relative humidity (`"SrfcRH"`). For example, if we wanted to ignore measurements taken between two dates, we can specify those dates as UTC timestamps and run

```
data_cuts = {"CTIO":
    {"SrfcPress": [
            [time_start, time_end]
        ]
    }
}

new_config = ConfigBuilder(
    site_name="cerro_tololo",
    primary_rec="CTIO",
    data_cuts=data_cuts)
```

Once a configuration file has been created, it can be permanently added to the locally installed `pwv_kpno` package by running

```
from pwv_kpno.package_settings import settings

settings.import_site_config("./cerro_tololo.ecsv")
```

This command only needs to be run once, after which `pwv_kpno` will retain the new model on disk, even in between package updates. The package can then be configured to use the new model by running:

```
settings.set_site("cerro_tololo")
```

After setting `pwv_kpno` to a model a specific site, the package will return atmospheric models and PWV data exclusively for that site. It is important to note that this setting is not persistent. When `pwv_kpno` is first imported into a new environment the package will always default to using the standard model for Kitt Peak, and the above command will have to be rerun.

A complete summary of package settings can be accessed using attributes of the `settings` object.

```
settings.set_site("kitt_peak")
print(settings.site_name)
> kitt_peak

print(settings.available_sites)
> ["kitt_peak", "cerro_tololo"]

print(settings.receivers)
> ["AZAM", "KITT", "P014", "SA46", "SA48"]

print(settings.primary_rec)
> KITT

print(settings.supplement_rec)
> ["AZAM", "P014", "SA46", "SA48"]
```

The configuration file for the currently modeled location can be exported by running

```
settings.export_site_config("./current_site_name.ecsv")
```

4.2.3 Package Demonstration

The `pwv_kpno` package can be used to correct both spectrographic and photometric observations. As an example, we use the `pwv_kpno` package to determine the atmospheric correction presented in Figure 21. We also demonstrate how to calculate the photometric correction factor defined in Equation 3.6 for a black body.

4.2.3.1 Correcting Spectra

Spectrographic observations are corrected by dividing observed spectra by the modeled atmospheric transmission function. To account for the spectral resolution function of the observing spectrograph, the modeled transmission is first binned to approximately match the observed spectra's resolution. Depending on the resolution of the observation, further smoothing can then be performed using a Gaussian kernel. Assume that the observed wavelength and flux values are stored in equal length arrays `obs_wavelength` and `obs_flux` respectively. Using the date, time, and airmass of the observation, the binned transmission function is found by running

```
import numpy as np

resolution = 16  # Angstroms
bins = np.arange(min(obs_wavelength), max(obs_wavelength + 1), resolution)

airmass = 1.2
obs_data = datetime(2010, 09, 19, 6, 29, tz_info=pytz.utc)
transm = pwv_atm.trans_for_date(obs_date, airmass, bins)
```

In order to divide the observed spectrum and modeled transmission, we linearly interpolate the binned transmission to the observed wavelength values. We then apply a Gaussian smoothing using an arbitrary standard deviation of 2 Å.

```
from scipy.ndimage.filters import gaussian_filter

interp_transm = np.interp(
    obs_wavelength,
    transm["wavelength"],
    transm["transmission"])

smoothed_transm = gaussian_filter(input=interp_transm, sigma=2)
```

The corrected spectrum is then given as the observed flux divided by the smoothed transmission function on a wavelength by wavelength basis.

```
corrected_spec = np.divide(obs_flux, smoothed_transm)
```

4.2.3.2 Correcting Photometry

The `pwv_kpno` package can also be used to correct photometric observations of objects with a known spectral type. To do so, it is necessary to evaluate Equation 3.6. Note that the product in the numerator $S(\lambda) \cdot T(\lambda)$ represents the

SED under the influence of atmospheric effects, while $S(\lambda)$ in the denominator represents the intrinsic SED. For a black body observed in the i band, these values can be found as

```
# S(lambda) * T(lambda)
sed_with_atm = bb_atm.sed(sed_temp, i_band, pwv)

# S(lambda)
intrinsic_sed = bb_atm.sed(sed_temp, i_band, 0)
```

In practice the SED of a photometrically observed object may not be available. In such a case it is sufficient to use spectral templates instead. For example, the SED of a star can be reasonably well parametrized by its observed color.

Using the above results, we evaluate Equation 3.6 by performing trapezoidal integration with the Numpy package.

```
numerator = np.trapz(sed_with_atm, i_band)
denominator = np.trapz(intrinsic_sed, i_band)
photo_corr = np.divide(numerator, denominator)
```

The corrected photometric flux of the black body is then found by dividing the observed flux by the correction factor `photo_corr`.

4.3 Egon: A Framework for Parallel Data Analysis Pipelines

The `egon` package[5] is a lightweight Python framework for building parallelized data analysis pipelines. The package is designed to support the development of modular systems with minimal added overhead, making it easy to implement parallel workflows without having to deal with complex constructs like message brokers or task managers. Some of the benefits of the `egon` framework are that it:

- Integrates easily with existing projects, including test suites and continuous-integration tasks.
- Includes automatic validation protocols for reducing unexpected errors and improving

[5]The package name is a play on the concept of data analyses being run as *pipelines*. The name "Egon" was inspired by Egon Perrefort, my grandfather who ran his own plumbing company for many years.

overall pipeline stability.

- Encourages modular design principles that promote the development of maintainable code that is easy to test, build, and deploy.

In this chapter we outline the package's core functionality and demonstrate how `egon` can be used to build a simple, three step analysis pipeline.

For the most up-to-date documentation of the `egon` package, see https://mwvgroup.github.io/egon/.

4.3.1 Installation

`egon` is available for Python 3.7 onward and can be installed using the `pip` package manager:

```
pip install egon
```

Alternatively, the package source code can be downloaded from GitHub using the website's graphical interface or the `git` command line utility:

```
git clone https://github.com/mwvgroup/Egon
cd egon
python setup.py install --user
```

4.3.2 Basic Framework Overview

`egon` works by breaking down an analysis pipeline into discrete, reusable units. Those units are then implemented and tested individually before being assembled and deployed as a single coherent pipeline. Consider for a moment the classic example of an Extract-Transform-Load (ETL) style pipeline. Shown in Figure 30, an ETL pipeline consists of three discrete steps. The *Extract* stage is responsible for loading data into the analysis pipeline for processing. Next, the *Transform* stage manipulates the data in some predefined way and calculates derived data products. Finally, the *Load* stage uploads the manipulated data into an endpoint storage solution.

Figure 30: The basic structure of an ETL pipeline includes loading data into the pipeline (*Extract*), manipulating the data in an arbitrary way (*Transform*), and loading the transformed data into a data storage solution (*Load*). In principle, a generic software pipeline is designed by outlining individual processing stages as independent abstractions. In practice, this also requires a robust mechanism for facilitating data transfer between each pipeline stage.

Figure 31 demonstrates how the same ETL structure can be achieved using `egon`. When building with `egon`, the ETL structure is represented as a collection of interconnected `Node` objects. `Node` objects can be connected in any order and are designed to run asynchronously. This means an individual analysis stage (or an entire pipeline) can be easily scaled by allocating additional processes. Nodes are also smart enough to automatically exit once they are no longer needed, freeing up resources on the host machine once they are no longer needed.

The first step in building a pipeline is to define the necessary nodes. Ideally, each node should reflect a single independent task. When processing a fixed amount of information, nodes should also be able to shutdown automatically once there is no more data for them to process. Once the individual nodes are defined, they are connected using a flexible, signal/slot style interface and executed as a coherent pipeline object.

4.3.2.1 Source Nodes

`Source` type nodes are special in that they have no incoming pipeline connections and can only output data to downstream nodes. This makes them ideal for the **extract** task of an ETL pipeline. We build the extract node as a class that inherits from the `Source` class and add a single output connector, which is responsible for passing on data to the rest of the pipeline. The core analysis logic of the node is defined in the `action` method.

```
from egon.connectors import Output
from egon.nodes import Source


class Extract(Source):
    """The extract step of an ETL style pipeline"""

    def __init__(self, num_processes=1):

        # Here we define an output connector for the class
        # This allows the node to send data further down the pipeline
        # For more complicated nodes, define as many connectors as needed.
        self.data_output = Output()
        super().__init__(num_processes)
```

The core analysis logic of the node is defined inside the `action` method. Any setup or teardown tasks should be defined in the `setup` and `teardown` methods respectively. All three of these functions should take no arguments and return `None`.

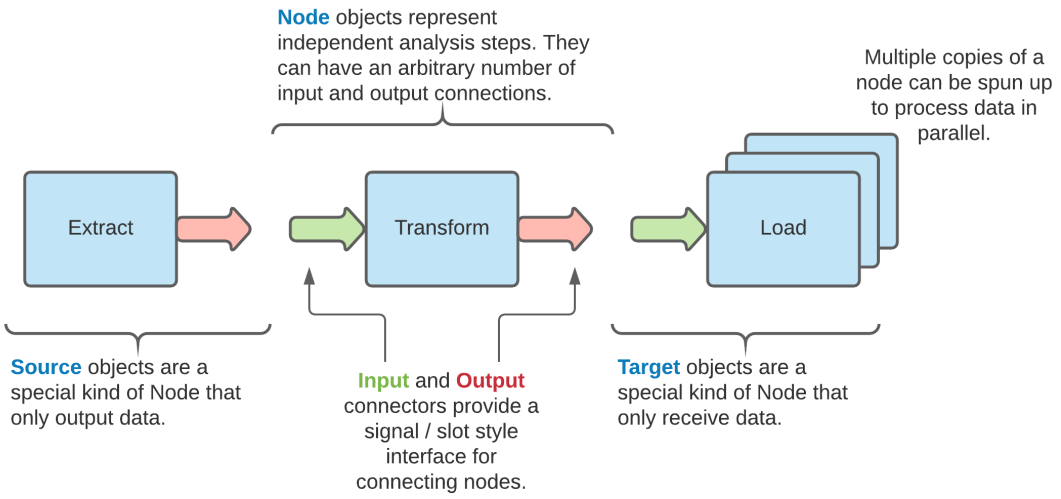


Figure 31: A visual representation of how the `egon` framework represents data analysis pipelines in an object oriented structure. Individual analysis stages are represented as distinct objects and connect using a signal/slot style interface. Nodes can leverage resources from a predefined number of processes and can take advantage of arbitrarily many input/output connections.

```
class Extract(Source):
    # Continued from the previous example

    def setup(self):
        """Setup tasks are optional and run setup tasks for the node"""

        self.database = connect_to_database()

    def action(self):
        """Action tasks reflect the core analysis logic the node"""

        for data in self.database.get_fixed_number_of_items():
            self.data_output.put(data)

    def teardown(self):
        """Teardown tasks are optional and act as clean up tasks"""

        self.database.disconnect()
```

It is typically undesirable for a pipeline to run indefinitely, and it is important to include a mechanism by which each node can shut down automatically (i.e., terminate any forked processes). A node will terminate it's associated processes as soon as the `setup`, `action`, and `teardown` methods have exited. In the above example we assume the iteration over database entries is fixed in size, thus ensuring the `action` method will eventually exit. In more complex situations, such as handling an incoming stream of real-time data, it is up to the user to implement a data access object (DAO) that is capable of appropriately terminating iteration.

4.3.2.2 Target Nodes

`Target` nodes are similar to `Source` nodes except they receive data instead of sending it. This makes them ideal for the **load** task. For example:

```
from egon.connectors import Input
from egon.nodes import Target


class Load(Target):
    """The load step of an ETL style pipeline"""

    def __init__(self, num_processes=1):

        # We define an input connector to receive data from other nodes
        self.data_input = Input()
        super().__init__(num_processes)

    # The rest of the class should include the same
    # 'setup', 'action', and 'teardown' style structure from earlier
```

It is important that pipeline nodes exit gracefully once there is no more data to process. Input connectors include the `iter_get` method which allows you to iterate over incoming data. `iter_get` is smart enough to automatically exit once there are no longer any upstream processes left to feed it data.

```
from egon.connectors import Input
from egon.nodes import Target


class Load(Target):
    """The load step of an ETL style pipeline"""

    def __init__(self, num_processes=1):

        self.data_input = Input()
        super().__init__(num_processes)

    def action(self):
        """Load data into a directory"""

        for data in self.data_input.iter_get():
            ...  # Logic to insert data into a target database goes here
```

4.3.2.3 Generic Nodes

Generic `Node` objects can have both input and output connectors. In principle we can define as many input or output connections as necessary. However for the ETL example we only need one of each.

```
from egon.connectors import Input, Output
from egon.nodes import Node


class Transform(Node):
    """The transform step of an ETL style pipeline"""

    def __init__(self, num_processes=1):

        # 'Node' subclasses can have an arbitrary number of input/outputs
        # Here we define input AND output connectors for the class so
        # That the node to communicate upstream and downstream
        self.data_in = Input()
        self.data_out = Output()
        super().__init__(num_processes)


    # The rest of the class should include the same
    # 'setup', 'action', and 'teardown' style structure from earlier
```

4.3.2.4 Building a Pipeline

The last step in constructing a pipeline is to connect together the individual nodes. `input` and `output` connectors provide a signal/slot style interface:

```
from egon.pipeline import Pipeline


class ETL(Pipeline):

    def __init__(self):

        # Define the nodes of the pipeline
        self.extract = Extract()
        self.transform = Transform(num_processes=4)
        self.load = Load()

        # Limit the number of objects the input connector stores in memory
        # This limits memory usage and forces the 'extract' node to wait
        # for the 'transform' node to catch up if it falls behind.
        self.transform.data_in.maxsize = 100

        # Connect each of the nodes together
        self.extract.data_out.connect(self.transform.data_in)
        self.transform.data_out.connect(self.load.data_in)
        super().__init__()  # Super should be called at the very end
```

The `validate` task handles process allocation and runs automated validation tasks (see the documentation below). The entire pipeline can then be launched in one line:

```
etl_pipeline = ETL()
etl_pipeline.validate()
etl_pipeline.run()
```

In general, an output connector can connect to multiple inputs. Each input will receive an identical copy of the data send by the output. Multiple output connectors can also connect to a single input. This allows you to consolidate data from multiple nodes into a single input.

4.3.3 Function Decorator Short-Hands

`egon` provides function decorators that make it easy to cast existing logic as pipeline objects. This allows developers to quickly build pipeline nodes without having to define an entire class. For example, a node designed to determine the sum of two numbers might look as follows:

```
from egon.decorators import as_node

@as_node
def add_two_numbers(x, y):
    return x + y
```

Wrapped functions maintain the same docstring and signature as the original function. They can also be called in exactly the same manner as their unwrapped counterparts. This allows existing logic to be easily ported into a pipeline structure without impacting the existing API design.

Separate decorators exist for casting functions as different kinds of pipeline nodes. They behave exactly the same as class based nodes, except they do not support setup or teardown tasks. Here is an example of a pipeline for adding two numbers using function decorators:

```
from egon.decorators import as_source, as_node, as_target
from egon.pipeline import Pipeline


@as_source
def generate_numbers():
    for i in range(100):
        yield i, i + 1


@as_node
def add(x, y):
    return x + y


@as_target
def print_sum(x):
    print(x)


class AddingPipeline(Pipeline):
    """A pipeline for generating and then adding numbers"""

    def __init__(self):
        # Define the nodes of the pipeline
        self.generate_numbers = generate_numbers
        self.add = add
        self.print_sum = print_sum

        # Nodes run on a single forked process by default.
        # Additional processes can be allocated for higher performance.
        self.add.num_processes = 4

```

```
        self.generate_numbers.output.connect(self.add.input)
        self.add.output.connect(self.print_sum.input)


pipeline = AddingPipeline()
pipeline.validate()
pipeline.run()
```

4.3.4 Pipeline Validation

As the size and complexity of your pipeline grows, you'll want to be sure you don't leave any loose ends. Nodes with unconnected inputs/outputs can leave you with a broken pipeline that consumes unnecessary resources. Fortunately, `egon` incorporates automated validation methods to help ensure you pipeline runs smoothly.

The `validation` method checks the inter-connectivity of a node to its neighbors. If a node has any input/output connectors that aren't connected to other nodes, an error will be raised. The error will indicate both the problem and problematic connector.

```
from egon.decorators import as_source

@as_source
def generate_numbers():
    for i in range(100):
        yield i, i + 1

generate_numbers.validate()
> egon.exceptions.MissingConnectionError:
>     Connector <egon.connectors.Input(name=None) object does
>     not have an established connection
```

Individual node validation is a useful tool when building unit tests (see Section 4.3.5 for further discussion). `Pipeline` objects also provide a `validation` method. Similarly, `Pipeline` objects also provide a `validate` method that checks all nodes in the parent pipeline. The recommended practice is to validate a pipeline at least once before running it. For example:

```
my_pipeline = MyCustomPipeline()
my_pipeline.validate()
my_pipeline.run()
```

4.3.5 Test Suite Support

`egon`'s object oriented design allows developers to test nodes individually before deploying them to a combined pipeline. The `mock` module provides mockup pipeline nodes useful for the construction of unit-tests or as a placeholders in early development. Mock nodes differ from the normal pipeline nodes described above in the following ways:

1. By default, mock nodes do not fork any processes and run entirely in the same process they were instantiated in.
2. Mock objects include predefined `action` methods so you don't need to write your own.

When building your test suite, be sure to run your nodes in the main thread by setting `num_process=0` for each instantiated node. This will ensure you don't end up with any annoying forked processes should you need to call the `execute` method during testing.

`MockSource` objects allow you to easily load data into a connected, downstream node. Any data stored in the `load_data` attribute will be sent to the `output` connector when the node executes.

```
from egon.mock import MockSource

# You can define the data to load at instantiation
mock_source = MockSource(load_data=[1, 2, 3])

# You can also modify the list after instantiation
mock_source.load_data.append(4)

# Any objects in 'load_data' above are sent to the 'output' connector
mock_source.output.connect(my_node_to_test.input)

mock_source.execute()
my_node_to_test.execute()
```

`MockTarget` objects act as accumulators. When executed they collect any data passed to their `input` connector attribute into a list.

```
from egon.mock import MockTarget

# The mock target will accumulate any data sent to it's input connector
mock_target = MockTarget()
my_node_to_test.output.connect(mock_target.input)

my_node_to_test.execute()
mock_target.execute()

# The mock target will store data in the 'accumulated_data' list
first_value = mock_target.accumulated_data[0]
```

Using the `unittest` testing framework, we outline an example setup method below. The example is designed to test a user defined node `MyCustomNode` that has a single pair of input and output connectors.

```
from unittest import TestCase

from egon.mock import MockSource, MockTarget


class MyTestClass(TestCase):

    def setUp(self) -> None:
        """Define and setup the nodes used in the test suite"""

        # Create the nodes to be used in testing
        self.source = MockSource([...])
        self.node = MyCustomNode(num_processes=0)
        self.target = MockTarget()

        # Connect the nodes
        self.source.output.connect(self.node.input)
        self.node.output.connect(self.target.input)

        # Run the nodes
        for node in (self.source, self.node, self.target):
            node.execute()

    def test_against_node_outputs(self):
        """Run a test based on the value outputted by 'MyCustomNode'"""

        node_output_value = self.target.accumulated_data[0]
```

The number of processes assigned to a given node can be changed at any point after instantiation. However, the number cannot be changed while the node is running. If any forked process assigned to a node is currently running, you must wait for it to exit.

5.0 Concluding Remarks

Full operations for LSST are currently scheduled to begin in 2023. The results presented in this work lay the foundation for future research with LSST and leave open multiple avenues for further exploration. Among the most significant benefits provided by LSST will be the number of observed SNe. Modern SN data sets include only a handful of objects classified in each peculiar subtype. At the time of writing, this includes on order 100 spectroscopically confirmed 91bg-like objects spanning multiple independent surveys (Guillochon et al. 2017). Using the classification technique presented in Chapter 2, we will be able to increase the number of photometrically identified 91bg-like objects significantly.

The most immediate benefit provided by a large sample of 91bg-like objects is the further analysis of populations statistics. However, having a large number of SNe Ia also means that the availability of dedicated spectroscopic follow-up observations will be limited. By identifying these objects photometrically, they can be flagged for further observation at other observatories. In this way, the classification technique can also be used to support the detailed study of individual 91bg-like events.

Future work is also intended to broaden the scope of the presented classification scheme. In its current form, the presented technique is restricted to the identification of 91bg-like events. However, by substituting our chosen 91bg-model with models for a different class of peculiar SNe, the same procedure can be used to identify other kinds of peculiar SNe. The way in which the spectroscopic template of each model is varied to fit photometric observations can also be examined to explore any potential biases introduced by the minimization procedure.

The simulated SN data set presented in Chapter 3 also provides a foundation for continuing LSST preparations. In this work, we studied the propagation of PWV effects on SN observations after the removal of the overall cadence. This provided key insight into how PWV absorption propagates into the fitting of SN distance estimates. A direct extension of this analysis is to re-run the estimation of SN distance using realistic cadences proposed for LSST. This allows for the investigation of potential effects introduced by the sampling

rate of the overall cadence on the minimization of light-curve fits (and the resulting distance estimates). By comparing results determined for different cadences, the collective set of potential LSST cadences can be compared to determine which, if any, provides the highest level of precision.

In the longer term, the same GPS modeling techniques applied in our simulations can be used to provide near real-time PWV measurements for LSST. The LSST will incorporate multiple calibration systems designed to estimate the atmospheric state and isolate systematic errors (Ingraham et al. 2016; Sebag et al. 2014). Among these systems is the Rubin Auxiliary Telescope[1], which will use state-of-the-art atmospheric models in conjunction with spectroscopic observations of bright, isolated stars to provide dedicated measurements of the atmospheric transmittance. Current plans for the LSST calibration system also include the use of a dual-band GPS receiver to determine the localized density of PWV. The use of a dedicated GPS system allows for the real-time monitoring of PWV separate from the other atmospheric components, which is important for the cross validation of results from the auxiliary telescope.

[1]See LSE-379: https://ls.st/LSE-379.

Appendix A Spectroscopic Classification of SDSS SNe

In order to optimize Equation 2.4, we require a set of spectroscopically subtyped SNe. To satisfy this requirement, we use the SN IDentification software[1] (SNID; Stéphane Blondin et al. 2007) to analyze all 938 transients in the S18 sample that were observed spectroscopically.

SNID works by cross-correlating input spectra against a set of observed spectral templates that have been blue-shifted into the rest-frame. The quality of the resulting classification is then reported in terms of an *rlap* value, which is a quality parameter that represents the strength of the correlation between the input spectrum and its best-matching spectra. For the duration of this work, we use version 2.0 of the default SNID template spectra. This includes corrected versions of the original SNID 1.0 templates supplemented with observations from the CfA SN Program. For more details, see S. Blondin et al. 2012.

Many of the SDSS spectra have strong emission lines indicative of a poor host galaxy subtraction and are dominated by noise at fringe wavelengths. This is particularly problematic since SNID results are sensitive to the presence of noisy ends in observed spectra. In order to reduce the impact of noise and possible light contamination, we truncate each spectrum to the wavelength range $4000 \leq \lambda \leq 9000$. We also restrict our analysis to targets with spectroscopic observations between -15 and 15 days of estimated B-band brightness, thus eliminating the faintest and noisiest spectra in our sample.

Unless stated otherwise, we enforce the use of SDSS determined values for the redshift (SNID argument `forcez`) and initial phase guess (`age`) of each spectrum while allowing SNID to search templates within a phase range of ± 10 days (`dage`). We also clip galaxy emission lines using the same redshift for the host, as was specified for the observed spectrum (`emclip`). Any other SNID parameters are left at their default values.

When determining spectral (sub)types, we follow the prescription of Silverman et al. 2012. In order to avoid any spurious correlation peaks, we first determine an initial spectral type for each object by running SNID with all available templates and a minimum rlap value of 10. A spectrum is considered to have an accepted type if at least half of the correlations

[1] Version 5.0: `https://people.lam.fr/blondin.stephane/software/snid/index.html`

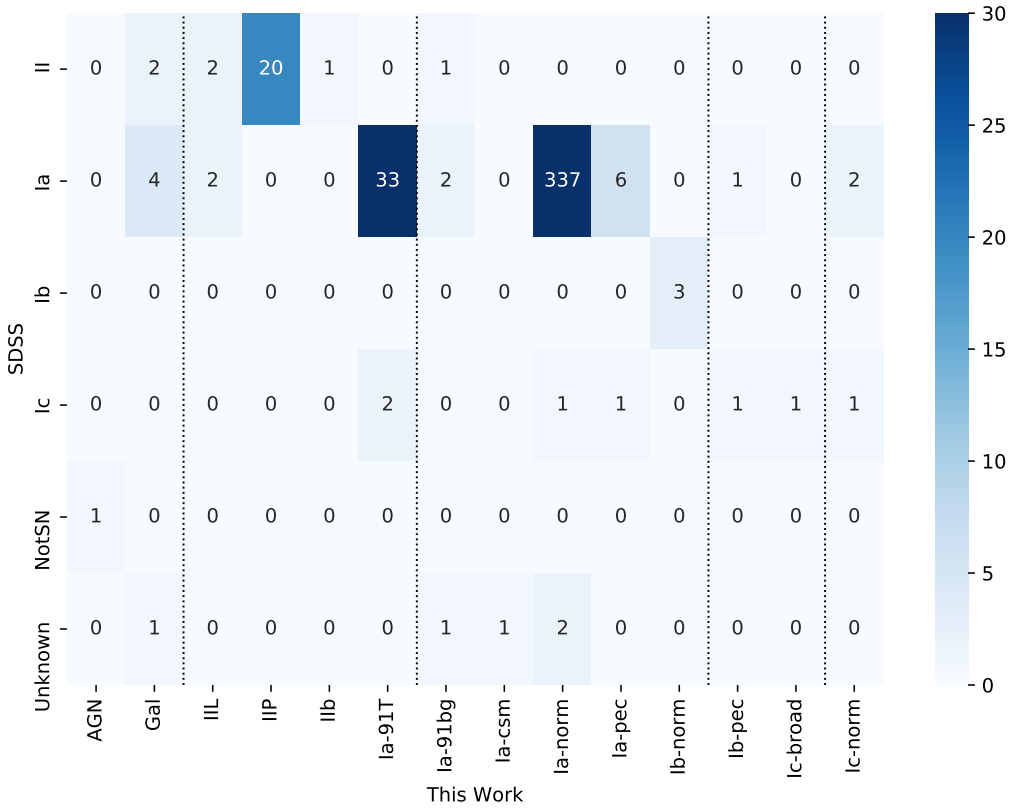


Figure 32: Spectroscopic classifications from Sako, Bassett, A. C. Becker, et al. 2018 are compared against spectroscopically determined subtypes in this work. Objects with disagreeing classifications were visually inspected and found to have spectra with poor host galaxy subtractions as indicated by strong emission features and noisy continua. Included are 5 objects for which spectra were published by SDSS without a classification.

correspond to the best matching template. For any spectra that do not pass this cut, we perform another SNID run using a minimum rlap value of 5.

Once an initial spectral type has been established, we re-run SNID on each spectrum using only the templates that correspond to the accepted type (e.g., a spectrum with an accepted type of `Ia` is re-analyzed using the `Ia-norm`, `Ia-91bg`, `Ia-91T`, `Ia-csm`, and `Ia-pec` spectral templates). As was done in the previous step, this process is performed once using a minimum rlap value of 10 and then again using a minimum rlap of 5. Resulting subtypes were accepted for targets where at least half of the correlations correspond to the best matching subtype.

Shown in Figure 32, we compare our subtyping results against spectroscopic classifications determined by Zheng et al. 2008 as presented in S18. Some disagreement is expected given the use of different template spectra in the classification process. However, we find excellent agreement between the two sets of classifications with only 18 of the considered targets being in disagreement. Visual inspection for many of these spectra show either a limited wavelength coverage or strong emission features and noisy continua, indicating a poor host galaxy subtraction.

We note that the number of spectroscopically typed 91T-like objects are in rough agreement with rate estimates from the literature (S. Blondin et al. 2012; W. Li, Leaman, et al. 2011; Silverman et al. 2012). Unfortunately, visual inspection of these and other SNe classified as peculiar subtypes show that their spectra suffer from a poor host galaxy subtraction. Furthermore, the classification result for these objects varies significantly with the wavelength range used in the analysis. While there exists a subset of well observed SNe for which the SNID results are more robust, we ultimately conclude that the collective classification results are not suitable for the training of a photometric classifier.

Appendix B Tables

Supplemental tables are provided in this section for reference. Tables that are too large to display in printed format have been limited to include only the first few rows. For a complete, machine readable copy of each table, please see the appropriate published work listed in the table caption. Tables for which an associated publication is not mentioned are presented here in their entirety.

CID	Classification
4524	SN 2002ci
6295	SN 1991bg
7017	SN 2002ci
8151	SN 2002cx
12979	SN 1991bg
13357	SN 2002cx
15340	SN 2000cx
17886	SN 1991bg
18890	SN 1991bg
20208	SN 2002cx

Table 1: Objects identified in the SDSS SN Survey as being potentially peculiar objects based on visual inspection of their spectrophotometric properties. Objects are listed using their Candidate Identifier from Sako, Bassett, A. C. Becker, et al. 2018

CID	Band	Model	z	z err	t_0	t_0 err	c	c err	x_1	x_1 err	E(B - V)	chisq	ndof	B_{max}	$\Delta M(B)_{15}$
					MJD	MJD					mag			mag	mag
679	all	Hsiao	0.124		53690.37	0.01			0.50	0.72	0.07	68.10	51.0	-15.80	0.52
679	all	sn91bg	0.124		53689.20	0.05	0.00	0.78	1.25	0.06	0.07	81.61	50.0	-15.72	1.45
679	blue	Hsiao	0.124		53690.37				0.33	0.30	0.07	16.90	19.0	-15.64	0.65
679	blue	sn91bg	0.124		53689.20		0.00	0.92	1.25	0.00	0.07	22.08	18.0	-16.14	1.45
679	g5	Hsiao	0.124		53690.37				0.30	0.31	0.07	7.75	9.0	-15.63	0.68
679	g5	sn91bg	0.124		53689.20		0.00	0.63	1.25	0.00	0.07	12.13	8.0	-16.12	1.45
679	i5	Hsiao	0.124		53690.37				0.50	0.15	0.07	20.59	9.0	-16.54	0.52
679	i5	sn91bg	0.124		53689.20		0.27	0.31	1.25	0.00	0.07	28.69	8.0	-15.18	1.45
679	r5	Hsiao	0.124		53690.37				0.50	0.17	0.07	11.83	9.0	-15.98	0.52
679	r5	sn91bg	0.124		53689.20		0.03	0.55	1.25	0.00	0.07	17.34	8.0	-15.67	1.45
679	red	Hsiao	0.124		53690.37				0.50	0.10	0.07	46.75	31.0	-16.14	0.52
679	red	sn91bg	0.124		53689.20		0.00	0.62	1.25	0.00	0.07	57.97	30.0	-15.62	1.45

Table 2: Fitted parameters are listed for a combination of supernova models and bandpass collections. *red* and *blue* bands indicate a collection of bandpasses where the rest-frame effective wavelength is redward or blueward of 5500 Å. Fits to individual bands are listed using the banpass name and SDSS CCD column number (see Doi et al. 2010). We note that the version of the Hsiao model used does not include a color parameter c and thus has one degree of freedom more than the SN 1991bg model. Any redshift values missing a reported error were specified using spectroscopic measurements. Any remaining missing entries for a particular fit indicate that parameter was not included in the given fit and the result from a fit to all data was used instead. Results are limited to the first 12 table entries. For a full version of the table, see Perrefort, Zhang, et al. 2020.

CID	xBand	yBand	xCollective	yCollective
2778	8.42	2.96	6.46	9.62
11570	-4.53	-33.21	4.34	13.74
12689			6.32	0.75
15204	-19.94	11.02	33.66	137.53
16215	-16.9	-23.6	7.14	29.66
16309	-0.11	-2.39	5.18	2.54
16692	0.88	-4.04	9.06	21.12
17094	1.34	15.82	3.25	13.37
17468	3.59	8.04	6.85	2.43
17886	0.03	6.92	10.64	13.55
18218	3.69	0.14	4.18	0.2
18751	12.28	3.04	12.28	4.59
18890	0.56	5.75	3.64	10.95
19065	-0.22	0.68	4.58	0.86
21678	10.87	8.38	24.23	10.2
21898	0.86	12.67	4.2	28.31

Table 3: Classification coordinates for objects classified as being photometrically similar to SN 1991bg. Included are coordinates calculated by fitting photometric bandpasses independently (*xBand*, *yBand*) and as collective red/blue sets (*xCollective*, *yCollective*). SNe are classified as being 91bg-like if they satisfy *xCollective* $x > 3$ and *yCollective* > 0. Missing entries indicate a set of fits where one or more fits failed to converge.

CID	PSNID	Data Sets	x	y	Class
1706	pSNII	8	-9.43	-4.02	Normal
2050	Unknown	7	-0.72	-0.26	Normal
2093	pSNII	2	-1.0	-1.91	Normal
2661	SNII	4	0.35	-0.52	Normal
2809	pSNII	2	-1.43	-0.52	Normal
4226	pSNII	2	-12.5	-0.84	Normal
4330	pSNII	5	-1.66	-1.28	Normal
4652	pSNII	4	-4.52	-2.0	Normal
5314	pSNII	8	-8.71	0.97	Peculiar
6992	pSNII	1	-1.64	-0.38	Normal
12868	pSNII	3	-2.62	-1.02	Normal
13112	pSNII	8	-1.31	-2.55	Normal
13291	pSNII	4	-47.91	-20.57	Normal
13461	pSNII	8	-1.8	-1.77	Normal
13589	pSNII	4	-2.62	-0.63	Normal
13725	pSNII	6	-11.87	-2.37	Normal
13741	pSNII	5	-1.13	-0.64	Normal
14170	pSNII	8	-2.09	-0.75	Normal
15048	pSNIa	3	-0.68	-0.04	Normal
15565	pSNII	8	0.97	0.16	Peculiar

Table 4: A comparison of objects classified as anomalies by the machine learning classifier published in Pruzhinskaya et al. 2019 and their corresponding classifications determined in this work. Objects with coordinates $x > 3.00, y > 0.00$ are classified as SN 1991bg-like SNe. Objects with coordinates $x < 3.00, y < 0.00$ are classified as normal SNe. Results are limited to the first 20 table entries. For a full version of the table, see Perrefort, Zhang, et al. 2020.

CID	**z**	**z err**	**t_0**	**t_0 err**	**c**	**c err**	**x_1**	**x_1 err**	**E(B - V)**	**chisq**	**ndof**	**B_{max}**	**$\Delta M(B)_{15}$**
			MJD	MJD					Mag			mag	mag
679	0.124		53689.20	1.25	0.50	0.19	5.00	9.46	0.07	62.20	43.0	-15.45	0.19
682													
685	0.327	0.016	53652.64	0.90	0.04	0.10	5.00	0.13	0.05	577.95	94.0	-19.37	0.18
688	0.067		53621.86	0.00	0.50	0.01	1.35	0.98	0.07	86.74	52.0	-15.48	0.80
691	0.130		53606.20	0.49	0.02	0.05	-0.46	0.31	0.06	32.98	32.0	-19.05	1.14
694	0.126		53622.89	0.00	0.25	0.01	1.11	0.15	0.06	364.52	96.0	-18.89	0.85
696	0.549	0.040	53623.64	1.13	-0.08	0.10	-0.45	0.67	0.03	77.69	71.0	-20.88	1.14
697	0.155		53617.20	1.24	0.50	0.01	1.56	0.66	0.07	193.63	96.0	-17.62	0.77
701	0.205		53609.93	0.00	0.10	0.04	-1.76	0.40	0.05	63.81	56.0	-19.20	1.43
703	0.296		53626.57	0.61	-0.01	0.04	0.72	0.59	0.05	64.60	81.0	-19.50	0.93
704	0.205		53607.94	0.00	0.21	0.07	-0.71	0.63	0.06	54.99	51.0	-18.87	1.18
706	0.174	0.023	53626.04	1.61	0.50	0.08	5.00	0.10	0.02	148.45	63.0	-17.68	0.19

Table 5: Parameters are listed for fit results of the Salt 2.4 model to SDSS SNe. Any redshift values missing a reported error were specified using spectroscopic measurements. For each fit the stretch parameter x_1 was bounded to the interval [-5, 5] and the color parameter c to the interval [-0.5, 0.5]. If a spectroscopic redshift was not available, the redshift was bounded to [0, 1] and fit for. Results are limited to the first 20 table entries. For a full version of the table, see Perrefort, Zhang, et al. 2020.

Survey Name	**Data Release**	**Data Type**
Carnegie Supernova Project	DR1	Spectroscopic
Carnegie Supernova Project	DR3	Photometric
Dark Energy Survey	SN3YR	Photometric
Equation of State: Supernovae Trace Cosmic Expansion	Narayan et al. 2016	Photometric
Joint Light-Curve Analysis	Betoule et al. 2014	Photometric
Lick Observatory Supernova Search	Ganeshalingam et al. 2013	Photometric
Sloan Digital Sky Survey	Sako et al. 2018	Photometric
Sloan Digital Sky Survey	Sako et al. 2018	Spectroscopic
Supernova Legacy Survey	Balland et al. 2009	Spectroscopic
Sweetspot	Weyant et al. 2018	Photometric

Table 6: A summary of supernova data releases available at the time of writing through the `sndata` Python package. Additional data releases can be included by submitting a request to the package maintainers.

Bibliography

[1] T. M. C. Abbott et al. “First Cosmology Results using Type Ia Supernovae from the Dark Energy Survey: Constraints on Cosmological Parameters”. In: *ApJL* 872.2, L30 (Feb. 2019), p. L30.

[2] Joseph P. Anderson et al. “Characterizing the V-Band Light-Curves of Hydrogen Rich Type II Supernovae”. In: *The Astrophysical Journal* 786.1 (Apr. 2014), p. 67.

[3] T. W. Anderson et al. “Asymptotic Theory of Certain ‘Goodness of Fit’ Criteria Based on Stochastic Processes”. In: *Ann. Math. Statist.* 23.2 (June 1952), pp. 193–212.

[4] W. David Arnett. “A Possible Model of Supernovae: Detonation of ^{12}C”. In: *Ap&SS* 5.2 (Oct. 1969), pp. 180–212.

[5] Ashley D. Baker et al. “Monitoring Telluric Absorption with CAMAL”. In: *PASP* 129.978 (Aug. 2017), p. 85002.

[6] Kyle Barbary et al. *sncosmo/sncosmo: v1.4.0*. Version v1.4.0. Nov. 2016.

[7] S. Benetti et al. “The Diversity of Type Ia Supernovae: Evidence for Systematics?” In: *ApJ* 623.2 (Apr. 2005), pp. 1011–1016.

[8] A. Berk et al. “MODTRAN6: a Major Upgrade of the MODTRAN Radiative Transfer Code”. In: *Proc. SPIE* 9088 (2014), pp. 9088–9088–7.

[9] Melina C. Bersten et al. “Light Curves of Type I Supernovae”. In: *Handbook of Supernovae*. Ed. by Athem W. Alsabti et al. Cham: Springer International Publishing, 2017, pp. 1–13.

[10] Betoule, M. et al. “Improved Photometric Calibration of The SNLS and The SDSS Supernova Surveys”. In: *A&A* 552 (2013), A124.

[11] M. Betoule et al. “Improved Cosmological Constraints from a Joint Analysis of the SDSS-II and SNLS Supernova Samples”. In: *A&A* 568, A22 (Aug. 2014), A22.

[12] Cullen H. Blake et al. “Measuring NIR Atmospheric Extinction Using a Global Positioning System Receiver”. In: *Publications of the Astronomical Society of the Pacific* 123.909 (Nov. 2011), pp. 1302–1312.

[13] S. Blondin et al. “The Spectroscopic Diversity of Type Ia Supernovae”. In: *AJ* 143.5, 126 (May 2012), p. 126.

[14] Stéphane Blondin et al. “Determining the Type, Redshift, and Age of a Supernova Spectrum”. In: *ApJ* 666.2 (Sept. 2007), pp. 1024–1047.

[15] David Branch et al. “Comparative Direct Analysis of Type Ia Supernova Spectra. II. Maximum Light”. In: *PASP* 118.842 (Apr. 2006), pp. 560–571.

[16] John Braun et al. “Recent Improvements in the Retrieval of Precipitable Water Vapor”. In: *Proceedings of the 18th International Technical Meeting of the Satellite Division of The Institute of Navigation*. Jan. 2001.

[17] Peter J Brown et al. “Ultraviolet Light Curves of Supernovae with the SWIFT Ultraviolet/Optical Telescope”. In: *The Astronomical Journal* 137.5 (2009), p. 4517.

[18] D. L. Burke et al. “Forward Global Photometric Calibration of the Dark Energy Survey”. In: *AJ* 155, 41 (Jan. 2018), p. 41.

[19] David L. Burke, T. Axelrod, et al. “Precision Determination of Atmospheric Extinction at Optical and Near-infrared Wavelengths”. In: *The Astrophysical Journal* 720.1 (2010), p. 811.

[20] David L. Burke, Abhijit Saha, et al. “All-Weather Calibration of Wide-Field Optical and NIR Surveys”. In: *The Astronomical Journal* 147.1 (2014), p. 19.

[21] D. H. Clark et al. “Which Historical New Stars were Supernovae?” In: *QJRAS* 17 (Sept. 1976), p. 290.

[22] Alex Conley et al. “SiFTO: an Empirical Method for Fitting SN Ia Light Curves”. In: *The Astrophysical Journal* 681.1 (2008), p. 482.

[23] Charlie Conroy and James E. Gunn. “The Propagation of Uncertainties in Stellar Population Synthesis Modeling. III. Model Calibration, Comparison, and Evaluation”. In: *ApJ* 712.2 (Apr. 2010), pp. 833–857.

[24] Charlie Conroy, James E. Gunn, and Martin White. “The Propagation of Uncertainties in Stellar Population Synthesis Modeling. I. The Relevance of Uncertain Aspects of Stellar Evolution and the Initial Mass Function to the Derived Physical Properties of Galaxies”. In: *ApJ* 699.1 (July 2009), pp. 486–506.

[25] C. Contreras et al. “The Carnegie Supernova Project: First Photometry Data Release of Low-Redshift Type Ia Supernovae”. In: *AJ* 139 (Feb. 2010), pp. 519–539.

[26] Mi Dai et al. “Photometric Classification and Redshift Estimation of LSST Supernovae”. In: *Monthly Notices of the Royal Astronomical Society* 477.3 (Apr. 2018), pp. 4142–4151.

[27] Dark Energy Survey Collaboration et al. “The Dark Energy Survey: more than dark energy - an overview”. In: *MNRAS* 460.2 (Aug. 2016), pp. 1270–1299.

[28] T. de Jaeger et al. “Observed Type II Supernova Colours from the Carnegie Supernova Project-I”. In: *MNRAS* 476.4 (June 2018), pp. 4592–4616.

[29] DES Collaboration, T. M. C. Abbott, A. Alarcon, et al. “Cosmological Constraints from Multiple Probes in the Dark Energy Survey”. In: *arXiv e-prints*, arXiv:1811.02375 (Nov. 2018), arXiv:1811.02375.

[30] DES Collaboration, T. M. C. Abbott, S. Allam, et al. “First Cosmology Results using Type Ia Supernovae from the Dark Energy Survey: Constraints on Cosmological Parameters”. In: *arXiv e-prints*, arXiv:1811.02374 (Nov. 2018), arXiv:1811.02374.

[31] M. Doi et al. “Photometric Response Functions of the Sloan Digital Sky Survey Imager”. In: *AJ* 139 (Apr. 2010), pp. 1628–1648.

[32] D. M. Dumont et al. “A Comparison of GPS-measured Precipitable Water at Bartlett, NH With Radiosonde Measurements in the Northeast”. In: *Proceedings of the Eleventh Symposium on Meteorological Observations and Instrumentation*. 2001, pp. 245–247.

[33] B. Efron. “Bootstrap Methods: Another Look at the Jackknife”. In: *The Annals of Statistics* 7.1 (1979), pp. 1–26.

[34] J. H. Elias et al. “Type I supernovae in the infrared and their use as distance indicators.” In: *ApJ* 296 (Sept. 1985), pp. 379–389.

[35] Joseph Felsenstein. “Confidence Limits on Phylogenies: An Approach Using the Bootstrap”. In: *Evolution* 39.4 (1985), pp. 783–791.

[36] Alexei V. Filippenko, Michael W. Richmond, David Branch, et al. “The Subluminous, Spectroscopically Peculiar Type 1a Supernova 1991bg in the Elliptical Galaxy NGC 4374”. In: *AJ* 104 (Oct. 1992), p. 1543.

[37] Alexei V. Filippenko, Michael W. Richmond, Thomas Matheson, et al. "The Peculiar Type IA SN 1991T: Detonation of a White Dwarf?" In: *ApJ* 384 (Jan. 1992), p. L15.

[38] M. Fioc et al. "PEGASE: a UV to NIR spectral evolution model of galaxies. Application to the calibration of bright galaxy counts." In: *A&A* 500 (Oct. 1997), pp. 507–519.

[39] Edward L. Fitzpatrick. "Correcting for the Effects of Interstellar Extinction". In: *PASP* 111.755 (Jan. 1999), pp. 63–75.

[40] G. Folatelli et al. "Spectroscopy of Type Ia Supernovae by the Carnegie Supernova Project". In: *ApJ* 773, 53 (Aug. 2013), p. 53.

[41] M. Fukugita et al. "The Sloan Digital Sky Survey Photometric System". In: *AJ* 111 (Apr. 1996), p. 1748.

[42] Avishay Gal-Yam. "Observational and Physical Classification of Supernovae". In: *Handbook of Supernovae*. Ed. by Athem W. Alsabti et al. Cham: Springer International Publishing, 2017, pp. 1–43.

[43] L. Galbany et al. "Evidence for a Chandrasekhar-Mass Explosion in the Ca-strong 1991bg-like Type Ia Supernova 2016hnk". In: *A&A* 630, A76 (Oct. 2019), A76.

[44] Lluís Galbany, Mario Hamuy, et al. "UBVRIz Light Curves of 51 Type II Supernovae". In: *AJ* 151.2, 33 (Feb. 2016), p. 33.

[45] Lluís Galbany, Ramon Miquel, et al. "Type Ia Supernova Properties as a Function of the Distance to the Host Galaxy in the SDSS-II SN Survey". In: *ApJ* 755.2, 125 (Aug. 2012), p. 125.

[46] Mohan Ganeshalingam, Weidong Li, Alexei V. Filippenko, Carmen Anderson, et al. "Results of the Lick Observatory Supernova Search Follow-Up Photometry Program: BVRI Light Curves of 165 Type Ia Supernovae". In: *The Astrophysical Journal Supplement Series* 190.2 (Sept. 2010), pp. 418–448.

[47] Mohan Ganeshalingam, Weidong Li, Alexei V. Filippenko, Jeffrey M. Silverman, et al. "The Low-velocity, Rapidly Fading Type Ia Supernova 2002es". In: *ApJ* 751.2, 142 (June 2012), p. 142.

[48] S. González-Gaitán et al. "Defining Photometric Peculiar Type Ia Supernovae". In: *The Astrophysical Journal* 795.2 (2014), p. 142.

[49] S. González-Gaitán et al. “Subluminous Type Ia Supernovae at High Redshift from the Supernova Legacy Survey”. In: *The Astrophysical Journal* 727.2 (Jan. 2011), p. 107.

[50] James Guillochon et al. “An Open Catalog for Supernova Data”. In: *ApJ* 835.1, 64 (Jan. 2017), p. 64.

[51] J. E. Gunn et al. “The 2.5 m Telescope of the Sloan Digital Sky Survey”. In: *AJ* 131 (Apr. 2006), pp. 2332–2359.

[52] J. Guy et al. “SALT2: Using Distant Supernovae to Improve the Use of Type Ia Supernovae as Distance Indicators”. In: *A&A* 466.1 (Apr. 2007), pp. 11–21.

[53] Stephan Hachinger et al. “Spectral Luminosity Indicators in Type Ia Supernovae. Understanding the (Si ii) Line-Strength Ratio and Beyond”. In: *Monthly Notices of the Royal Astronomical Society* 389.3 (2008), pp. 1087–1096.

[54] M. Hamuy et al. “The Carnegie Supernova Project: The Low-Redshift Survey”. In: *PASP* 118 (Jan. 2006), pp. 2–20.

[55] A. Heger et al. “How Massive Single Stars End Their Life”. In: *ApJ* 591.1 (July 2003), pp. 288–300.

[56] Jon A. Holtzman et al. “The Sloan Digital Sky Surbey-II: Photometry and Supernova Ia Light Curves from the 2005 Data”. In: *The Astronomical Journal* 136.6 (Nov. 2008), pp. 2306–2320.

[57] D. Andrew Howell et al. “The Type Ia Supernova SNLS-03D3bb From a Super – Chandrasekhar – Mass White Dwarf Star”. In: *Nature* 443 (Sept. 2006), pp. 308–311.

[58] E. Y. Hsiao et al. “K-Corrections and Spectral Templates of Type Ia Supernovae”. In: *ApJ* 663.2 (July 2007), pp. 1187–1200.

[59] T. -O. Husser et al. “A new extensive library of PHOENIX stellar atmospheres and synthetic spectra”. In: *A&A* 553, A6 (May 2013), A6.

[60] Patrick Ingraham et al. “The LSST calibration hardware system design and development”. In: *Ground-based and Airborne Telescopes VI*. Ed. by Helen J. Hall et al. Vol. 9906. International Society for Optics and Photonics. SPIE, 2016, pp. 250–259.

[61] Ž. Ivezić et al. “Sloan Digital Sky Survey Standard Star Catalog for Stripe 82: The Dawn of Industrial 1% Optical Photometry”. In: *AJ* 134 (Sept. 2007), pp. 973–998.

[62] Željko Ivezić et al. “LSST: From Science Drivers to Reference Design and Anticipated Data Products”. In: *ApJ* 873.2, 111 (Mar. 2019), p. 111.

[63] Saurabh Jha et al. “Late-Time Spectroscopy of SN 2002cx: The Prototype of a New Subclass of Type Ia Supernovae”. In: *The Astronomical Journal* 132.1 (June 2006), pp. 189–196.

[64] D. O. Jones et al. “The Foundation Supernova Survey: Measuring Cosmological Parameters with Supernovae from a Single Telescope”. In: (2018).

[65] N. V. Karpenka et al. “A Simple and Robust Method for Automated Photometric Classification of Supernovae Using Neural Networks”. In: *Monthly Notices of the Royal Astronomical Society* 429.2 (Dec. 2013), pp. 1278–1285.

[66] K. Krisciunas et al. “The Carnegie Supernova Project. I. Third Photometry Data Release of Low-redshift Type Ia Supernovae and Other White Dwarf Explosions”. In: *AJ* 154, 211 (Nov. 2017), p. 211.

[67] Bruno Leibundgut et al. “SN 1991bg: A Type IA Supernova With a Difference”. In: *AJ* 105 (Jan. 1993), p. 301.

[68] Dan Li et al. “Temporal Variations of Telluric Water Vapor Absorption at Apache Point Observatory”. In: *Publications of the Astronomical Society of the Pacific* 130.983 (Dec. 2017), p. 14501.

[69] T. S. Li et al. “Assessment of Systematic Chromatic Errors that Impact Sub-1% Photometric Precision in Large-Area Sky Surveys”. In: *The Astronomical Journal* 151.6 (May 2016), p. 157.

[70] Ting Li et al. “Monitoring the atmospheric throughput at Cerro Tololo Inter-American Observatory with aTmCam”. In: *Proc. SPIE*. Vol. 9147. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series. 2014, 91476Z.

[71] Weidong Li, Alexei V. Filippenko, et al. “SN 2002cx: The Most Peculiar Known Type Ia Supernova”. In: *Publications of the Astronomical Society of the Pacific* 115 (Apr. 2003), pp. 453–473.

[72] Weidong Li, Jesse Leaman, et al. “Nearby Supernova Rates From the Lick Observatory Supernova Search – II. The Observed Luminosity Functions and Fractions of Supernovae in a Complete Sample”. In: *Monthly Notices of the Royal Astronomical Society* 412.3 (Apr. 2011), pp. 1441–1472.

[73] Michelle Lochner, Jason D. McEwen, et al. “Photometric Supernova Classification with Machine Learning”. In: *The Astrophysical Journal Supplement Series* 225.2 (Aug. 2016), p. 31.

[74] Michelle Lochner, Dan Scolnic, et al. “The Impact of Observing Strategy on Cosmological Constraints with LSST”. In: *arXiv e-prints*, arXiv:2104.05676 (Apr. 2021), arXiv:2104.05676.

[75] LSST Science Collaboration et al. “LSST Science Book, Version 2.0”. In: *arXiv e-prints*, arXiv:0912.0201 (Dec. 2009), arXiv:0912.0201.

[76] Eugene. A. Magnier et al. “Pan-STARRS Photometric and Astrometric Calibration”. In: *ApJS* 251.1, 6 (Nov. 2020), p. 6.

[77] Shilpa Manandhar et al. “A Potential Low Cost Remote Sensing Using GPS Derived PWV”. In: *arXiv e-prints*, arXiv:1805.01961 (May 2018), arXiv:1805.01961.

[78] X. Meng et al. “Do SN 2002cx-like and SN Ia-CSM Objects Share the Same Origin?” In: *ApJ* 861.2, 127 (July 2018), p. 127.

[79] R. Minkowski. “Spectra of Supernovae”. In: *PASP* 53.314 (Aug. 1941), p. 224.

[80] A. Möller et al. “Photometric Classification of Type Ia Supernovae in the SuperNova Legacy Survey with Supervised Learning”. In: *Journal of Cosmology and Astroparticle Physics* 2016.12 (Dec. 2016), pp. 008–008.

[81] J. Mosher et al. “A Precision Photometric Comparison between SDSS-II and CSP Type Ia Supernova Data”. In: *AJ* 144, 17 (July 2012), p. 17.

[82] Daniel Muthukrishna et al. “RAPID: Early Classification of Explosive Transients using Deep Learning”. In: *arXiv e-prints* (Mar. 2019), arXiv:1904.00014.

[83] Michael H. Nahmias et al. “A statistical examination and case study of integrated precipitable water (IPW) in Northern New England”. In: *BAMS* (Jan. 2004).

[84] K. Nomoto et al. “Carbon Deflagration Supernova, an Alternative to Carbon Detonation”. In: *Ap&SS* 39 (Feb. 1976), p. L37.

[85] Peter Nugent, Alex Kim, et al. “K-Corrections and Extinction Corrections for Type Ia Supernovae”. In: *Publications of the Astronomical Society of the Pacific* 114.798 (Aug. 2002), pp. 803–819.

[86] Peter Nugent, Mark Phillips, et al. “Evidence for a Spectroscopic Sequence among Type 1a Supernovae”. In: *ApJL* 455 (Dec. 1995), p. L147.

[87] F. Ochsenbein et al. “The VizieR Database of Astronomical Catalogues”. In: *A&AS* 143 (Apr. 2000), pp. 23–32.

[88] D. E. Osterbrock. “Who Really Coined the Word Supernova? Who First Predicted Neutron Stars?” In: *American Astronomical Society Meeting Abstracts*. Vol. 199. American Astronomical Society Meeting Abstracts. Dec. 2001, p. 15.01.

[89] Rüdiger Pakmor et al. “Sub-luminous type Ia supernovae from the mergers of equal-mass white dwarfs with mass ~0.9M_{solar}”. In: *Nature* 463.7277 (Jan. 2010), pp. 61–64.

[90] Pasquet, Johanna et al. “PELICAN: deeP architecturE for the LIght Curve ANalysis”. In: *A&A* 627 (2019), A21.

[91] S. Perlmutter et al. “Measurements of Omega and Lambda from 42 High-Redshift Supernovae”. In: *ApJ* 517 (June 1999), pp. 565–586.

[92] Daniel Perrefort, W. M. Wood-Vasey, et al. “pwv_kpno: A Python Package for Modeling the Atmospheric Transmission Function due to Precipitable Water Vapor”. In: *PASP* 131.996 (Feb. 2019), p. 025002.

[93] Daniel Perrefort, Yike Zhang, et al. “A Template-based Approach to the Photometric Classification of SN 1991bg-like Supernovae in the SDSS-II Supernova Survey”. In: *ApJ* 904.2, 156 (Dec. 2020), p. 156.

[94] S. E. Persson et al. “The Las Campanas Infrared Survey Camera”. In: *The Astronomical Journal* 124.1 (July 2002), pp. 619–634.

[95] M. M. Phillips. “The Absolute Magnitudes of Type IA Supernovae”. In: *ApJL* 413 (Aug. 1993), pp. L105–L108.

[96] M. M. Phillips, Paulina Lira, et al. “The Reddening-Free Decline Rate Versus Luminosity Relationship for Type IA Supernovae”. In: *AJ* 118 (Oct. 1999), pp. 1766–1776.

[97] M. M. Phillips, Lisa A. Wells, et al. “SN 1991T: Further Evidence of the Heterogeneous Nature of Type IA Supernovae”. In: *AJ* 103 (May 1992), p. 1632.

[98] J. D. R. Pierel et al. “Extending Supernova Spectral Templates for Next-generation Space Telescope Observations”. In: *PASP* 130.993 (Nov. 2018), p. 114504.

[99] A. M. Price-Whelan et al. “The Astropy Project: Building an Open-Science Project and Status of the v2.0 Core Package”. In: *AJ* 156, 123 (Sept. 2018), p. 123.

[100] M. V. Pruzhinskaya et al. “Anomaly Detection in the Open Supernova Catalog”. In: *MNRAS* 489.3 (Nov. 2019), pp. 3591–3608.

[101] R. R. Querel et al. “All-Sky Homogeneity of Precipitable Water Vapour over Paranal”. In: *Ground-based and Airborne Instrumentation for Astronomy V*. Vol. 9147. Proc. SPIE. Aug. 2014, p. 914792.

[102] A. Rest et al. “Cosmological Constraints from Measurements of Type Ia Supernovae Discovered during the First 1.5 yr of the Pan-STARRS1 Survey”. In: *ApJ* 795, 44 (Nov. 2014), p. 44.

[103] Joseph W. Richards et al. “Construction of a Calibrated Probabilistic Classification Catalog: Application to 50k Variable Sources in the All-Sky Automated Survey”. In: *The Astrophysical Journal Supplement Series* 203.2 (Dec. 2012), p. 32.

[104] A. G. Riess, A. V. Filippenko, et al. “Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant”. In: *AJ* 116 (Sept. 1998), pp. 1009–1038.

[105] A. G. Riess, W. H. Press, et al. “A Precise Distance Indicator: Type IA Supernova Multicolor Light-Curve Shapes”. In: *ApJ* 473 (Dec. 1996), p. 88.

[106] Masao Sako. Private Communication. Philadelphia, PA USA, 2020.

[107] Masao Sako, Bruce Bassett, Andrew C. Becker, et al. “The Data Release of the Sloan Digital Sky Survey-II Supernova Survey”. In: *Publications of the Astronomical Society of the Pacific* 130.988 (May 2018), p. 64002.

[108] Masao Sako, Bruce Bassett, Andrew Becker, et al. “The Sloan Digital Sky Survey-II Supernova Survey: Search Algorithm and Follow-up Observations”. In: *AJ* 135.1 (Jan. 2008), pp. 348–373.

[109] Masao Sako, Bruce Bassett, Brian Connolly, et al. “Photometric Type Ia Supernova Candidates from the Three-year SDSS-II SN Survey Data”. In: *ApJ* 738.2, 162 (Sept. 2011), p. 162.

[110] M. Sasdelli et al. “Exploring the Spectroscopic Diversity of Type Ia Supernovae with DRACULA: a machine learning approach”. In: *MNRAS* 461 (Sept. 2016), pp. 2044–2059.

[111] David J. Schlegel et al. “Maps of Dust Infrared Emission for Use in Estimation of Reddening and Cosmic Microwave Background Radiation Foregrounds”. In: *ApJ* 500.2 (June 1998), pp. 525–553.

[112] F. W. Scholz et al. “K-Sample Anderson-Darling Tests”. In: *Journal of the American Statistical Association* 82.399 (1987), pp. 918–924.

[113] D. M. Scolnic et al. “The Complete Light-curve Sample of Spectroscopically Confirmed SNe Ia from Pan-STARRS1 and Cosmological Constraints from the Combined Pantheon Sample”. In: *ApJ* 859.2, 101 (June 2018), p. 101.

[114] D. Scolnic et al. “Systematic Uncertainties Associated with the Cosmological Analysis of the First Pan-STARRS1 Type Ia Supernova Sample”. In: *ApJ* 795, 45 (Nov. 2014), p. 45.

[115] Jacques Sebag et al. “Calibration systems for LSST”. In: *Ground-based and Airborne Telescopes V*. Ed. by Larry M. Stepp et al. Vol. 9145. International Society for Optics and Photonics. SPIE, 2014, pp. 1335–1345.

[116] Jeffrey M. Silverman et al. “Berkeley Supernova Ia Program – I. Observations, data reduction and spectroscopic sample of 582 low-redshift Type Ia supernovae”. In: *Monthly Notices of the Royal Astronomical Society* 425.3 (Sept. 2012), pp. 1789–1818.

[117] M. D. Stritzinger et al. “The Carnegie Supernova Project: Second Photometry Data Release of Low-redshift Type Ia Supernovae”. In: *AJ* 142, 156 (Nov. 2011), p. 156.

[118] Stefan Taubenberger. “The Extremes of Thermonuclear Supernovae”. In: *Handbook of Supernovae, ISBN 978-3-319-21845-8. Springer International Publishing AG, 2017, p. 317*. 2017, p. 317.

[119] The LSST Dark Energy Science Collaboration et al. “The LSST Dark Energy Science Collaboration (DESC) Science Requirements Document”. In: *arXiv e-prints*, arXiv:1809.01669 (Sept. 2018), arXiv:1809.01669.

[120] The PLAsTiCC team et al. “The Photometric LSST Astronomical Time-series Classification Challenge (PLAsTiCC): Data set”. In: *arXiv e-prints*, arXiv:1810.00001 (Sept. 2018), arXiv:1810.00001.

[121] Doug Tody. “IRAF in the Nineties”. In: *Astronomical Data Analysis Software and Systems II*. Ed. by R. J. Hanisch et al. Vol. 52. Astronomical Society of the Pacific Conference Series. Jan. 1993, p. 173.

[122] Robert Tripp et al. “Determination of the Hubble Constant Using a Two-Parameter Luminosity Correction for Type Ia Supernovae”. In: *The Astrophysical Journal* 525.1 (Nov. 1999), pp. 209–214.

[123] M. Turatto et al. “The Properties of the Peculiar Type Ia Supernova 1991bg. I. Analysis and Discussion of Two Years of Observations.” In: *MNRAS* 283 (Nov. 1996), pp. 1–17.

[124] Melvin M. Varughese et al. “Non-parametric Transient Classification Using Adaptive Wavelets”. In: *Monthly Notices of the Royal Astronomical Society* 453.3 (Sept. 2015), pp. 2848–2861.

[125] R. A. Wade et al. “The Radial Velocity Curve and Peculiar TiO Distribution of the Red Secondary Star in Z Chamaeleontis”. In: *ApJ* 324 (Jan. 1988), pp. 411–430.

[126] X. Wang et al. “Improved Distances to Type Ia Supernovae with Two Spectroscopic Subclasses”. In: *ApJL* 699.2 (July 2009), pp. L139–L143.

[127] Randolph H. Ware et al. “SuomiNet: A Real-Time National GPS Network for Atmospheric Research and Education”. In: *Bulletin of the American Meteorological Society* 81.4 (2000), pp. 677–694.

[128] R. F. Webbink. “Double white dwarfs as progenitors of R Coronae Borealis stars and type I supernovae.” In: *ApJ* 277 (Feb. 1984), pp. 355–360.

[129] J. C. Wheeler and R. P. Harkness. “Type I supernovae”. In: *Reports on Progress in Physics* 53.12 (Dec. 1990), pp. 1467–1557.

[130] J. C. Wheeler and R. Levreault. “The peculiar type I supernova in NGC 991.” In: *ApJL* 294 (July 1985), pp. L17–L20.

[131] John Whelan et al. “Binaries and Supernovae of Type I”. In: *ApJ* 186 (Dec. 1973), pp. 1007–1014.

[132] W. M. Wood-Vasey et al. “GPS Measurements of Precipitable Water Vapor Can Improve Survey Calibration: A Demonstration from KPNO and the Mayall z-band Legacy Survey”. In: *arXiv e-prints*, arXiv:2010.14240 (Oct. 2020), arXiv:2010.14240.

[133] D. G. York et al. “The Sloan Digital Sky Survey: Technical Summary”. In: *AJ* 120 (Sept. 2000), pp. 1579–1587.

[134] Chen Zheng et al. “First-Year Spectroscopy for the Sloan Digital Sky Survey-II Supernova Survey”. In: *AJ* 135.5 (May 2008), pp. 1766–1784.