Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **536,** 624–663 (2025) https://doi.org/10.1093/mnras/stae2501
Advance Access publication 2024 November 7

Hawai'i Supernova Flows: a peculiar velocity survey using over a Thousand Supernovae in the near-infrared

Aaron Do [1,2★] Benjamin J. Shappee [2] John L. Tonry [2] R. Brent Tully [2] Thomas de Jaeger [2,3] David Rubin [2,4] Chris Ashall [2,5] Christopher R. Burns [6] Dhvanil D. Desai [2] Jason T. Hinkle [2] Willem B. Hoogendam [2] Mark E. Huber [2] David O. Jones [2] Kaisey S. Mandel [1] Anna V. Payne [7] Erik R. Peterson [8] Dan Scolnic [8] and Michael A. Tucker [9,10,11]

[1]*Institute of Astronomy and Kavli Institute for Cosmology, Madingley Road, Cambridge, CB3 0HA, UK*
[2]*Institute for Astronomy, University of Hawai'i, 2680 Woodlawn Dr, Honolulu, HI 96822, USA*
[3]*CNRS/IN2P3, Sorbonne Université, Université Paris Cité, Laboratoire de Physique Nucléaire et de Hautes Énergies, F-75005 Paris, France*
[4]*Department of Physics and Astronomy, University of Hawai'i, Honolulu, HI 96822, USA*
[5]*Department of Physics, Virginia Tech, Blacksburg, VA 24061, USA*
[6]*The Observatories of the Carnegie Institution for Science, 813 Santa Barbara Street, Pasadena, CA 91101, USA*
[7]*Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA*
[8]*Department of Physics, Duke University, Durham, NC 27708, USA*
[9]*Center for Cosmology and Astroparticle Physics, The Ohio State University, 191 West Woodruff Avenue, Columbus, OH 43210, USA*
[10]*Department of Astronomy, The Ohio State University, 140 West 18th Avenue, Columbus, OH 43210, USA*
[11]*Department of Physics, The Ohio State University, 191 West Woodruff Avenue, Columbus, OH 43210, USA*

Accepted 2024 October 30. Received 2024 October 15; in original form 2024 March 8

ABSTRACT

We introduce the Hawai'i Supernova Flows project and present summary statistics of the first 1217 astronomical transients observed, 668 of which are spectroscopically classified Type Ia Supernovae (SNe Ia). Our project is designed to obtain systematics-limited distances to SNe Ia while consuming minimal dedicated observational resources. To date, we have performed almost 5000 near-infrared (NIR) observations of astronomical transients and have obtained spectra for over 200 host galaxies lacking published spectroscopic redshifts. In this survey paper, we describe the methodology used to select targets, collect/reduce data, calculate distances, and perform quality cuts. We compare our methods to those used in similar studies, finding general agreement or mild improvement. Our summary statistics include various parametrizations of dispersion in the Hubble diagrams produced using fits to several commonly used SN Ia models. We find the lowest dispersions using the `SNooPy` package's EBV_model2, with a root mean square deviation of 0.165 mag and a normalized median absolute deviation of 0.123 mag. The full utility of the Hawai'i Supernova Flows data set far exceeds the analyses presented in this paper. Our photometry will provide a valuable test bed for models of SN Ia incorporating NIR data. Differential cosmological studies comparing optical samples and combined optical and NIR samples will have increased leverage for constraining chromatic effects like dust extinction. We invite the community to explore our data by making the light curves, fits, and host galaxy redshifts publicly accessible.

Key words: catalogues – galaxies: distances and redshifts – (cosmology:) large-scale structure of Universe – transients: supernovae.

1 INTRODUCTION

Hawai'i Supernova Flows is an ongoing effort to map the distribution of mass in the local universe ($z < 0.1$) using near-infrared (NIR) observations of Type Ia Supernovae (SNe Ia) in combination with untargeted optical surveys. In this paper, we provide an overview of the Hawai'i Supernova Flows project to support future papers examining detailed science cases using data from Hawai'i Supernova Flows.

The paper is organized as follows. In this Section, we review the connection between large-scale structure and peculiar velocities, describe the largest contemporary peculiar velocity surveys, and motivate our use of SNe Ia. Section 2 describes the individual components of the project: including the target selection process, the observing facilities used, the photometric calibration and analysis, the identification of host galaxies, and the determination of their redshifts. Section 3 describes the three SN Ia fitting procedures we employ and how each set of fitting parameters is converted to distance moduli. We validate our fitting methodology and photometry using data from the Dark Energy, H_0, and peculiar Velocities using Infrared Light from Supernovae survey (DEHVILS; Peterson et al. 2023) and

★ E-mail: ajmd6@cam.ac.uk

the Carnegie Supernova Project's third data release (CSP-I DR3; Krisciunas et al. 2017) in Section 4. Section 5 lists the quality cuts used to define the various samples we describe, analyse, and discuss in Section 6.

1.1 Peculiar velocities and the state-of-the-art

While the LSS contains some luminous, baryonic matter, the majority of its mass may only be studied through its gravitational effects (for a review of dark matter, consider Bertone & Hooper 2018). In comoving coordinates, objects accelerate towards denser regions of LSS and away from voids. This motion is called peculiar velocity and its projection on our line of sight may be calculated as (Davis & Scrimgeour 2014)

$$v = c\left(\frac{z_{\rm obs} - z_{\rm cos}(d_L)}{1 + z_{\rm cos}(d_L)}\right), \tag{1}$$

where $z_{\rm obs}$ is the observed redshift and $z_{\rm cos}(d_L)$ is the redshift at luminosity distance d_L due to universal expansion in a given cosmological model with deceleration parameter q_0 and Hubble constant H_0 (Peebles 1993)

$$z_{\rm cos} \approx \frac{1}{1-q_0}\left[-1 + \sqrt{1 + \frac{2H_0 d_L}{c}(1-q_0)}\right]. \tag{2}$$

Peculiar velocities have been used to infer the distribution of LSS through a variety of approaches like the POTENT method (Bertschinger & Dekel 1989; Dekel, Bertschinger & Faber 1990; Dekel et al. 1999), the Wiener Filter and constrained realizations method (Ganon & Hoffman 1993; Zaroubi et al. 1995; Zaroubi, Hoffman & Dekel 1999; Courtois et al. 2012), the unbiased minimal variance estimator (Zaroubi 2002), and various Bayesian hierarchical approaches (Lavaux 2016; Graziani et al. 2019; Valade et al. 2022). These methods commonly assume the LSS formed through gravitational instabilities, and is thus irrotational on large scales ($\nabla \times \vec{v} = 0$) (Peebles 1980). Variations between the methods typically represent different approaches to minimizing the systematic effects of smoothing, uneven sky coverage, and biases in peculiar-velocity surveys. Modern cosmographic surveys are not limited by analytical tools, but by the number and precision of distance measurements.

Many peculiar velocity surveys use either the fundamental plane (FP; Djorgovski & Davis 1987; Dressler et al. 1987) or the Tully–Fisher relation (TF; Tully & Fisher 1977) to measure distances because these methods can be applied to a significant fraction of all galaxies, whereas other methods require relatively rare phenomena like a gravitational lens (Refsdal 1964), a maser (Herrnstein et al. 1999) or megamaser (Gao et al. 2016), a gravitational wave event (Holz & Hughes 2005), or a supernova (SN). However, while the FP and TF methods have significant advantages in target availability, the resulting distance measurements are often five to ten times less precise than measurements from more narrowly applicable distance probes. The FP and TF methods, along with most photometric measures of distance, produce errors in distance modulus, which causes error to increase with the distance. This proportionality is directly passed on to the uncertainties in peculiar velocity. While independent peculiar velocity measurements of N neighbouring galaxies can be combined to reduce the statistical uncertainty by a factor of $\sqrt{N}$, galaxies have a finite amount of neighbours. A volume-limited peculiar velocity survey will always find a noise floor that scales with the uncertainty in the distance-measuring technique and inversely with the root of galaxy number density. Put another way, a survey with an explicit precision requirement has a maximum effective range that cannot be extended without more precise measures of distance.

For this reason, two of the three largest homogeneous collections of peculiar velocities extend no farther than a cosmic microwave background (CMB) rest-frame redshift of $z_{\rm CMB} = 0.05$. These are the Cosmicflows-IV TF catalogue (CF4-TF, 9792 galaxies; Kourkchi et al. 2020) and the FP-based 6-degree Field Galaxy Survey peculiar velocity sample (6dFGSv, 8885 galaxies; Campbell et al. 2014; Springob et al. 2014). The Sloan Digital Sky Survey (SDSS) peculiar velocity catalogue (SDSS-PV, 34 059 galaxies; Howlett et al. 2022) is the first FP- or TF-based survey to extend to $z_{\rm CMB} = 0.1$, but so far only covers the SDSS North Galactic Cap contiguous area (7016 deg^2). The largest compilation of extragalactic distances is the heterogeneous catalogue Cosmicflows-IV (CF4, 55 877 galaxies; Tully et al. 2023), which consolidates these and other surveys and uses both FP and TF measurements, as well as surface brightness fluctuations (Tonry & Schneider 1988), core-collapse SNe (Hamuy & Pinto 2002), and SNe Ia (Phillips 1993).

The SDSS-PV sample has not yet produced any detailed cosmographic studies, but the authors measured a bulk flow in mild excess ($p \sim 0.06$–0.20 depending on cuts) of what a fiducial dark energy and cold dark matter (ΛCDM) model would suggest. This excess has been suggested before using various independent data sets (e.g. Pike & Hudson 2005; Feldman & Watkins 2008; Kashlinsky et al. 2008; Feldman, Watkins & Hudson 2010; Lavaux et al. 2010). Contemporary analyses extend the scale of the issue, with Watkins et al. (2023) finding that CF4 data indicate excess bulk flows on scales of 200 h^{-1} Mpc that have a 1.5×10^{-6} chance of occurring in the standard cosmological model using CMB-derived parameters. Howlett et al. (2022) theorize that the Shapley Supercluster as seen in the 2M++ redshift compilation (Carrick et al. 2015) could be responsible, but because it is not in the SDSS-PV survey footprint it will be difficult to test. A survey that trades depth for sky coverage will still struggle to constrain the effects of the Shapley Supercluster, as Carrick et al. (2015) find their bulk flow measurements prefer a contribution from sources at $z > 0.067$ at a 5.1σ level.

Thus far, the Hawai'i Supernova Flows project has obtained peculiar velocity measurements over three quarters of the sky to a depth of $z \sim 0.1$. This encompasses the gravitational sources thought to dominate local dynamics, including the Shapley supercluster, the dipole repeller (Hoffman et al. 2017), and the cold spot repeller (Courtois et al. 2017). Equation (1) shows that peculiar velocities require an assumed cosmology and two measurements: an observed redshift and a proper distance. The redshift can be measured to high precision with a single spectrum, but measuring distances is more difficult. Techniques have been developed and refined to excel in various niches of a parameter space spanning applicability, maximum range, and precision. Our project uses optical and NIR observations of SNe Ia to measure distances.

1.2 Type Ia Supernovae

Following the discovery that SNe Ia could be used as standardizable candles (Pskovskii 1977; Phillips 1993; Tripp 1998) there have been continuous efforts to improve the accuracy and precision of SNe Ia distance inference. These efforts include refining theoretical models of SNe Ia progenitors and explosions (reviewed in Liu, Röpke & Han 2023); increasing the sample of well-studied SNe Ia (e.g. Amanullah et al. 2010; Phillips et al. 2019; Scolnic et al. 2018; Brout et al. 2022); and empirically identifying correlations between SNe Ia luminosities and observable parameters like host-galaxy mass (Kelly et al. 2010; Lampeitl et al. 2010; Sullivan et al. 2010), host-galaxy specific star

formation rate (Uddin et al. 2017), local H $_\alpha$ surface brightness (Rigault et al. 2013), host-galaxy metallicity (Moreno-Raya et al. 2016), host-galaxy colours (Roman et al. 2018), ejecta velocity (Léget et al. 2020), and more. Accompanying these efforts are improvements to fitting and modelling techniques (`BayeSN`, Mandel et al. 2009; Mandel, Narayan & Kirshner 2011; Thorp et al. 2021; Mandel et al. 2022; Grayling et al. 2024; `MLCS2k2`[1], Jha, Riess & Kirshner 2007; `SALT`[2], Guy et al. 2005, 2007, 2010; Kenworthy et al. 2021; Pierel et al. 2022; `SiFTO`, Conley et al. 2008; `SNEMO`[3], Saunders et al. 2018; `SNooPy`[4], Burns et al. 2011, 2014; `SUGAR`[5], Léget et al. 2020). This body of work has established SNe Ia as excellent probes of distance. We choose to use them over competing distance measuring techniques for three reasons.

First, SNe Ia are abundant. With modern surveys across the globe constantly scanning the sky, SNe are no longer rare targets of opportunity, but are discovered every night. Desai et al. (2024) use data from the All-Sky Automated Survey for SuperNovae (ASAS-SN; Shappee et al. 2014; Kochanek et al. 2017; Hart et al. 2023) to find an SN Ia volumetric rate of $\sim 2.3 \times 10^4$ yr^{-1} Gpc$^{-3}h_{70}^3$, which amounts to about 20 each night within $z < 0.1$. Wiseman et al. (2021) use results from the Dark Energy Survey (DES) to calculate a rate of SNe Ia per galaxy between one every 3000 yr to more than one every 100 yr depending on host-galaxy properties. This means that although the number of usable galaxies in an SNe Ia-based peculiar velocity survey is relatively low compared to TF or FP surveys, it scales with time and can exceed competing methods with enough observational support.

Secondly, SNe Ia are bright enough to be used at the distances we require. The demonstration of accelerating expansion relied on measurements of SNe Ia at redshifts near unity (Riess et al. 1998; Perlmutter et al. 1999). Our interests are more local, extending to redshifts $z < 0.1$. The mean absolute magnitude of SNe Ia before correcting for host-galaxy extinction is about -18.6 mag in B and -18.7 mag in V (e.g. Ashall et al. 2016). At $z = 0.1$ this corresponds to an apparent magnitude of about 19.6 or 19.5 mag, within the limiting magnitude of two of the all-sky surveys described in Section 2.1.

Lastly, SNe Ia-based distance measurements are far more precise than those of competing methodologies. This is not to say that SNe Ia are the most precise of all distance indicators. Distances based on Cepheid period–luminosity relations (Leavitt & Pickering 1912) or the Tip of the Red Giant Branch (Freedman et al. 2020; Anand et al. 2021) are typically more precise than those based on SNe Ia, but the objects of study for these probes are about 13–16 mag fainter than SNe Ia. This restricts them to $z < 0.023$ even with 22 HST orbits per galaxy (PI D. Jones; proposal 16269). Hawai'i Supernova Flows extends about 4 times farther. The SDSS-PV sample has used the FP method to measure distances at $z \sim 0.1$, but these distances are only precise to around 20 per cent. SNe Ia-based distances can be systematically corrected to a root mean square (RMS) scatter between 4 and 7 per cent (Burns et al. 2018; Scolnic et al. 2018). This means that it would take several dozens of independent TF or FP measurements to reach the precision of a single SN Ia distance measurement.

Optical SNe Ia light curves have been used as standardizable candles for several decades (e.g. Phillips 1993; Hamuy et al. 1995; Riess et al. 1998; Perlmutter et al. 1999), but a growing body of evidence (e.g. Kasen 2006; Wood-Vasey et al. 2008; Burns et al. 2011; Dhawan, Jha & Leibundgut 2018; Avelino et al. 2019) suggests that the NIR may offer substantial advantages.

1.2.1 SNe Ia in the near-infrared

NIR bandpasses like Y, J, H, and K are 5–11 times less affected by dust than the traditionally used B band (Cardelli, Clayton & Mathis 1989; O'Donnell 1994; Fitzpatrick 1999). The total-to-selective extinction parameter R_V is known to vary based on the properties of dust, even in our own galaxy (Draine 2003). Brout & Scolnic (2021) and Popovic et al. (2023) proposed that the dispersion in Hubble residuals of red SNe Ia may be largely due to the uncertain properties of extragalactic dust, which varies as a function of position in the host galaxy. The effects of dust correlate with the colour of any given SN Ia, making any added uncertainty a systematic issue that may not be resolved with a larger sample. Studying SNe Ia in the NIR suppresses the systematic error associated with dust.

Additionally, SNe Ia have been claimed to be more uniform in the NIR (e.g. Wood-Vasey et al. 2008; Barone-Nugent et al. 2012; Kattner et al. 2012; Stanishev et al. 2018; Avelino et al. 2019; Galbany et al. 2023; Jones et al. 2022). Avelino et al. (2019) used NIR light curves to determine distances consistent with those determined using optical light curves. Notably, Avelino et al. (2019) did not apply the typical standardizations to the NIR light curves, but did correct the optical light curves for decline rate, host-galaxy extinction, and host-galaxy mass. The empirical regularity of SNe Ia peak magnitudes in the NIR is supported by theory (Kasen 2006), with radiative transfer calculations showing how decreases in bolometric flux are balanced by increases in relative emission at longer wavelengths. The remarkable uniformity of SNe Ia peak absolute magnitudes in the NIR makes any distance measurement much more robust against systematic uncertainties.

2 PROJECT COMPONENTS AND OBSERVATIONAL FACILITIES

Initial testing showed that SNe Ia observations spanning the NIR-peak produce RMS dispersions in Hubble residuals $\sim$ 10–30 per cent lower than values obtained for SNe Ia only observed after the peak. Thus, to obtain distances to SNe Ia and recessional velocities for their host galaxies, we require three types of data: high-cadence photometry to find SNe Ia before they reach their NIR peaks, NIR photometry of each SN Ia near their peaks, and spectroscopically determined redshifts of their host-galaxies.

Fig. 1 illustrates the various components of the program, delineating what is supplied from the community and what requires dedicated observing resources.

2.1 Triggers from All-Sky Surveys

The entire sky is imaged multiple times per night by All-Sky Surveys like the Asteroid Terrestrial-impact Last Alert System (ATLAS; Tonry et al. 2018), the Zwicky Transient Facility (ZTF; Bellm et al. 2019), and ASAS-SN. These surveys operate with different cadences and depths to cover a range of science cases, but they all search the sky for objects that vary on time-scales of hours, days, or months. SNe Ia are in this class of astronomical objects, with light curves that increase in brightness for a few weeks before peaking, declining over a month, and then exponentially decaying. Here, we describe

[1]Multicolour light-curve shapes.
[2]Spectral adaptive light-curve template.
[3]SuperNova Empirical MOdels.
[4]SuperNovae in Object Oriented Python.
[5]SUpernova Generator And Reconstructor.

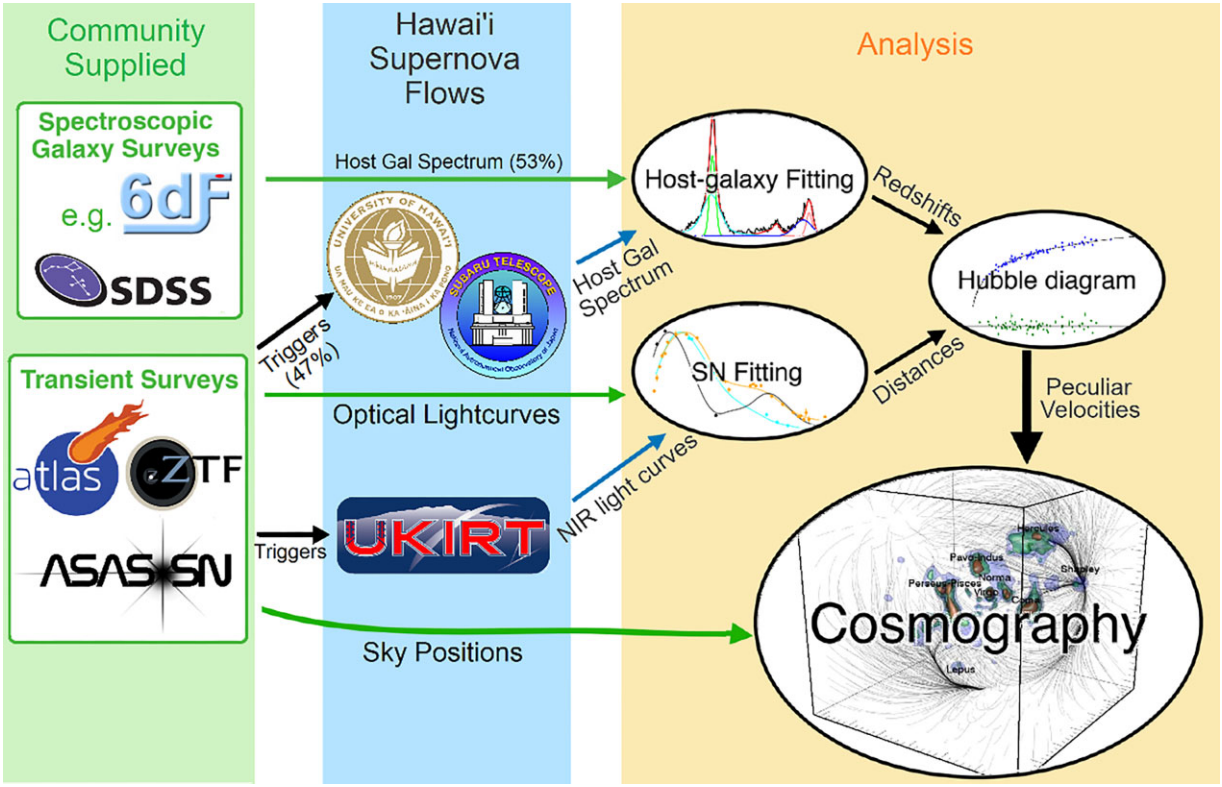

Figure 1. Our project uses archival and survey data as a foundation for supplementary observations. Whenever a new transient is reported, we collect optical light curves from ATLAS, ASAS-SN, and ZTF. We trigger NIR observations for targets that are either classified as SNe Ia or are unclassified and have a SN Ia-like light curve. About 53 per cent of targets we observe are associated with galaxies that have spectroscopic redshifts, and we pursue spectroscopic observations for the remaining 47 per cent with either the University of Hawai'i (UH) 2.2 m telescope or Subaru depending on their surface brightness profiles. The optical and NIR light curves allow us to infer luminosity distances, which we combine with host-galaxy redshifts to derive peculiar velocities.

the archival and observational facilities used, and how we access, store, and process the data.

2.1.1 *The Transient Name Server*

The Transient Name Server (TNS)[6] is the official International Astronomical Union repository for extragalactic transients. Large observational campaigns such as Pan-STARRS (Chambers et al. 2016), GaiaAlerts[7] (Gaia Collaboration 2016, 2018), the surveys described in the following sections, and many more automatically generate reports within minutes to hours of exposure read-out. Averaging overall reports from TNS, about 10 per cent of transients receive observational follow-up and spectroscopic classification, and of these, about 70 per cent are SNe Ia.[8] The majority of transients fade and become unobservable without being classified.

The Hawai'i Supernova Flows project uses the TNS-provided PYTHON code[9] to solicit new and recently updated reports every half hour, and uses these reports to generate a list of SNe Ia candidates. We ignore transients that are classified as anything other than an SN Ia or non-peculiar sub-type, but still consider unclassified transients as potential SNe Ia. This leads to some NIR observations of targets that are later classified as non-SN Ia, but we cannot afford to wait for spectroscopic classification of each target, which often occurs after the NIR-peak as seen in Fig. 2.

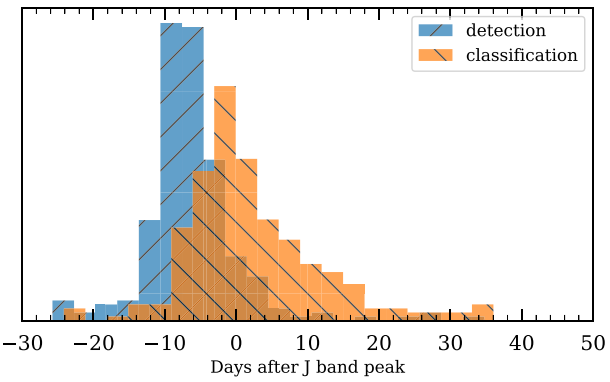

Figure 2. The distribution of detection dates and public classification dates for SNe Ia relative to the J-band maximum light. About 40 per cent of all SNe Ia are classified more than a day before the NIR peak.

The reduction in efficiency can be mitigated in several ways. The Hawai'i Supernova Flows team relays targets of interest to the Spectroscopic Classification of Astronomical Transients (SCAT) program (Tucker et al. 2022). The SCAT team classifies astronomical transients using spectra primarily from the University of Hawai'i (UH) 2.2 m telescope (instrumentation described in more detail in Section 2.3.3), but has recently expanded to the Australian National University 2.3 m telescope through a collaboration with Melbourne

[6] https://www.wis-tns.org/
[7] http://gsaweb.ast.cam.ac.uk/alerts
[8] https://www.wis-tns.org/stats-maps
[9] https://www.wis-tns.org/sites/default/files/api/tns_api_search.py.zip

Table 1. The spectroscopic classifications of our observed targets show that a strong majority of our targets receive classification, and most of those classifications are some kind of SN Ia. We need to select targets before they are classified in order to observe the NIR peak, which precedes the optical peak by several days. This results in some observations of non-SN Ia objects, but these targets are promptly removed from the observing queue.

Type	Number
SN Ia-norm	637
SN Ia-91T-like	25
SN Ia-91bg-like	6
Unclassified	327
SN Ia-pec	3
SN Ia-CSM	2
SN Iax[02cx-like]	2
SN Ia-SC	1
SN II	93
SN IIn	15
SN Ic	12
SN Ib	11
CV	7
SN IIP	6
SN	4
SN Ibn	4
SN Ib/c	3
SN I	3
SN Ic-BL	3
Nova	2
SLSN-II	2
LRN	1
AGN	1
SN Ib-Ca-rich	1
Varstar	1
SLSN-I	1
Impostor-SN	1
ILRT	1

University. In a random sampling of TNS objects, one would expect 10 per cent to be classified, but by providing SCAT with a list of candidates to observe, we increase the fraction of classified transients in our observed sample to about 73 per cent. Additionally, Möller & de Boissière (2020) demonstrated that using whole light curves, SNe Ia and non-SNe Ia can be identified with up to 95 per cent accuracy, or 98 per cent accuracy when including host-galaxy information. Even when restricting the light curves to early times, the difference in light-curve shape between various SNe allows us to avoid observing unclassified targets that are unlikely to be SNe Ia. The demographics of Hawai'i Supernova Flows targets are presented in Table 1.

The following sections describe three untargeted surveys with publicly available light-curve generation services that we use to improve our triggering process, and as later detailed in Section 3, improve our distance determinations.

2.1.2 ATLAS

ATLAS consists of four fully robotic, 0.5 m *f*/2 Wright Schmidt telescopes that image the entire night sky about once every two days (Tonry 2011; Tonry et al. 2018). This system was designed to identify potentially hazardous asteroids, and optimizations for that purpose affect the utility of ATLAS in studying astrophysical transients.

An object's orbital elements are fairly decoupled from its spectral properties, so to increase throughput, ATLAS uses two non-standard broad filters, a 'cyan' filter covering 420–650 nm and an 'orange' filter covering 560–820 nm. This aids its primary science mission by increasing 'survey speed' (Tonry 2011), but presents unique challenges for integrating observations with other filter systems, which we describe in Section 3.1.

Additionally, to specialize in moving object detection, the telescope system observes each field of view with four 30-s exposures over a 1-h interval. Under nominal conditions, each 30-s exposure reaches a median 5σ detection limit of $o \sim 19.1$ AB mag and $c \sim 19.6$ AB mag. For stationary targets, these exposures can be co-added to improve depth by about 0.75 AB mag and increase the signal-to-noise-ratio (SNR) at a given brightness by a factor of 2. However, we found that interobservational variation in point spread function (PSF), pointing, and atmospheric conditions made combining multiple exposures difficult. Instead, we combine the four photometric measurements of each object using an inverse variance weighted median, excluding any measurement more than three times its uncertainty away from the median flux. Additionally, we ignore measurements where the object is within 40 pixels of a chip edge or has an axis ratio greater than 1.5 and measurements where the sky brightness is under 16.

Although ATLAS specializes in astronomy at the Solar system scale, it is a leading source of high-cadence data for studying astrophysical transients. Smith et al. (2020) describe the utility of ATLAS in this context and how to access data using the ATLAS Forced Photometry server.[10] Hawai'i Supernova Flows continues to use the proprietary channel we developed to access light curves before the forced photometry server came online, but the data collected exactly match the publicly available data.

2.1.3 ASAS-SN

ASAS-SN is a globally distributed system of 20 fully robotic telescopes focused on discovering bright, nearby SNe (Shappee et al. 2014; Kochanek et al. 2017; Hart et al. 2023). Each of the five ASAS-SN sites employs four 14 cm telescopes sharing a common mount. The original two sites used the Johnson *V*-bandpass, but since 2019 all observations use the Sloan *g*-bandpass (Holoien et al. 2020). Each pointing consists of three dithered 90 s exposures, reaching median 5σ detection limits of 17.8 AB mag each (Kochanek et al. 2017). These exposures can be co-added to improve depth by about 0.6 AB mag and increase SNR by a factor of $\sqrt{3}$. The system images the entire sky about once every 20 h, with few losses due to weather because of the numerous sites.

The ASAS-SN light curve server described in Kochanek et al. (2017) has grown into the ASAS-SN Sky Patrol,[11] which serves light curves for any position on the sky. As with ATLAS, we access this publicly available data using a proprietary channel to minimize overheads.

2.1.4 ZTF

ZTF uses the Palomar 48-in Schmidt telescope to pursue science objectives across a range of cadences, depths, and areas, with an emphasis on SNe (Bellm et al. 2019; Graham et al. 2019). Through the public surveys, ZTF covered the night sky North of $\delta = -31^\circ$ once every 3 d, increasing to once every 2 d with ZTF-II.

ZTF uses custom *g*-, *r*-, and *i*- band filters designed to avoid prominent sky lines at the Palomar site. These filters reach 30-s exposure 5σ limiting magnitudes of 20.8, 20.6, and 19.9 mag,

[10] https://fallingstar-data.com/forcedphot/
[11] https://asas-sn.osu.edu/

respectively. Each field of view is typically imaged twice, once in ZTF-*g* and once in ZTF-*r* (Bellm et al. 2019).

The ZTF alert distribution system produces over a million alerts each night, which feed into brokers that parse the data and make it publicly available. We access ZTF light curves through the Automatic Learning for the Rapid Classification of Events (ALeRCE) broker's Python client[12] (Förster et al. 2021).

2.1.5 *Triggering criteria*

When our half-hourly sync with TNS reveals a new target, we obtain light curves from ATLAS and ZTF, and if the target is brighter than 18 mag in any filter we also obtain an ASAS-SN light curve. We then attempt to fit the data to an SN Ia model using `SNooPy` (Contreras et al. 2010; Burns et al. 2011) and SALT3-NIR (Pierel et al. 2022) (our fitting procedure is discussed further in Section 3). We manually inspect the light curves and fits to address two points: is the candidate consistent with an SN Ia and is it possible to obtain observations at or before the NIR peak? If the candidate does not have spectroscopic classification, we assess the quality of successful fits. If the residuals indicate a poor fit to the data, or if the reduced χ^2 is greater than 2, we reject the candidate or defer judgment until more photometry becomes available. We estimate the time of peak brightness in the NIR using the best-fitting SALT3-NIR parameters. If the candidate is either classified as an SN Ia or is photometrically consistent with one, and if it has not yet reached its NIR peak, we pursue NIR observations as described in the following section.

2.2 Hawai'i Supernova Flows NIR photometry

2.2.1 *UKIRT – WFCAM*

For NIR observations, Hawai'i Supernova Flows uses the Wide Field Camera (WFCAM) mounted on the UKIRT 3.8 m telescope owned and operated by the University of Hawai'i[13] (Hodapp et al. 2018). UKIRT is a 3.8-m Cassegrain telescope on the summit of Maunakea. It has a declination limit of $-40^\circ < \delta < 60^\circ 07'$, granting access to about 3/4 of the sky. The Cambridge Astronomical Survey Unit (CASU) continues to provide data processing services and the Wide Field Astronomy Unit at the University of Edinburgh maintains the WFCAM Science Archive (Hambly et al. 2008) through which data are distributed.

WFCAM is a NIR imager developed specifically for large-scale surveys (Casali, M. et al. 2007). Its four detectors are Rockwell Hawaii-II (HgCdTe 2,048×2,048) arrays (Hodapp et al. 2004) each covering $13.65' \times 13.65'$ at a scale of about $0''.4$ per pixel. With its 0.9° diameter focal plane, WFCAM enabled the UKIRT Infrared Deep Sky Survey (Lawrence et al. 2007) and the UKIRT Hemisphere Survey (Dye et al. 2018). Hodgkin et al. (2009) explain that an astrometric distortion causes the pixel scale to vary radially, with per cent level differences in pixel area between the centre and edge of the focal plane. This changes the flux from the sky in each pixel, but their equation (1) provides a method for correcting this effect. We confirm this spatial variation and its resolution through the provided correction.

WFCAM uses a set of five broad-band filters, ZYJHK, and two narrow-band filters, H2 1–0 S1 and 1.644 Fe II. Each detector is equipped with its own set of filters, with inter-detector filter variations leading to photometric differences of no more than 0.01 mag (Hewett et al. 2006). The performance of WFCAM in the above filters was analysed in Hodgkin et al. (2009), who compared instrumental magnitudes against the Two Micron All Sky Survey (2MASS) Point Source Catalogue (Skrutskie et al. 2006). We use the *J*-band colour equation they derive to convert 2MASS *J* and *H* magnitudes to WFCAM *J* magnitudes, which we use to calculate zero-points for each image.

Hodgkin et al. (2009) also identified spatially correlated photometric variability, even when accounting for the astrometric distortion mentioned previously. The exact cause of the issue is unknown, but CASU provides an empirically derived table of corrections on a monthly basis. We address this spatial correlation independently by treating each image's zero-point as a second-order 2D polynomial centred on the SN candidate, inferred with the probabilistic programming language Stan (implemented through PyStan Riddell; Hartikainen & Carter 2021) for each image (Carpenter et al. 2017; Stan Development Team 2024). Stan provides a framework for specifying fully Bayesian statistical models and conditioning them on data using a no-U-turn sampler (NUTS; Hoffman & Gelman 2011; Betancourt 2013; Hoffman & Gelman 2014), an adaptive variant of Hamiltonian Monte Carlo sampling (HMC; Duane et al. 1987; MacKay 2003; Neal 2011). The scale of the effect is ∼0.021 mag from the centre to the edges of the image, comparable to the tables provided by CASU.

2.2.2 *Source characterization and galaxy subtraction*

The data distributed through the WFCAM Science Archive include catalogues of photometric parameters for sources extracted with the program `imcore`.[14] Initial testing highlighted issues in the catalogues when point sources coincided with extended sources. This compromised the photometry of most SNe Ia that were not exceptionally well separated from their host-galaxy.

Leveraging the multiplicity of our observations, we analysed each supernova and host-galaxy image series as an ensemble using the forward-model (or scene-model) code from Rubin et al. (2021). In short, this procedure assumes a series of images contains a time-independent 2D surface (modelled with splines) and a time-varying point-source. This allows for degeneracies when 'sharp' features in the galaxy (such as the nucleus) coincide with the SN Ia, but late-time observations of the galaxy taken after the SN Ia has faded resolve this issue by essentially providing a traditional reference image for subtraction. We manually determine which host-galaxies require late-time observations using diagnostic images such as those in Fig. 3. We pursue late-time observations if the galaxy model exhibits sharp features at the site of the SN, or if the residuals after subtracting either the galaxy or the galaxy and SN appear to have spatial structure.

We use the subsample of targets with late-time observations to validate our methodology against an independent data reduction process using traditional image subtraction performed with `ISIS` (Alard & Lupton 1998, 1999) and source characterization using `tphot` (Sonnett et al. 2013). The differences between the forward-modelled and image-subtracted photometry have a median of 0.008 mag and a standard deviation of 0.07 mag. We also examine how the forward-modelling code performs without late-time observations, and find the median difference remains low at 0.02 mag, but the standard deviation increases to 0.826 mag. This increase is driven

[12] https://alerce.readthedocs.io/en/latest/
[13] https://about.ifa.hawaii.edu/ukirt/
[14] http://casu.ast.cam.ac.uk/surveys-projects/software-release/imcore

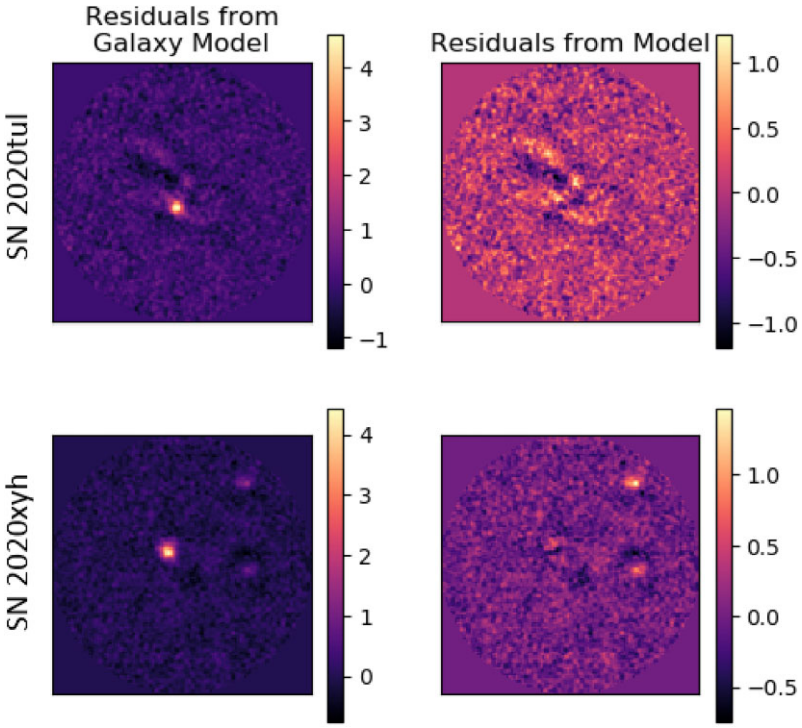

Figure 3. The forward-modelling code used to make photometric measurements produces diagnostic images showing the observed flux, the galaxy model, the residuals after subtracting the galaxy model ('Residuals from Galaxy Model'), and the residuals after subtracting both the galaxy and supernova models ('Residuals from Model'). We present a few examples representative of bad subtractions. SN 2020tul shows spatially correlated structure after galaxy subtraction, indicating the galaxy was not accurately modelled. Additionally, the supernova appears oversubtracted. This effect is more clearly seen in SN 2020xyh, which occurred near the nucleus of its host galaxy. The images on the left seem to show the galaxy has been subtracted, leaving only point sources at the location of the supernova and two nearby galaxies, but the images on the right show that SN 2020xyh appears to leave a small hole in some unmodelled structure.

by a few cases where the forward-modelling code struggled to separate the galaxy and the transient. Fig. 4 shows the average difference in a galaxy's forward-modelled photometry with and without late-time observations as a function of projected separation between the supernova and host-galaxy nucleus. The histogram shows that in the majority of cases, late-time observations do not result in significantly different photometry. In a few cases, the observations break degeneracies in the forward-modelling process, resulting in photometry up to a few magnitudes different. These cases are visually conspicuous, as seen in Fig. 3. In Appendix A, we fit a Gaussian mixture-model to the photometric differences (Δm) using Stan (Carpenter et al. 2017; Stan Development Team 2024) and find 74.0 per cent of the differences appear tightly dispersed ($\Delta m \sim \mathcal{N}(0.01 \pm 0.004 \text{ mag}, (0.08 \pm 0.005 \text{ mag})^2)$), and the remaining 26.0 per cent vary much more dramatically ($\Delta m \sim \mathcal{N}(0.33 \pm 0.050 \text{ mag}, (0.68 \pm 0.037 \text{ mag})^2)$). The fraction of targets reliant upon late-time observations for accurate photometry is vastly exaggerated in this analysis because the subsample comprises only targets manually determined to potentially benefit from late-time observations. Forward-modelled photometry is thus as accurate as traditional image subtraction, and more economical in that it often does not require a late-time observation.

2.3 Host galaxy redshifts

Although dozens of surveys have collectively measured redshifts for millions of galaxies, about half of the SNe Ia in our sample have host galaxies with no publicly available redshifts. Furthermore, the redshift measurements that are publicly available come from heterogeneous methodologies and at times are inconsistent with other measurements of the same galaxy. Here, we describe how we identify host-galaxies for each SN Ia, incorporate data from extant surveys, and obtain redshifts for galaxies that do not have publicly available spectroscopic redshifts.

2.3.1 Identifying host galaxies

All SN host galaxies in our survey were identified manually. This decision introduces an unquantified systematic error in our final peculiar velocity measurements due to the possibility of inaccurate host galaxy identification. Without a detailed simulation, it is unclear how often we misidentify host galaxies. However, the error rate is definitively lower than an algorithmic approach we tested, which produced obvious misidentifications. This alternative approach is detailed in Appendix B.

The SN Ia-galaxy associations produced manually were flagged if the host galaxy was ambiguous or otherwise problematic. These manual flags allow us to exclude these SNe Ia in our analyses, but introduce a hard-to-quantify bias (Gupta et al. 2016), and will not scale well if operations significantly expand. Recent work (e.g. Aggarwal et al. 2021; Qin et al. 2022) has formalized various methods of associating transient events with their host-galaxies using objective parameters, but still critically depends on the completeness and accuracy of galaxy catalogues. Automatic association will become necessary when our sample expands, but we will continue to associate SNe Ia and their host galaxies manually while it remains accurate and practical.

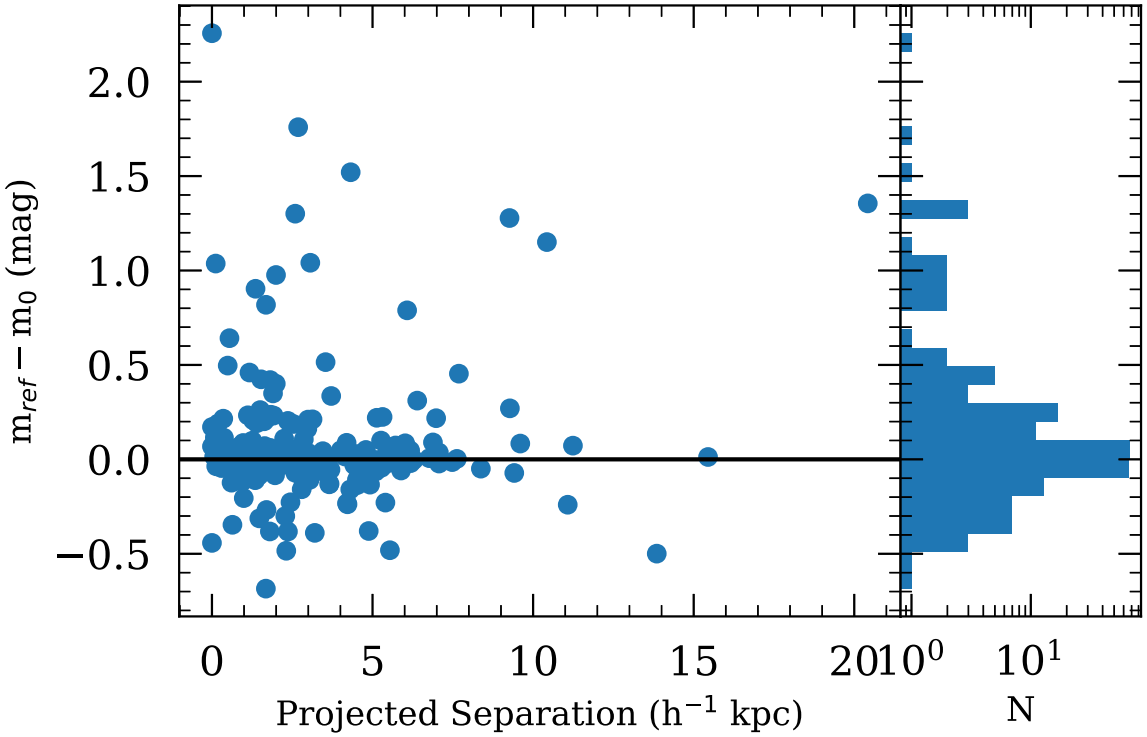

Figure 4. The differences between measurements made with and without late-time observations are minimal for a large number of targets, indicating accurate reconstruction of the galaxy surface profile. However, there are many targets where a late-time observation is crucial for decoupling the SN and host galaxy. The abundance of targets needing late-time observation is biased high in the plotted data because all targets were manually determined to potentially benefit from late-time observations.

2.3.2 Incorporating redshifts from literature

Before we pursue spectroscopic observations to find each host-galaxy's redshift, we search for existing measurements in the HyperLEDA data base (Paturel et al. 2003a, b; Makarov et al. 2014), which is based on the Lyon-Meudon Extragalactic Database (LEDA; Paturel et al. 1988) and Hypercat (Prugniel & Simien 1996). This significantly reduces our observational needs, but the variety of measurement techniques necessitates the careful handling of systematic differences. HyperLEDA uses a system of quality flags[15] to hierarchically combine optical and radio redshift measurements, and applies corrections on a reference by reference level to minimize systematic offsets between data sources (Paturel et al. 1997). If a host galaxy does not have a radial velocity in HyperLEDA, we pursue spectroscopic observations.

2.3.3 UH 2.2 m – SNIFS

The primary instrument we use for measuring host-galaxy redshifts is the Supernova Integral Field Spectrograph (SNIFS; Lantz et al. 2004) on the UH 2.2 m Telescope. SNIFS samples a $6'' \times 6''$ field with $0\farcs4 \times 0\farcs4$ spaxels, each of which produces two spectra, one blue (320–560 nm, R(430 nm) $\sim$ 1000) and one red (520–1000 nm, R(760 nm) $\sim$ 1300). Our exposure times are manually chosen based on galaxy surface brightness, atmospheric conditions, and galaxy spectral type, with late-type galaxies typically featuring emission lines and thus requiring less integration. The average exposure time was 1800 s. We use the data reduction pipeline described in Tucker et al. (2022) to produce 1D spectra. Absolute wavelength calibration is provided by arc-lamp exposures taken immediately after each science exposure. We include the average discrepancies between the arc spectra and their models when calculating redshift uncertainties, though the contribution is typically sub-dominant at $\sim$1 km s^{-1}. All galaxy spectra are converted to the heliocentric rest frame.

2.3.4 Subaru – FOCAS

When a galaxy is too faint for SNIFS, we use the 8.2 m Subaru telescope's Faint Object Camera and Spectrograph (FOCAS; Kashikawa et al. 2002) with its 300B grating with no filter (365–830 nm, R(550 nm) $\sim$ 700) and a $0\farcs6$ or $0\farcs8$ wide slit depending on the atmospheric conditions (Ebizuka et al. 2011). Subaru's mirror has over 13 times more light-gathering power than the UH 2.2 m mirror. This allows us to increase our limiting magnitude from $r < 19.1$ to $r < 22.9$ mag using comparable exposure times.

In addition to the increased light-gathering power, FOCAS's slit spectroscopy has proven necessary for very diffuse galaxies. Our reduction pipeline for SNIFS spectra struggles with sky subtraction if the entire $6'' \times 6''$ microlens array is filled. In such a case, we would need to obtain a sky observation for proper subtraction, doubling the exposure time required per object. For each galaxy, we perform a 900 s exposure and examine the summit-pipeline-reduced spectrum. If the galaxy has no strong emission lines, we pursue one or two additional 900 s exposures as deemed necessary by the observer. We perform bias subtraction and flat-fielding data using the routines described in the FOCAS Cookbook.[16] We use skylines for relative wavelength calibration, and use Subaru's location, the time of each

[15] http://leda.univ-lyon1.fr/a110/

[16] https://subarutelescope.org/Observing/DataReduction/Cookbooks/FOCAS_cookbook_2010jan05.pdf

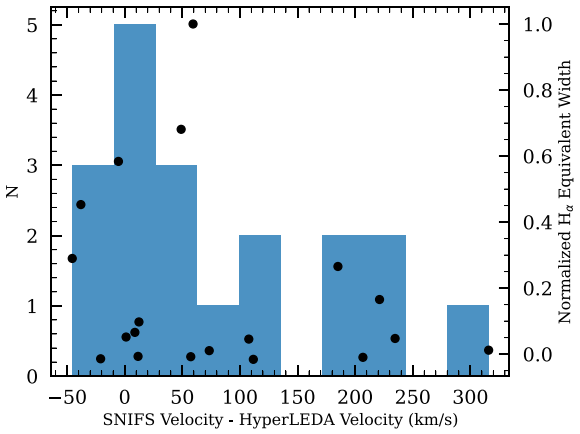

Figure 5. We compare the differences between our measured SNIFS velocities and HyperLEDA's aggregated velocities, finding two distinct groups. The 12 galaxies on the left side of the histogram have an average difference of 27 km s^{-1} and a standard deviation of 48 km s^{-1}. The four galaxies on the right are offset by several hundred km s^{-1}. Three of the four HyperLEDA velocities come from relatively older sources, and could be the result of inaccurate methodologies. Each galaxy in the histogram also has a marker with a *y*-value corresponding to its rescaled H_α equivalent width. The presence of weak equivalent widths in the sample with low velocity differences suggests that the four discrepancies are not due to weak spectral features.

exposure, and the position of each target to transform all spectra to a heliocentric rest frame.

2.3.5 *Redshift determination and uncertainties*

Once we have spectra from either SNIFS or FOCAS, we compare them with spectral templates from SDSS DR5[17] (Adelman-McCarthy et al. 2007) using the weighted cross-correlation routine in the SeeChange Tools[18] (Hayden et al. 2021). We tested the accuracy of this method by calculating redshifts for 158 galaxies using spectra from SDSS DR12, removing cross-correlations with an *r*-value less than 5 (as defined in Tonry & Davis 1979), and comparing our recession velocities with those in HyperLEDA. The differences averaged to ∼7 km s^{-1} with a standard deviation of ∼45 km s^{-1}. Thus, we include a 45 km s^{-1} uncertainty when inferring host-galaxy redshifts using this cross-correlation technique.

Additionally, we looked for systematic differences in absolute wavelength calibration between redshifts from literature and redshifts from our SNIFS and FOCAS spectra. We observed 24 galaxies with redshifts available in HyperLEDA using SNIFS, and 4 using FOCAS. Five of our SNIFS spectra had insufficient SNR and are not included in this analysis. The 19 remaining spectra yielded redshifts within about 100 km s^{-1} of their HyperLEDA values, with a few exceptions. We measure five galaxies to have redshifts several hundred km s^{-1} greater than their literature values. In descending order of discrepancy, these galaxies are PGC 40363, 4579, 29889, 13428, and 1033041, shown in the right side of Fig. 5. These galaxies include early and late-type morphologies, emission and absorption spectra, and their colours are not at the extremes of the 19 galaxy sample. The only unifying theme is that HyperLEDA sources the PGC 40363, 4579, and 13 428 from relatively older sources (Eastmond & Abell 1978; Sakai, Giovanelli & Wegner 1994; Thoraval, Boissé & Duvert 1999), whereas PGC 29 889 and 1 033 041 have more recent measurements, such as those from SDSS or 6dF. HyperLEDA aggregates and weights various sources, which should privilege more accurate observations, but these galaxies have only been spectroscopically observed once or twice before our observations with SNIFS. It is unclear why our measured redshifts are uniformly greater than their literature values. Disregarding these five exceptions, the average difference between the SNIFS-derived and HyperLEDA redshifts is ∼27 km s^{-1} with a standard deviation of ∼48 km s^{-1}. Including them, the average and standard deviation rise to ∼81 and ∼102 km s^{-1}, respectively. We subtract ∼27 km s^{-1} from our SNIFS-derived redshifts and interpret the ∼48 km s^{-1} standard deviation as a rough confirmation of the previously identified ∼45 km s^{-1} uncertainty. We also note that redshifts in HyperLEDA that have not been verified through repeated observations could benefit from additional measurements. Fig. 6 shows the distributions of heliocentric redshifts and *g*-band Kron magnitudes for the values taken from the literature and the values calculated with either SNIFS or FOCAS.

We note that galaxies in larger groups will have an additional velocity term due to intracluster dynamics, and that using the group redshift would likely probe large-scale flows more robustly, as done in Peterson et al. (2022). However, pursuing spectroscopic observations for all members of an associated group would reduce the number of SNe Ia host galaxies we could observe. We note that our analysis will benefit from future large spectroscopic surveys such as the Multi-Object Spectroscopy of Transient Hosts survey (MOST Hosts; Soumagnac et al. 2024) Dark Energy Spectroscopic Instrument (DESI; Collaboration 2022).

All redshift uncertainties are converted to uncertainties in distance modulus via the distance–redshift relation for an empty universe presented in Kessler et al. (2009b):

$$\sigma_\mu^z = \sigma_z \left(\frac{5}{\ln 10}\right) \frac{1+z}{z(1+z/2)}. \quad (3)$$

Different cosmological models produce negligible differences in σ_μ^z, which is already subdominant compared to other sources of uncertainty in the distance modulus.

3 DISTANCE DETERMINATION

In this section, we describe the specific methodology used to convert our data into distance moduli using `SNooPy` and `SALT3-NIR` as they were the only publicly available fitting programs that can utilize optical and NIR observations when our analyses began. We only intend to describe our fitting procedures to contextualize the results presented in Section 6, and as such we will not be claiming one program is more accurate or more appropriate for our use case. We leave such an analysis for future work, where we will also incorporate fits from BayeSN, which was made public with Mandel et al. (2022), and has been updated with Grayling et al. (2024).

3.1 SNooPy

`SNooPy` is a Python package designed for fitting light-curves of SNe Ia from the Carnegie Supernova Project (CSP; Contreras et al. 2010; Burns et al. 2011). It estimates luminosity distances by comparing data spanning flux, phase, and a shape parameter to filter-specific 3D models (Burns et al. 2011). These models were produced using high-cadence observations of SNe Ia in the CSP photometric system (Hamuy et al. 2006). We use version 2.6.0, which

[17] https://classic.sdss.org/dr5/algorithms/spectemplates/spectemplatesDR2.tar.gz
[18] https://zenodo.org/record/4064139#.YHkLvC1h2X0

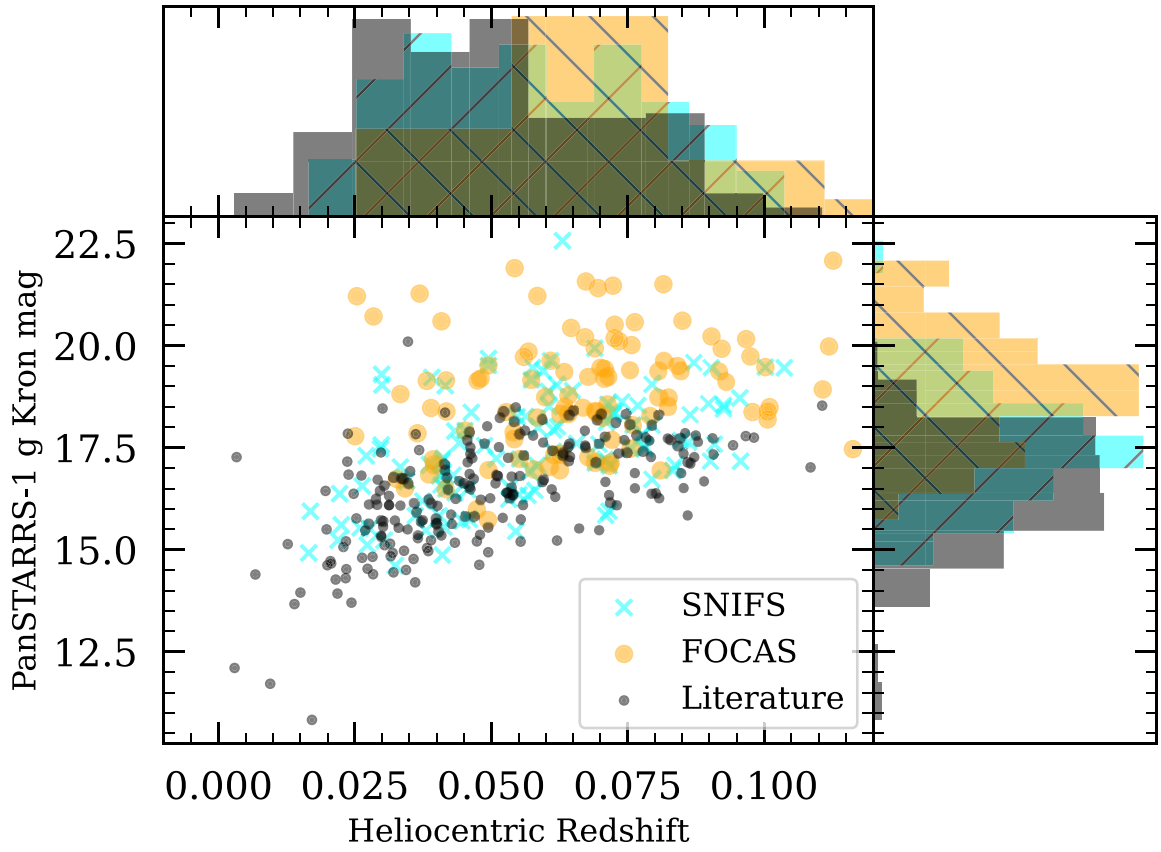

Figure 6. The normalized histograms of targets from SNIFS, FOCAS, and literature in z show distinct redshift distributions. Each galaxy's Pan-STARRS g Kron mag is plotted against redshift to show that the distribution in magnitudes are also distinct. Triggering only on transients in galaxies with redshifts in the literature biases the sample towards lower redshifts and brighter galaxies. By triggering on galaxies regardless of redshift availability, we mitigate this issue.

does not yet include the improved models of Lu et al. (2023). We look forward to reprocessing our sample when SNooPy incorporates these templates. Decreased systematic uncertainties in the NIR SED could increase the weight of J-band photometry in a multiband fit.

SNooPy is described by the CSP[19] as 'not a fixed algorithm for fitting light curves, but rather a collection of tools that are useful for building your own fitter (or fitting interactively)'. As such, there are a variety of non-trivial decisions that influence the distance moduli inferred using SNooPy. In version 2.6.0, there are three primary decisions:

(i) Parametrizing shape with Δm_{15} or s_{BV}.
(ii) Selecting one of the available models: EBV_model, EBV_model2, max_model, max_model2, Rv_model, color_model, SALT_model, and MLCS_model.
(iii) Selecting a 'calibration' to describe the correlation between SN parameters and absolute magnitude.
(iv) Selecting a reddening law.

We describe and qualitatively justify our choices here, but refer the reader to Appendix C for a quantitative analysis exploring alternative decisions.

3.1.1 *Choice of shape parameter*

SNooPy offers two distinct ways to characterize the shape of an SN Ia light curve; one being the decline rate parameter (Δm_{15}; Phillips et al. 1999), and the other being the colour-stretch parameter (s_{BV}; Burns et al. 2014). The latter is less sensitive to changes in reddening (varying ~ 1 per cent across $A_V = 3$ mag) and does not become degenerate for fast-declining SNe Ia ($s_{BV} < 0.7$), as seen with Δm_{15} (Burns et al. 2014). As such, we use s_{BV} when characterizing light curves with SNooPy.

3.1.2 *Choice of model*

The SNooPy models are described more comprehensively in the online documentation[20], but we summarize them here to provide context for our decision.

The EBV_model and EBV_model2 use light curves in numerous filters to infer four parameters of each SN Ia: the shape, the time of B-band maximum, the colour excess of the host galaxy ($E(B-V)_{\rm host}$), and distance modulus ($\mu_{\rm cos}$). The EBV_model is restricted to using Δm_{15} while the EBV_model2 can use that or s_{BV} to parametrize shape. They also differ in that the former model approximates the luminosity–shape correlation as a linear function using the six calibrations presented in Folatelli et al. (2010), whereas the latter

[19] https://csp.obs.carnegiescience.edu/data/snpy

[20] https://users.obs.carnegiescience.edu/cburns/SNooPyDocs/html/models.html

model uses a quadratic function calibrated with additional CSP data. In the EBV_model2, the cosmological distance modulus $\mu_{\cos}$ is related to the observed apparent magnitude in observer-frame bandpass OF at time t since B-band maximum ($m_{OF}(t)$) using a template in rest-frame bandpass RF with shape factor s_{BV} at de-redshifted time since B-band maximum t' ($T_{RF}(t', s_{BV})$) with the following equation:

$$\begin{aligned} m_{OF}(t) = {} & T_{RF}(t', s_{BV}) + P_0 + P_1(s_{BV} - 1) + P_2(s_{BV} - 1)^2 \\ & + \mu_{\cos} + R_{OF}E(B - V)_{\rm MW} + R_{RF}E(B - V)_{\rm host} \\ & + K_{OF,RF}(T_{RF}(t', s_{BV}), z, R_{OF}, R_{RF}), \end{aligned} \quad (4)$$

where P_0, P_1, and P_2 are polynomial coefficients defined by the calibration, R_{OF} and R_{RF} are the total-to-selective absorptions of the Galactic and host galaxy dust, and $K_{OF,RF}$ is the cross-band k-correction (described in Section 3.1.6). R_{OF} and R_{RF} are calculated by using an R_V-dependent reddening law to compute synthetic extinction values. We assume the Galactic average of $R_V = 3.1$ (Schlafly & Finkbeiner 2011) for calculating R_{OF}, and calculate R_{RF} with the calibration-provided value for host galaxy R_V. If parametrizing shape with Δm_{15}, the template term changes to $T_{RF}(t', \Delta m_{15})$ and the shape polynomial's $(s_{BV} - 1)$ terms change to $(\Delta m_{15} - 1.1)$.

The max_model and max_model2 also simultaneously fit light curves in multiple bandpasses, but relax the requirement that the photometry follows a well-characterized reddening law. Like the previous models, the two max_models fit for a global shape parameter and time of B-band maximum, but rather than fit for a distance modulus and host galaxy colour excess, these models fit for a peak apparent magnitude in each rest-frame bandpass (m_{RF}).

$$\begin{aligned} m_{OF}(t) = {} & T_{RF}(t', s_{BV}) + m_{RF} + R_{OF}E(B - V)_{\rm MW} \\ & + K_{OF,RF}(T_{RF}(t', s_{BV}), z) \end{aligned} \quad (5)$$

m_{RF} is not necessarily equal to $m_{RF}(t = 0)$ because the evolution of SN Ia SEDs produces peaks in different bandpasses at different times (e.g. Kasen, Thomas & Nugent 2006; Krisciunas et al. 2009; Phillips 2012; Burns et al. 2014). While distance moduli are not fitting parameters, they can be calculated based on each apparent maximum using a Tripp-like formula, such as the one presented in equation (4) of Burns et al. (2018):

$$m_X = P_0 + P_1(s_{BV} - 1) + P_2(s_{BV} - 1)^2 + \mu_{\cos} + \beta_{\max}(m_Y - m_Z), \quad (6)$$

where m_X, m_Y, and m_Z are the peak apparent magnitudes determined by the max_model fit in the bandpasses X, Y, and Z (these arbitrary labels are not to be confused with the Y or Z bandpasses). The difference between the max_model and max_model2 is that the latter allows for each bandpass to correspond to a unique time of B-band maximum light.

The Rv_model is similar to the EBV_model, in that it uses Δm_{15} and models the luminosity–shape correlation as a linear function. The primary difference is that the total-to-selective extinction of the host galaxy is a fitting parameter rather than a global constant taken from the calibration. Additionally, rather than using the Folatelli et al. (2010) values for calibrating luminosity, shape, and colour, this model uses values from Burns et al. (2011).

The color_model infers the shape parameter (only s_{BV}), the time of B-band maximum, and the peak apparent B magnitude, but also uses the difference between the observed colours and the intrinsic colours found in the 81 SNe Ia in Burns et al. (2014) to infer the host galaxy colour excess and R_V. Unfortunately, this model requires observations in the rest-frame B band, and as such we cannot evaluate this model. It is possible to use cross-band k-corrections to infer a rest-frame B-band light curve, but doing so would increase our vulnerability to differences between the real and modelled SED.

Lastly, the SALT_model and MLCS_model are wrappers for running the SALT2 and MLCS2k2 fitters in the `SNooPy` framework. Neither SALT2 (Guy et al. 2007) nor MLCS2k2 support NIR bandpasses (Jha et al. 2007), so we do not consider these models for Hawai'i Supernova Flows.

We decide to use EBV_model2 and the max_model because they support the use of s_{BV} and quadratic luminosity–shape correlations, both of which are favoured over their alternatives (Burns et al. 2014). Furthermore, recent work has made use of both the EBV_model2 (Jones et al. 2022; Phillips et al. 2022; Pierel et al. 2022; Peterson et al. 2023) and the max_model (Burns et al. 2018; Phillips et al. 2022; Lu et al. 2023; Uddin et al. 2023). We do not use the color_model because our observed bandpasses do not overlap with rest-frame B band in the majority of the redshift range we cover. We do not use the max_model2 because our J-band light curves are sparse and often times insufficient for estimating the time of B-band maximum alone.

3.1.3 Choice of calibration

The choice of 'calibration' refers to the values parametrizing the correlation between luminosity, shape, and colour (e.g. P_0, P_1, P_2, and $\beta_{\max}$ in equation 6). These values come from fits to samples of SNe Ia observed by CSP. While the method of fitting varies (χ^2 minimization in Folatelli et al. (2010) and Markov chain Monte Carlo (MCMC) methods in Burns et al. (2011, 2014, 2018)), differences in calibration values are primarily driven by variation in the samples used. As an example, the first calibration from Burns et al. (2018) was produced from 137 SNe Ia, but there are alternative calibrations, one excluding SNe Ia with s_{BV} values less than 0.5, one excluding those with m_B–m_V pseudo-colours greater than 0.5 mag, and one excluding those that meet either criteria. For our EBV_model2 fits, we use the calibration from Burns et al. (2018) based on the full sample since the reduced χ^2 values of the fits using the Burns et al. (2018) calibrations are typically lower than those using the Folatelli et al. (2010) calibrations (details in Appendix C) and because there are SNe Ia in our sample that have s_{BV} values less than 0.5 and m_B–m_V colours greater than 0.5 mag. For our max_model fits we use Stan (Carpenter et al. 2017; Stan Development Team 2024) to infer the nuisance parameters P_0, P_1, P_2, and β_{max} using our photometry. We omit the term correlating luminosity and host-galaxy mass to maintain consistency with EBV_model2, which does not factor in galaxy mass.

3.1.4 Choice of reddening law

The final decision point is the choice of reddening law. `SNooPy`'s default reddening law (O94; O'Donnell 1994) is a corrected version of the CCM89 reddening law (Cardelli et al. 1989). It also natively supports the original, uncorrected version, as well as the reddening laws F99 (Fitzpatrick 1999) and FM07 Fitzpatrick & Massa (2007). We have performed minor modifications to the `SNooPy` source code[21] to accommodate the reddening laws provided in the `dust_extinction` package (Gordon 2024). After comparing fits produced with the O94, F99, and F19 (also referred to as F20; Fitzpatrick et al. 2019) reddening laws (details in Appendix C) we find that the reduced χ^2 values are typically lowest when using

[21] Modified version available at https://github.com/ado8/snpy

the F19 reddening law. Thus we use the F19 reddening law for all SNooPy fits, which applies to both host galaxy and Galactic extinction in the EBV_model2, but only Galactic extinction in the max_model. We assume the total-to-selective extinction parameter for Galactic dust is $R_V = 3.1$ and use the R_V value defined in the Burns et al. (2018) calibration for host galaxy dust. Galactic colour excess values comes from the SFD dust map (Schlegel, Finkbeiner & Davis 1998) with the 0.86 scaling factor described in Schlafly & Finkbeiner (2011) (hereafter, the rescaled SFD dust map).

3.1.5 *Estimating uncertainties*

SNooPy provides estimates of statistical uncertainty in all inferred parameters following either frequentist or Bayesian conventions. Initial fits without priors produce statistical errors using the standard frequentist convention of inverting the Hessian matrix at the best-fitting parameters to produce a covariance matrix.[22] When this matrix is singular, as can happen with undersampled light curves or for light curves of non-SNe Ia, the model becomes insensitive to one or more parameters and will not infer values for any of them. After the initial fit, SNooPy offers an MCMC method which samples their posterior distributions with the package emcee (Foreman-Mackey et al. 2013). The default priors are based on previous work with the CSP sample, but can be overwritten with arbitrary functions.

In addition to providing statistical errors, SNooPy provides an uncertainty floor for each parameter. The floor in the distance modulus reflects the uncertainty in the various terms used to standardize SNe Ia luminosities. These terms depend on the model used, but generally include filter-specific measurements of peak absolute magnitude and how that changes with s_{BV}. Thus, the distance modulus accuracy has a systematic floor determined by the sample used to calibrate it and becomes less accurate as the shape factor deviates from its normal value. The other floors have constant values derived from various analyses. The uncertainty floor in s_{BV} is 0.03, and comes from the dispersion around a quadratic fit of s_{BV} to the SALT x_1 parameter (discussed in Section 3.2) (Burns et al. 2014). The host galaxy colour excess floor is 0.06 mag, coming from the intrinsic dispersion of the m_B–m_V colours in the CSP sample after correcting for reddening. In the max_model, the peak magnitudes in each bandpass are presented with uncertainty floors based on Folatelli et al. (2010). Lastly, the time of B-band maximum is fixed to have an uncertainty floor of 0.34 d. We define the uncertainty on each parameter estimate as the quadrature sum of the statistical uncertainty and the floor.

3.1.6 *K- and S-corrections*

Observations of SNe Ia at significant redshift can lead to a mismatch between the observed and rest-frame spectral energy distribution (SEDs). One could almost trivially account for this issue in spectral observations if the redshift is known (telluric corrections aside), but photometric observations require some knowledge of the underlying SED to determine what is shifted into and out of the effective bandpass. The adjustments needed to compensate for the mismatches between observed and emitted SEDs are called 'K-corrections' (Humason, Mayall & Sandage 1956; Oke & Sandage 1968).

[22] https://users.obs.carnegiescience.edu/cburns/SNooPyDocs/html/fitting_LM.html

Similarly, variations in an optical system's transmission function leads to differences in instrumental magnitudes that depend on the SED observed. SNooPy models are defined in the CSP photometric system, and using data from other bandpasses would introduce systematic errors in the parameter inferences. The typical treatment for managing multiple filter sets is to observe a range of standard stars and perform linear fits of colour terms to transform one set to the other. Using stellar standards produces equations capable of converting stellar observations between filter sets, but SNe Ia have non-stellar SEDs, and there are no perennially available standard SNe Ia. The solution is to apply an 'S-correction' (Burns et al. 2011).

SNooPy applies both of these corrections simultaneously by calculating a 'cross-band K-correction' (Kim, Goobar & Perlmutter 1996) using the spectral library from Hsiao et al. (2007), which combines ~ 600 heterogeneous spectra of ~ 100 SNe Ia. Although the library covers a wide breadth, the available spectra cannot represent every kind of SN Ia at every possible epoch. To account for levels of reddening and intrinsic colours not seen in the spectral library, Hsiao et al. (2007) describe a 'mangling' process by which template spectra can be multiplied by a smoothly varying spline to match observed colours. The statistical error on each K-correction and mangling varies between about 0.01 and 0.04 mag depending the amount of overlap between the redshifted rest-frame CSP bandpass and the observed bandpass. Pairs with little overlap rely on extrapolation, and are more sensitive to the spectral template used (Hsiao et al. 2007), whereas a rest-frame bandpass that maps exactly on to an observed bandpass would be completely insensitive to the underlying spectrum. The ATLAS c and o bandpasses are wider than those in the CSP photometric system, and so they necessarily belong to the former category.

3.2 SALT

SALT fits SNe Ia light curves using a different approach (Guy et al. 2005, 2007, 2010). Roughly speaking, where SNooPy attempts to fit observed light curves to well studied light curves, SALT attempts to fit observed light curves to a spectral time series. This model is built from a term that describes the phase-independent effect of the colour law ($CL(\lambda)$) and two or more surfaces spanning flux, phase (p), and wavelength (λ), whose combinations describe the spectral flux and evolution of all SNe Ia:

$$F(p,\lambda) = x_0[M_0(p,\lambda) + x_1 M_1(p,\lambda) + \ldots] \times \exp\left[c \times CL(\lambda)\right], \tag{7}$$

where M_i is the ith surface, x_i scales how much that surface contributes to the spectral flux, and c scales the colour law (Guy et al. 2007). The surfaces are empirically derived, with M_0 encapsulating the 'standard' SN Ia spectral time series while the remaining surfaces describe all other modes of variation. This means the surfaces themselves may not correlate exactly with the physical parameters of SNe Ia, but instead may be understood as principal components. With that said, x_1 is often considered a shape factor like s_{BV} or Δm_{15} since light-curve shape seems to be the dominant mode of variation. Each combination of x terms defines an SED and evolution that can be further sculpted by c, the colour law, and redshift. At any observational epoch, a filter set's transmission function is used to make synthetic magnitudes, which can be compared to real photometry. Thus one can infer the most likely SALT parameters and their uncertainties given observations of a particular SN Ia. These parameters provide a distance modulus (μ) by the equation

$$\mu = m_B^* - M + \alpha x_1 - \beta c, \tag{8}$$

where m_B^* is the rest-frame Bessell B-band magnitude (Perlmutter et al. 1997), M is the absolute magnitude of an SN Ia with $x_1 = c = 0$, and α and β are standardization coefficients. While m_B^* can be approximated by $-2.5\log_{10}(x_0) + \text{const.}$, we calculate its value using synthetic photometry based on model parameters.

Rubin (2020) suggested that SNe Ia luminosity variability may consist of three to five independent parameters. Attempts to standardize SNe Ia luminosities using one or two parameters report an 'intrinsic scatter' that cannot be explained by measurement error alone (e.g. Scolnic et al. 2018; Brout et al. 2022). Rose et al. (2020) explored the differences between two and seven-component fits using SNEMO (Saunders et al. 2018), and found that only CSP data had the SNR and coverage to constrain the additional parameters. Put another way, a two-component fit with SALT compares to a seven-component fit with SNEMO for all but the most extensively covered light curves. With that in mind, we use the two-component fits of SALT3-NIR (Pierel et al. 2022). The only other SALT model that can process NIR light curves is SALT2-Extended, but it was trained on optical data extrapolated to the NIR and is thus insensitive to correlations between SALT parameters and NIR light-curve properties (Pierel et al. 2018). SALT3-NIR was jointly trained on the optical sample of 1083 SNe Ia from Kenworthy et al. (2021) and 166 SNe Ia with NIR data (Pierel et al. 2022). We access the SALT3-NIR model through the Python package `SNCosmo` version 2.10.4 (Barbary et al. 2022), and utilize the convenience functions therein to account for Galactic extinction using the rescaled SFD dust map and the reddening law from Fitzpatrick et al. (2019) with $R_V = 3.1$. Notably, we use `SNCosmo` to calculate model fluxes given a set of SN Ia parameters, but do not use the built-in functions to estimate those parameters. Instead, we use the fitting methodology of Rubin et al. (2023), defining a χ^2 function and using a downhill-simplex algorithm to iteratively identify the SALT parameters that minimize that function.

3.2.1 Estimating uncertainties

The covariance matrices we obtain for each object's best-fitting SALT parameters (time of B-band maximum light, x_0, x_1, and c) reflect three sources of uncertainty. Our NIR photometric methods produce correlation matrices, but we assume the measurements and errors from ATLAS, ASAS-SN, and ZTF are completely independent. We incorporate the SALT3-NIR model uncertainties during our fitting process. Lastly, we repeat each fit with slightly varied inputs to calculate derivatives between the fitting parameters and quantities like redshift, Galactic colour excess, and the photometric zero-point in each bandpass.

The error explicitly associated with K-corrections and S-corrections is ostensibly removed due to SALT's use of spectra when fitting. However, if the intrinsic SED of an SN Ia differs from the form of equation (7) truncated after $i = 1$, the synthetic photometry will be inaccurate. We assume these errors are encapsulated in the model uncertainties.

The distance modulus in equation (8) requires specifying the standardization coefficients α and β, which are typically calibrated empirically. Fitting for α and β by minimizing dispersion in the Hubble residuals introduces a form of Eddington bias due to uncertainties in x_1 and c. We estimate the standardization coefficients using a Bayesian framework called UNITY (Unified Nonlinear Inference for Type-Ia cosmologY; Rubin et al. 2015, 2023). UNITY assumes a Gaussian and skew normal distribution for the population distributions of the true value of each SN's x_1 and c, respectively, and uses flat hyperpriors for the means of each distribution and the log of their standard deviations. This approach avoids Eddington bias, which would suppress both coefficients. Although UNITY can model α and β as broken-linear functions, we assume the coefficients are constants. In Section 6.2, we identify and discuss a systematic issue tied to this decision.

4 VALIDATING DATA AND METHODOLOGY

In this section, we validate our data reduction and modelling techniques by partially reproducing the analysis of the DEHVILS survey (Peterson et al. 2023) using our NIR photometry and fitting methodologies. To evaluate the differences produced by these variations, we compare each inferred distance modulus ($\mu_{\rm fit}$) and the theoretical distance modulus at its corresponding redshift in a fiducial cosmology ($\mu_{\rm cos}$). These Hubble residuals are calculated as

$$\Delta\mu = \mu_{\rm fit} - \mu_{\rm cos} \tag{9}$$

$$\mu_{\rm cos} = 5\log_{10}\left[\left(\frac{cz_{\rm CMB}}{H_0}\right)\left(\frac{1+z_{\rm hel}}{1+z_{\rm CMB}}\right)\left(1+\frac{1-q_0}{2}z_{\rm CMB}\right)\right] + 25, \tag{10}$$

where H_0 is the Hubble constant and q_0 is the cosmic deceleration parameter, which we take as -0.53 (Planck Collaboration et al. 2020). As stated in Burns et al. (2018), the factor of $(1+z_{\rm hel})/(1+z_{\rm CMB})$ accounts for observational effects which should be corrected in a heliocentric rest frame. In each sample we define H_0 such that the inverse-variance weighted average of the Hubble residuals is 0 mag.

The dispersion in $\Delta\mu$ is typically characterized through RMS (e.g. Blondin, Mandel & Kirshner 2011; Foley et al. 2017; Avelino et al. 2019; Jones et al. 2022; Pierel et al. 2022; Peterson et al. 2023); inverse-variance weighted RMS (WRMS; e.g. Blondin et al. 2011; Foley et al. 2017; Avelino et al. 2019), or normalized median absolute deviation (NMAD; e.g. Boone et al. 2021; Peterson et al. 2023). SNe Ia analyses repeatedly find that measurement uncertainty alone cannot explain the observed dispersion, indicating that SNe Ia luminosities include some unmodelled variance commonly called intrinsic scatter ($\sigma_{\rm int}$; e.g. Blondin et al. 2011; Burns et al. 2018; Scolnic et al. 2018).

Lastly, we validate our treatment of max_model parameters by using photometry from CSP-I DR3 (Krisciunas et al. 2017) to re-derive the Tripp calibration parameters in table 1 of Burns et al. (2018).

4.1 Comparisons with DEHVILS

The DEHVILS survey collected data in tandem with Hawai'i Supernova Flows, also using UKIRT's WFCAM to collect NIR observations of SNe Ia (Peterson et al. 2023). Our programs differ in that DEHVILS collected photometry in the Y, J, and H bands and pursued more observations (median 6 epochs per bandpass) for fewer SNe ($N = 96$). We shared J-band observations near peak to avoid redundancy, but reduced the data through independent photometric pipelines. The DEHVILS analysis employs the following quality cuts: $|x_1| < 3$, $\sigma_{x_1} < 1$, $\sigma_{t_0} < 2$, $E(B-V)_{\rm MW} < 0.2$ mag, and Type Ia LC fit probability $P_{\rm fit} > 0.01$. σ_{x_1} and σ_{t_0} refer to the uncertainty in the SALT parameter x_1 and the estimated time of maximum light, while $P_{\rm fit}$ is defined in `SNANA` as the fraction of the χ^2 distribution with k degrees of freedom above a given χ^2 value (Kessler et al. 2009a):

$$P_{\rm fit}(k, \chi^2) = \frac{1}{\Gamma(k/2)}\int_{\chi^2/2}^{\infty} t^{k/2-1}e^{-t}dt. \tag{11}$$

Further, the target's host galaxy must have a spectroscopic redshift. There are differences between the redshifts in the DEHVILS sample and the redshifts assembled following the methods described in Section 2.3. In this section, we use the published DEHVILS redshifts for a fairer comparison. The sample analysed in Peterson et al. (2023) comprises 47 of the 83 spectroscopically classified normal SNe Ia with DEHVILS photometry. Using fitting parameters to define cuts means differences in fitting methods may lead to differences in the objects cut. When we apply the same cuts using our implementation of the SALT3-NIR model we find that 56 of the optical-only (ATLAS *co*) fits pass all cuts, 47 fits using optical (ATLAS *co*) and DEHVILS-reduced NIR[23] (*YJH*) photometry pass, and 30 DEHVILS-reduced NIR-only (*YJH*) fits pass. Refitting the 83 SNe Ia assuming the reddening law from Fitzpatrick (1999) for Milky Way extinction does not lead to any difference in the objects cut.

4.1.1 *Varying sample selection and fitting methodology*

The 47 SNe Ia analysed in Peterson et al. (2023) do not exactly match the 47 that pass the same cuts in our analysis, indicating a difference between our methodologies. Identifying the exact point of departure is of intrinsic interest, but more immediately concerning are the consequences of such a difference. Our goal in this section is to compare the dispersion of Hubble residuals found in Peterson et al. (2023) to our values calculated with the same photometry but different methods. We assume there are negligible differences in our ATLAS photometry and that there are no unstated quality cuts in the DEHVILS analysis.

We use DEHVILS photometric measurements for all NIR data and fit each of the normal SNe Ia in host galaxies with spectroscopic redshifts using SNooPy's EBV_model2, SNooPy's max_model, and SALT3-NIR using the bandpass combinations *co*, *coYJH*, and *YJH* for all three fitters. The DEHVILS cuts are based on the SALT fitting parameter x_1, which we approximate in the SNooPy fits as $0.65 < s_{BV} < 1.40$ using a conversion we empirically determine in equation (14) of Section 5.3.1. We also use this equation to convert the $\sigma_{x_1} < 1$ cut to $\sigma_{s_{BV}} < 0.125$.

The 83 normal SNe Ia are defined as such based on their spectroscopic classification as SN Ia on TNS, but after visually inspecting the light curves we believe there are nine non-normal SNe Ia in this sample. We find four underluminous candidates (SNe 2020jsa, 2020rlj, 2020unl, and 2021mim), four SN 2006bt-like candidates (SNe 2020naj, 2020sme, 2020mbf, 2020tkp), and one Ia-pec candidate (SN 2020kzn). These targets are eliminated by the $P_{\rm fit}$ cut in the DEHVILS analysis and are excluded from the analyses in this section based on their suspected classification. These targets are not explicitly excluded in the main sample selection process described in Section 5, but all are removed by either quality cuts or the outlier detection methods described in Section 5.3 except for SN 2021mim.

The SNooPy-based fits produce higher χ^2 values than the SALT3-NIR fits for targets that were successfully fit and passed quality cuts (excluding the $P_{\rm fit}$ cut). Andrae, Schulze-Hartung & Melchior (2010) review the inherent problems with using χ^2 values and degrees of freedom to assess model performance (especially non-linear models), but the $P_{\rm fit}$ cut is a function of those parameters. The median ratio between χ^2/DoF values from EBV_model2 fits and SALT3-NIR fits using the *coYJH* bandpasses is about 2.22. However, when excluding model variance in both sets of χ^2 calculations, the median ratio is 0.88. This reversal when excluding model variance applies to fits using other bandpass combinations, with the ratio falling from 1.37 to 0.94 in the optical only fits, and from 2.31 to 0.8 in the NIR only fits. This indicates the uncertainties in the SALT3-NIR model may be overestimated, at least compared to the SNooPy model uncertainties which may themselves be underestimated. This finding is similar to that of Taylor et al. (2023) who compared SALT2 and SALT3 models trained on identical data and found that the SALT3 model showed multiple indicators of overestimated model-plus-data uncertainties. However, this is inconsistent with Peterson et al. (2023) who performed a visual inspection of their fits and suspected the model uncertainties may be underestimated, leading to significant cuts due to fit probability. Regardless of the reason for the different χ^2 values found by different models, the application of a $P_{\rm fit} > 0.01$ cut will produce imbalanced sample sizes cut at different quantiles of $P_{\rm fit}$, which will skew the comparison of Hubble residual dispersions. Instead, we define model-specific samples based on the 47 highest $P_{\rm fit}$ values from each model's fits to the *coYJH* bandpasses. This is consistent with the DEHVILS analysis, which analysed the same 47 SNe Ia when fit with optical-only, optical and NIR, or NIR-only photometry. The χ^2/DoF cut values are 1.31 for SALT3-NIR fits, 4.14 for EBV_model2 fits, and 4.51 for max_model fits.

With the differences in sample selection defined, we now describe the differences in fitting methodology. By equation (6), calculating distance moduli using the max_model requires specifying a bandpass (m_X) and a colour (m_Y–m_Z), which makes comparisons between max_model fits subject to systematic discrepancies when the bandpasses differ. There is no bandpass and colour common to the bandpass combinations we examine, but we may still compare each implementation of the max_model against the DEHVILS results. For the *co* combination, we use the V bandpass and the $V - r$ colour; for *coYJH*, we use J and $V - r$; and for *YJH*, we use J and $Y - J$. We calculate SALT-based distance moduli using α and β parameters derived with UNITY (Rubin et al. 2015), except for the *YJH* sample which encountered numerous problems during modelling and produced an anomalously low and noisy $\beta = 0.14 \pm 1.80$. For this sample, we calculate the α and β values that minimize the standard deviation of the Hubble residuals. The standardization coefficients for the *co*, *coYJH*, and *YJH* samples are (α, β) = (0.155, 3.3), (0.138, 3.702), and (0.111, 2.475), respectively. For comparison, Peterson et al. (2023) used standardization coefficients of (α, β) = (0.145, 2.359) and (0.075, 2.903) for the *co* and *coYJH* samples, with no standardization applied to the *YJH* sample. They characterize the dispersion in Hubble residuals using NMAD and standard deviation (STD), so we use the same statistics in this section.

Our methods noticeably differ in fitting one of the bandpass combinations. In the DEHVILS analysis, the fit parameters x_1 and c were held fixed at 0 for the NIR-only sample. Our methodology does not hold these parameters fixed, and we found greater dispersion. This is consistent with their finding that keeping c constant while allowing x_1 to vary led to increased scatter. For the other *co* bandpass combination, we found dispersions in Hubble residuals roughly consistent with the DEHVILS values and errors presented in Peterson et al. (2023) and reproduced in Table 2. Our NMAD values were lower and our STD values higher, implying our Hubble residuals are heavier-tailed than a Gaussian distribution. This could be an effect of different sample selection, different treatment of ATLAS photometry, or different standardization coefficients. For the *coYJH* bandpass combination, our analysis with SNooPy's EBV_model2 is consistent with the DEHVILS values, but the other two models tend to produce higher dispersion values. We note that in our SALT3-NIR analysis, if we use the α and β values that minimize the standard deviation

[23] Available at https://github.com/erikpeterson23/DEHVILSDR1

Table 2. We highlight any differences due to methodology by using DEHVILS photometry and approximating their quality cuts. We did not replicate their findings when using only YJH photometry, which is where our methodologies differ the most. The DEHVILS team fixed x_1 and c to 0 for those fits and we allowed them to vary. For the *co* bandpass combination, our methodology produced Hubble residual dispersions consistent with the values reported by DEHVILS. We find mixed results with the $coYJH$ combination, with our max_model analysis producing larger dispersions, and our SALT3-NIR analysis producing consistent results only if we solve for the α and β values that minimize dispersion in the Hubble residuals.

Model	Filters	N	NMAD (mag)	STD (mag)
DEHVILS	*co*	47	0.177(029)	0.221(043)
DEHVILS	*coYJH*	47	0.132(025)	0.175(034)
DEHVILS	*YJH*	47	0.139(026)	0.172(027)
EBV_model2	*co*	55	0.177(041)	0.327(065)
EBV_model2	*coYJH*	47	0.126(023)	0.131(014)
EBV_model2	*YJH*	50	0.152(025)	0.165(022)
max_model	*co*	51	0.215(039)	0.234(032)
max_model	*coYJH*	47	0.159(027)	0.153(015)
max_model	*YJH*	47	0.181(034)	0.182(023)
SALT3-NIR	*co*	56	0.225(043)	0.246(025)
SALT3-NIR	*coYJH*	47	0.184(030)	0.186(023)
SALT3-NIR	*YJH*	30	0.164(036)	0.161(020)

(0.100 and 3.052, respectively), we find a value of 0.162 mag and a NMAD of 0.124 mag, which is consistent with the DEHVILS values. Our max_model analysis is also not optimized against dispersion. We use the J-band peak magnitude and $V-r$ pseudo-colour to infer distances because that is the methodology we apply to our own photometry, which does not include Y- or H-band observations.

The consistency between the dispersion values we measure and the values reported in Peterson et al. (2023) suggests that our methodology is comparable for fits when using optical data or optical and NIR data. Our methodology is inferior for fits using only NIR photometry, and max_model fits using $coYJH$ photometry, indicating that we would need to adapt our methodology if we were to collect Y- and H-band data like the DEHVILS team and produce NIR-only samples. The samples we produce using our own J-band data always include optical data.

Our samples are distinct from the one analysed in Peterson et al. (2023). However, the effects of a few mismatched SNe should be suppressed after bootstrap resampling the Hubble residuals. As in the DEHVILS analysis, for each sample we perform 5000 iterations of randomly choosing 47 residuals with replacement. The dispersion values and uncertainties presented in Table 2 are the averages and standard deviations of the values measured across the 5000 iterations (shown in Figs 7 and 8).

4.1.2 *Varying photometry*

We repeat the comparative analysis of the previous section, this time isolating the effects of differing photometry. We fit ATLAS and either our J-band data or that of the DEHVILS survey to create two sets of fits for each of our three models. We apply the model-specific χ^2/DoF cuts based on the greater value between the fits using our photometry or that of DEHVILS.

Once more, we bootstrap resample the Hubble residuals to estimate the uncertainties in our dispersion measurements, but we include an additional set of statistics. When varying methodology, we could only compare the distributions of our resampled dispersion measurements with the values reported in Peterson et al. (2023), but in this analysis we can make pairwise comparisons between individual iterations of the resampling process. For each iteration, we randomly choose SNe Ia with replacement, record the NMAD and STD of their Hubble residuals in our six samples, and additionally calculate the differences in dispersion between each model's sample using our J-band photometry and using DEHVILS photometry ($\Delta D = D_{\rm HSF} - D_{\rm DEHVILS}$ where D is either NMAD or STD). Thus, we not only produce distributions of NMAD and STD, but also distributions of ΔNMAD and ΔSTD.

Table 3. We use our methodology to calculate Hubble residuals using ATLAS photometry and either our J-band photometry or that of the DEHVILS survey. Bootstrap resampling these residuals 5000 times shows the dispersion measurements are insensitive to any differences between our photometry. In SNooPy's EBV_model2 and max_model and in SALT3-NIR, the change in dispersion is consistent with 0.

Model	J data	N	NMAD (mag)	STD (mag)
EBV_model2	HSF	48	0.109(022)	0.143(020)
EBV_model2	DEHVILS	48	0.142(024)	0.152(017)
max_model	HSF	50	0.144(032)	0.196(029)
max_model	DEHVILS	50	0.165(029)	0.161(015)
SALT3-NIR	HSF	49	0.180(034)	0.197(022)
SALT3-NIR	DEHVILS	49	0.218(044)	0.227(021)

The averages and standard deviations of these values are presented in Table 3 and the histograms of dispersions and differences are plotted in Figs 9 and 10. None of the distributions indicate that using our photometry instead of DEHVILS photometry leads to increased dispersion measurements. The averages are within one standard deviation of each other, and the differences within one standard deviation of no change in dispersion.

4.2 Comparison with CSP

The EBV_model2 produces Hubble residuals with lower dispersion than those produced by either SALT3-NIR or the max_model. The greater dispersion in the max_model was unexpected since the EBV_model2 is calibrated to CSP observations of 36 SNe, whereas in this analysis we derived standardization coefficients for the max_model using our observations of 47 SNe.

4.2.1 *Validating Tripp calibration*

To test our derivation process, we used photometry from CSP-I DR3 (Krisciunas et al. 2017) to solve for the calibration coefficients presented in table 1 of Burns et al. (2018). We fit all CSP photometry with the SNooPy max_model, parametrizing light-curve shape with s_{BV}. We use the heliocentric redshifts provided in the data release rather than redshifts from HyperLEDA to focus on differences due to methodology. Our equation (6) does not include a term for host-galaxy mass, but in order to match the CSP derivation methodology we reintroduce this term:

$$m_X = P_0 + P_1(s_{BV}-1) + P_2(s_{BV}-1)^2 + \mu_{\rm cos} + \beta_{\rm max}(m_Y - m_Z) + \alpha_{\rm M}\left(\log\left(M_*/M_\odot\right) - M_0\right), \quad (12)$$

where $\alpha_{\rm M}$ is the coefficient correlating magnitude and host-galaxy stellar mass (M_*) and M_0 is an arbitrary mass zero point, taken as $10^{11}M_\odot$. We follow the methodology in appendix B of Burns et al. (2018) for assembling host-galaxy stellar masses, primarily drawing from the 2MASS Extended Source Catalogue (Jarrett et al. 2000),

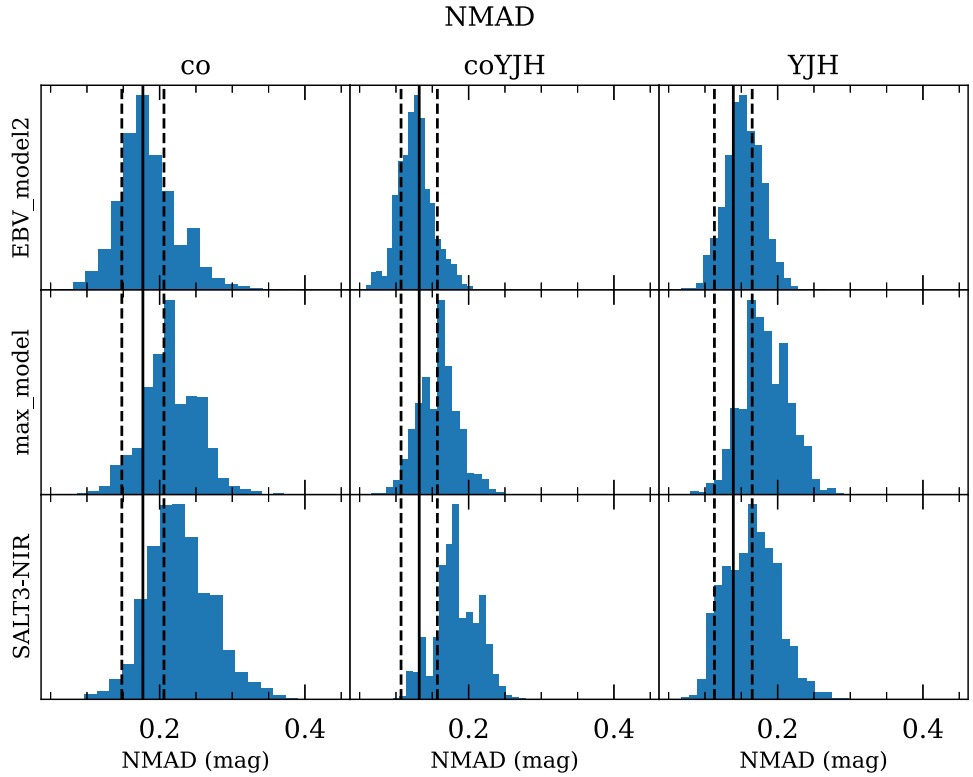

Figure 7. We show the distributions of NMAD in 5000 bootstrap resamplings of each set of Hubble residuals. The columns correspond to the fitting model used the rows to the bandpass combination. The solid and dashed vertical lines show the values and uncertainties of the NMAD reported in the DEHVILS survey. Our methodology produces Hubble residuals with dispersions consistent with the values reported by the DEHVILS survey using `SNooPy`'s EBV_model2, but not when using SALT3-NIR or `SNooPy`'s max_model.

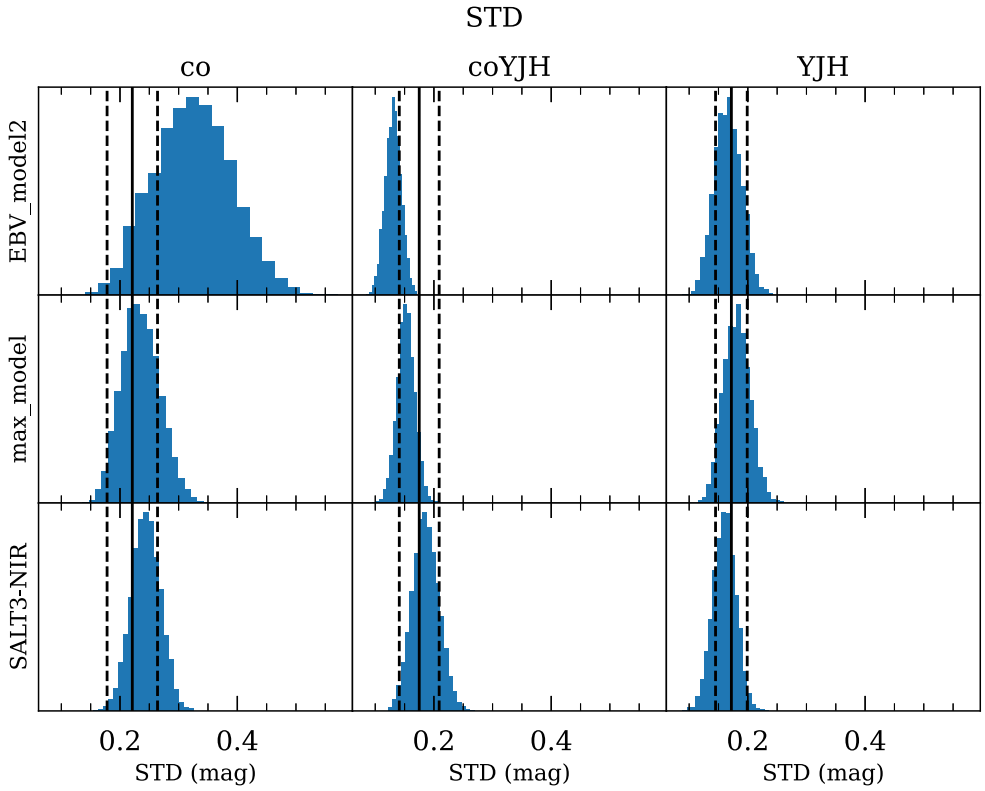

Figure 8. The same set of plots as Fig. 7, but showing standard deviation instead of NMAD.

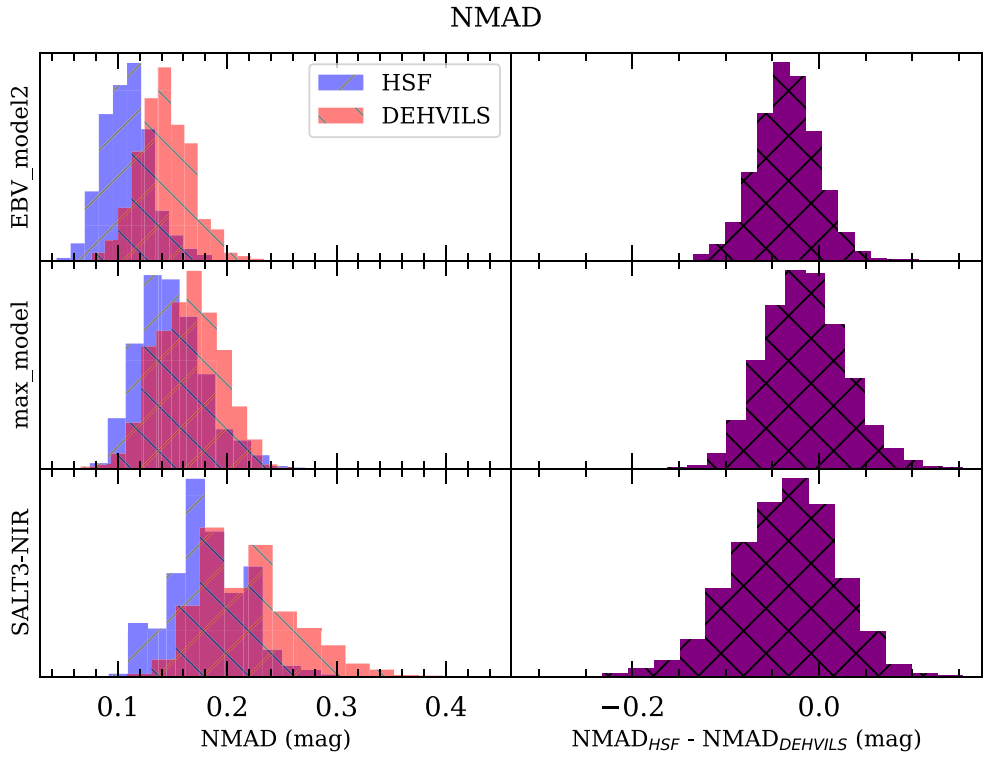

Figure 9. The histograms in the left panels show the distributions of NMAD dispersion measurements after 5000 iterations of bootstrap resampling Hubble residuals. The histograms in the right panels show the distributions of differences in dispersion between the samples using HSF photometry and the samples using DEHVILS photometry in each iteration. Using our measurements instead of DEHVILS photometry may lead to a decrease in the dispersion of the Hubble residuals, but it is not statistically significant.

which we convert from K-band apparent magnitudes to stellar masses assuming a constant mass-to-light ratio.

$$\log_{10}(M_*/M_\odot) = -0.4\,(m_K + \mu) + C, \tag{13}$$

where μ is the distance modulus and C is a constant which CSP determined to be 1.04 dex by comparing masses from the 2MASS catalogue with mass estimates from Neill et al. (2009). We verify that this is the best-fitting value from a simple least-squares regression. When there is no K-band magnitude available, we use estimates from Neill et al. (2009) and Chang et al. (2015) when possible, as Burns et al. (2018) did.

The coefficients in equation (6) derived in Burns et al. (2018) and re-derived with our methods are presented in Table 4. The average deviation between the two sets of coefficients is 0.582 times the quadrature sum of the uncertainties. Additionally, we derive a set of coefficients while not accounting for host-galaxy mass. As expected, the average difference between this set and the original values is greater, albeit only slightly at 0.598 times the combined uncertainty.

We conclude that our methodology for calibrating the Tripp method is consistent with the method used in Burns et al. (2018). The difference in dispersion in Hubble residuals between the max_model and EBV_model2 seen in Section 4.1.1 is not due to errors in determining the calibration coefficients. Additionally, we do not find a significant difference in dispersion between the two models when examining the CSP data. Using the max_model, the Hubble residuals have an NMAD dispersion of 0.163 mag and a standard deviation of 0.233 mag, which is only marginally greater than the same values using EBV_model2: 0.157 and 0.227 mag.

5 SAMPLE SELECTION

We have NIR observations of 1217 unique transients, but only about a quarter of those are presently useful for cosmology. Our final sample is comprised of targets that pass three sets of cuts: one based on observational data, one based on fitting parameters, and one based on several outlier detection algorithms. The number of targets discarded and remaining after each cut are presented in Tables 5 and 6.

5.1 First cut: observational data

The set of all our observed transients includes unclassified or misclassified non-SNe Ia, galaxies with photometric or unknown redshifts, and SNe Ia missing coverage near maximum light in one or more all-sky survey bandpasses. In future work we intend to incorporate the unclassified transients that are photometrically consistent with SN Ia light curves, but for this paper, we do not include them in our analysis. Vincenzi et al. (2023) describe the magnitude of biases in cosmological measurements when using photometrically classified samples and discuss various methods for mitigating them to sub- per cent levels when estimating the dark energy equation of state parameter w.

Of the 1217 observed transients, 668 have been spectroscopically classified as usable SNe Ia. This number does not include SNe Ia subtypes that are unsuitable for distance inference using SALT3-NIR or `SNooPy`: 2002cx-like SNe (sometimes called SNe Iax, Li et al. 2003), 2002ic-like SNe (sometimes called SNe Ia-CSM, Hamuy et al. 2003), 2003fg-like SNe (formerly called super-Chandrasekhar SNe or SNe Ia-SC, Howell et al. 2006; Hicken et al. 2007; Ashall et al. 2021), and generally peculiar SNe Ia (Ia-pec). This number

Table 4. We show some of the Tripp calibration coefficients presented in Table 1 of Burns et al. (2018) and our derivations using the same data with our methodology. Our values differ from the original values by an average of 0.582 times the combined uncertainty. When not accounting for host-galaxy masses, the average difference slightly increases to 0.598 times the combined uncertainty.

Derivation	P_0 (mag)	P_1 (mag)	P_2 (mag)	$\beta_{\max}$	$\alpha_{\rm gal}$ (mag/dex)	$\sigma_{\rm int}$ (mag)	$v_{\rm pec}$ (km s^{-1})
CSP	−18.633(062)	−0.37(12)	0.61(32)	0.36(10)	−0.056(029)	0.11	336
This Work	−18.626(028)	−0.407(128)	−0.021(344)	0.292(096)	−0.044(032)	0.083(033)	384(57)
This Work (No Masses)	−18.607(024)	−0.352(126)	0.102(346)	0.270(097)	N/A	0.093(032)	384(57)

Table 5. Our first set of cuts is independent of the fitting model used and applies to all targets observed.

Cuts	Number cut	Remaining sample
Spec. classification	327	890
SN Ia	222	668
Spec. redshift	65	603
$E(B-V)_{\rm MW} < 0.3$	8	595
$N_{\rm obs} \geq 5$	0	595
Successful spectroscopic reduction	76	519
Successful photometric reduction	15	504

Table 6. The second set of cuts is based on the fitting model used, the SNPY_EBV sample using `SNooPy`'s EBV_model2, the SNPY_Max sample using the max_model, and the SALT sample using SALT3-NIR. We calculate the χ^2/DoF thresholds based on our comparison to the DEHVILS cut based on `SNANA`'s fit probability parameter. Our final two cuts are based on outlier detection. d_M refers to the Mahalanobis distance described in Section 5.3.1.

Cuts	SNPY_EBV Cut	Remaining	SNPY_Max Cut	Remaining	SALT Cut	Remaining
Successful fit	2	502	2	502	1	503
Rest frame m_J, m_V, and m_r			100	402		
$0.6 < s_{BV} < 1.3$	24	478	41	361		
$\sigma_{s_{BV}} < 0.2$	2	476	2	359		
$E(B-V)_{\rm host} < 0.3$ mag	75	401				
$\lvert x_1 \rvert < 3$					20	483
$\sigma_{x_1} < 1.5$					4	479
$\lvert c \rvert < 0.3$					43	436
$\sigma_c < 0.2$					1	435
Phase requirements	0	401	1	358	2	433
Reduced $\chi^2 < 4.14/4.51/1.31$	38	363	28	330	65	368
$d_M < 5\sigma_{d_M}$	4	359	4	326	4	364
UNITY outlier	2	357	2	324	2	362

does include several 2006bt-like candidates, which we discuss in Section 5.3.

Spectroscopic host-galaxy redshifts are available or have been successfully measured for 603 of these 668 SNe Ia. The remaining 65 include galaxies scheduled for spectroscopic observation, galaxies with spectral features manually judged to be too weak for accurate redshift determination, and galaxies with exceptionally low surface brightness, such that spectroscopic observation is prohibitively expensive. We remove additional eight targets that have Galactic reddening greater than 0.3 mag according to Schlafly & Finkbeiner (2011). As the last cut in this set, we remove targets with fewer than 5 optical and NIR observations, counting each quartet of ATLAS exposures as a single observation. Of the remaining 595 SNe Ia, 76 are in galaxies for which we have unreduced spectroscopic observations, and 15 encountered errors during photometric analysis, leaving 504 SNe Ia.

5.2 Second cut: fitting parameters

Removing targets based on fitting parameters necessarily requires successfully running each model's fitting procedure, which is not guaranteed for each possible permutation of input data. Without sufficient phase coverage in photometry, the shape parameter of a SN Ia becomes underconstrained. The same is true for insufficient wavelength coverage and the colour parameter or host-galaxy extinction. These produce singular covariance matrices, indicating degeneracy in the fitting parameters. Additionally, the models span finite combinations of phase and wavelength, making comparisons to some observations interpolative at best and often times extrapolative. The fit is unsuccessful if all data in a given bandpass lie outside the model domain. However, the phase of any observation is dependent on the estimated time of maximum light, which itself is a fitting parameter. This means that the success of a fit is partially dependent on how the fitting parameters are initialized. When a fit fails because one of the bandpasses has no data in a model's domain, we attempt to perform the same fit without data from the behaviour bandpass. If that succeeds, we use those fitting parameters to initialize a new fit, reintroducing the excluded data. Sometimes this leads to a successful fit using all available bandpasses, at other times a subset of available bandpasses, and occasionally the fit cannot be salvaged. The success or failure of a fit acts as a cut. We now define three distinct samples based on the three fitting models: SNPY_EBV with 502 fits from `SNooPy`'s EBV_model2, SNPY_Max with 502 fits from `SNooPy`'s max_model, and SALT with 503 fits from SALT3-NIR.

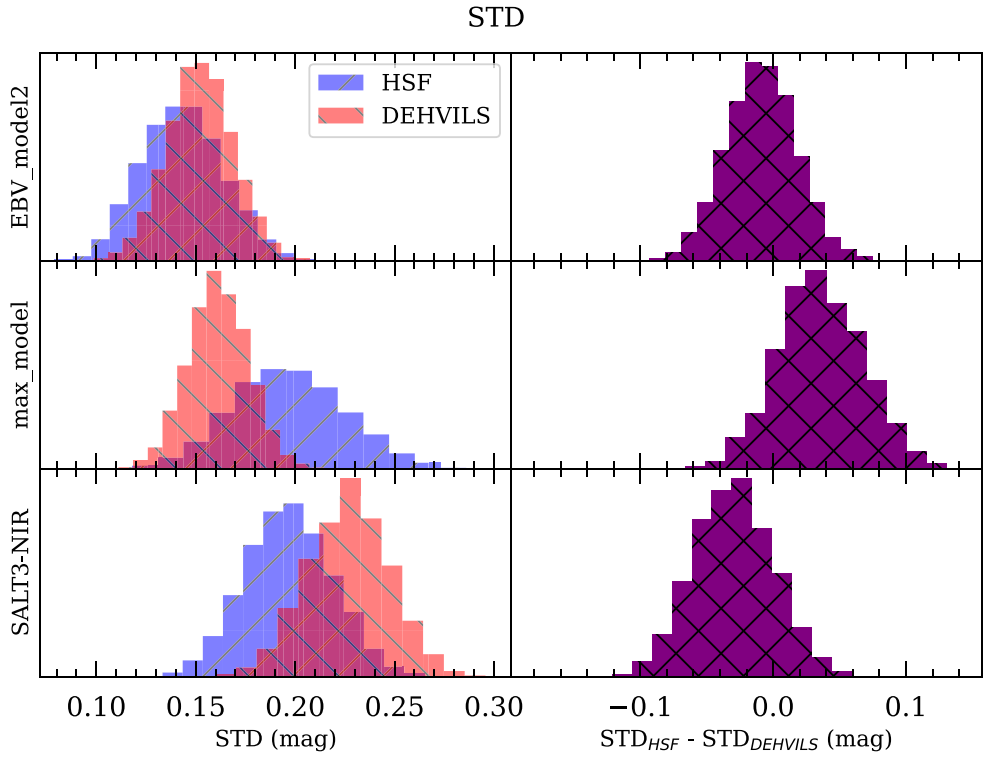

Figure 10. The same set of histograms as Fig. 9, but applied to standard deviation rather than NMAD. Again, using our measurements instead of DEHVILS photometry does not lead to a statistically significant difference in the dispersion of the Hubble residuals. The closest case is the standard deviation in the max_model, which increases by an average of 0.014 ± 0.022 mag. The more robust NMAD decreases by 0.015 ± 0.038 mag, indicating the increase in standard deviation is due to a few discrepant values rather than a systematically preferred set of photometry.

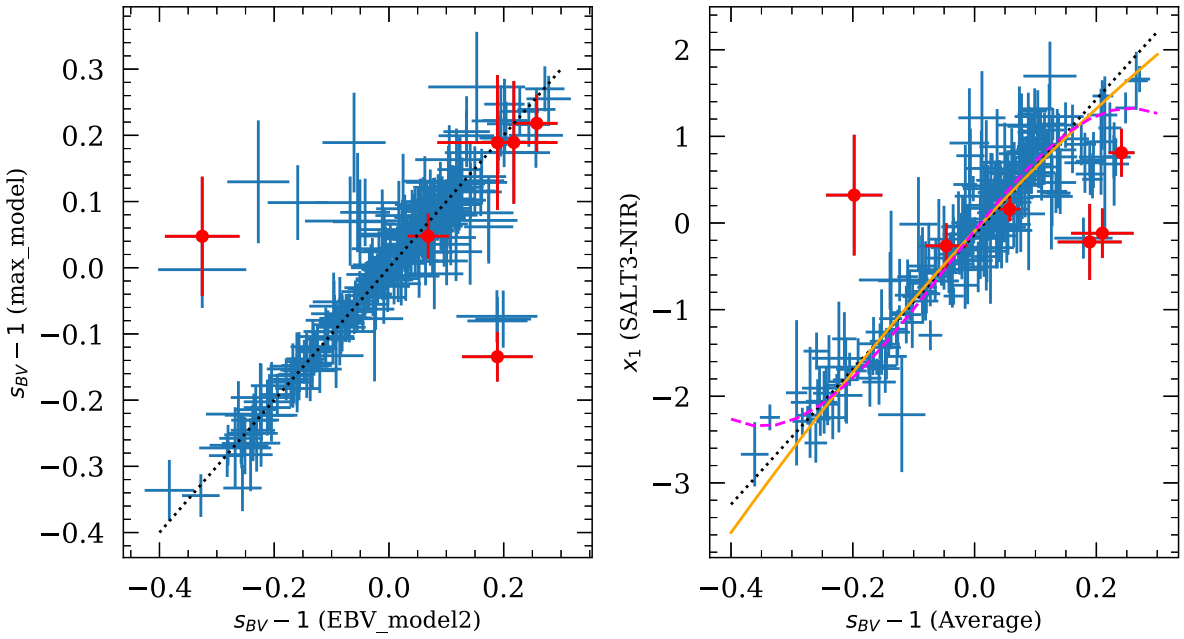

Figure 11. Left: The s_{BV} values of each SN are inferred using the same data, and either SNooPy's EBV_model2 or max_model. The one-to-one line is plotted in dotted black. Right: Each SN's inverse-variance weighted average s_{BV} value is compared to the x_1 value inferred by SALT3-NIR. We plot the linear (dotted black), quadratic (solid orange), and cubic (dashed magenta) polynomial fits determined through orthogonal distance regression. The Bayesian information criterion favours the cubic fit. In both plots, outliers identified through divergent model inferences are red and have circle markers. These outliers are ignored when calculating the parameter transformation equations.

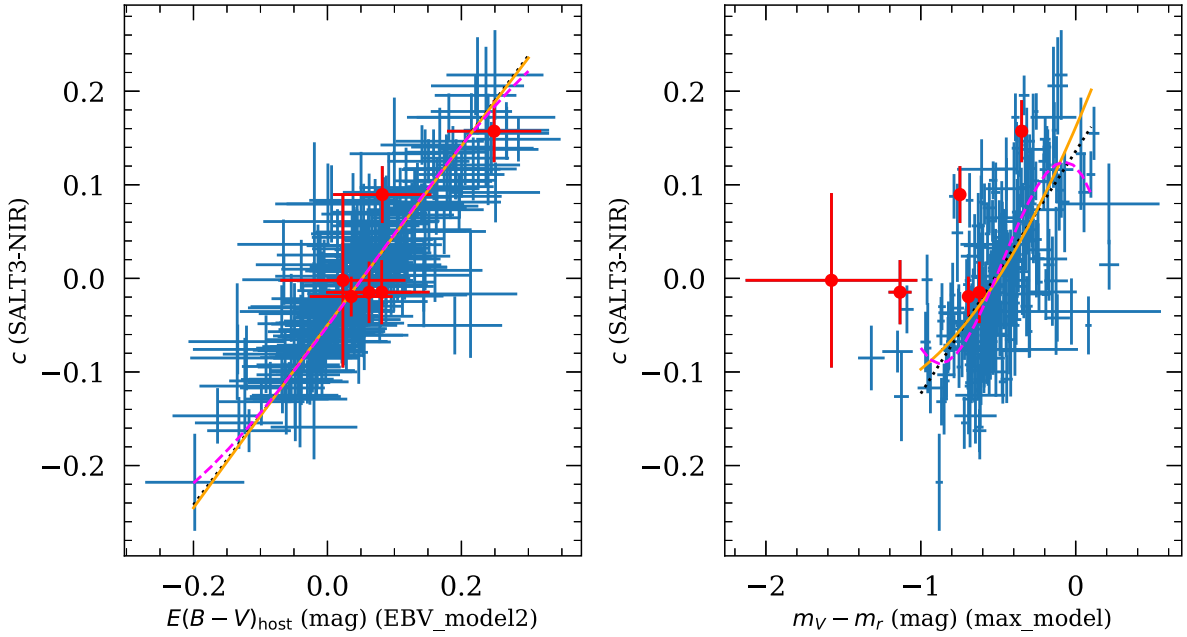

Figure 12. Left: The $E(B-V)_{\rm host}$ values inferred by SNooPy's EBV_model2 are plotted against the c values inferred by SALT3-NIR. As in Fig. 11, we show the linear (dotted black), quadratic (solid orange), and cubic (dashed magenta) polynomial fits determined through orthogonal distance regression. The linear and quadratic relationships overlap. Right: For each SN, the pseudo-colour between the peak apparent magnitudes in V and J determined by SNooPy's max_model are plotted against c. Once again, outliers identified through divergent model inferences are red and have circle markers. These outliers are ignored when calculating the parameter transformation equations.

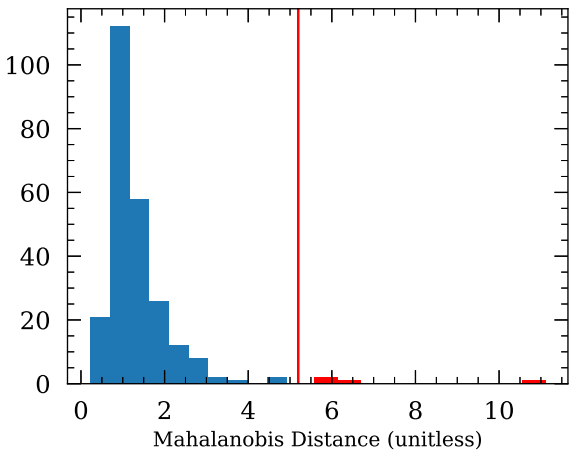

Figure 13. We identify outliers (red) based on disagreement between the three models inferences for an SN's time of maximum, shape, and colour. After transforming the SNooPy parameters s_{BV}, $E(B-V)_{\rm host}$, and $m_V - m_J$ to x_1, c, and c, we calculate the standard deviations in each SN's three inferred values for t_0, x_1, and c. We then parametrize disagreement using the Mahalanobis distance between the standard deviations ($\sigma_{t_0,i}$, $\sigma_{x_1,i}$, and $\sigma_{c,i}$) and a distribution centred at the origin. When the three models produce consistent estimates the standard deviation is relatively low, but when they diverge the standard deviation increases.

After fitting, we apply the following cuts. In the SNPY samples we use quality cuts from Jones et al. (2022), rejecting fits with shape factors outside the interval $0.6 < s_{BV} < 1.3$ (their 'loose' cut) or with uncertainty $\sigma_{s_{BV}} > 0.2$, and for SNPY_EBV, rejecting fits with host-galaxy $E(B-V)_{\rm host} > 0.3$. In the SNPY_Max sample, the rest-frame bandpasses used for calculating distances depend on both the observed bandpasses and the redshift. Since we infer distances using the J band and the $V-r$ colour, we cut SNe from SNPY_Max whenever the max_model does not provide inferences for the maximum apparent magnitudes in those bandpasses. While it is possible to force SNooPy to map to these bandpasses, the cross-band K-corrections required become much more sensitive to differences between the assumed and actual SED. This acts as a cut based on redshift. In the SALT sample we reject fits where $|x_1| > 3, \sigma_{x_1} > 1.5$, $|c| > 0.3$, or $\sigma_c > 0.2$ (Foley et al. 2017; Scolnic et al. 2018; Scolnic et al. 2022). We use the temporal coverage cut from Rubin et al. (2023), which is based on the calculated time of maximum light (t_0), the phase of the initial observation (t_i), and the phase of the final observation (t_f). Given that t_0 can vary between the three samples, we apply this cut to each sample independently. Adequately observed SNe Ia meet at least one of two sets of criteria. The first set requires t_i no more than 2 d after t_0, t_f at least 8 d after t_0, and $t_f - t_i$ must span at least 10 d. The second set allows for a later t_i, up to 6 d after t_0, as long as $t_f - t_i$ spans at least 15 d. Lastly, we remove fits with reduced χ^2 values above 4.14, 4.51, or 1.31 for fits in the SNPY_EBV, SNPY_Max, and SALT samples, respectively. These cut values come from the comparison to the DEHVILS sample in Section 4.1. This leaves our three samples with 363/502 objects in SNPY_EBV, 330/502 in SNPY_Max, and 368/503 in SALT.

5.3 Third cut: outlier detection

There are many vectors for outliers to appear in our sample: spectroscopic misclassification of core-collapse SNe, incorrectly assigned host-galaxy redshifts, errors in photometric reduction, or errors in fitting. Even with 'perfect' data and methods, an outlier could arise from anomalous astrophysical properties (e.g. an exotic progenitor system or detonation mechanism) or unclassified Type-Ia peculiarity. In particular, 2006bt-like SNe are difficult to identify without i band or NIR observations (Stritzinger et al. 2011; Phillips 2012). There are several objects in our sample that are classified as

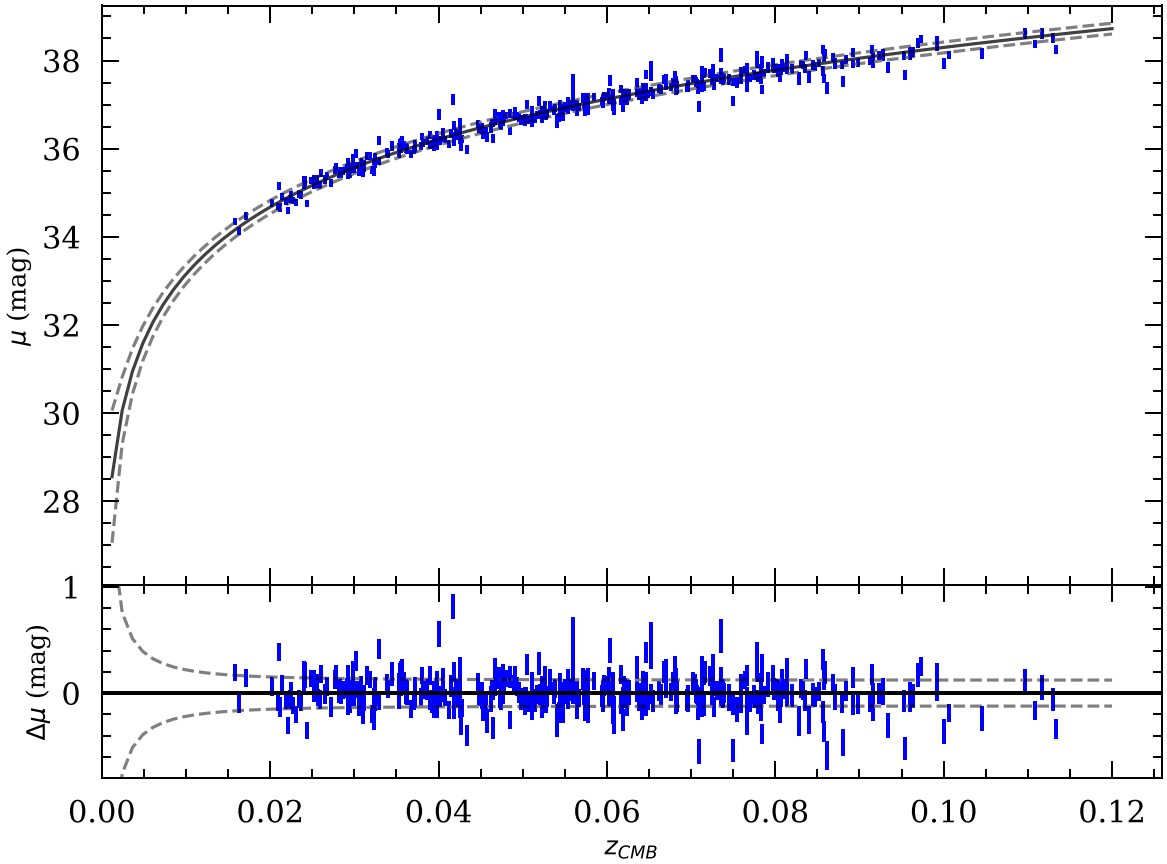

Figure 14. The top panel shows the Hubble diagram of the SNPY_EBV sample with residuals plotted below. The value of H_0 is degenerate with the absolute magnitude of SNe Ia, amounting to a constant vertical offset. The solid black line shows the ΛCDM model that zeros the inverse-variance weighted residuals. The dashed lines show the combined uncertainty due to the sample's $\sigma_{\rm int}$ and 250 km s^{-1} of uncertainty in peculiar velocity converted to uncertainty in distance modulus via equation (3).

SNe Ia on TNS, but have NIR light-curves suggestive of 2006bt-like SNe: SN 2020naj, SN 2020tkp, SN 2020mbf, and SN 2020sme. We employ two kinds of outlier detection methods. The first compares inferred parameters for common targets between the samples, and the second is based on the mixture model of Kunz, Bassett & Hlozek (2007) as implemented through UNITY (Rubin et al. 2015).

5.3.1 Divergent model inferences

In a Bayesian framework, the physical parameters inferred by each fitting model should draw from the same posterior distribution of 'true' physical parameters. This common quantity allows for simple error detection in the 246 SNe common to all samples. Where the estimates of the same parameter vary significantly, at least one model is likely to have converged on a local maximum in likelihood and is not reliable for inferring other parameters. The SNPY and SALT samples share a common definition for the time of maximum light, but differ in exactly how they quantify light-curve shape and colour. Burns et al. (2018) described a linear transformation between the x_1 parameter in SALT2 and the s_{BV} parameter in SNooPy. We use orthogonal distance regression and find a slightly different relationship, potentially due to differences between SALT2 and SALT3-NIR. After testing linear, quadratic, and cubic polynomial fits, the Bayesian information criterion favours a cubic relationship (102.6, 103.5, 77.2):

$$\begin{aligned} x_1 = &-0.09(02) + 8.97(27)(s_{BV} - 1) \\ &- 4.73(92)(s_{BV} - 1)^2 \\ &- 34.35(04)(s_{BV} - 1)^3. \end{aligned} \tag{14}$$

Here, s_{BV} is the average between the values inferred by SNooPy's two models. The relationship between s_{BV} values from the two SNooPy models as well as the relationship between their average and the SALT x_1 parameter is shown in Fig. 11.

The c parameter in SALT represents both intrinsic colour variation in SNe Ia and reddening from dust, while the $E(B-V)_{\rm host}$ fitting parameter in SNooPy is strictly concerned with the latter. However, Brout & Scolnic (2021) found that the correlation between intrinsic colour and luminosity may be weak, and that dust can provide the observed diversity of colours. We test linear, quadratic, and cubic fits, and the Bayesian information criterion supports a linear fit (−245.4, −239.9, −234.7):

$$c = -0.05(00) + 0.96(03)E(B-V)_{\rm host}\ {\rm mag}^{-1}. \tag{15}$$

Our colour information in the SNPY_Max sample comes from the differences in apparent maxima. To more effectively parametrize dust, we use the $m_V - m_J$ pseudo-colour. We test the same polynomial

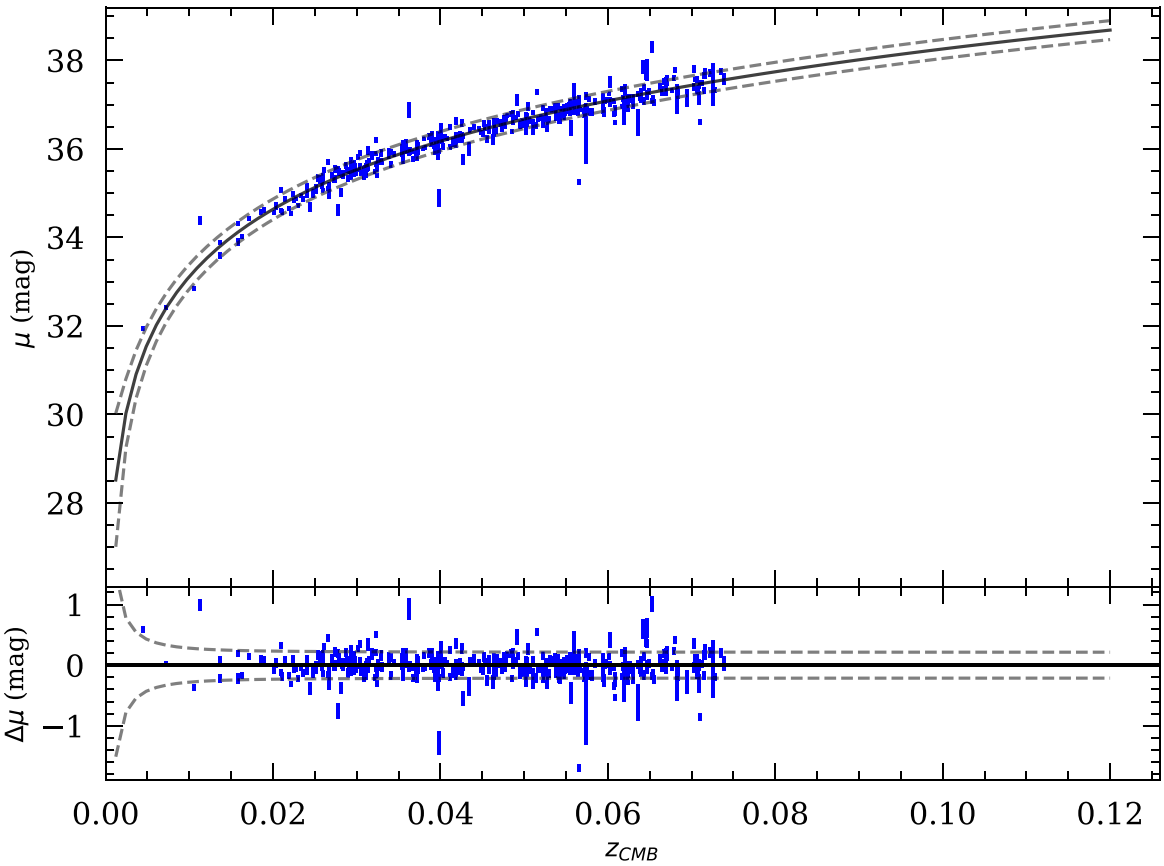

Figure 15. Same as Fig. 14, but using the SNPY_Max sample. The limited redshift range is due to the cut requiring the observed filters to map to the CSP V and r filters in the rest frame.

fits, and find support for a cubic fit (364.2, 369.4, 340.9):

$$\begin{aligned} c = {} & 0.12(02) - 0.17(10)(m_V - m_J)\ \mathrm{mag}^{-1} \\ & - 1.21(23)(m_V - m_J)^2\ \mathrm{mag}^{-2} \\ & - 0.85(16)(m_V - m_J)^3\ \mathrm{mag}^{-3} \end{aligned} \tag{16}$$

The relationships between the colour parameters are shown in Fig. 12.

After converting the SNooPy parameters into SALT parameters, we can directly compare each model's inferences for each SN to find where they disagree. We define $\sigma_{t_0,i}$, $\sigma_{x_1,i}$, and $\sigma_{c,i}$ as the standard deviation between the transformed fitting parameters of SN i in the SNPY samples its parameters in the SALT sample. We account for correlations between the differences by calculating the Mahalanobis distance between each point $m_i = (\sigma_{t_0,i}, \sigma_{x_1,i}, \sigma_{c,i})$ and a distribution D centred at the origin with covariance matrix Σ (Mahalanobis 1930). We approximate Σ by bootstrap resampling the parameter differences 5000 times, calculating each sample covariance S, and defining each element $\Sigma_{j,k}$ as the average of all sample elements $S_{j,k}$. Each distance $d_i(m_i, D) = \sqrt{m_i \Sigma^{-1} m_i^T}$, and can be understood as the number of standard deviations between point m_i and distribution D. The quadrature sum of the standard deviations is a similar metric if all dimensions are normalized to have unit variance, but does not account for correlations. Fig. 13 shows the histogram of distances. There are 4 SNe with distances greater than 5 times the standard deviation in d, indicating significant disagreement between the models. We recalculated the parameter transformation equations and Mahalanobis distances excluding these 4, and identified no additional outliers. The equations and figures presented are the recalculated versions. Disagreement alone leaves room for one or two of the models to have accurately fit the data, but while manual inspection often reveals which models fit the data well and which do not, we err on the side of caution by removing these 4 SNe from all three samples.

5.3.2 *Mixture-model analysis*

The mixture model introduced as BEAMS (Bayesian Estimate Applied to Multiple Species; Kunz et al. 2007) posits that an imperfect SNe Ia survey will lead to measurements following the sum of multiple distributions. Measurements of real SNe Ia should feature relatively low dispersion whereas measurements of survey contaminants will be more dispersed, and may have a different mean. As implemented in UNITY, both populations are assumed to be Gaussian around a common mean, and the outlier population is assumed to have variances of one in m_B^*, x_1, and c (Rubin et al. 2015). UNITY's use of SALT parameters means in its present version it can only process the SALT sample. Since the mixture-model framework is generalizable to arbitrary fitting parameters, future work could allow UNITY to process results from SNooPy. At such a time comparing outliers between samples could indicate whether the SNe Ia is astrophysically exceptional, or whether one of the models is unreliable. For now we are limited to examining SALT.

As UNITY sifts the data through its Bayesian hierarchical model, it produces a population level estimate of the fraction of outliers in the sample (with a prior of $\log f^{outl} \sim \mathcal{N}(-3, 0.5^2)$), and a pair

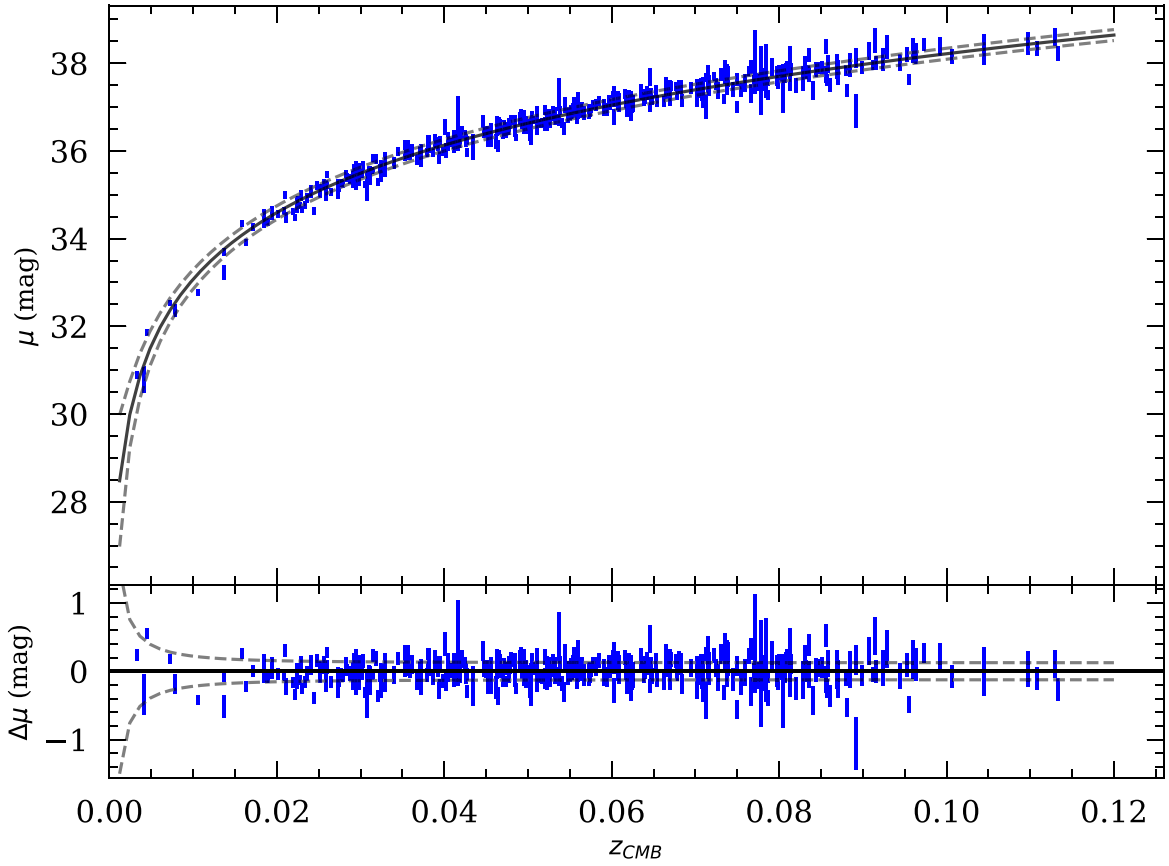

Figure 16. Same as Fig. 14, but using the SALT sample.

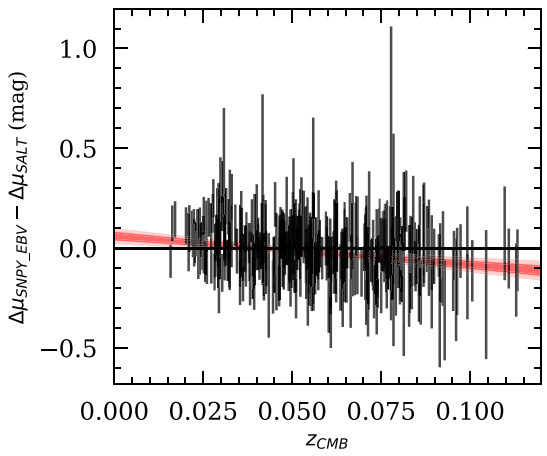

Figure 17. The difference between Hubble residuals from the SNPY_EBV sample ($\Delta_{\rm SNPY\_EBV}$) and the SALT sample ($\Delta_{\rm SALT}$) is anticorrelated with redshift. We fit a linear trend to the data, finding a slope of -1.4(3) mag. The contours show the 1σ and 2σ uncertainties.

of estimates for each object describing the likelihood it belongs to the normal or outlier population. Our sample has an estimated outlier fraction of 0.012±0.004 and there are 2 SNe more likely to be outliers than a part of the normal population. As before, we take the conservative approach of removing these objects from all samples. Some of the two may have been eliminated from the SNPY samples by previous cuts, which is why the number cut at this stage may vary across the samples.

5.4 Final Samples

After all three sets of cuts the sizes of our samples are 357, 324, and 362 for the SNPY_EBV, SNPY_Max, and SALT samples, respectively. Tables 5 and 6 enumerate the effect of each cut.

6 RESULTS

6.1 Hubble diagrams

We now present measurements of dispersion in the Hubble residuals of our three samples and their inferred intrinsic dispersions. We do not list the value of H_0 used in each sample because it is not a direct result of the data as explained in Section 3. To reiterate, H_0 is degenerate with the absolute magnitude of SNe Ia and we do not use alternative distance probes to estimate that magnitude.

The Hubble diagrams of each sample are presented in Figs 14, 15, and 16 and their dispersion measurements in Table 7. The same measurements of the targets common to all three samples are presented in Table 8.

6.2 Trend with redshift

There is a trend between the Hubble residuals and redshift in the SNPY_EBV and SALT samples. We perform linear fits using the Bayesian approach detailed in Jaynes & Crow (1999), using flat priors in sin(slope) and intercept, and a Jeffreys prior on scatter (Jeffreys 1946).

$$\Delta\mu_{\rm SNPY\_EBV} = -0.6(3) z_{\rm CMB}\ {\rm mag} + 0.03(02)\ {\rm mag} \quad (17)$$

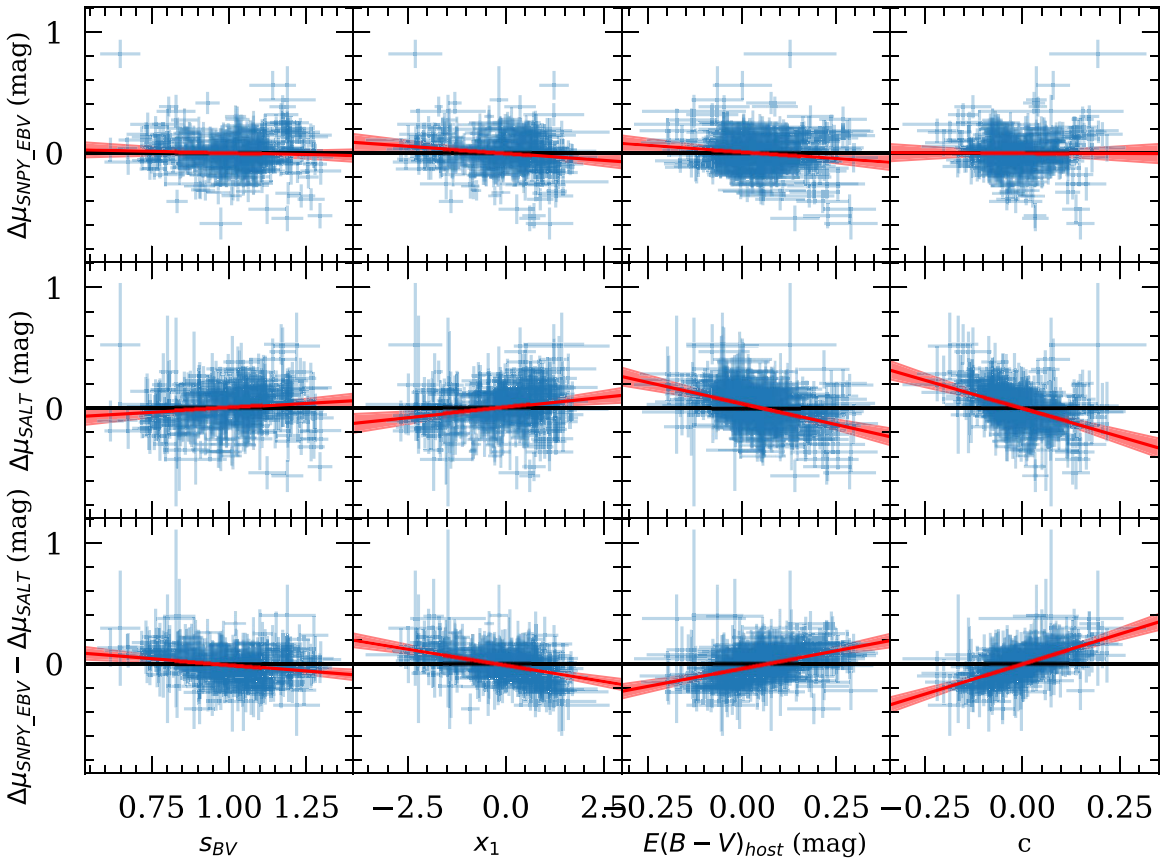

Figure 18. We compare the Hubble residuals in the SNPY_EBV sample, the residuals in the SALT sample, and their differences to the SNooPy fitting parameters s_{BV} and $E(B-V)_{\rm host}$ and SALT fitting parameters x_1 and c. The best-fitting linear relation is plotted with a solid red line and the 95 per cent confidence interval in light red contours. The strong correlation with colour parameters implies the standardization coefficient β is not calibrated correctly for our sample, and that the reddening law assumed in the EBV_model2 may be inappropriate for our sample.

$$\Delta\mu_{\rm SNPY\_Max} = 0.0(5)z_{\rm CMB}\ {\rm mag} + 0.01(02)\ {\rm mag} \quad (18)$$

$$\Delta\mu_{\rm SALT} = 0.4(4)z_{\rm CMB}\ {\rm mag} - 0.02(02)\ {\rm mag}. \quad (19)$$

The Hubble residuals in the SNPY_Max sample do not appear to trend with redshift. The differing signs in the slopes of the SNPY_EBV and SALT samples indicate that the issue is due to differences in the fitters rather than a real trend in the data or an issue in the estimation of $\mu_{\rm cos}$. A review of the SNooPy and SNCosmo code revealed no issue with the programmatic implementation of the methods in the literature (Guy et al. 2005, 2007; Burns et al. 2011, 2014; Kenworthy et al. 2021; Pierel et al. 2022).

The difference between SNPY_EBV and SALT Hubble residuals is seen most clearly in Fig. 17. Comparing residuals accounts for the zero-point offset in inferred μ in each sample and suppresses astrophysical properties that should affect both inferences equally, such as peculiar velocity or intrinsic variation in luminosity. We calculate uncertainties for the differences using the Pearson correlation between the distance modulus uncertainties in each sample.

We tested the SNooPy EBV_model2 and SALT3-NIR by fitting the r- and i-band photometry of an SN Ia at $z_{\rm CMB} \approx 0.72$ using both programs.[24] Using the H_0 and M values from the SNPY_EBV and SALT samples, we found the corresponding fitters produced residual distance moduli of 0.095 ± 0.167 and -0.315 ± 0.167. This indicates that any trend between residual distance modulus and redshift does not continue at higher redshifts.

The trend with redshift could be the result of differences in how the fitters account for shape or colour, which could both evolve with redshift due to selection effects. We investigated whether the SNPY_EBV residuals, SALT residuals, or their differences were correlated with the fitting parameters s_{BV}, $E(B-V)_{\rm host}$, x_1, and c, plotting the results in Fig. 18. Correlations imply the fitter is not properly accounting for the effect shape or colour has on the luminosity. In SALT this would mean the standardization parameters α or β are improperly calibrated. In SNooPy's EBV_model2 a correlation between the Hubble residuals and s_{BV} would imply there is a systematic difference between the light curves in our sample and the light curves used for interpolation. A correlation with $E(B-V)_{\rm host}$ would imply that the reddening law assumed in the EBV_model2 does not fully capture the dust properties affecting our observations.

The residuals appear correlated with the colour parameters $E(B-V)_{\rm host}$ and c, with the correlation most obvious in the SALT residuals and the residual differences. This calls our β coefficient into question, which comes from an analysis of the SALT sample performed in UNITY. The coefficients are not inferred by minimizing dispersion in the Hubble residuals, but by maximizing a likelihood in a Bayesian hierarchical model. UNITY models the 'true' x_1 and c parameters of an SN Ia as latent variables to account for Eddington bias. The

[24] SN 05D4ev from Guy et al. (2010).

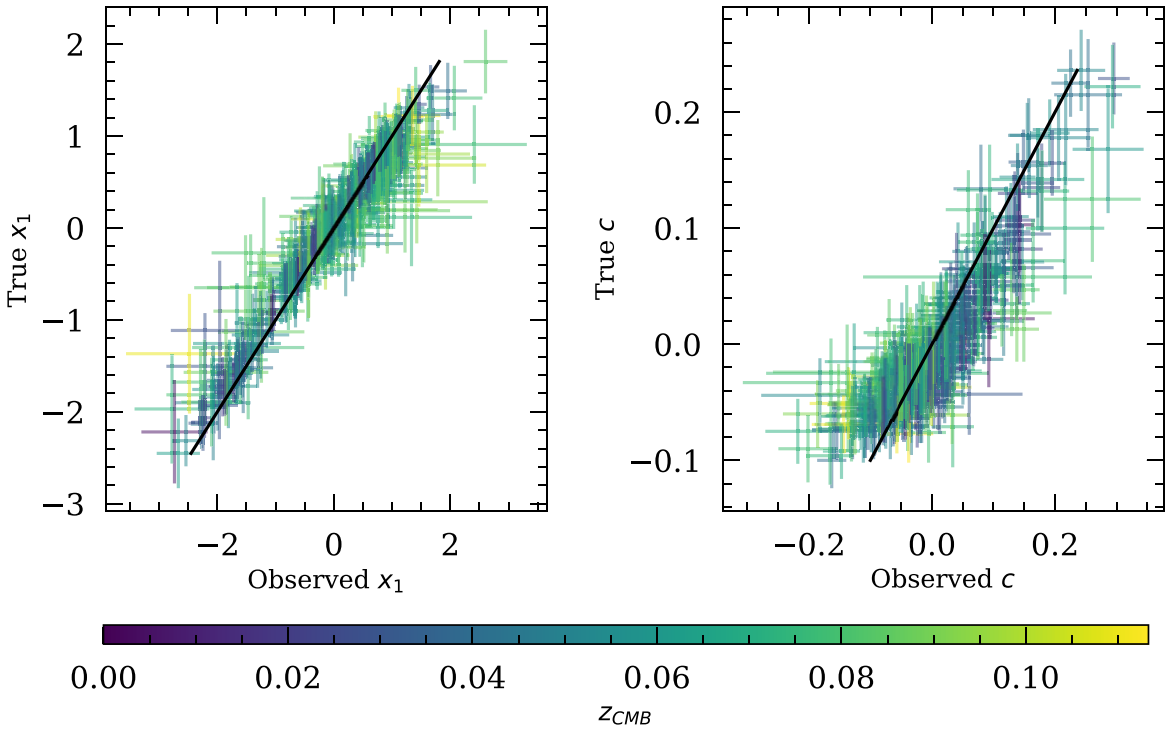

Figure 19. UNITY models the 'true' values of x_1 and c as latent variables in its hierarchical model. We compare these against the 'observed' values that come from SALT3-NIR fits, with one-to-one correspondence lines plotted in black. The standardization coefficients from UNITY are calibrated against the true values, and will not minimize Hubble residuals when used with the observed values.

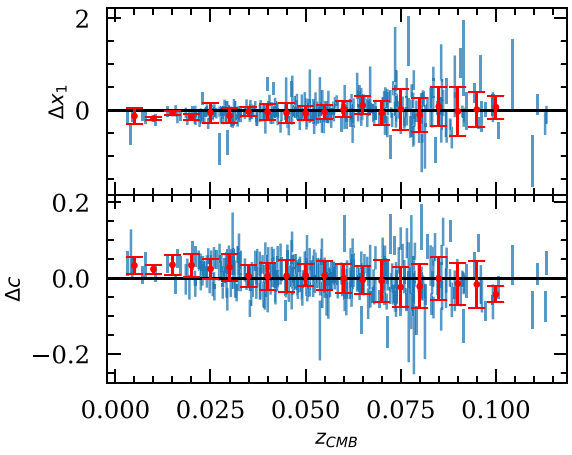

Figure 20. We parametrize the difference between the 'true' and 'observed' x_1 and c parameters as Δx_1 and Δc. The errors in each difference are estimated using the sample Pearson correlation coefficient. The red markers and errorbars show the medians and standard deviations of differences in bins 0.005 wide in redshift space. The standard deviations generally increase with redshift until $z \sim 0.09$, where data is relatively sparse. Δc appears offset from 0, especially at lower redshifts. This could imply that the hyperparameters UNITY uses to describe the distribution of c may require more flexibility to accurately model our observations.

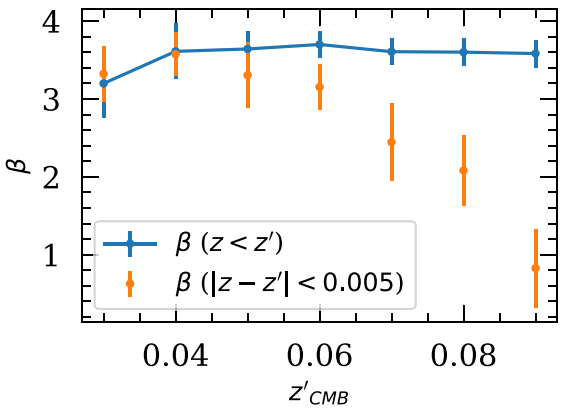

Figure 21. We analyse subsets of the SALT sample using UNITY and show the inferred β values as a function of redshift cut value z'_{CMB}. The data connected by the line show the results from subsets defined by a maximum redshift $z < z'$. β is relatively stable as the sample expands to include higher redshift targets. The unconnected data are from redshift bins centred on z' and 0.01 wide, such that $|z - z'| < 0.005$. SNe in higher redshift bins prefer lower β values, which could be a result of Eddington bias becoming more significant at greater distances.

standardization coefficients operate on these 'true' values rather than the 'observed' values that come from a light-curve fit. We compare the 'true' and 'observed' x_1 and c parameters in Fig. 19. Deviations from one-to-one correspondence come from both statistical error and Eddington bias, which manifests as 'observed' parameters scattering away from 0.

Fig. 20 shows that the bias seems more prevalent at higher redshifts. Parameter differences in bins at higher redshifts generally have larger standard deviations, with sample incompleteness heavily affecting bins beyond $z \sim 0.09$. This could indicate that the uncertainties in x_1 or c are underestimated in fits to noisier data, but verifying such a claim would require simulations beyond the scope of this paper (such as those in Peterson et al. 2024). Additionally, the median differences in c appear non-zero, especially in lower redshift bins. We have not identified a definitive cause for this behaviour,

but speculate that our choice of hyperparameters in UNITY does not allow for the flexibility needed to model the distribution of c over the parameter space spanned by our sample. This could produce a correlation between SALT Hubble residuals and redshift or c independent of β.

To ascertain the validity of using a single β value over the entire redshift range, we analysed subsets of the SALT sample in UNITY. We examined subsets consisting of targets within redshift bins of width 0.01, as well as a cumulative sum including all targets below a series of redshifts. Fig. 21shows that the inferred β decreases in higher redshift bins, but that it remains relatively stable in the cumulative case. This suggests the hyperparameters describing the distribution of c are robust against higher redshift SNe Ia, but that these SNe prefer a lower value for β. This assumes the c parameter follows a single distribution over the whole sample, rendering measurements that may suggest otherwise an effect of bias. UNITY permits alternative parametrizations, such as a broken-linear form for β, but adding such complexity is beyond the scope of this paper. Similarly, more sophisticated analyses of dust properties using `SNooPy` have been performed with its color_model (e.g. Burns et al. 2018; Johansson et al. 2021) or by using more of the colour information in the max_model (e.g. Uddin et al. 2020, 2023). Implementing and evaluating these approaches will be necessary before using our data for robust cosmological analyses.

Thus, we do not find a satisfactory solution to eliminate the correlations between Hubble residuals and redshift or colour parameter. Empirical corrections are possible using equations (17), but such an approach is neither physically motivated nor statistically rigorous. More detailed analyses are required to fully understand and rectify this issue.

6.3 The effect of NIR photometry on distance measurements

As mentioned in Section 1.2.1, NIR photometry offers two key benefits when analysing SNe Ia. The effects of dust are suppressed and SNe Ia demonstrate less variable peak luminosities in the NIR. We examine the benefits of NIR photometry by comparing fits using NIR and optical photometry to fits using only optical photometry. We split our three samples into six: SNPY_EBV_OJ, SNPY_Max_OJ, and SALT_OJ (the OJ samples), which include optical and J-band photometry and SNPY_EBV_O, SNPY_Max_O, and SALT_O (the O samples), which are their optical-only counterparts. Unlike the SNPY_EBV, SNPY_Max, and SALT samples, which vary in size, target selection, and even bandpasses used to fit a given target, we enforce parity between the OJ and O samples. To do this we prepare the O samples following the same methodology used to produce SNPY_EBV, SNPY_Max, and SALT except without J-band photometry. The number of SNe discarded at each cut and the size of the final samples are listed in Table 9. The six samples are made of the SNe common to both the O and OJ samples.

We modify the outlier detection method described in Section 5.3.1 to highlight disagreement between the OJ and O samples rather than between SNPY_EBV, SNPY_Max, and SALT. This precludes the need to transform `SNooPy` fitting parameters into SALT parameters. The fitting parameters of SNe in the OJ and O samples produce difference vectors: $m = (\delta t_0, \delta s_{BV}, \delta E(B-V)_{\rm host})$ for differences between SNPY_EBV_OJ and SNPY_EBV_O, $m = (\delta t_0, \delta s_{BV}, \delta(V-r))$ for differences between SNPY_Max_OJ and SNPY_Max_O, and $m = (\delta t_0, \delta m_B^*, \delta x_1, \delta c)$ for differences between SALT_OJ and SALT_O. We use the Mahalanobis distance $d(m, D) = \sqrt{m\Sigma^{-1}m^T}$ to identify outliers, once again approximating the covariance matrix Σ by bootstrap resampling the parameter differences 5000 times and averaging the sample covariances. The distributions of Mahalanobis distances for samples the three pairs of samples are given in Fig. 22.

Measurements of dispersion in each sample's Hubble residuals are presented in Table 10. We characterize the differences between the OJ and O samples with the bootstrapping method we used when varying photometry in Section 4.1.2. The values and uncertainties in Table 10 are the averages and standard deviations of this process. Histograms of the resampled dispersion differences are plotted in Fig. 23. The various dispersion estimators show a general decrease when adding J-band photometry to the O samples, but most of the differences are within one standard deviation of no change. The exceptions are the NMAD in the SALT and SNPY_EBV samples and the RMS in the SALT sample. Interpreting these exceptions as indicators of decreased dispersion while ignoring the other measures is a classic case of the multiple comparisons problem. To control the family-wise error rate we use the sequentially rejective Bonferroni test (Holm 1979). None of the distributions are far enough from 0 to claim that including J-band photometry leads to statistically significant decreases in dispersion. This does not imply other methodologies do not or cannot benefit from the J-band photometry, but that with our samples, cuts, and methods, we cannot definitively say NIR photometry leads to smaller Hubble residuals.

7 DISCUSSION

The RMS of the Hubble residuals is 0.165 mag in the SNPY_EBV sample ($N = 357$), 0.245 mag in the SNPY_Max sample ($N = 324$), and 0.186 mag in the SALT sample ($N = 362$). Our result goes against a number of studies that support the use of NIR photometry in deriving distances to SNe Ia, but is not alone in finding relatively large dispersions. Stanishev et al. (2018) combine optical and NIR light curves from numerous sources, including new observations, and find RMS values of ∼0.15 mag ($N \sim 120$) while Johansson et al. (2021) did the same and found an RMS of 0.19 mag ($N = 165$). Notably, the sub-sample of 16 new SNe Ia presented in Stanishev et al. (2018) were only imaged once or twice in the NIR, and the RMS of their residuals is ∼0.2. Sparsity may play a role in the greater dispersion, but Müller-Bravo et al. (2022) found that the effect is relatively small, bringing an RMS of 0.166 mag to 0.180 mag ($N = 36$) when removing all but one epoch from J-band light curves.

One key difference between our work and those which find smaller dispersions is that our analysis does not force SNe Ia to be standard candles in the NIR. It may be possible that variation in SN Ia NIR luminosity, if it does vary, is not parametrized by the correlations observed in the optical. In our comparison with Peterson et al. (2023) we found that our fitting methods applied to their YJH data resulted in large dispersion, whereas they fixed the shape and colour parameters in the NIR-only fits to be 0, removing any variation in luminosity between SNe, and measured lower dispersion than when using bandpass combinations including optical data from ATLAS. This is similar to Avelino et al. (2019), who treated SNe Ia as standard candles in the NIR and found that smaller Hubble residuals than those from optical-only fits using `SNooPy` or SALT2. That said, not all studies favour this approach. Jones et al. (2022) measured Hubble residual scatter over several analyses of 79 SNe Ia and measured an RMS of ∼0.17 mag using `SNooPy` to fit only NIR data, and ∼0.14 mag when including optical data with $R_V = 1.52$.

Table 7. We present various parametrizations of the dispersion in Hubble residuals in the SNPY_EBV, SNPY_Max, and SALT samples, as well as the intrinsic dispersion needed to reconcile the propagated uncertainties and measured dispersion. The samples overlap significantly, but they are not identical. For comparison purposes, we provide measurements of the common subset in Table 8.

Sample	N	RMS (mag)	WRMS (mag)	NMAD (mag)	σ_{int} (mag)
SNPY_EBV	357	0.165(010)	0.152(008)	0.123(010)	0.121(011)
SNPY_Max	324	0.245(024)	0.214(028)	0.164(011)	0.212(028)
SALT	362	0.186(011)	0.174(009)	0.153(010)	0.123(011)

Table 8. We present the values from Table 7 derived from the intersection between the SNPY_EBV, SNPY_Max, and SALT samples. All samples are comprised of the same 240 objects.

Sample	RMS (mag)	WRMS (mag)	NMAD (mag)	σ_{int} (mag)
SNPY_EBV	0.137(008)	0.133(007)	0.116(011)	0.098(010)
SNPY_Max	0.171(011)	0.148(009)	0.147(012)	0.135(010)
SALT	0.146(007)	0.150(008)	0.142(012)	0.103(010)

Table 9. Similar to Table 6, we list the number of SNe discarded at each cut for our optical-only samples. We begin after the survey wide cuts of Table 5, starting with the number of successful fits in SNooPy's EBV_model2 (SNPY_EBV_O), max_model (SNPY_Max_O), and SALT3-NIR (SALT_O). Our final two cuts are based on outlier detection.

Cuts	SNPY_EBV_O Cut	Remaining	SNPY_Max_O Cut	Remaining	SALT_O Cut	Remaining
Passed sample wide cuts		504		504		504
Successful fit	2	502	2	502	1	503
Rest frame m_V and m_r			92	410		
$0.6 < s_{BV} < 1.3$	23	479	40	370		
$\sigma_{s_{BV}} < 0.2$	3	476	7	363		
$E(B-V)_{\rm host} < 0.3$ mag	66	410				
$\lvert x_1\rvert < 3$					22	481
$\sigma_{x_1} < 1.5$					5	476
$\lvert c\rvert < 0.3$					38	438
$\sigma_c < 0.2$					5	433
Phase requirements	1	409	1	362	2	431
Reduced $\chi^2 < 4.14/4.51/1.31$	21	388	21	341	40	391
Also In OJ sample	173	215	166	175	36	355
$d_M < 5\sigma_{d_M}$	2	213	1	174	5	350
UNITY outlier	1	212	0	174	2	348

The optical and NIR RMS increased to ∼0.18 mag when using a Milky Way-like value of $R_V = 3.1$, emphasizing the importance of assumptions regarding dust. Similarly, Pierel et al. (2022) examined the performance of SALT3-NIR, SALT3, and SNooPy over various bandpass combinations and model parametrizations. They found a Hubble residual RMS of ∼0.12 mag ($N = 24$) when using SALT3-NIR with optical and YJH data, fitting for shape and colour, and a greater RMS of ∼0.13 mag for the same sample using only YJH data and keeping the shape and colour parameters fixed at 0.

As explored in Section 6.2, there is a trend between redshift and the Hubble residuals in SNPY_EBV and SALT. Correcting this trend empirically will decrease the dispersion of the Hubble residuals, but such an a posteriori treatment invites bias. There are several other obvious ways to decrease the measured dispersion. One could calibrate α or β by minimizing Hubble residuals, employ corrections by the redshift bin, or use cuts based on Hubble residuals such as Chauvanet's crietrion or σ-clipping. There is ample motivation for using such techniques. Our robust measure of dispersion, NMAD, is consistently lower than RMS and WRMS in all three samples, which suggests that there are SNe Ia in our samples could be considered outliers. However, we choose to present our results as we found them to avoid contaminating them with ad hoc corrections.

The study of SNe Ia in the NIR has advanced as more data have become available, but there are still challenges that must be met to maximize the potential benefits. At the moment it is unclear whether SNe Ia are standard candles in the NIR or simply require less standardization than in the optical. Similarly, it is not clear if the shape–luminosity correlation observed in the optical is still the primary mode of variation in the NIR. Answering these questions will require various kinds of data. Spectral time series provide unique views into the physical mechanisms of SNe Ia, while also improving the accuracy of K-corrections. High-cadence, multiband observations like those pursued by the DEHVILS survey are vital for building standardization models. The Hawai'i Supernova Flows project provides a valuable test bed for SN Ia research through its unprecedented sample size.

8 CONCLUSION

This paper introduces the Hawai'i Supernova Flows project, a peculiar velocity survey designed to obtain systematics-limited distances to SNe Ia while consuming minimal dedicated observational resources. We review the observational components of our project: optical photometry from public all-sky surveys, NIR photometry from UKIRT, and optical spectroscopy from the UH 2.2 m and

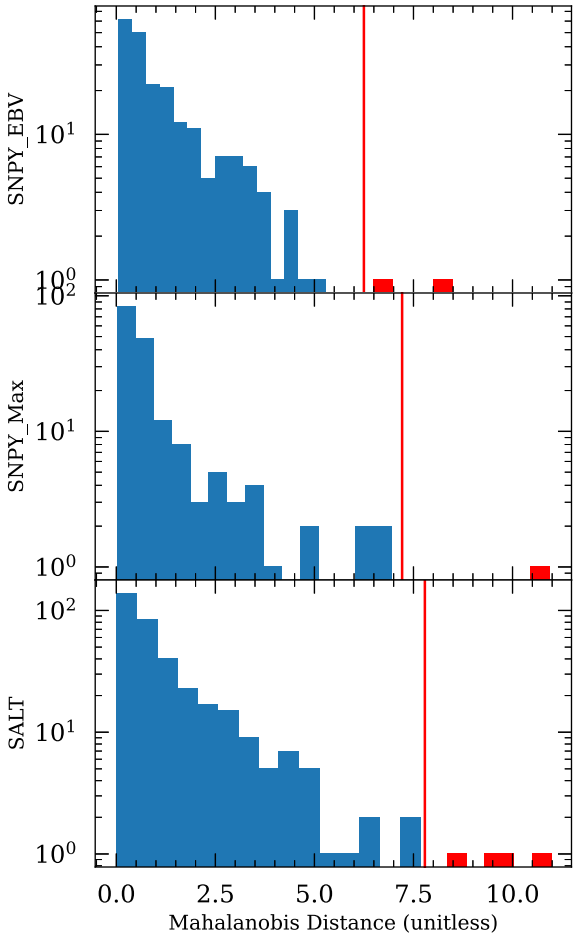

Figure 22. Our first outlier detection algorithm is based on agreement between models fit with and without J-band photometry as quantified by the Mahalanobis distance between parameter differences and the origin, representing an SN with identical estimates in the OJ and O samples. A greater distance indicates greater disagreement between fits, with significant disagreement indicating at least one of the models is unreliable.

Subaru. We validate our methods and data against external sources. We use SDSS spectra to find that our redshift inferences are typically within 45 km s^{-1} of the values in HyperLEDA. The absolute wavelength calibration of our spectroscopic observations introduces minimal error, producing redshifts typically within 48 km s^{-1} of their published values. Using our methodology to fit data from our partner program DEHVILS, we found no increase in the dispersion of Hubble residuals when using only optical data, or using optical and NIR data with SNooPy's EBV_model2. The increase in dispersion seen when using SALT3-NIR seems to come from our choice to calibrate the standardization coefficients with UNITY instead of only using the Hubble residuals. Our independent photometric reductions of the same J-band observations are consistent, producing similar dispersions. Given cuts on both SNooPy and SALT fitting parameters, our three final samples include 357, 324, and 362 SNe. The RMS values of their Hubble residuals are 0.165, 0.245, and 0.186 mag.

The Nancy Grace Roman space telescope will obtain rest-frame NIR observations of SNe Ia within $z \sim 0.7$ (Hounsell et al. 2018; Rose et al. 2021), necessitating the maturation of SN Ia cosmology in the NIR. Thus far, the majority of publicly available NIR SN Ia light-curves have come from CSP-I ($N = 123$, Krisciunas et al. 2017), CfAIR2 ($N = 94$; Friedman et al. 2015), or recent work from our partner program DEHVILS ($N = 96$; Peterson et al. 2023). Data from CSP-II ($N = 214$; Hsiao et al. 2019; Phillips et al. 2019), the SIRAH program ($N = 24$, HST-GO 15889; Jha et al. 2019), and other exciting projects are expected in the near future. Upon publication of this work, we will release NIR observations of 1217 transients, including 668 spectroscopically classified SNe Ia, 437 of which are in at least one of our final samples, and 215 spectroscopic redshifts for SN Ia host-galaxies that have not been previously measured. The NIR photometry of the Hawai'i Supernova Flows project is the largest homogeneous collection of its kind in terms of unique SNe Ia. This growing sample will provide increasing resolution into peculiar velocities as a function of position on the sky and redshift, permitting us to map the structure of dark matter.

ACKNOWLEDGEMENTS

This work, AD, and BS are supported by National Science Foundation grant AST-1911074.

AD and KSM are supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant Agreement No. 101002652).

This publication makes use of data products from the Two Micron All Sky Survey, which is a joint project of the University of Massachusetts and the Infrared Processing and Analysis Center/California Institute of Technology, funded by the National Aeronautics and Space Administration and the National Science Foundation.

The Pan-STARR Surveys (PS1) and the PS1 public science archive have been made possible through contributions by the Institute for Astronomy, the University of Hawaii, the Pan-STARRS Project Office, the Max-Planck Society and its participating institutes, the Max Planck Institute for Astronomy, Heidelberg and the Max Planck Institute for Extraterrestrial Physics, Garching, The Johns Hopkins University, Durham University, the University of Edinburgh, the Queen's University Belfast, the Harvard–Smithsonian Center for Astrophysics, the Las Cumbres Observatory Global Telescope Network Incorporated, the National Central University of Taiwan, the Space Telescope Science Institute, the National Aeronautics and Space Administration under Grant No. NNX08AR22G issued through the Planetary Science Division of the NASA Science Mission Directorate, the National Science Foundation Grant No. AST-1238877, the University of Maryland, Eotvos Lorand University (ELTE), the Los Alamos National Laboratory, and the Gordon and Betty Moore Foundation.

This research has made use of NASA's Astrophysics Data System.

We acknowledge the usage of the HyperLeda data base (http://leda.univ-lyon1.fr).

This research has made use of the Set of Identifications, Measurements and Bibliography for Astronomical Data (SIMBAD) data base, operated at Centre de Données astronomiques de Strasbourg (CDS), Strasbourg, and France.

This research has made use of the NASA/IPAC Extragalactic Database (NED), which is funded by the National Aeronautics and Space Administration and operated by the California Institute of Technology.

UKIRT is owned by the University of Hawaii (UH) and operated by the UH Institute for Astronomy. When (some of) the data reported here were obtained, the operations were enabled through the cooperation of the East Asian Observatory.

Table 10. Adding NIR photometry does not lead to statistically significant decreases in the various measures of dispersion. Each estimator is calculated after bootstrap resampling the Hubble residuals 5000 times. The value is the average and the uncertainty is the standard deviation.

Sample	RMS (mag)	WRMS (mag)	NMAD (mag)	σ_{int} (mag)
SNPY_EBV_OJ	0.166(012)	0.154(010)	0.120(011)	0.122(014)
SNPY_EBV_O	0.171(011)	0.162(010)	0.149(012)	0.127(014)
SNPY_Max_OJ	0.281(037)	0.245(047)	0.173(019)	0.237(044)
SNPY_Max_O	0.276(028)	0.227(021)	0.188(022)	0.248(028)
SALT_OJ	0.170(008)	0.171(009)	0.146(010)	0.122(011)
SALT_O	0.185(009)	0.184(010)	0.162(011)	0.129(012)

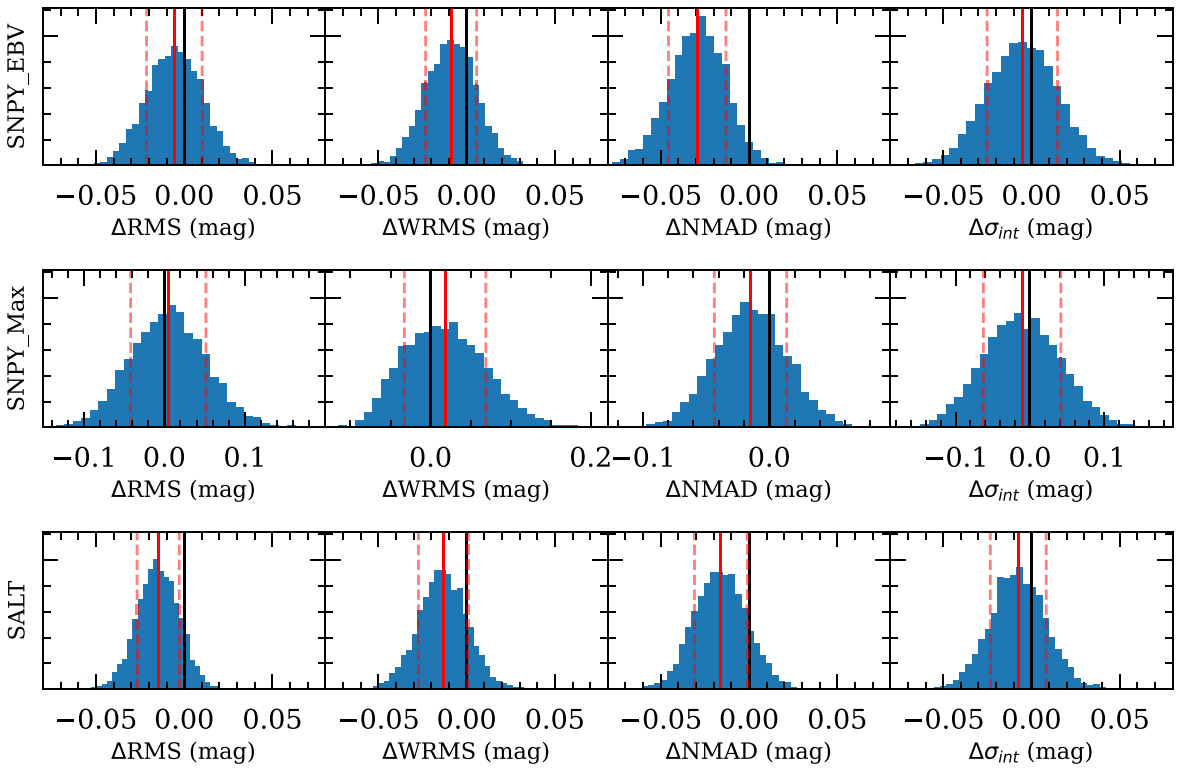

Figure 23. We recorded various measures of dispersion over 5000 iterations of bootstrap resampling, and show the distributions of differences between the OJ and O samples with the averages given by the solid red lines. Including J-band photometry with the optical data typically leads to decreases in the three tested measures of dispersion in Hubble residuals and the inferred intrinsic dispersion, but those differences are usually within one standard deviation (red dashed lines) of 0 mag (solid black lines).

The Zwicky Transient Facility (ZTF) forced-photometry service was funded under the Heising–Simons Foundation grant #12540303 (PI: Graham).

This work has made use of data from the European Space Agency (ESA) mission *Gaia* (https://www.cosmos.esa.int/gaia), processed by the *Gaia* Data Processing and Analysis Consortium (DPAC, https://www.cosmos.esa.int/web/gaia/dpac/consortium). Funding for the DPAC has been provided by national institutions, in particular the institutions participating in the *Gaia* Multilateral Agreement.

We acknowledge ESA Gaia, DPAC, and the Photometric Science Alerts Team (http://gsaweb.ast.cam.ac.uk/alerts).

Funding for the Sloan Digital Sky Survey (SDSS) has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Aeronautics and Space Administration, the National Science Foundation, the U.S. Department of Energy, the Japanese Monbukagakusho, and the Max Planck Society. The SDSS Web site is http://www.sdss.org/.

The SDSS is managed by the Astrophysical Research Consortium (ARC) for the Participating Institutions. The Participating Institutions are The University of Chicago, Fermilab, the Institute for Advanced Study, the Japan Participation Group, The Johns Hopkins University, the Korean Scientist Group, Los Alamos National Laboratory, the Max-Planck-Institute for Astronomy (MPIA), the Max-Planck-Institute for Astrophysics (MPA), New Mexico State University, University of Pittsburgh, University of Portsmouth, Princeton University, the United States Naval Observatory, and the University of Washington.

DATA AVAILABILITY

The following data are available at https://www.github.com/ado8/hsf_DR1:

(i) J-band light curves of all observed targets regardless of spectroscopic classification.

(ii) Weighted cross-correlation results for all galaxies we observed with SNIFS or FOCAS.

(iii) Multiple sets of fitting parameters and uncertainties for all spectroscopically confirmed SNe Ia, with sets covering all combinations of fitting model (EBV_model2, max_model, SALT3-NIR) and data used (optical only or optical and NIR).

(iv) Additional sets of fitting parameters for SNe Ia used in our comparisons with DEHVILS and CSP.

The code used in our analysis can be found at https://www.github.com/ado8/hsf_code. Data such as images and spectra may be available upon reasonable request.

REFERENCES

Adelman-McCarthy J. K. et al., 2007, ApJS, 172, 634
Aggarwal K., Budavári T., Deller A. T., Eftekhari T., James C. W., Prochaska J. X., Tendulkar S. P., 2021, ApJ, 911, 95
Alard C., Lupton R. H., 1998, ApJ, 503, 325
Alard C., Lupton R., 1999, Astrophysics Source Code Library, record ascl:9909.003
Amanullah R. et al., 2010, ApJ, 716, 712
Anand G. S. et al., 2021, AJ, 162, 80
Andrae R., Schulze-Hartung T., Melchior P., 2010, preprint (arXiv:1012.3754)
Ashall C., Mazzali P., Sasdelli M., Prentice S. J., 2016, MNRAS, 460, 3529
Ashall C. et al., 2021, ApJ, 922, 205
Avelino A., Friedman A. S., Mandel K. S., Jones D. O., Challis P. J., Kirshner R. P., 2019, ApJ, 887, 106
Barbary K. et al., 2022, SNCosmo, Zenodo, available at: https://doi.org/10.5281/zenodo.7117347
Barone-Nugent R. L. et al., 2012, MNRAS, 425, 1007
Bellm E. C. et al., 2019, PASP, 131, 018002
Bertone G., Hooper D., 2018, Rev. Mod. Phys., 90, 045002
Bertschinger E., Dekel A., 1989, ApJ, 336, L5
Betancourt M. J., 2013, preprint (arXiv:1304.1920)
Blondin S., Mandel K. S., Kirshner R. P., 2011, A&A, 526, A81
Boggess A. et al., 1978a, Nature, 275, 377
Boggess A. et al., 1978b, Nature, 275, 372
Boone K. et al., 2021, ApJ, 912, 71
Brout D., Riess A., 2023, preprint (arXiv:2311.08253)
Brout D., Scolnic D., 2021, ApJ, 909, 26
Brout D. et al., 2022, ApJ, 938, 110
Burns C. R. et al., 2011, AJ, 141, 19
Burns C. R. et al., 2014, ApJ, 789, 32
Burns C. R. et al., 2018, ApJ, 869, 56
Campbell L. A. et al., 2014, MNRAS, 443, 1231
Cardelli J. A., Clayton G. C., Mathis J. S., 1989, ApJ, 345, 245
Carpenter B. et al., 2017, J. Stat. Softw., 76, 1
Carrick J., Turnbull S. J., Lavaux G., Hudson M. J., 2015, MNRAS, 450, 317
Casali M. et al., 2007, A&A, 467, 777
Chambers K. C. et al., 2016, preprint (arXiv:1612.05560)
Chang Y.-Y., van der Wel A., da Cunha E., Rix H.-W., 2015, ApJS, 219, 8
Conley A. et al., 2008, ApJ, 681, 482
Contreras C. et al., 2010, AJ, 139, 519
Courtois H. M., Hoffman Y., Tully R. B., Gottlöber S., 2012, ApJ, 744, 43
Courtois H. M., Tully R. B., Hoffman Y., Pomarède D., Graziani R., Dupuy A., 2017, ApJ, 847, L6
Dálya G. et al., 2022, MNRAS, 514, 1403
Davis T. M., Scrimgeour M. I., 2014, MNRAS, 442, 1117
Dekel A., Bertschinger E., Faber S. M., 1990, ApJ, 364, 349
Dekel A., Eldar A., Kolatt T., Yahil A., Willick J. A., Faber S. M., Courteau S., Burstein D., 1999, ApJ, 522, 1
Desai D. D. et al., 2024, MNRAS, 530, 5016
DESI Collaboration, 2022, AJ, 164, 207
Dhawan S., Jha S. W., Leibundgut B., 2018, A&A, 609, A72
Djorgovski S., Davis M., 1987, ApJ, 313, 59
Draine B. T., 2003, ApJ, 598, 1017
Dressler A., Lynden-Bell D., Burstein D., Davies R. L., Faber S. M., Terlevich R., Wegner G., 1987, ApJ, 313, 42
Duane S., Kennedy A., Pendleton B. J., Roweth D., 1987, Phys. Lett. B, 195, 216
Dye S. et al., 2018, MNRAS, 473, 5113
Eastmond T. S., Abell G. O., 1978, PASP, 90, 367
Ebizuka N. et al., 2011, PASJ, 63, S613
Feldman H. A., Watkins R., 2008, MNRAS, 387, 825
Feldman H. A., Watkins R., Hudson M. J., 2010, MNRAS, 407, 2328
Fitzpatrick E. L., 1999, PASP, 111, 63
Fitzpatrick E. L., Massa D., 1986, ApJ, 307, 286
Fitzpatrick E. L., Massa D., 1988, ApJ, 328, 734
Fitzpatrick E. L., Massa D., 2007, ApJ, 663, 320
Fitzpatrick E. L., Massa D., Gordon K. D., Bohlin R., Clayton G. C., 2019, ApJ, 886, 108
Folatelli G. et al., 2010, AJ, 139, 120
Foley R. J. et al., 2017, MNRAS, 475, 193
Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, PASP, 125, 306
Förster F. et al., 2021, AJ, 161, 242
Freedman W. L. et al., 2020, ApJ, 891, 57
Friedman A. S. et al., 2015, ApJS, 220, 9
Gaia Collaboration, 2016, A&A, 595, A1
Gaia Collaboration, 2018, A&A, 616, A1
Galbany L. et al., 2023, A&A, 679, A95
Ganon G., Hoffman Y., 1993, ApJ, 415, L5
Gao F. et al., 2016, ApJ, 817, 128
Gelman A., Rubin D. B., 1992, Stat. Sci., 7, 457
Gontcharov G. A., 2016, Astrophysics, 59, 548
Gordon K., 2024, dust_extinction, Github, available at: https://github.com/karllark/dust_extinction
Gordon K. D. et al., 2016, ApJ, 826, 104
Graham M. J. et al., 2019, PASP, 131, 078001
Grayling M. et al., 2024, MNRAS, 531, 953
Graziani R., Courtois H. M., Lavaux G., Hoffman Y., Tully R. B., Copin Y., Pomarède D., 2019, MNRAS, 488, 5438
Gupta R. R. et al., 2016, AJ, 152, 154
Guy J., Astier P., Nobili S., Regnault N., Pain R., 2005, A&A, 443, 781
Guy J. et al., 2007, A&A, 466, 11
Guy J. et al., 2010, A&A, 523, A7
Hambly N. C. et al., 2008, MNRAS, 384, 637
Hamuy M., Pinto P. A., 2002, ApJ, 566, L63
Hamuy M., Phillips M. M., Maza J., Suntzeff N. B., Schommer R. A., Aviles R., 1995, AJ, 109, 1
Hamuy M. et al., 2003, Nature, 424, 651
Hamuy M. et al., 2006, PASP, 118, 2
Hart K. et al., 2023, preprint (arXiv:2304.03791)
Hayden B. et al., 2021, ApJ, 912, 87
Helou G., Madore B. F., Schmitz M., Bicay M. D., Wu X., Bennett J., 1991, in Albrecht M. A., Egret D., eds, Astrophysics and Space Science Library, Vol. 171, Databases and On-line Data in Astronomy. Springer-Verlag, Berlin, p. 89
Herrnstein J. R. et al., 1999, Nature, 400, 539
Hewett P. C., Warren S. J., Leggett S. K., Hodgkin S. T., 2006, MNRAS, 367, 454
Hicken M., Garnavich P. M., Prieto J. L., Blondin S., DePoy D. L., Kirshner R. P., Parrent J., 2007, ApJ, 669, L17
Hodapp K. W. et al., 2004, in Amico P., Beletic J. W., Beletic J. E., eds, Scientific Detectors for Astronomy. Springer Netherlands, Dordrecht, p. 501
Hodapp K. W. et al., 2018, in Marshall H. K., Spyromilio J., eds, Proc. SPIE Conf. Ser. Vol. 10700, Ground-based and Airborne Telescopes VII. SPIE, Bellingham, p. 107002Z
Hodgkin S. T., Irwin M. J., Hewett P. C., Warren S. J., 2009, MNRAS, 394, 675
Hoffman M. D., Gelman A., 2011, preprint (arXiv:1111.4246)

Hoffman M. D., Gelman A., 2014, J. Mach. Learn. Res., 15, 1593
Hoffman Y., Pomarède D., Tully R. B., Courtois H. M., 2017, Nat. Astron., 1, 0036
Holm S., 1979, Scand. J. Stat., 6, 65
Holoien T. W. S. et al., 2020, ApJ, 898, 161
Holz D. E., Hughes S. A., 2005, ApJ, 629, 15
Hounsell R. et al., 2018, ApJ, 867, 23
Howell D. A. et al., 2006, Nature, 443, 308
Howlett C., Said K., Lucey J. R., Colless M., Qin F., Lai Y., Tully R. B., Davis T. M., 2022, MNRAS, 515, 953
Hsiao E. Y., Conley A., Howell D. A., Sullivan M., Pritchet C. J., Carlberg R. G., Nugent P. E., Phillips M. M., 2007, ApJ, 663, 1187
Hsiao E. Y. et al., 2019, PASP, 131, 014002
Humason M. L., Mayall N. U., Sandage A. R., 1956, AJ, 61, 97
Jarrett T. H., Chester T., Cutri R., Schneider S., Skrutskie M., Huchra J. P., 2000, AJ, 119, 2498
Jaynes E. T., Crow W. J., 1999. https://api.semanticscholar.org/CorpusID:18814871(Accessed March 2024)
Jeffreys H., 1946, Proc. R. Soc. Lond. Ser. A, 186, 453
Jha S., Riess A. G., Kirshner R. P., 2007, ApJ, 659, 122
Jha S. W. et al., 2019, Supernovae in the Infrared avec Hubble, HST Proposal. Cycle 27, ID. #15889
Johansson J. et al., 2021, ApJ, 923, 237
Jones D. O. et al., 2022, ApJ, 933, 172
Kasen D., 2006, ApJ, 649, 939
Kasen D., Thomas R. C., Nugent P., 2006, ApJ, 651, 366
Kashikawa N. et al., 2002, PASJ, 54, 819
Kashlinsky A., Atrio-Barandela F., Kocevski D., Ebeling H., 2008, ApJ, 686, L49
Kattner S. et al., 2012, PASP, 124, 114
Kelly P. L., Hicken M., Burke D. L., Mandel K. S., Kirshner R. P., 2010, ApJ, 715, 743
Kenworthy W. D. et al., 2021, ApJ, 923, 265
Kessler R. et al., 2009a, PASP, 121, 1028
Kessler R. et al., 2009b, ApJS, 185, 32
Kim A., Goobar A., Perlmutter S., 1996, PASP, 108, 190
Kochanek C. S. et al., 2017, PASP, 129, 104502
Kourkchi E. et al., 2020, ApJ, 902, 145
Krisciunas K. et al., 2009, AJ, 138, 1584
Krisciunas K. et al., 2017, AJ, 154, 211
Kunz M., Bassett B. A., Hlozek R. A., 2007, Phys. Rev. D, 75, 103508
Lampeitl H. et al., 2010, ApJ, 722, 566
Lantz B. et al., 2004, in Mazuray L., Rogers P. J., Wartmann R., eds, Proc. SPIE Conf. Ser. Vol. 5249, Optical Design and Engineering. SPIE, Bellingham, p. 146
Lavaux G., 2016, MNRAS, 457, 172
Lavaux G., Tully R. B., Mohayaee R., Colombi S., 2010, ApJ, 709, 483
Lawrence A. et al., 2007, MNRAS, 379, 1599
Leavitt H. S., Pickering E. C., 1912, Harv. Coll. Obs. Circ., 173, 1
Léget P. F. et al., 2020, A&A, 636, A46
Li W. et al., 2003, PASP, 115, 453
Liu Z.-W., Röpke F. K., Han Z., 2023, Res. Astron. Astrophys., 23, 082001
Lu J. et al., 2023, ApJ, 948, 27
MacKay D., 2003, Information Theory, Inference and Learning Algorithms. Cambridge Univ. Press, Cambridge, available at: https://books.google.co.uk/books?id = AKuMj4PN_EMC
Mahalanobis P. C., 1930, J. Proc. Asiatic Soc. Bengal New Ser., 26, 541
Makarov D., Prugniel P., Terekhova N., Courtois H., Vauglin I., 2014, A&A, 570, A13
Mandel K. S., Wood-Vasey W. M., Friedman A. S., Kirshner R. P., 2009, ApJ, 704, 629
Mandel K. S., Narayan G., Kirshner R. P., 2011, ApJ, 731, 120
Mandel K. S., Thorp S., Narayan G., Friedman A. S., Avelino A., 2022, MNRAS, 510, 3939
Masci F. J. et al., 2019, PASP, 131, 018003
Mazzarella J. M., NED Team, 2007, in Shaw R. A., Hill F., Bell D. J., eds, ASP Conf. Ser. Vol. 376, Astronomical Data Analysis Software and Systems XVI. Astron. Soc. Pac., San Francisco, p. 153
McCall M. L., 2004, AJ, 128, 2144
Möller A., de Boissière T., 2020, MNRAS, 491, 4277
Moreno-Raya M. E., Mollá M., López-Sánchez Á. R., Galbany L., Vílchez J. M., Rosell A. C., Domínguez I., 2016, ApJ, 818, L19
Müller-Bravo T. E. et al., 2022, A&A, 665, A123
Neal R., 2011, in Brooks S., Gelman A., Jones G., Meng X., eds, Handbook of Markov Chain Monte Carlo. Chapman and Hall/CRC, New York, p. 113
Neill J. D. et al., 2009, ApJ, 707, 1449
O'Donnell J. E., 1994, ApJ, 422, 158
Oke J. B., Sandage A., 1968, ApJ, 154, 21
Paturel G., Bottinelli L., Fouque P., Gouguenheim L., 1988, in Murtagh F., Heck A., Benvenuti P., eds, European Southern Observatory Conference and Workshop Proceedings. European Southern Observatory, Garching bei München, p. 435
Paturel G. et al., 1997, A&AS, 124, 109
Paturel G., Petit C., Prugniel P., Theureau G., Rousseau J., Brouty M., Dubois P., Cambrésy L., 2003a, A&A, 412, 45
Paturel G., Theureau G., Bottinelli L., Gouguenheim L., Coudreau-Durand N., Hallet N., Petit C., 2003b, A&A, 412, 57
Peebles P. J. E., 1980, in Wightman A.S., Anderson P.W., eds, The Large-Scale Structure of the Universe. Princeton Univ. Press, Princeton, New Jersey, p. 49
Peebles P. J. E., 1993, in Anderson P.W., Wightman A.S., Treiman S.B., eds, Principles of Physical Cosmology. Princeton Univ. Press, Princeton, New Jersey, p. 320
Perlmutter S. et al., 1997, ApJ, 483, 565
Perlmutter S. et al., 1999, ApJ, 517, 565
Peterson E. R. et al., 2022, ApJ, 938, 112
Peterson E. R. et al., 2023, MNRAS, 522, 2478
Peterson E. R. et al., 2024, A&A, 690, A56
Phillips M. M., 1993, ApJ, 413, L105
Phillips M. M., 2012, PASA, 29, 434
Phillips M. M., Lira P., Suntzeff N. B., Schommer R. A., Hamuy M., Maza J., 1999, AJ, 118, 1766
Phillips M. M. et al., 2019, PASP, 131, 014001
Phillips M. M. et al., 2022, ApJ, 938, 47
Pierel J. D. R. et al., 2018, PASP, 130, 114504
Pierel J. D. R. et al., 2022, ApJ, 939, 11
Pike R. W., Hudson M. J., 2005, ApJ, 635, 11
Planck Collaboration VI, 2020, A&A, 641, A6
Popovic B., Brout D., Kessler R., Scolnic D., 2023, ApJ, 945, 84
Prugniel P., Simien F., 1996, A&A, 309, 749
Pskovskii I. P., 1977, Soviet Ast., 21, 675
Qin Y.-J. et al., 2022, ApJS, 259, 13
Refsdal S., 1964, MNRAS, 128, 307
Riddell A., Hartikainen A., Carter M., 2021, pystan (3.0.0), PyPI
Riess A. G. et al., 1998, AJ, 116, 1009
Rigault M. et al., 2013, A&A, 560, A66
Roman M. et al., 2018, A&A, 615, A68
Rose B. M. et al., 2020, ApJ, 890, 60
Rose B. M. et al., 2021, preprint (arXiv:2111.03081)
Rubin D., 2020, ApJ, 897, 40
Rubin D. et al., 2015, ApJ, 813, 137
Rubin D., Cikota A., Aldering G., Fruchter A., Perlmutter S., Sako M., 2021, PASP, 133, 064001
Rubin D. et al., 2022, ApJS, 263, 1
Rubin D. et al., 2023, preprint (arXiv:2311.12098)
Sakai S., Giovanelli R., Wegner G., 1994, AJ, 108, 33
Sako M. et al., 2018, PASP, 130, 064002
Saunders C. et al., 2018, ApJ, 869, 167
Schlafly E. F., Finkbeiner D. P., 2011, ApJ, 737, 103
Schlegel D. J., Finkbeiner D. P., Davis M., 1998, ApJ, 500, 525
Scolnic D. M. et al., 2018, ApJ, 859, 101
Scolnic D. et al., 2022, ApJ, 938, 113
Shappee B. J. et al., 2014, ApJ, 788, 48
Skrutskie M. F. et al., 2006, AJ, 131, 1163
Smith K. W. et al., 2020, PASP, 132, 085002

Sonnett S., Meech K., Jedicke R., Bus S., Tonry J., Hainaut O., 2013, PASP, 125, 456
Soumagnac M. T. et al., 2024, ApJS, 275, 22
Springob C. M. et al., 2014, MNRAS, 445, 2677
Stan Development Team, 2024, Stan Modeling Language Users Guide and Reference Manual, Version 2.35. Available at:https://mc-stan.org
Stanishev V. et al., 2018, A&A, 615, A45
Stritzinger M. D. et al., 2011, AJ, 142, 156
Sullivan M. et al., 2006, ApJ, 648, 868
Sullivan M. et al., 2010, MNRAS, 406, 782
Taylor G. et al., 2023, MNRAS, 520, 5209
Thoraval S., Boissé P., Duvert G., 1999, A&A, 351, 1051
Thorp S., Mandel K. S., Jones D. O., Ward S. M., Narayan G., 2021, MNRAS, 508, 4310
Tonry J. L., 2011, PASP, 123, 58
Tonry J., Davis M., 1979, AJ, 84, 1511
Tonry J., Schneider D. P., 1988, AJ, 96, 807
Tonry J. L. et al., 2018, PASP, 130, 064505
Tripp R., 1998, A&A, 331, 815
Tucker M. A. et al., 2022, PASP, 134, 124502
Tully R. B., Fisher J. R., 1977, A&A, 500, 105
Tully R. B. et al., 2023, ApJ, 944, 94
Uddin S. A., Mould J., Lidman C., Ruhlmann-Kleider V., Zhang B. R., 2017, ApJ, 848, 56
Uddin S. A. et al., 2020, ApJ, 901, 143
Uddin S. A. et al., 2023, preprint (arXiv:2308.01875)
Valade A., Hoffman Y., Libeskind N. I., Graziani R., 2022, MNRAS, 513, 5148
Vincenzi M. et al., 2023, MNRAS, 518, 1106
Watkins R. et al., 2023, MNRAS, 524, 1885
Wenger M. et al., 2000, A&AS, 143, 9
White D. J., Daw E. J., Dhillon V. S., 2011, Class. Quantum Gravity, 28, 085016
Wiseman P. et al., 2021, MNRAS, 506, 3330
Wood-Vasey W. M. et al., 2008, ApJ, 689, 377
Zaroubi S., 2002, MNRAS, 331, 901
Zaroubi S., Hoffman Y., Fisher K. B., Lahav O., 1995, ApJ, 449, 446
Zaroubi S., Hoffman Y., Dekel A., 1999, ApJ, 520, 413

APPENDIX A: NIR PHOTOMETRY AND LATE-TIME OBSERVATIONS

Late-time observations are not always critical for accurate forward-modelled photometry. When the surface-brightness profile of the host galaxy varies smoothly or has sharp features (e.g. galaxy nucleus, spiral arms) that do not overlap with the supernova, the forward-modelling code is able to cleanly separate the flux from the supernova

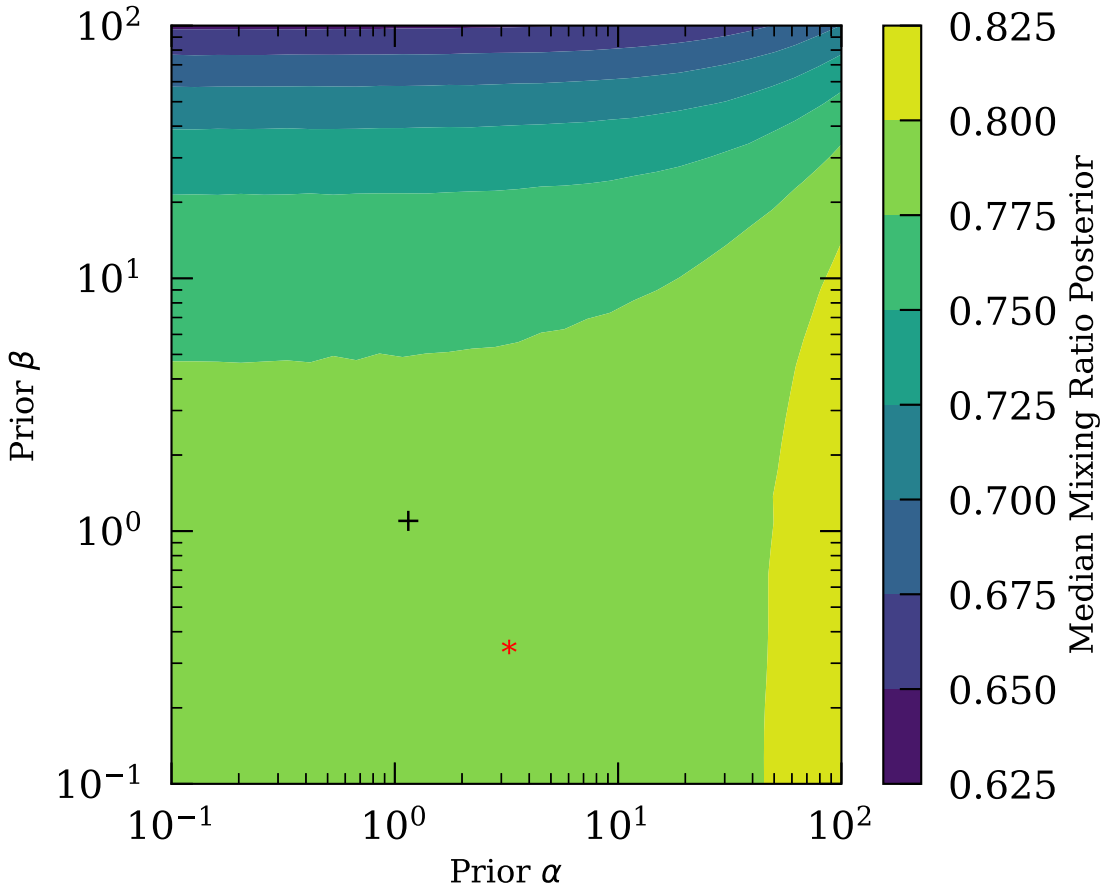

Figure A1. We resample our model using a grid of α and β parameters for the prior beta distribution of the mixing ratio and plot the median values of the mixing ratio posteriors. The set of values used in our analysis is marked with a red star, and the set that produces a flat prior is marked with a black +. The inferred mixing ratio is not sensitive to variations in the initial beta distribution unless extremely strong priors are assumed.

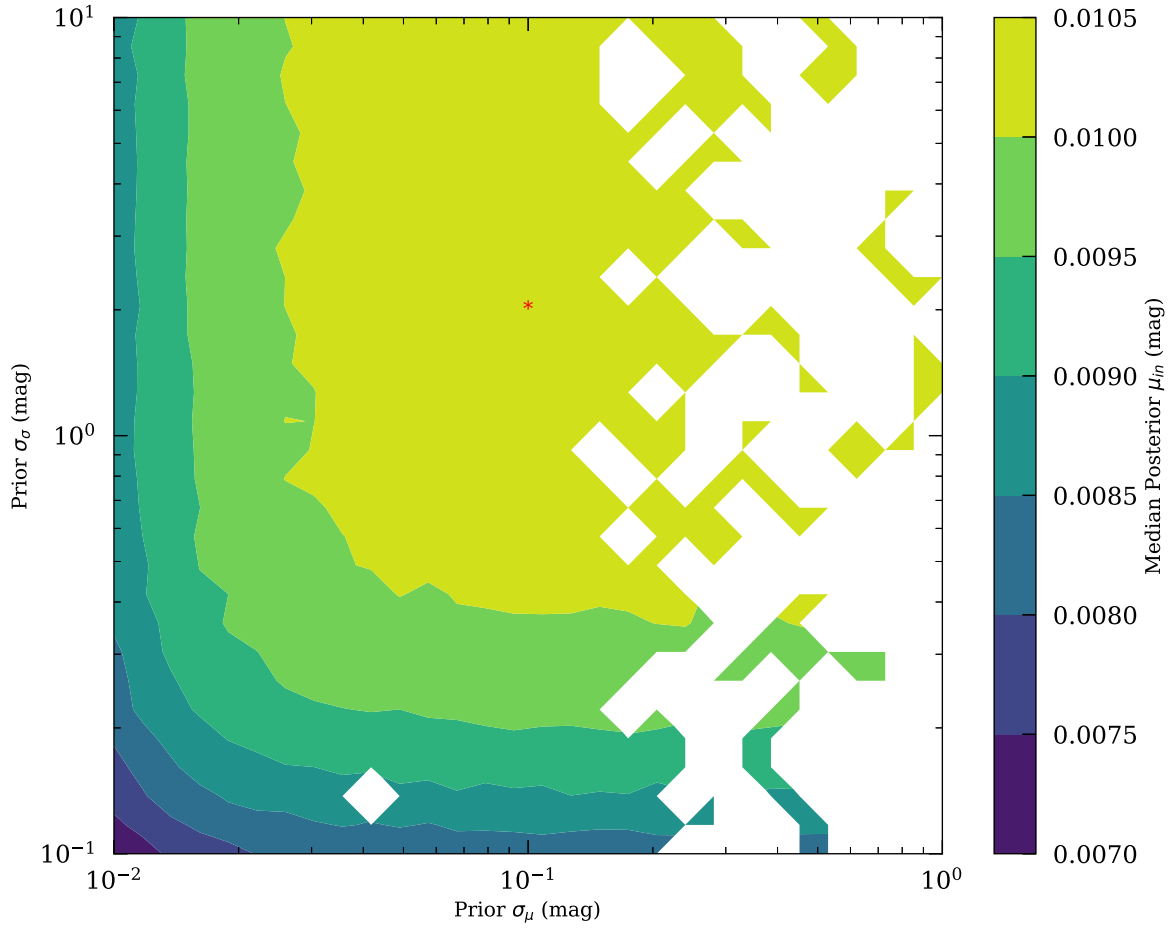

Figure A2. We resample our model using a grid of standard deviation values for the prior distributions of $\mu_{\rm in}$ and σ, and plot the median values of the posterior $\mu_{\rm in}$ samples. The set of values used in our analysis is marked with a red star. Increasing σ_μ leads to the convergence issues we observed when using a flat prior, so we mask the runs where the average Gelman–Rubin $\widehat{R}$ value across the sampled parameters is greater than 1.05.

and galaxy. In these cases, a late-time observation does not provide new information, and the inferred photometry does not change. However, there are supernova-galaxy configurations that critically dependent on late-time observations for accurate modelling. This motivates the use of a mixture model to simultaneously infer the parameters of both populations.

For each SN, we create two ensembles of images; one with and one without late-time observations. This provides two values for the magnitude of the SN at each observed epoch, $m_{\rm ref}$ and m_0, respectively. The differences $\Delta m = m_{\rm ref} - m_0$ should be distributed about 0 mag if $m_{\rm ref}$ and m_0 are normally distributed about the same mean. We use Stan to infer the population means, standard deviations, and the mixing ratio.

Our priors are based on a crude analysis where we consider the subset $|\Delta m| \leq 0.5$ mag and $|\Delta m| > 0.5$ mag, where 0.5 was chosen arbitrarily. The subset near 0 comprises 752 of our 832 observations. Our prior on the mixing ratio is a Beta distribution with $\alpha = 3$ and $\beta = 0.3$ such that the mean expectation value $\frac{\alpha}{\alpha+\beta} \approx 752/832$. The scale of α and β was chosen to create a moderately informed prior. Our priors on the population means (in magnitudes) are $\mu_{\rm in} \sim \mathcal{N}(0\ {\rm mag}, 0.1^2\ {\rm mag}^2)$ for the tightly dispersed population, and $\mu_{\rm out} \sim \mathcal{U}(-\infty, \infty)$ for the late-time sensitive population. Lastly, our priors on the standard deviations (in magnitudes) are $\sigma \sim \mathcal{N}(0\ {\rm mag}, 2^2\ {\rm mag}^2)$, with $0 < \sigma_{\rm in} < \sigma_{\rm out}$.

We fit a Gaussian mixture-model to the photometric differences using Stan (Carpenter et al. 2017) and find 74.0 ± 2.3 per cent of the differences appear tightly dispersed ($\Delta m \sim \mathcal{N}(0.01 \pm 0.004\ {\rm mag}, (0.08 \pm 0.005\ {\rm mag})^2)$), and the remaining 26.0 per cent vary much more dramatically ($\Delta m \sim \mathcal{N}(0.33 \pm 0.050\ {\rm mag}, (0.68 \pm 0.037\ {\rm mag})^2)$). The fraction of targets reliant upon late-time observations for accurate photometry is vastly exaggerated in this analysis because the subsample comprises only targets manually determined to need late-time observations. The critical information is the distribution of the tightly dispersed population, which describes the effect late-time observations have on typical photometric measurements.

We perform sensitivity analyses on the priors used for the mixing ratio and the population parameters of the tightly dispersed group. For testing the former, we tested prior beta distributions parametrized by α and β parameters drawn from a 30 by 30 grid spaced logarithmically between 0.1 and 100. Fig. A1 shows that the recovered posterior estimate is not affected unless extreme values for α and β are assumed, corresponding to an extremely strong prior. More specifically, the recovered mixing ratio is within the joint uncertainty of the value inferred when using our the original priors ($\alpha = 3$, $\beta = 0.3$) unless $\alpha \approx 100$ while $\beta \lesssim 5$ or $\alpha \lesssim 3\beta - 50$ while $\beta \gtrsim 20$. This implies that our inference of the mixing ratio is driven by data rather than the moderately informative prior we used.

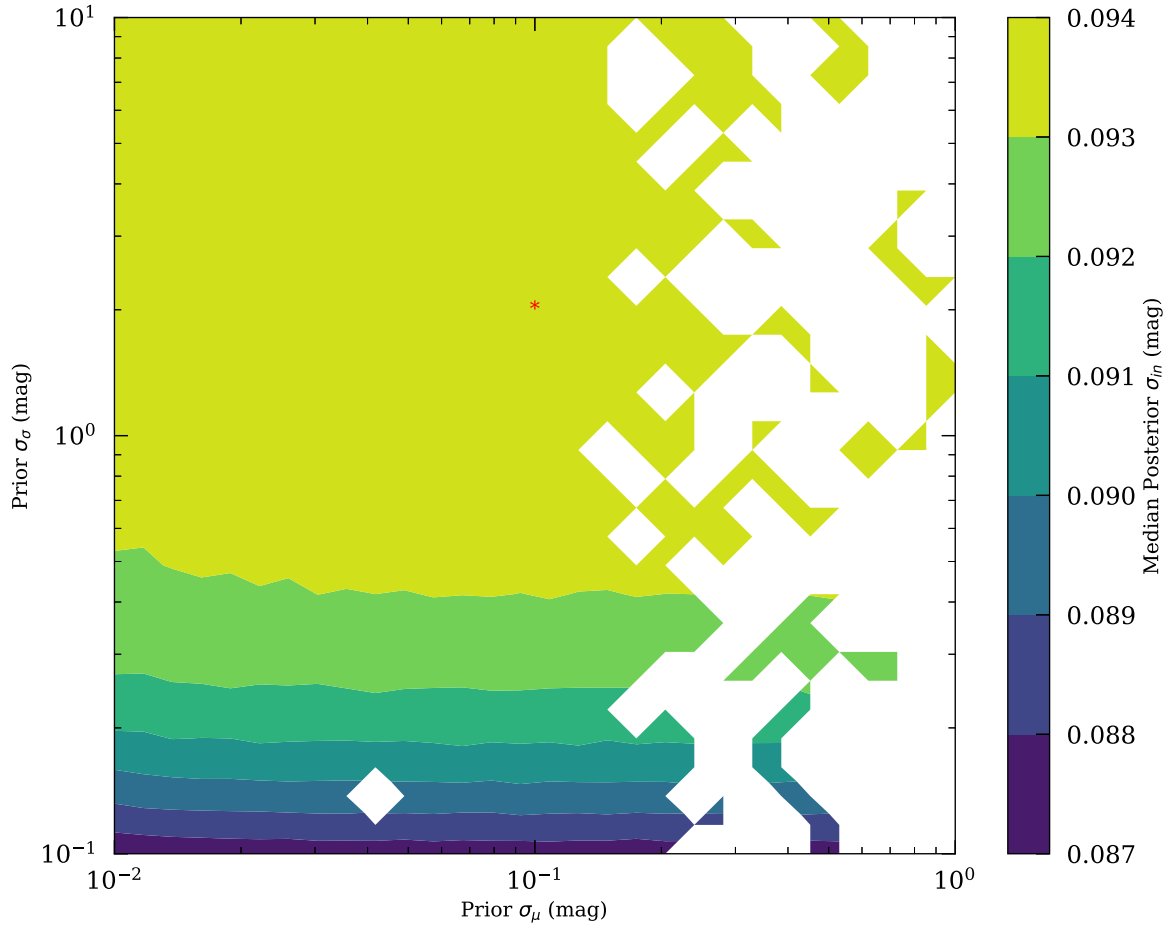

Figure A3. Similar to Fig. A2, but plotting the median values of the posterior $\sigma_{\rm in}$ samples rather than the $\mu_{\rm in}$ samples.

The priors on the population parameters of the tightly dispersed group encode the assumption that the magnitude differences Δm should be 0 mag if the late-time observations are not providing new information to break model degeneracies. Alternatively, one could assume that there is a systematic offset that must be estimated, which could make a flat uniform prior more appropriate. However, this leads to convergence issues when sampling our model with 7 chains and 5000 steps using Stan's no-U-Turn Hamiltonian Monte Carlo sampler. Instead, we examine the sensitivity of the posterior estimates to different levels of variance in the priors $\mu_{\rm in} \sim \mathcal{N}(0\ {\rm mag}, \sigma_\mu^2\ {\rm mag}^2)$ and $\sigma \sim \mathcal{N}(0\ {\rm mag}, \sigma_\sigma^2\ {\rm mag}^2)$ while maintaining $\sigma_{\rm in} < \sigma_{\rm out}$. We sample our model using a 30 by 30 grid of σ_μ and σ_σ values logarithmically spaced between 0.01 and 1 mag and between 0.1 and 10 mag, respectively. As σ_μ increases, the distribution approaches a flat prior, and we observe similar convergence issues as σ_μ approaches 1 mag. We mask the runs where the average Gelman–Rubin $\widehat{R}$ value (Gelman & Rubin 1992) across all parameters is greater than 1.05. Figs A2 and A3 show the medians of the posterior $\mu_{\rm in}$ and $\sigma_{\rm in}$ samples, respectively, as a function of different values for the priors σ_μ and σ_σ. Convergence issues aside, the inferred population parameters describing the tightly dispersed group appear robust against variations in the priors.

A joint sensitivity analysis examining the effects of varying more than two scalar priors at a time is possible, but given the insensitivity demonstrated in the above analyses and the computational expense of increasing the dimensions of the problem, we deem that a joint analysis is not currently necessary.

APPENDIX B: HOST GALAXY IDENTIFICATION ALGORITHM

Before choosing to proceed with manual host galaxy identification, we investigated the Directional Light Radius method (Sullivan et al. 2006; Sako et al. 2018) which normalizes angular separation by the elliptical radius of a galaxy in the direction of the transient. The morphological data came from the NASA/IPAC Extragalactic Database[25] (NED; Helou et al. 1991; Mazzarella & NED Team 2007), the Set of Identifications, Measurements and Bibliography for Astronomical Data[26] (SIMBAD; Wenger et al. 2000), and the GLADE + Galaxy Catalogue[27] (GLADE+; Dálya et al. 2022), which itself consolidates galaxies from the Gravitational Wave Galaxy Catalogue (White, Daw & Dhillon 2011), HyperLEDA[28] (Makarov et al. 2014), and the 2MASS Extended Source Catalogue (Jarrett et al. 2000; Skrutskie et al. 2006). Unfortunately, the heterogeneity and sparsity of the available data presented several failure modes. First, the correct host galaxy could not be identified if it was not included in at least one of the aforementioned data bases or catalogues. Similarly, if a galaxy's morphological data was not available, the DLR method could not be applied. Lastly, if multiple galaxies have categorically

[25] https://ned.ipac.caltech.edu/
[26] https://simbad.u-strasbg.fr/simbad/
[27] https://glade.elte.hu/
[28] http://leda.univ-lyon1.fr/

distinct morphological measurements, either using different filters or different metrics, the DLR method would have been biased.

APPENDIX C: ALTERNATIVE SNOOPY CONFIGURATIONS

In Section 3.1, we describe the decisions that affect the inference of distance moduli given a set of photometry. Those decisions are:

(i) Parametrizing shape with s_{BV} instead of Δm_{15}.
(ii) Performing fits with EBV_model2 and the max_model and not with the EBV_model, max_model2, Rv_model, color_model, SALT_model, or MLCS_model.
(iii) Using the calibration from the full sample of Burns et al. (2018) to describe the correlation between SN parameters and absolute magnitude.
(iv) Using the F19 reddening law.

In this section, we present the quantitative analyses that led to our choices of shape parameter, calibration, and reddening law.

C1 Calibration

The calibration used in the SNPY_Max sample is defined based on our sample, but the calibration used in the SNPY_EBV sample is selected from a list of available calibrations. We reproduce fits of our SNPY_EBV sample using the calibrations from Tables C1 and C2. We also reproduce our fits using Δm_{15} to parametrize shape and the calibrations from Table C3. Each fit is performed assuming the default O94 reddening law (O'Donnell 1994). In the EBV_model2, the calibration values are P_0, P_1, P_2, and the host galaxy R_V, with a fit dispersion of $\sigma_{\rm int}$ mag.

We compare the χ^2 values of the resultant fits to determine which calibration to use. All χ^2 values are calculated as the square of the data-model residual divided by the quadrature sum of the uncertainties in the data and the model. Table C4 lists the median χ^2/DoF in the SNPY_EBV sample fit using the listed calibrations. The table also includes the median $\chi^2_{\rm bp}/N_{\rm bp}$ for each bandpass, where $\chi^2_{\rm bp}$ is the sum of χ^2 values in that bandpass and $N_{\rm bp}$ is the corresponding number of photometric epochs. These are not reduced χ^2 values because the four fitting parameters are not removed from the DoF. As such, χ^2/DoF is not an average of the bandpass specific values $\sum_{\rm bp} \chi^2_{\rm bp}/N_{\rm bp}$, but rather $\sum_{\rm bp} \chi^2_{\rm bp}/(\sum_{\rm bp} N_{\rm bp} - 4)$.

The χ^2/DoF values are consistently lowest in the calibrations sourced from table 1 of Burns et al. (2018). Calibration 8, which was calculated without SNe Ia with $s_{BV} < 0.5$, has the lowest χ^2/DoF of all. However, given that our uncut sample contains SNe Ia with $s_{BV} < 0.5$ we choose to use calibration 6 for all EBV_model2 fits.

The ZTF g- and r-bandpasses and the ATLAS o-bandpass have $\chi^2_{\rm bp}/N_{\rm bp}$ values below 1 across all calibrations, suggesting that the uncertainties in the photometry or in the model may be overestimated. Our decision to combine photometry from each ATLAS quad into a single measurement based on a weighted median (see Section 2.1.2) could produce such an overestimate, but it is not clear why χ^2_c/N_c would be consistently larger than χ^2_o/N_o.

C2 Reddening Law

A thorough review of the effects of dust (e.g. McCall 2004; Gontcharov 2016), even limited to studies of SNe Ia (e.g. Brout & Riess 2023), is beyond the scope of this work, but we will review a few definitions to contextualize the present analysis and our decision.

Extinction is parametrized as a function of wavelength, where the observed flux at wavelength λ is decreased by $A(\lambda)$ mag due to dust. The extinction curve $A(\lambda)$ is roughly inversely proportional to wavelength, meaning the intrinsic colour of an object is reddened. This reddening, or colour excess, is conventionally defined as the differential or selective extinction between the Johnson–Cousins B and V bands ($E(B-V) = A(B) - A(V)$). The total-to-selective extinction parameter R is defined as the ratio between the total extinction at a given wavelength and the colour excess ($R_\lambda = A(\lambda)/E(B-V)$). Both the total and selective extinction are linearly proportional to the amount of dust along the line of sight, which leaves R constant across different amounts of interposing dust. However, the scattering cross-section of dust varies with the shape and size of the dust grains, producing diverse extinction curves and R values. Cardelli et al. (1989) found that the stellar samples from Fitzpatrick & Massa (1986, 1988) permitted normalized extinction curves from the ultraviolet (UV) to the NIR that depend on only one parameter chosen to be R_V. This is an example of a 'reddening law' or 'extinction law' which is a function that uses R_V (or additional parameters, e.g. Gordon et al. 2016) to infer R at a given wavelength or integrated across a given bandpass.

The definition of SNooPy's EBV_model2 (equation 4) involves two terms that describe an R value multiplied by an estimate of colour excess; one to account for Galactic extinction and one for host galaxy extinction. The rescaled SFD dust map provides estimates of Galactic colour excess while the colour excesses of the host galaxies are inferred during the fitting process. The inference of colour excess is largely degenerate with the inference of R values, so the model requires the assumption of a reddening law and R_V values for the Milky Way and the host galaxies.

We reproduce the fits of SNPY_EBV sample using three reddening laws: O94 (O'Donnell 1994), F99 (Fitzpatrick 1999), and F19 (Fitzpatrick et al. 2019). These specific reddening laws are chosen for the following reasons. O94 is the default reddening law used in SNooPy and in the derivation of the Folatelli et al. (2010) calibrations. The analysis of Schlafly & Finkbeiner (2011) found that the reddening measured in SDSS stellar spectra agreed with the rescaled SFD dust map better when using the F99 reddening law than when using the O94 law. The F19 reddening law presents several improvements over the F99 law. The foundational data used to derive the F19 law are spectrophotometric, which allows for normalization based on a single wavelength (4400 and 5500 Å) rather than broadband photometry (B and V bands). Additionally, the new data set spans the gap between the UV and optical regimes with homogeneous coverage whereas other reddening laws extrapolate or interpolate between qualitatively different data sets to cover this gap.

The three reddening laws we examine were defined using data spanning specific ranges in wavelength and R_V. O94 is based on data spanning wavelengths between about 3030 and 9090 Å and R_V values between 2.85 and 5.6. F99 uses spectra from the International Ultraviolet Explorer (IUE; Boggess et al. 1978a, b) and photometry in the Johnson broadband *UBVRIJHKLM* system and intermediate-band Strömgren *uvby* system, effectively spanning wavelengths between 1150 Å and 6 μm. The fit assumes that $A(\lambda)$ approaches 0 as wavelength approaches infinity, but the author cautions that the curve 'should be treated as very approximate' beyond 6 μm. The R_V values of the data range between about 2 and 6. The F19 law uses spectra spanning 1150 to 10000 Å and 2MASS photometry in the JHK-bandpasses which extends the red end to about 2.2μm. This fit also assumes that $A(\lambda)$ approaches 0 as wavelength approaches infinity. The R_V values of the data span a slightly smaller range than the data used to define the F99 law, spanning about 2.5–6. The J-

Table C1. The calibrations available for the EBV_model2 in SNooPy version 2.6.0. All bandpasses are from the natural CSP photometric system. The shape factor used to calculate these values is s_{BV}.

Calibration number	Sample description	Bandpass	P_0 (mag)	P_1 (mag)	P_2 (mag)	R_V	$\sigma_{\rm int}$ (mag)
0	u-band $s_{BV} > 0.5$	B	−19.310(025)	−0.675(110)	3.415(387)	1.518(082)	0.072
		V	−19.264(022)	−0.727(092)	2.161(341)	1.518(082)	0.067
		u	−18.945(035)	−1.077(163)	4.066(533)	1.518(082)	0.148
		g	−19.345(023)	−0.719(102)	2.760(364)	1.518(082)	0.067
		r	−19.146(023)	−0.619(094)	1.968(347)	1.518(082)	0.076
		i	−18.529(024)	−0.541(102)	0.705(382)	1.518(082)	0.092
		Y	−18.532(025)	−0.387(112)	0.232(416)	1.518(082)	0.105
		J	−18.646(026)	−0.719(126)	−0.714(462)	1.518(082)	0.117
		H	−18.470(034)	−0.456(171)	−0.192(622)	1.518(082)	0.172
1	u-band $(m_B - m_V) < 0.3$	B	−19.317(025)	−0.655(108)	3.500(378)	1.746(180)	0.071
		V	−19.278(024)	−0.718(094)	2.249(341)	1.746(180)	0.068
		u	−18.972(031)	−1.028(143)	4.416(418)	1.746(180)	0.124
		g	−19.349(024)	−0.710(100)	2.782(352)	1.746(180)	0.065
		r	−19.162(025)	−0.613(095)	2.049(351)	1.746(180)	0.077
		i	−18.550(026)	−0.530(103)	0.848(384)	1.746(180)	0.092
		Y	−18.547(026)	−0.378(114)	0.320(427)	1.746(180)	0.106
		J	−18.665(028)	−0.697(129)	−0.538(475)	1.746(180)	0.119
		H	−18.490(036)	−0.431(176)	−0.005(637)	1.746(180)	0.175
2	u-band all objects	B	−19.325(022)	−0.676(103)	3.804(262)	1.531(081)	0.069
		V	−19.277(020)	−0.732(088)	2.422(222)	1.531(081)	0.065
		u	−18.969(032)	−1.123(142)	4.742(253)	1.531(081)	0.153
		g	−19.359(021)	−0.719(095)	3.098(243)	1.531(081)	0.065
		r	−19.154(020)	−0.637(089)	2.048(221)	1.531(081)	0.074
		i	−18.555(021)	−0.510(099)	1.378(240)	1.531(081)	0.092
		Y	−18.560(022)	−0.350(112)	0.975(269)	1.531(081)	0.107
		J	−18.686(026)	−0.639(143)	0.460(313)	1.531(081)	0.140
		H	−18.499(030)	−0.416(165)	0.637(354)	1.531(081)	0.168
3	No u-band $s_{BV} > 0.5$	B	−19.271(024)	−0.753(116)	2.928(411)	1.699(089)	0.078
		V	−19.246(021)	−0.791(093)	1.867(343)	1.699(089)	0.066
		g	−19.315(022)	−0.785(105)	2.369(375)	1.699(089)	0.070
		r	−19.134(021)	−0.678(094)	1.728(346)	1.699(089)	0.075
		i	−18.518(023)	−0.599(100)	0.476(374)	1.699(089)	0.090
		Y	−18.528(023)	−0.415(108)	0.123(398)	1.699(089)	0.102
		J	−18.638(025)	−0.743(122)	−0.827(445)	1.699(089)	0.112
		H	−18.462(032)	−0.513(168)	−0.374(606)	1.699(089)	0.169
4	No u-band $(m_B - m_V) < 0.3$	B	−19.276(025)	−0.730(116)	3.053(412)	1.716(202)	0.078
		V	−19.247(022)	−0.780(095)	1.909(351)	1.716(202)	0.068
		g	−19.311(023)	−0.782(106)	2.363(383)	1.716(202)	0.070
		r	−19.134(023)	−0.672(095)	1.744(354)	1.716(202)	0.076
		i	−18.524(024)	−0.589(101)	0.557(381)	1.716(202)	0.092
		Y	−18.529(025)	−0.409(111)	0.146(418)	1.716(202)	0.104
		J	−18.646(026)	−0.728(126)	−0.720(465)	1.716(202)	0.115
		H	−18.477(035)	−0.489(174)	−0.188(644)	1.716(202)	0.175
5	No u-band all objects	B	−19.304(022)	−0.682(113)	3.916(321)	1.729(089)	0.077
		V	−19.270(019)	−0.751(092)	2.460(254)	1.729(089)	0.065
		g	−19.344(021)	−0.727(102)	3.166(292)	1.729(089)	0.067
		r	−19.154(019)	−0.655(092)	2.155(238)	1.729(089)	0.074
		i	−18.553(020)	−0.536(099)	1.409(238)	1.729(089)	0.089
		Y	−18.561(021)	−0.360(107)	1.024(244)	1.729(089)	0.103
		J	−18.687(025)	−0.633(139)	0.639(296)	1.729(089)	0.139
		H	−18.495(028)	−0.456(161)	0.594(331)	1.729(089)	0.164

band data used in our project is redder than the data used to calculate the O94 law, and the host galaxy R_V values in calibration 6 (1.1–1.9) are all below the minimum R_V values used to define the O94, F99, and F19 laws. The low R_V values are likely due to the conflation of intrinsic SN Ia colour and the effects of host galaxy extinction in the EBV_model2. We edit the allowed wavelength and R_V ranges in the `dust_extinction` package (Gordon 2024) to allow for the extrapolations we require.

The χ^2/DoF values presented in Table C5 are similar across the three reddening laws, which implies that the choice of reddening law does not significantly impact the accuracy of the EBV_model2. The fits using the F99 law have the lowest χ^2/DoF value, but this is

Table C2. Similar to Table C1, except the values are those presented in table 1 of Burns et al. (2018).

Calibration number	Sample description	Bandpass	P_0 (mag)	P_1 (mag)	P_2 (mag)	R_V	$\sigma_{\rm int}$ (mag)
6	Full sample	B	−19.182(062)	−0.89(11)	−0.02(30)	1.65(08)	0.13
		V	−19.181(061)	−0.89(11)	−0.02(30)	1.65(08)	0.13
		u	−18.818(097)	−1.28(17)	0.32(44)	1.13(52)	0.22
		g	−19.229(084)	−0.90(11)	−0.13(31)	1.57(09)	0.13
		r	−19.099(059)	−0.74(10)	0.38(27)	1.78(08)	0.12
		i	−18.523(059)	−0.48(10)	0.41(27)	1.85(09)	0.12
		Y	−18.517(077)	−0.07(11)	1.19(30)	1.34(21)	0.12
		J	−18.633(062)	−0.37(12)	0.61(32)	1.27(36)	0.11
		H	−18.431(062)	−0.05(12)	1.18(31)	1.28(57)	0.11
0	$(m_B - m_V) < 0.5$	B	−19.161(062)	−0.94(11)	−0.36(43)	1.54(14)	0.13
		V	−19.161(061)	−0.94(11)	−0.37(44)	1.54(14)	0.13
		u	−18.793(095)	−1.35(18)	−0.47(64)	1.12(51)	0.21
		g	−19.206(082)	−0.97(11)	−0.57(43)	1.48(14)	0.13
		r	−19.081(060)	−0.77(10)	0.12(41)	1.67(13)	0.13
		i	−18.501(060)	−0.52(10)	−0.10(41)	1.79(17)	0.13
		Y	−18.497(076)	−0.10(11)	0.34(41)	1.69(35)	0.12
		J	−18.601(062)	−0.43(11)	−0.42(45)	1.51(58)	0.11
		H	−18.400(062)	−0.10(12)	0.17(47)	1.33(85)	0.11
0	$s_{BV} > 0.5$	B	−19.159(062)	−0.93(12)	−0.61(43)	1.64(09)	0.13
		V	−19.159(061)	−0.94(11)	−0.62(43)	1.64(09)	0.13
		u	−18.790(097)	−1.32(18)	−0.35(70)	1.10(45)	0.22
		g	−19.204(084)	−0.96(12)	−0.80(43)	1.56(09)	0.13
		r	−19.081(060)	−0.77(11)	−0.05(39)	1.76(08)	0.12
		i	−18.499(059)	−0.52(10)	−0.21(38)	1.82(10)	0.12
		Y	−18.480(076)	−0.11(11)	0.32(42)	1.18(22)	0.11
		J	−18.593(060)	−0.44(12)	−0.35(45)	1.02(36)	0.11
		H	−18.394(061)	−0.10(12)	0.13(47)	0.82(52)	0.11
0	$s_{BV} > 0.5 (m_B - m_V) < 0.5$	B	−19.162(061)	−0.94(11)	−0.30(46)	1.55(14)	0.13
		V	−19.163(061)	−0.94(11)	−0.31(46)	1.55(14)	0.13
		u	−18.796(095)	−1.35(17)	−0.42(69)	1.12(51)	0.21
		g	−19.207(083)	−0.96(11)	−0.53(46)	1.48(15)	0.13
		r	−19.083(060)	−0.77(10)	0.17(42)	1.68(13)	0.13
		i	−18.501(061)	−0.52(10)	−0.10(43)	1.78(17)	0.13
		Y	−18.489(075)	−0.10(10)	0.15(42)	1.59(35)	0.12
		J	−18.598(063)	−0.43(12)	−0.48(47)	1.48(57)	0.11
		H	−18.395(061)	−0.11(12)	0.03(48)	1.24(84)	0.11

Table C3. Similar to Table C1, except the shape parameter used is Δm_{15}.

Calibration number	Sample description	Bandpass	P_0 (mag)	P_1 (mag)	P_2 (mag)	R_V	$\sigma_{\rm int}$ (mag)
10	u-band $s_{BV} > 0.5$	B	−19.360(030)	0.433(090)	2.356(293)	1.533(084)	0.076
		V	−19.282(027)	0.540(080)	1.253(260)	1.533(084)	0.070
		u	−18.979(042)	0.751(136)	2.526(433)	1.533(084)	0.150
		g	−19.380(028)	0.473(084)	1.958(275)	1.533(084)	0.069
		r	−19.171(027)	0.489(081)	1.187(262)	1.533(084)	0.077
		i	−18.548(027)	0.379(088)	0.383(278)	1.533(084)	0.093
		Y	−18.549(027)	0.082(091)	0.372(282)	1.533(084)	0.095
		J	−18.662(028)	0.175(100)	0.152(311)	1.533(084)	0.107
		H	−18.475(037)	0.143(136)	0.035(418)	1.533(084)	0.165
11	u-band $(m_B - m_V) < 0.3$	B	−19.369(030)	0.419(089)	2.425(297)	1.589(155)	0.076
		V	−19.292(028)	0.528(080)	1.322(265)	1.589(155)	0.070
		u	−19.016(039)	0.717(120)	2.788(393)	1.589(155)	0.126
		g	−19.384(029)	0.464(082)	1.992(274)	1.589(155)	0.066
		r	−19.182(029)	0.478(081)	1.256(266)	1.589(155)	0.079
		i	−18.568(029)	0.362(088)	0.511(283)	1.589(155)	0.093
		Y	−18.560(028)	0.070(091)	0.446(285)	1.589(155)	0.094
		J	−18.679(029)	0.156(100)	0.274(312)	1.589(155)	0.108
		H	−18.494(038)	0.122(136)	0.176(424)	1.589(155)	0.165

Table C3 – *continued*

Calibration number	Sample description	Bandpass	P_0 (mag)	P_1 (mag)	P_2 (mag)	R_V	$\sigma_{\rm int}$ (mag)
12	u-band all objects	B	−19.394(029)	0.370(089)	2.820(267)	1.593(084)	0.074
		V	−19.320(025)	0.472(077)	1.736(214)	1.593(084)	0.067
		u	−19.056(042)	0.579(144)	3.793(373)	1.593(084)	0.173
		g	−19.409(027)	0.424(082)	2.306(243)	1.593(084)	0.067
		r	−19.205(024)	0.430(078)	1.591(204)	1.593(084)	0.075
		i	−18.598(025)	0.279(085)	1.070(207)	1.593(084)	0.093
		Y	−18.594(024)	−0.010(088)	1.013(207)	1.593(084)	0.095
		J	−18.707(027)	0.087(103)	0.825(228)	1.593(084)	0.123
		H	−18.509(032)	0.066(128)	0.559(275)	1.593(084)	0.161
13	No u-band $s_{BV} > 0.5$	B	−19.281(030)	0.516(095)	1.768(306)	1.727(097)	0.086
		V	−19.235(026)	0.613(081)	0.840(259)	1.727(097)	0.071
		g	−19.314(028)	0.547(087)	1.454(282)	1.727(097)	0.074
		r	−19.134(026)	0.558(081)	0.833(258)	1.727(097)	0.077
		i	−18.516(026)	0.445(088)	0.064(276)	1.727(097)	0.092
		Y	−18.529(026)	0.123(090)	0.180(277)	1.727(097)	0.095
		J	−18.643(027)	0.202(098)	−0.000(301)	1.727(097)	0.107
		H	−18.453(036)	0.211(136)	−0.212(418)	1.727(097)	0.168
14	No u-band $(m_B - m_V) < 0.3$	B	−19.287(031)	0.512(096)	1.824(314)	1.544(181)	0.087
		V	−19.228(027)	0.610(081)	0.840(265)	1.544(181)	0.073
		g	−19.308(028)	0.554(088)	1.430(286)	1.544(181)	0.075
		r	−19.125(027)	0.552(081)	0.826(262)	1.544(181)	0.078
		i	−18.516(028)	0.435(087)	0.112(278)	1.544(181)	0.092
		Y	−18.523(027)	0.121(090)	0.173(282)	1.544(181)	0.095
		J	−18.644(028)	0.205(098)	0.016(307)	1.544(181)	0.108
		H	−18.465(038)	0.207(138)	−0.129(434)	1.544(181)	0.169
15	No u-band all objects	B	−19.335(028)	0.410(092)	2.559(273)	1.824(096)	0.084
		V	−19.288(023)	0.515(076)	1.549(209)	1.824(096)	0.067
		g	−19.360(025)	0.463(083)	2.065(247)	1.824(096)	0.070
		r	−19.182(022)	0.469(076)	1.447(190)	1.824(096)	0.074
		i	−18.582(022)	0.308(082)	1.013(184)	1.824(096)	0.093
		Y	−18.586(022)	0.002(084)	1.017(178)	1.824(096)	0.096
		J	−18.704(025)	0.070(100)	0.961(203)	1.824(096)	0.125
		H	−18.512(031)	0.072(128)	0.737(250)	1.824(096)	0.171

Table C4. Various χ^2 metrics are presented for the SNPY_EBV sample fit with the 16 calibrations listed in Tables C1, C2, and C3. After the χ^2/DoF column, each column lists the sum of χ^2 values in the subscripted bandpass divided by the corresponding number of photometric epochs. Each listed value is the median across all SNe Ia in the SNPY_EBV sample fit with the calibration in the first column. The c- and o-bandpasses are from ATLAS and the g- and r-bandpasses are from ZTF.

Calibration	χ^2/DoF	χ^2_g/N_g	χ^2_c/N_c	χ^2_r/N_r	χ^2_o/N_o	χ^2_J/N_J
0	1.051	0.491	1.021	0.782	0.876	1.234
1	1.036	0.491	1.013	0.760	0.866	1.189
2	1.045	0.471	0.980	0.777	0.858	1.233
3	1.015	0.489	1.010	0.753	0.867	1.126
4	1.012	0.491	1.013	0.746	0.865	1.148
5	1.011	0.470	0.978	0.731	0.848	1.108
6	0.955	0.484	1.105	0.636	0.846	1.052
7	0.953	0.467	0.963	0.667	0.866	1.002
8	0.937	0.460	0.974	0.637	0.867	1.032
9	0.947	0.458	0.963	0.657	0.874	0.997
10	1.009	0.570	1.115	0.670	0.892	1.181
11	1.002	0.576	1.102	0.657	0.891	1.157
12	0.991	0.559	1.115	0.648	0.891	1.142
13	0.996	0.548	1.085	0.592	0.887	1.123
14	1.008	0.567	1.108	0.613	0.888	1.162
15	0.968	0.555	1.101	0.588	0.871	1.115

Table C5. Similar to Table C4, but presenting various χ^2 metrics for the O94, F99, and F19 reddening laws rather than the calibrations.

Reddening law	χ^2/DoF	χ^2_g/N_g	χ^2_c/N_c	χ^2_r/N_r	χ^2_o/N_o	χ^2_J/N_J
O94	0.955	0.484	1.105	0.636	0.846	1.052
F99	0.947	0.464	1.139	0.609	0.844	1.135
F19	0.961	0.493	1.089	0.633	0.837	1.004

not the case over all bandpasses. The F99 law produces the lowest $\chi^2_{\rm bp}/N_{\rm bp}$ values for the ZTF g and r bands, while simultaneously producing the highest values the ATLAS c band and the J band. The ATLAS, ZTF, and Hawai'i Supernova Flows observing strategies produce more epochs of photometry in the former two bandpasses than the latter two, which accounts for the lower χ^2/DoF in F99 via weighting. However, we do not presently understand why the $\chi^2_{\rm bp}/N_{\rm bp}$ values are so much lower than 1 for the gro-bands, and note that the F19 law produces the lowest values for the coJ-bands. Even though the median χ^2/DoF value is highest in the fits assuming the F19 law, it is still indicative of a good set of fits. Thus, we choose to adopt the F19 reddening law for our EBV_model2 fits.

APPENDIX D: ALL-SKY SURVEY INDEPENDENCE

When multiple observers record photometric time series of a single source, the correlation between the resultant light curves is based primarily on the time-evolution of the astrophysical source, but is also affected by correlated observational or instrumental effects. For example, the orbital motion of the Earth Doppler shifts the SED of any observed target, leading to slight annual correlations for non-flat SEDs. Unmodelled variability in reference stars used by multiple surveys could lead to common errors in zero-point calibration. We assume independence between the ATLAS, ASAS-SN, and ZTF photometry in the sense that we consider any correlated observational or instrumental effects as insignificant.

To justify this assumption, we analyse forced photometry of CALSPEC stars in the footprint of all surveys and fainter than 15 mag in V to avoid saturation. To account for proper motion, we access ASAS-SN lightcurves from the ASAS-SN Sky Patrol[29] (Shappee et al. 2014; Hart et al. 2023) and ZTF lightcurves from the ZTF DR 21 archive (Masci et al. 2019) hosted at the NASA/IPAC Infrared Science Archive.[30] The CALSPEC targets in NGC 6681 are excluded due to crowding. The list of CALSPEC stars and their synthetic magnitudes in the bandpasses of the three surveys are presented in Table D1 CALSPEC stars demonstrate minimal stellar variability (Rubin et al. 2022), which we use to exclude astrophysical time-evolution as a source of correlation between light curves. As mentioned in Section 2.1.2, we combine ATLAS data from the same nights with a weighted median.

For the 10 bandpass pairs possible using ATLAS c, ATLAS o, ASAS-SN g, ZTF g, and ZTF r, we identify observations where a given star was observed in both bandpasses within 12 h. This makes our analysis sensitive to correlated effects on characteristic time-scales greater than half a day, but insensitive to effects that vary on shorter time-scales. We calculate observed-synthetic magnitude residuals and normalize by the recorded uncertainties to produce z-scores. We do not include pairs where either observation is in the bottom or top 5 per cent of z-scores for that bandpass and star. We concatenate the rest of the z-score pairs into equal length sets for both bandpasses. Table D2 shows the calculated Pearson correlation coefficients between those sets.

All bandpass pairs besides ZTF g and ZTF r are consistent with a correlation coefficient of 0 at the 95 per cent level. The correlation between the two ZTF bandpasses implies there is at least one observational or instrumental effect that applies to both sets of observations, but the magnitude of such a correlation is small at about 0.05. This is to be expected since both sets of observations come from the Palomar 48 inch Schmidt telescope. Perhaps more surprising is that the ATLAS c and o data do not appear correlated. This could be due to the distribution of observations across multiple sites (Haleakala and Maunaloa in Hawai'i, El Sauce Observatory in Chile, Sutherland Observing Station in South Africa), or due to the low number of observations in both bandpasses.

[29] http://asas-sn.ifa.hawaii.edu/skypatrol/

[30] https://irsa.ipac.caltech.edu/Missions/ztf.html

Table D1. This table lists the CALSPEC stars used in our analysis, which were selected to be fainter than 16 mag in V to avoid saturation issues. We do not include the stars in NGC 6681 due to crowding. The columns list the synthetic magnitudes in the five bandpasses examined.

Name	ATLAS *c* (mag)	ATLAS *o* (mag)	ASAS-SN *g* (mag)	ZTF *g* (mag)	ZTF *r* (mag)
C26202	16.55	16.32	16.73	16.69	16.34
HS2027 + 0651	16.56	16.96	16.39	16.44	16.89
NGC2506 – G31	17.99	17.66	18.24	18.18	17.69
SDSS132811	17.05	17.33	17.01	16.99	17.28
SDSSJ151421	15.81	16.23	15.66	15.68	16.16
SF1615 + 001A	16.82	16.48	17.07	17.01	16.52
SNAP-2	16.29	15.98	16.53	16.47	16.01
VB8	17.04	14.2	17.85	17.57	15.58
WD0947 + 857	15.66	16.14	15.47	15.51	16.06
WD1026 + 453	16.03	16.5	15.85	15.88	16.43
WD1657 + 343	16.35	16.83	16.15	16.2	16.75

Table D2. We present correlation measurements between the 10 bandpass pairs between the five survey bandpasses. For each pair we assemble all N observations of common targets performed on the same date and calculate the Pearson correlation coefficient r. We present the 95 per cent confidence intervals (CI 95 per cent) and p-values, finding that all combinations but ZTF g and ZTF r are consistent with no correlation and are not significant at the $p<0.05$ level.

Bandpass 1	Bandpass 2	N	r	CI 95 per cent	p-value
ATLAS *c*	ATLAS *o*	53	0.011	(−0.26, 0.28)	0.940
ATLAS *c*	ASAS-SN *g*	345	−0.057	(−0.16, 0.05)	0.291
ATLAS *c*	ZTF *g*	235	−0.086	(−0.21, 0.04)	0.188
ATLAS *c*	ZTF *r*	244	−0.031	(−0.16, 0.09)	0.628
ATLAS *o*	ASAS-SN *g*	922	−0.012	(−0.08, 0.05)	0.719
ATLAS *o*	ZTF *g*	629	0.020	(−0.06, 0.1)	0.613
ATLAS *o*	ZTF *r*	686	0.035	(−0.04, 0.11)	0.359
ASAS-SN *g*	ZTF *g*	804	0.030	(−0.04, 0.1)	0.396
ASAS-SN *g*	ZTF *r*	808	−0.012	(−0.08, 0.06)	0.741
ZTF *g*	ZTF *r*	1669	0.053	(0.01, 0.1)	0.030

This paper has been typeset from a TEX/LATEX file prepared by the author.