Testing the Consistency of Dust Laws in SN Ia Host Galaxies: A BayeSN Examination of Foundation DR1

Stephen Thorp,1* Kaisey S. Mandel,1,2 David O. Jones,3† Sam M. Ward,1 and Gautham Narayan4,5
1Institute of Astronomy and Kavli Institute for Cosmology, Madingley Road, Cambridge, CB3 0HA, UK
2Statistical Laboratory, DAMTP, University of Cambridge, Wilberforce Road, Cambridge, CB3 0WA, UK
3Department of Astronomy and Astrophysics, University of California, Santa Cruz, CA 95064, USA
4University of Illinois at Urbana-Champaign, 1003 W. Green St., IL 61801, USA
5Center for Astrophysical Surveys, National Center for Supercomputing Applications, Urbana, IL 61801, USA

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

We apply BayeSN, our new hierarchical Bayesian model for the SEDs of Type Ia supernovae (SNe Ia), to analyse the g,r,i,z light curves of 157 nearby SNe Ia (0.015 < z < 0.08) from the public Foundation DR1 dataset. We train a new version of BayeSN, continuous from 0.35–0.95 μm, which we use to model the properties of SNe Ia in the rest-frame z-band, study the properties of dust in their host galaxies, and construct a Hubble diagram of SN Ia distances determined from full g,r,i,z light curves. Our g,r,i,z Hubble diagram has a low total RMS of 0.13 mag using BayeSN, compared to 0.16 mag using SALT2. Additionally, we test the consistency of the dust law R_v between low- and high-mass host galaxies by using our model to fit the full time- and wavelength-dependent SEDs of SNe Ia up to moderate reddening (peak apparent B – V ≤ 0.3). Splitting the population at the median host mass, we find R_v = 2.84 ± 0.31 in low-mass hosts, and R_v = 2.58 ± 0.23 in high-mass hosts, both consistent with the global value of R_v = 2.61 ± 0.21 that we estimate for the full sample. For all choices of mass split we consider, R_v is consistent across the step within ± 1.2σ. Modelling population distributions of dust laws in low- and high-mass hosts, we find that both subsamples are highly consistent with the full sample’s population mean μ(R_v) = 2.70 ± 0.25 with a 95% upper bound on the population σ(R_v) < 0.61. The R_v population means are consistent within ± 1.2σ. We find that simultaneous fitting of host-mass-dependent dust properties within our hierarchical model does not account for the conventional mass step.

Key words: supernovae: general – distance scale – dust, extinction – methods: statistical

1 INTRODUCTION

From the landmark discovery of the Universe’s accelerating expansion (Riess et al. 1998; Perlmutter et al. 1999), to state-of-the-art efforts measuring the Hubble constant (Dhawan et al. 2018; Burns et al. 2018; Riess et al. 2019) and dark energy equation-of-state parameter (Scolnic et al. 2018; Abbott et al. 2019), Type Ia supernovae (SNe Ia) have been a key pillar in our understanding of cosmology. The ongoing development of a homogeneous low redshift SNAP Ia sample (Foley et al. 2018; Jones et al. 2019, 2021), as well as future high redshift data from the Nancy Grace Roman Space Telescope (Spergel et al. 2015; Hounsell et al. 2018) and the Vera C. Rubin Observatory’s Legacy Survey of Space and Time (Ivezić et al. 2019), will enable a new era of precision cosmology. To make the most of the incoming data, however, SN Ia models must be improved, and subtleties such as SN–host correlations (e.g. Kelly et al. 2010; Sullivan et al. 2010) must be better understood. In this paper, we apply the state-of-the-art hierarchical model, BayeSN (Mandel et al. 2020), to analyse current low-redshift data from the Foundation Supernova Survey (Foley et al. 2018; Jones et al. 2019), and to explore the relation between the dust properties and stellar masses of SN Ia host galaxies.

The host galaxy ‘mass step’ – a tendency for post-standardisation SN Ia magnitudes to be brighter in more massive galaxies (Kelly et al. 2010; Sullivan et al. 2010; Smith et al. 2020) – is well established, and routinely corrected for, but its cause remains uncertain. Another question is the proper interpretation of the SN Ia apparent colour–luminosity relation, and the value(s) of the R_v parametrizing the extinction and reddening law of host galaxy dust affecting the light of SNe Ia. For some highly reddened SNe Ia with peak apparent (B – V) > 0.3 (e.g. SN 2003hg, SN 2002cv, Elias-Rosa et al. 2006, 2008; SN 2006X, Wang et al. 2008; SN 2014J, Amanullah et al. 2014; Hoang 2017), unusually low individual R_v ≈ 1.5 – 2 values have been inferred – far from the mean Galactic value of ≈ 3.1. However, such reddened SNe are typically excluded from cosmological analyses, due to a standard colour cut (B – V) ≤ 0.3. Still, there is ongoing debate over whether ‘normal’ samples of SNe Ia with (B – V) ≤ 0.3 are consistent with a low R_v ≈ 1 – 2 (e.g. Nobili & Goobar 2008; Stanishev et al. 2018), a higher R_v of ≈ 2.5 – 3 (e.g. Folatelli et al. 2010; Chotard et al. 2011; Foley & Kasen 2011; Mandel et al. 2011, 2017, 2020; Burns et al. 2014; Sasdelli et al. 2016; Léget et al. 2020; Arima et al. 2021), or a wider range of values (e.g. Amanullah et al. 2015). These studies are made difficult by the limited range of reddening within the (B – V) ≤ 0.3 colour cut, and limited wavelength range of conventional optical analyses. The confounding of intrinsic variation with extrinsic dust effects in the apparent colours of
SNe Ia also complicates the interpretation of the data (e.g. Freedman et al. 2009). For example, Mandel et al. (2017) showed that the convolution of an intrinsic SN colour-luminosity trend with dust reddening-extinction effects generically results in a nonlinear effective mean apparent colour-luminosity curve, and the conventional linear fit approximates an average of the physically-distinct intrinsic and dust slopes.

Recently, it was proposed by Brout & Scolnic (2021) that a difference in the dust properties of low- and high-mass host galaxies is the root cause of the mass step. They use survey simulations with an input $B-V$ colour distribution consisting of an intrinsic and extrinsic component, the latter obeying a colour–luminosity relation parameterized by $R_V + 1$. A population distribution of these $R_V$ values is included, and a version of their model where this is permitted to differ between low and high host galaxy stellar masses is also considered. They iterate over survey simulations, attempting to find the values of the input population parameters for which the simulation outputs best agree with observed supernova data. In particular, for each simulation they compare its colour distribution, its Hubble residual bias and scatter in different colour bins, and its overall fitted colour–luminosity trend, to the equivalent quantities obtained from SALT2 (Guy et al. 2007, 2010; Betoule et al. 2014) fits to a compilation of supernova light curves. At low redshift, they used data from the CfA Supernova Program (CfA; Hicken et al. 2009a, 2012), Carnegie Supernova Project (CSP; Contreras et al. 2010; Folatelli et al. 2010; Stritzinger et al. 2011; Krisciunas et al. 2017) and Foundation Supernova Survey (Foley et al. 2018; Jones et al. 2019), with high redshift data coming from the Pan-STARRS-1 Medium Deep Survey (MDS; Rest et al. 2014; Scolnic et al. 2018), Supernova Legacy Survey (SNLS; Astier et al. 2006; Betoule et al. 2014), SDSS-II Supernova Survey (Frieman et al. 2008; Sako et al. 2011, 2018) and the first three years of Dark Energy Survey data (DESyr3, Dark Energy Survey Collaboration et al. 2016; Brout et al. 2019), for a total of 1445 SNe. Comparing their simulations with SALT2 parameter fits to these data, they infer an $R_V$ population distribution peaking at about $2.75 \pm 0.35$ in low-mass hosts ($M_\star < 10^{10} M_\odot$), and $1.50 \pm 0.25$ in high-mass hosts, with wide standard deviations of $1.3 \pm 0.2$ in both mass bins. They conclude that including these different $R_V$ distributions eliminates the need for a 0.06 mag mass step which they otherwise observed in the data (see also Popovic et al. 2021).

However, the observation of a mass step in near-infrared (NIR) data, where sensitivity to dust should be reduced, disfavours a difference in host dust properties as an explanation of the step. Recently, Ponder et al. (2020) used separate SNooPy (Burns et al. 2011) fits to optical and $H$-band light curves of SNe Ia to explore how correlations in host mass change between the optical and NIR. They analysed the $H$-band (optical) light curves of 99 (71) supernovae from the CfA/CFAIR2 (Wood-Vasey et al. 2008; Friedman et al. 2015), CSP, SweetSpot (Weyant et al. 2018), and elsewhere in the literature (Jha et al. 1999; Hernandez et al. 2000; Krisciunas et al. 2000, 2003, 2004a,b, 2007; Valenti et al. 2003; Phillips et al. 2006; Pastorello et al. 2007a,b; Stanishev et al. 2007; Pignata et al. 2008), finding an $H$-band mass step of $0.18 \pm 0.05$ mag (0.10 $\pm$ 0.04 mag after an outlier cut) at a best fit location of $10^{10.48} M_\odot$, with a comparably sized $0.17 \pm 0.05$ mag step in the optical. Contemporary work by Uddin et al. (2020) recovered a similar result in an analysis of the CSP-I sample. Using independent SNooPy fits to each of the $\alpha$BvVrIrYJH light curves of every supernova, they measure comparable steps for all passbands at the median host mass of $10^{10.48} M_\odot$, with step sizes of between $0.07 \pm 0.03$ (V-band) and $0.15 \pm 0.04$ mag ($\alpha$-band). They also computed (see Uddin et al. 2020, fig. 13) the step size vs. wavelength behaviour that would be expected from the $R_V$ and $E(B - V)$ population distributions implied by the Brout & Scolnic (2021) results, finding this to be in poor agreement with the behaviour observed in the CSP data.

Recently, however, Johansson et al. (2021) analysed a sample from the literature (including CfA/CFAIR, CSP-I, and others from Barone-Nugent et al. 2012; Stanishev et al. 2018; Amanullah et al. 2015), along with 42 SNe Ia from their own intermediate Palomar Transient Factory (iPTF; Rau et al. 2009) survey whose NIR data were obtained with the Reionization and Transients InfraRed camera (RATIR; Butler et al. 2012). They estimated NIR mass steps consistent with zero (although not entirely inconsistent with previous literature estimates), and claimed that fitting for $R_V$ on a supernova-by-supernova basis eliminated the mass step in the optical.

Other recent work by González-Gaitán et al. (2020) studied the mass dependence of the apparent SN Ia colour–luminosity relation. They fit extensions of the Tripp (1998) standardisation formula to the SALT2 parameters of 740 supernovae from the Joint Lightcurve Analysis (JLA; Betoule et al. 2014). The Tripp (1998) formula,

$$\mu_s = m_B^{\text{sn}} - M_B + ax^{\text{1}B} + \beta c^s,$$

expresses the distance modulus, $\mu_s$, of a supernova, $s$, as a linear combination of its $B$-band apparent magnitude $m_B^{\text{sn}}$, light curve stretch $x^{\text{1}B}$, and apparent $B-V$ colour $c^s$, with stretch–luminosity and colour–luminosity coefficients $\alpha$ and $\beta$, and an absolute magnitude constant $M_B$. The mean magnitude offset, or mass step, between SNe Ia in low- and high-mass host galaxies is typically accounted for with two different values of $M_B$ for each subsample. In their analysis, González-Gaitán et al. (2020) allow for different values of the apparent colour–luminosity coefficient, $\beta$, in different mass and colour bins. They find a significant relation between $\beta$ and host galaxy mass. Since $\beta$ averages the intrinsic colour–luminosity relation and extrinsic dust law (Mandel et al. 2017), this could be explained by either a difference in $R_V$, or a difference in intrinsic properties. They also see a significant relation between $\beta$ and apparent colour – likely driven by a tendency for the reddest supernova to follow an $R_V$ driven colour–luminosity trend, with the bluest following an intrinsic trend (expected from Mandel et al. 2017). Unlike Brout & Scolnic (2021), González-Gaitán et al. (2020) find that the mass step is preserved even when allowing for different apparent colour–luminosity relations in different mass bins.

In the near future, the Vera C. Rubin Observatory’s Legacy Survey of Space and Time (LSST; Ivezić et al. 2019), and supernova programs on the Nancy Grace Roman Space Telescope (Roman Space Telescope; Spergel et al. 2015; Hounsell et al. 2018), will massively expand the high-redshift SN Ia dataset which informs dark energy analyses. However, these projects are likely to yield small or suboptimal low redshift samples (Foley et al. 2018; Hounsell et al. 2018; Jones et al. 2021), meaning these data must come from elsewhere. The low-redshift sample used in many previous analyses (e.g. Betoule et al. 2014; Rest et al. 2014; Scolnic et al. 2014b, 2018) has come from a variety of sources – particularly the CfA (Riess et al. 1999; Jha et al. 2006; Wood-Vasey et al. 2008; Hicken et al. 2009a, 2012; Friedman et al. 2015), CSP (Contreras et al. 2010; Folatelli et al. 2010; Stritzinger et al. 2011; Krisciunas et al. 2017), and Calán/Tololo Survey (Hamuy et al. 1996b). This heterogeneity gives rise to calibration and cross-calibration systematics which contribute greatly to the error budget (Conley et al. 2011; Scolnic et al. 2014b, 2018; Brout et al. 2019), in spite of efforts to remedy this (e.g. Scolnic et al. 2015; Currie et al. 2020). Since these data are often critical to training SN Ia models (e.g. BAYESN, Mandel et al. 2011, 2020; SALT2, Guy et al. 2007, 2010; Betoule et al. 2014), the uncertainty associated with their calibration can bleed into cosmological analy-
samples, which form a large fraction of the current low-\(z\) SNe in non-trivial ways. An additional complication is that the CfA piling cosmology-ready photometry of 180 SNe Ia at the Foundation Supernova Survey (Foundation DR1; Foley et al. 2018; Kaiser et al. 2010; Chambers et al. 2016). The first data release of the Foundation Supernova Survey (Foundation DR1; Foley et al. 2018; Jones et al. 2019) represents the current result of these efforts, compiling cosmology-ready photometry of 180 SNe Ia at \(z \lesssim 0.1\). By following up SN Ia discoveries that mainly come from ‘untargeted’ surveys – primarily the All-Sky Automated Search for Supernovae (ASAS-SN; Shappee et al. 2014) and Pan-STARRS Survey for Transients (PSST; Huber et al. 2015) – this sample should be considerably less biased in terms of host galaxy properties than existing low-redshift data. Future data releases of the Foundation program and the Young Supernova Experiment (YSE; Jones et al. 2021) will enlarge this further. This will provide a homogeneous, well-calibrated low-\(z\) SN Ia sample to anchor the LSST and Roman Space Telescope datasets of the future, and accompany the considerable high-\(z\) sample already observed in the PS1 MDS (Rest et al. 2014; Scolnic et al. 2018; Villar et al. 2020).

The excellent calibration properties, and internal consistency, of the Foundation data make them an ideal training set for SN Ia light curve and spectral energy distribution (SED) models. These models are fundamental to any cosmological analysis, as the route by which distances are estimated from photometric light curves. The conventional model used in most recent analyses is SALT2 (Guy et al. 2007, 2010; Betoule et al. 2014), whose parameter estimates (peak apparent \(B\)-band magnitude \(m_B\); stretch \(x_1\); and apparent colour \(c\)) are converted to a distance via a linear Tripp (1998) standardisation. This paradigm has a number of known weaknesses, including its limited wavelength coverage, confounding of intrinsic colour and extrinsic dust effects (see Mandel et al. 2017), and uncertain characterisation of residual scatter (a key systematic, see Scolnic et al. 2014a). Moreover, the most recent SALT2 training (Betoule et al. 2014) was carried out using a heterogeneous sample of SNe Ia with known calibration and cross-calibration issues, baking these systematics into the model. In their recent Pantheon analysis, Scolnic et al. (2018) found that these inherited systematics in the SALT2 model (which cannot easily be corrected) likely dominate over the calibration systematics in the data they were fitting (which can be at least partly alleviated, Scolnic et al. 2015).

Some of the limitations of the SALT2 SED model have been addressed by alternatives. In particular, the BayesSN model (Mandel et al. 2020) is a coherent framework for optical and near-infrared (NIR) SEDs, continuous in wavelength from 0.35–1.8 \(\mu\)m, with distinct treatments of dust and intrinsic variability. In Mandel et al. (2020), this model was trained on a compilation (Avelino et al. 2019) of 79 low-\(z\) optical and NIR SNe Ia, mainly from the CSP (Contreras et al. 2010; Stritzinger et al. 2011; Krisciunas et al. 2017) and CfA (Jha et al. 1999; Wood-Vasey et al. 2008; Hicken et al. 2009a, 2012; Friedman et al. 2015) surveys, as well as earlier data from the Las Campanas Observatory (LCO; Krisciunas et al. 2004a, b) and elsewhere in the literature (Krisciunas et al. 2003, 2007; Stanishev et al. 2007; Pignata et al. 2008; Leloudas et al. 2009). Whilst this dataset was over two times larger than those used in training earlier optical and NIR light curve models (Mandel et al. 2009, 2011; Burns et al. 2011), the heterogeneity of data sources is subject to cross-calibration issues. Additionally, host galaxy mass estimates (from e.g. Ponder et al. 2020; Uddin et al. 2020), indicate a considerable bias towards high masses (around 80% have host galaxy stellar masses \(> 10^{10} M_\odot\)).

The BayesSN framework is well suited to the incorporation and analysis of correlations of SN Ia properties with host galaxies. Additionally, the supernovae in Foundation DR1 benefit from calibration homogeneity, consistently determined host galaxy masses (Jones et al. 2019), and a less biased host distribution than other low-\(z\) SN Ia samples. This makes them an ideal dataset for studying SN–host correlations amongst the low–moderate reddening \((B - V \lesssim 0.3)\) SNe Ia typical of cosmological samples. Therefore, in this work, we present a first BayesSN analysis of Foundation DR1. As part of this study, we repeat the mass-agnostic Mandel et al. (2020) analysis on the Foundation data, verifying BayesSN’s performance, and modelling for the first time the statistical properties of SNe Ia in the rest-frame \(z\)-band. Additionally, we extend the BayesSN model to allow population dust properties to be split by host galaxy mass. By including this alongside a conventional step-like brightness offset when training our SED model, we are able to test if dust properties differ significantly in low and high mass SN Ia host galaxies, and if there is any interplay between dust properties and the mass step. In previous SALT2-based analyses of this problem, all inferences about \(R_V\) have been made from the apparent colour–luminosity relation between extinguished absolute magnitude in \(B\) and apparent \(B - V\) colour at the time of maximum light. In contrast, the hierarchical framework of BayesSN embeds a distinct dust law within the time- and wavelength-dependent SED model, inferring \(R_V\) by drawing on the colour–luminosity information contained in the full \(gri\) light curves. This properly leverages the fact that the dust law impacts the SED at all times and wavelengths, enabling us to better discern intrinsic SN variations from the effects of dust. Performing this inference within a Bayesian framework enables us to marginalise over other effects in the problem in order to draw robust conclusions about the dust and probabilistically quantify our uncertainties about its properties.

In Section 2, we recap the details of the BayesSN model (fully described in Mandel et al. 2020), and discuss the modifications we employ in our treatment of host galaxy mass. In Section 3 we describe the Foundation dataset, and the data cuts we applied. This section also discusses several choices of host galaxy mass split that we consider throughout. Section 4 presents and discusses our results, covering SN Ia \(z\)-band behaviour, the mass step and residual scatter, population dust properties, the distribution of light curve shape parameters, intrinsic colour and extrinsic dust, and Hubble diagram scatter in turn throughout Sections 4.1–4.6. Finally, in Section 5, we summarise our results, and discuss the outlook for future work.

2 THE BAYESSN MODEL

Hierarchical Bayes provides a natural probabilistic framework for modelling and inference of populations and their constituent individuals (e.g. Gelman et al. 2013), and was first applied to SN Ia modelling by Mandel et al. (2009). Throughout this work, we carry out our analysis in several different configurations of the BayesSN hierarchical model (Mandel et al. 2020, hereafter M20), detailed fully in Sections 2.1–2.3. Our No-Split model (Section 2.1) treats all SNe Ia as a single population with parameters agnostic of host galaxy mass. Our Partial-Split model (Section 2.2) allows key
parameters to differ between the low- and high-mass subsamples, whilst keeping others common to all supernovae irrespective of host mass. Finally, our Full-Split model (Section 2.3) treats the subsamples in the two mass bins as entirely separate populations. Within the No-Split configuration, we carry out parallel analyses where different assumptions are made about the RV value(s) parametrizing host galaxy dust laws. The first of these (Section 2.1.1) assumes a single value of RV for all SNe in the sample, whilst the second (Section 2.1.2) assumes a population distribution of RV, with individual values permitted for each SN. Within the Partial-Split configuration, we run equivalent analyses where either a separate single RV, or a separate population distribution of RV, is allowed in each mass bin.

As in M20, the models are all implemented in the Stan probabilistic programming language (Carpenter et al. 2017; Stan Development Team 2020), with the joint posterior over all global and individual SN parameters (eqs. 27-28 in M20, with modifications as stated below) sampled using Stan’s advanced Hamiltonian Monte-Carlo algorithm (Hoffman & Gelman 2014; Betancourt 2016). During training, we adopt external distance constraints from the distance-redshift relation based on a fiducial ΛCDM cosmology ($H_0 = 73.24$, $Ω_M = 0.28$, $Ω_Λ = 0.72$; Riess et al. 2016), with external distance uncertainties derived from peculiar velocity and spectroscopic redshift uncertainties. We assume a peculiar velocity uncertainty of $σ_{pec} = 150$ km s$^{-1}$ (Carrick et al. 2015).

2.1 No-Split Model

2.1.1 Global Dust Law

As a starting point, we consider the configuration of the BayesSN model described in M20 that is agnostic with respect to host galaxy mass. For full details of the model, we refer the reader to the M20 paper. We briefly summarise the essentials here. Throughout the paper, we will refer to this as the No-Split model, since it ignores host galaxy mass and treats all SNe in the sample jointly as a single population.

The BayesSN forward model specifies a time- and wavelength-varying surface describing the intrinsic SED of a Type Ia supernova. This is extinguished by host galaxy dust, redshifted, dimmed by distance, extinguished by Milky Way dust, integrated through photometric passbands at the observation times, and finally perturbed by measurement error, to yield observed SN Ia light curves. The host-dust-extinguished SED $S_I(t,λ_I)$ of a supernova $s$, as a function of rest-frame phase $t$ and wavelength $λ_I$, is modelled as:

$$-2.5 \log_{10} \left[ \frac{S_I(t,λ_I)}{S_0(t,λ_I)} \right] = M_0 + W_0(t,λ_I) + \delta M_s + \theta^1_s W_1(t,λ_I) + \epsilon_s(t,λ_I) + A^s_V(λ_I;RV) \tag{2}$$

(as in M20 Eq. 12). Here, $S_0(t,λ_I)$ is an optical-NIR SN Ia baseline spectral template (Hsiao et al. 2007; Hsiao 2009); $M_0 \equiv -19.5$ is a normalisation constant; $W_0(t,λ_I)$ is a broad warping of the baseline template, common to all SNe, to model the average intrinsic SED.

The next three terms model the variations of individual SN Ia SEDs in the absence of dust ($A^s_V = 0$). $\delta M_s$ is a ‘grey’ (time- and wavelength-independent) brightness offset specific to supernova $s$; $W_1(t,λ_I)$ is the first functional principal component, describing the primary SED features that vary between SNe; $\theta^1_s$ is a scalar coefficient describing the extent to which $W_1(t,λ)$ affects the $s$th supernova; $\epsilon_s(t,λ_I)$ is a time- and wavelength-varying residual function, describing SED perturbations particular to SN $s$ that are otherwise unexplained. The statistical properties of these terms are learnt during training and model the covariance structure of the population of intrinsic SN Ia SEDs. Since colours and luminosities are quantities derived from integrals over the SEDs, intrinsic colour and luminosity correlations across phase and wavelength are captured here.

The last term describes the effect of dust along the line-of-sight in the SN host galaxy on the SN Ia SED. The parameter $A^s_V$ is the $V$-band host dust extinction affecting supernova $s$; and $\xi(λ_I;RV)$ is the Fitzpatrick (1999) dust law, parametrized by RV. Initially, we assume RV has a single global value for the sample, but we later explore models where this is split by host galaxy mass, or has a population distribution within a (sub)sample.

In practice, we represent the $W_0(t,λ_I)$, $W_1(t,λ_I)$, and $\epsilon_s(t,λ_I)$ functions with natural cubic splines, parametrized by matrices of knots, $W_0$, $W_1$, and $E_s$. The main difference from M20 (which models a wider wavelength range of 0.35–1.8 μm) is the exact configuration of these spline knots. In rest-frame phase, we use the same knot locations as in M20, with knots every 10 d between −10 and 40 d. In wavelength, we place knots at $I = (3500, 4900, 6200, 7700, 8700, 9500)$ Å, corresponding to the centres of the griz-bands, plus the outer edges of $g$ and $z$.

After redshifting the host-dust-extinguished SED given by Eq. (2), and accounting for distance, we apply Milky Way dust reddening and extinction using the Fitzpatrick (1999) law, with $R_{MW} = 3.1$ and $A^V_{MW} = R_{MW} \times E(B-V)_{MW}$ computed for each SN using the Schlafly & Finkbeiner (2011) reddening map. The SED is then integrated through photometric bandpasses to compute model fluxes. This generative model forms the basis of our methodology (see M20 for exact details, and fig. 1 therein for a graphical representation).

We model the population distributions of the latent parameters of individual SNe as:

$$A^s_V \sim\text{Exponential}(τ_A), \tag{3}$$

$$θ^1_s \sim N(0, 1), \tag{4}$$

$$δM_s \sim N(0, σ^2_0), \tag{5}$$

$$\epsilon_s \sim N(0, Σ_ε), \tag{6}$$

where $ε_s = vec(E_s)$ is the vectorised $E_s$ matrix, and $τ_A$, $σ_0$ and $Σ_ε$ are hyperparameters describing the population mean host galaxy extinction, population standard deviation of the $δM_s$ offsets, and population covariance of the residual $ε_s$ perturbations, respectively. Our weak hyperpriors are as in M20.

2.1.2 Population Distribution of Dust Law RV

Modelling parameters of individuals as being drawn from a population distribution with unknown hyperparameters to be jointly inferred from the data is known as partial pooling (Gelman et al. 2013). Population distribution models for RV were first introduced by Mandel et al. (2009, 2011). To enable a more direct comparison with the results of Brout & Sciclun (2021), we consider an extension of our current model in which the RV values for individual SNe are drawn from a population distribution, rather than all having a single global value. In this model variant, the $R^s_V$ of an individual SN $s$ is now a latent variable, drawn from a truncated normal (Gaussian)$^1$ population

---

$^1$ Here, $X \sim \text{Trunc-N}(μ, σ^2, a, ∞)$ denotes that the random variable $X$ is drawn from a one-sided truncated normal distribution, with truncation of the lower tail at $a$. The probability density function for $X \geq a$ is $P(X|μ, σ, a) = \frac{1}{σ \sqrt{2π}} \phi(\frac{a-μ}{σ})$, and is zero for $X < a$. Here, $ξ = (X - μ)/σ$, $α = (a - μ)/σ$, and $φ(ξ)$ and $Φ(ξ)$ are respectively the PDF and CDF of a standard normal random variable $ξ$. 
distribution, 
\[ R_V \sim \text{Trunc-N}(\mu_R, \sigma_R^2, 0.5, \infty), \]  
with a truncation of the lower tail to force \( R_V \geq 0.5 \). A physically-motivated choice of lower bound would be \( R_V^* \geq 1.2 \) (the Rayleigh scattering limit, see Draine 2003), but we opt for the more liberal assumption of a Gaussian population distribution for \( R_V^* \) somewhat motivated by the empirical distribution of \( R_V \) values found along different sightlines through the Milky Way (Schlafly et al. 2016), which is well-described by a narrow Gaussian with mean 3.32 and standard deviation 0.18. While this may not be directly analogous to sightlines to SNe Ia in external galaxies, it is a working assumption that can be elaborated on in future work. Given that our results ultimately suggest a preference for narrow \( R_V \) distributions (see §4.3.2), we expect our results to be robust to the exact shape of the assumed distribution. One would expect that a Gaussian population distribution would be more sensitive to outliers than a heavier-tailed alternative, making an overestimate of the dispersion in the population (which does not appear to have happened here) more of a risk than an underestimate. In future work, we will be able to test alternative possible population distributions (e.g. skew-normal, Student’s t, Gaussian mixtures).

We place a uniform hyperprior on \( \mu_R \sim U(1,5) \), and a half-Normal hyperprior on \( \sigma_R \sim \text{Half-N}(0, 2^2) \). The latter is selected so as to avoid a hard limit on \( \sigma_R \), but to place relatively little prior probability at excessively high values of \( \sigma_R > 4 \). Sensitivity analysis for alternate \( \sigma_R \) hyperpriors is demonstrated in Appendix B.

### 2.2 Partial-Split Model

We also train a version of the model where several key parameters, previously treated as common to all supernovae, are split by host galaxy mass at a critical value, \( M_{\text{s,split}} \), to fit for a low- and high-mass version of each. We will refer to this as the Partial-Split model. By splitting only selected parameters, the model is able to probe salient differences between the high- and low-mass subsamples, whilst still utilising all of the data for determining the mean intrinsic SED, primary mode of intrinsic SED variation, and intrinsic residual covariance.

The parameters that we split in this way are the \( R_V \) (or the population distribution parameters, \( \mu_R \) and \( \sigma_R \), in the case where supernovae have individual \( R_V \) values) parametrizing the Fitzpatrick (1999) host galaxy dust extinction law; the population mean extinction \( \tau_A \); and \( \sigma_0 \), the population standard deviation of the ‘grey’ \( \delta M_s \) component of the intrinsic SED. We also introduce a parameter, \( \Delta M_0 \), which shifts the population mean of the \( \delta M_s \) for supernovae in low-mass hosts — this is functionally equivalent to allowing a constant ‘mass step’. This means that equation (5) becomes

\[ \delta M_s \sim \begin{cases} 
N(0, \sigma_{0,\text{HM}}^2) & \text{if } M_{s,s} \geq M_{s,\text{split}}, \\
N(\Delta M_0, \sigma_{0,\text{LM}}^2) & \text{if } M_{s,s} < M_{s,\text{split}}, 
\end{cases} \]  

where \( M_{s,s} \) is the host galaxy stellar mass of SN \( s \), and the subscript LM and HM denote the values of \( \sigma_0 \) below and above the mass split. The hyperpriors on the low- and high-mass values of \( R_V \) (or \( \mu_R, \sigma_R \)), \( \tau_A \), and \( \sigma_0 \) are taken to be the same as in the No-Split model. For the low- vs. high-mass offset, \( \Delta M_0 \), we assume a uniform hyperprior within 0.0 ± 0.2 mag, since previous works investigating dependence on host mass (e.g. Kelly et al. 2010; Sullivan et al. 2010; Roman et al. 2018; Uddin et al. 2020; Smith et al. 2020) have found mass step sizes safely within this range. If dust differences fully explain the mass step, we expect \( \Delta M_0 \approx 0 \) with the Partial-Split model.

### 2.3 Full-Split Model

Finally, we also consider a mode of analysis where entirely separate models (each configured exactly like the No-Split model) are trained for the low- and high-mass subsamples, with no shared parameters or information. We will refer to this scheme as the Full-Split model. This configuration provides a worthwhile sanity check of the Partial-Split model, since it allows all model components to vary between host-mass subsets. If this is not the case, however, this will lead to weaker inferences, since this effectively doubles the number of parameters and fewer supernovae are available on either side of the mass split.

### 3 DATA

#### 3.1 Foundation DR1

The first data release of the Foundation Supernova Survey (Foundation DR1; Foley et al. 2018; Jones et al. 2019) presents griz light curves for around 180 cosmologically useful Type Ia supernovae at \( z \lesssim 0.1 \), all observed on the Pan-STARRS-1 system. This sample is far more homogeneous than existing low-redshift SNe Ia samples, with all data having been obtained on a single instrument which is extremely well-calibrated, with accurately determined instrumental properties (Stubbs et al. 2010; Schlafly et al. 2012; Magnier et al. 2013, 2020). Unlike previous low-z samples, the SNe followed up by Foundation are mainly discovered by untargeted surveys like ASAS-SN (Shappee et al. 2014) and the PSST (Huber et al. 2015), meaning the population of host galaxies probed will be less biased (although it is still not totally reflective of the distribution in high-z samples).

#### 3.2 Pre-Training Cuts

Foley et al. (2018) already applied several data cuts to construct the Foundation DR1 cosmology sample of 180 SNe Ia, as detailed in their §5.3. In particular, a standard cut on the SALT2 apparent colour parameter \( c < 0.3 \) (equivalent to peak apparent \( B - V \)), has been applied, ensuring a colour range consistent with the cosmological sample. To define the sample for our present analysis, we additionally restrict the sample to redshifts \( 0.015 < z < 0.08 \). The lower cut here is to select only SNe Ia in the smooth Hubble flow, limiting the impact of peculiar velocity errors, with the upper limit being the redshift beyond which Foley et al. (2018) expect observations to be more vulnerable to Malmquist bias\(^2\). Requiring \( z > 0.015 \) cuts 8 supernovae, bringing the sample size down to 172, with the \( z < 0.08 \) cut removing a further 5, giving a sample of 167 supernovae. Requiring reliable host masses of \( M_s > 10^8M_\odot \) for all SNe Ia eliminates ten further supernovae, leaving a final sample of 157.

\(^2\)To ensure that our conclusions are not vulnerable to any redshift-dependent selection effects, we also repeat our analysis with a more conservative redshift cut of \( z < 0.05 \) (which gives a sample of 125 SNe). This does not change any of our conclusions, with the inferred \( R_V \) values being only slightly different (higher by about 0.1–0.15, not a statistically significant shift). The Hubble diagram scatter from training and resubstitution on the \( z < 0.05 \) subsample is also not significantly changed compared to the full sample. From a No-Split analysis under the \( z < 0.05 \) cut, we estimate a Hubble diagram RMS (\( \sigma_{pv} \)) of 0.131 (0.119) mag (c.f. the first row of Table 4).
3.3 Choosing a Mass Split

We explore several different possibilities for setting the critical host mass at which to split the training set. These are summarised in Table 1. Analyses using SALT2 (e.g. Betoule et al. 2014) typically allow for a step at $M_{\text{s,crit}} = 10^{10} M_\odot$ by convention, so this is one option we use. However, it leads to a fairly unbalanced divide, with only 48 of the 157 supernovae in our Foundation sample falling on the low-mass side. As a balanced alternative, we try setting $M_{\text{s,crit}}$ to the median host mass of our sample, $10^{10.33}M_\odot$. Additionally, we choose a split point that is favoured as a mass step location based on Hubble residuals obtained from the No-Split BaysE5 analysis, where parameters were not separated by host mass. Requiring that the step be located somewhere within the interquartile range (IQR) of host masses, the preferred step location (based on a maximum likelihood estimate – see Appendix A for details) is $10^{10.286} M_\odot$, with a fairly balanced low-:high-mass ratio of 75:82. Figure 1 illustrates the step location we derive in this way, along with the Hubble residuals it is inferred from. There is a second strong peak in the likelihood at $10^{9.6815} M_\odot$, which lies outside of the sample IQR (indicated with a grey band on the plot) and gives a highly unbalanced low-:high-mass ratio of 30:127. Although we do not feature it as one of our main choices, repeating our analysis with this split choice does not change our main conclusions.

4 RESULTS

4.1 Primary Intrinsic SED Component

The primary mode of intrinsic SED variation within the population is captured by our first functional principal component, $W_1$. The extent to which this is present for a particular supernova in the sample is controlled by that supernova’s $\theta_1$ coefficient. To visualise the contribution of $W_1$ to the intrinsic SED, we synthesise rest-frame light curves in the Pan-STARRS-1 griz passbands, varying $\theta_1$, and with all other effects ($A_V$, $\delta M$, $e$) set to zero. Figure 2 (left panel) shows this for the No-Split model.

In the g-band, we observe a width-luminosity relation as is typical in optical SN Ia light curves (Phillips 1993), with increasing $\theta_1$ corresponding to a faster decline. In the r band, we see a weaker version of the same behaviour in the primary maximum. We also see an effect similar to that seen in M20, with the brighter, slower-declining, supernovae having a delayed, more extended, and more prominent secondary bump/maximum – something more noticeable in the i-band. An interesting and surprising difference from the behaviour seen in M20 is in the effect of $W_1$ on the z-band light curve. The 0.35–1.8 μm M20 model was trained only on $BVri'YJH$ data, meaning its z-band behaviour (see M20, fig. 6) is an interpolation between the i and Y bands. Indeed, this is apparent, with the M20 z-band model looking quite similar to the Y-band model, with a secondary maximum similar in prominence to (or even brighter than) the first, and a temporary crossing over of the slow and fast declining light curves around the dip between NIR peaks. Here, however, where the model is trained on z-band data, we see somewhat different behaviour. For the dimmest supernovae, the z-band light curve almost follows a plateau, declining slowly until around 20 d (approximately in line with the second i-band maximum), at which point it begins to fall off more rapidly. For the brighter supernovae, the initial decline appears faster, with a more definite minimum at around 15 d before a small rise to a secondary peak which falls off later than in the dimmer supernovae. Although this is closer in shape to what we would expect in the NIR, the minimum is still very shallow – far more so than seen for the i-, Y-, or interpolated z-band in M20, or in the i-band here.

We also synthesise colour curves, showing the effect of $W_1$ and $\theta_1$ on the intrinsic colour evolution of the supernovae in the sample. The right hand panel of Figure 2 shows rest-frame $g-r$, $g-i$, and $g-z$ intrinsic colour curves for a range of $\theta_1$. An interesting result is that we see very little sensitivity of the $g-r$ and $g-z$ colours around peak to varying $\theta_1$. This suggests that there is little correlation between light curve brightness or decline rate, and these intrinsic colours. For all of the chosen colours, sensitivity to $\theta_1$ increases considerably after peak, particularly around 10–25 days. Additionally, at around 30 days, each colour converges for all values of $\theta_1$, before (most
noticeable in $g-z$ flipping over in behaviour (with the supernovae which were previously bluest becoming the reddest, and vice-versa).

### 4.2 Residual SED Variation

Within the BayesN model, residual SED variation not explained by the primary intrinsic component, $W_1$, or dust extinction, is captured by two components: the population of ‘grey’ $\delta M_s$ offsets, which are constant in time and wavelength; and the time- and wavelength-varying $e_\gamma(t, \lambda_r)$ realisations, whose correlation structure is captured by a population covariance matrix, $\Sigma^\gamma$.

Taking the two components together, we can study the total level of residual SED scatter within the population. Figure 3 isolates the effect of varying $\eta_a(t, \lambda_r) = \delta M_s + e_\gamma(t, \lambda_r)$ on rest-frame $griz$ light curves and $g-r$, $r-i$, and $g-z$ colour curves. The light curve residual variance is depicted in Figure 3 (left hand panel). The widths of the $\pm 1 \sigma$ envelopes, showing the residual variance within the population, are relatively constant near peak, generally showing only $\sim 0.1$ mag of scatter, and increasing somewhat towards later phases. The colour curve residuals are, by definition, independent of the grey residuals $\delta M_s$ and their variance is depicted in the right hand panel. The residual colour scatter is generally small, but shows especially little scatter near peak (particularly in $g-r$ and $g-z$). Given the relative insensitivity of peak intrinsic $g-r$ and $g-z$ colour to the principal mode of intrinsic SED variation, it seems that these intrinsic colours at peak have small variance across this sample.

When considering the relation to host mass, we will focus on the time- and wavelength-independent residual component. This is summarised by the variance, $\sigma^2_\gamma$, of the Gaussian population distribution of $\delta M_s$, and the shift, $\Delta M_0$, in the mean of this distribution between low- and high-host galaxy masses (see equation 8).

For the No-Split analysis where we ignore host galaxy mass, we find that $\sigma_0 = 0.10 \pm 0.01$. For the Partial-Split analyses where we split $\sigma_0$, $\tau_A$ and $R_V$ by host mass, the outcome depends somewhat on the exact choice of split point, but is qualitatively similar for all of the choices tried here (see Table 1 for a summary of these). For all choices, positive values of $\Delta M_0$ are at least somewhat favoured – suggesting lower average brightness in low-mass galaxies. For the median host mass split, we estimate $\Delta M_0 = 0.054 \pm 0.025$ mag, around $2 \sigma$ greater than zero. The difference of $\Delta M_0$ from zero is most significant ($3 \sigma$) when the sample is split at the median SED mass split preferred as a step location by the Hubble residuals. For all choices of mass split, the posterior estimate of $\Delta M_0$ is consistent with the step size seen at that mass in Hubble residuals computed from the No-Split analysis. For the split choice of $10^{10} M_\odot$, our inferred $\Delta M_0 = 0.042 \pm 0.028$ mag is consistent with, but somewhat less significant than, the mass step of $0.060 \pm 0.024$ mag found by Jones et al. (2019) for the Foundation sample. The final two columns of Table 1 summarise the different Partial-Split $\Delta M_0$ estimates and No-Split step sizes side-by-side.

For all choices of split point (see Table 2, which also repeats the $\Delta M_0$ values), our inferences of $\sigma_0$ are fairly consistent between low- and high-mass hosts. For the median and MLE split points, the posterior mean estimates of $\sigma_0$ are marginally higher in low-mass host galaxies. For the median mass split point, the joint posterior distribution of $\Delta M_0$ and the high- and low-mass values of $\sigma_0$ is shown in Figure 4. Our inferences of $\Delta M_0$ and $\sigma_0$ under the No-Split and Partial-Split models are consistent whether a single $R_V$ or a population distribution is assumed in the full sample or within each subsample.

For the Full-Split configuration, where entirely separate models are trained either side of the split point (see Section 2.3), our $\sigma_0$ estimates (see Table 2) are generally consistent with their counterparts from the Partial-Split analyses. For the $10^{10} M_\odot$ mass split, we estimate $\sigma_0$ in low-mass hosts to be around $2.2 \sigma$ higher than in high-mass hosts. For the other splits, the difference is insignificant. There is no explicit $\Delta M_0$ included in the Full-Split analysis, since an effective shift in $M_0$ is entirely exchangeable with a shift in the mean intrinsic SED $W_0(t, \lambda_r)$ (see Equation 2). Since, in Full-Split mode, the models are trained completely independently on each subsample, we fix each $M_0 = -19.5$ to its default value, and allow each $W_0(t, \lambda_r)$ to absorb any relative magnitude offset.

### Table 1. Summary of the different host galaxy stellar masses at which we split the supernova sample.

| Split Point | Host Galaxy Stellar Mass ($M_\odot$) | Low:High Mass Ratio$^a$ | HR Step Size in No-Split Run$^b$ | $\Delta M_0$ Offset in Partial-Split Run$^c$ |
|-------------|------------------------------------|-------------------------|-------------------------------|----------------------------------|
| $10^{10} M_\odot$ | $10^{10}$ | 48:109 | 0.050 ± 0.022 | 0.042 ± 0.028 |
| Median | $10^{10.331}$ | 78:79 | 0.055 ± 0.020 | 0.054 ± 0.025 |
| MLE | $10^{10.286}$ | 75:82 | 0.066 ± 0.020 | 0.072 ± 0.024 |

$^a$ Number of supernovae (out of the 157 in our Foundation cut) above and below each mass split.$^b$ Step size derived for each step location based on the Hubble residuals from a training and photometric distance estimation (resubstitution) under the No-Split model (see Sections 2.1 and 3.3).$^c$ The low – high host mass magnitude offset, $\Delta M_0$, estimated from the Partial-Split posterior analysis (see Sections 2.2 and 4.2).
Table 2. Posterior means and standard deviations of population parameters for different choices of host mass split.

| Model       | Split Point | ΔM0     | σ0) | τA     | R_V    |
|-------------|-------------|---------|------|--------|--------|
|             |             | LM      | HM   | LM     | HM     | LM     | HM     |
| No-Split    | -           | -       | 0.10 ± 0.01 | 0.19 ± 0.02 | 2.61 ± 0.21 |
| Partial-Split | 10^{10}M_⊙ | 0.042 ± 0.028 | 0.10 ± 0.02 | 0.10 ± 0.01 | 0.17 ± 0.03 | 0.22 ± 0.03 | 3.33 ± 0.50 | 2.67 ± 0.21 |
| Medieval    | -           | 0.054 ± 0.025 | 0.12 ± 0.02 | 0.08 ± 0.01 | 0.19 ± 0.03 | 0.21 ± 0.03 | 2.84 ± 0.31 | 2.58 ± 0.23 |
| MLE         | -           | 0.072 ± 0.024 | 0.11 ± 0.02 | 0.09 ± 0.01 | 0.18 ± 0.03 | 0.22 ± 0.03 | 2.86 ± 0.33 | 2.66 ± 0.24 |
| Full-Split  | 10^{10}M_⊙ | -       | 0.15 ± 0.02 | 0.10 ± 0.01 | 0.16 ± 0.04 | 0.22 ± 0.03 | 3.07 ± 0.64 | 2.43 ± 0.22 |
| Medieval    | -           | 0.10 ± 0.02 | 0.09 ± 0.01 | 0.21 ± 0.03 | 0.22 ± 0.03 | 2.83 ± 0.35 | 2.46 ± 0.25 |
| MLE         | -           | 0.10 ± 0.02 | 0.09 ± 0.01 | 0.19 ± 0.03 | 0.22 ± 0.03 | 2.86 ± 0.36 | 2.54 ± 0.25 |

Figure 4. Posterior distribution of the low- vs. high-host-mass magnitude offset, ΔM0, and the low- and high-host-mass values of σ0 – the population standard deviation of the ‘grey’ component of the intrinsic SED. The median host mass split was used here. Darker contours indicate the regions which contain 68 and 95 per cent of the posterior probability.

4.3 Population Dust Properties

4.3.1 Global Dust Law

In our baseline model, the population host galaxy dust properties are characterised by the global R_V value, and a population mean extinction value, τ_A. When host galaxy stellar mass is ignored, and a single value for each of these parameters is assumed, we find R_V = 2.61 ± 0.21 and τ_A = 0.19 ± 0.02 – see the upper left panel of Figure 7 for a plot of the joint posterior. This R_V value is consistent with the global value of R_V = 2.89 ± 0.20 found by M20 for the low-z optical–NIR supernova sample from Avelino et al. (2019). The mean extinction value we infer for the Foundation sample is somewhat lower than was found for the sample analysed by M20.

Our model constrains the dust properties by fitting time-dependent SEDs to the full griz light curves of the SN Ia sample, thereby leveraging the maximal information in the data. However, it is still useful to derive insights from low-dimensional visualisations of the full inference. Figure 5 shows estimated extinguished rest-frame g-band absolute magnitudes and apparent g – r colours at time of maximum for the unsplit sample, obtained from training the No-Split model. The SNe Ia in both low- and high-mass hosts exhibit an apparent colour–luminosity trend consistent with the expected reddening vector for our posterior mean R_V = 2.61 (red arrow). The more reddened SNe Ia are clearly inconsistent on average with a lower R_V = 1.5 (yellow arrow). In particular, for a given red apparent colour, the SNe Ia are on average dimmer than would be expected with an R_V = 1.5 dust law.

Figure 5 also shows how our model (similarly to Mandel et al. 2017) decomposes the apparent colour-luminosity distribution into two distinct physical effects: dust reddening-extinction (red arrow, with a slope determined by R_V) and an intrinsic colour-luminosity
trend (blue line with a shallower slope). This is in contrast to the conventional Tripp formula (Eq. 1), which models the apparent colour-magnitude relation with a single linear slope $\beta$. The blue contours indicate the finite extent of the inferred intrinsic colour-luminosity distribution. Proper accounting of this intrinsic colour distribution is especially critical for reliable inference of the dust law distribution (c.f. Section 4.3.2).

When we retrain our model with separate $R_V$ and $\tau_A$ values permitted for low- and high-mass hosts, we see little evidence that the extinction distribution is different between the two subsamples. Our estimates of $\tau_A$ are consistent to within $\lesssim 1.2\sigma$ between low- and high-mass hosts – Table 2 lists the inferred values for different choices of mass cut. Figure 6 shows the exponential extinction population distributions implied by the inferred $\tau_A$ values when the median host mass split is used, along with histograms of the individual posterior mean $A_V$ estimates for the two samples.

We see little evidence for significantly different $R_V$ values either side of the split point. For all choices of split point, the low- and high-mass $R_V$ values are consistent with one another, the global value found when training without a host mass split, and the value inferred from the completely independent Avelino et al. (2019) sample by M20. Taking the median split point, for example, we find $R_V = 2.84 \pm 0.31$ for the low-mass hosts, and $R_V = 2.58 \pm 0.23$ for the high-mass hosts. This consistency of $R_V$ across the mass split aligns with expectation from the apparent colours and extinguished absolute magnitudes estimated in our No-Split analysis (see Fig. 5). It also aligns with the results of Jones et al. (2018, fig. 4) who found consistency in the apparent colour-luminosity relation of supernovae in low- and high-mass host galaxies on a sample of 273 SNe Ia (including a subset of Foundation DR1). The joint posterior of $R_V$ and $\tau_A$ for the median mass split is shown in the upper right hand panel of Fig. 7. The posterior means and standard deviations for other choices of mass split are listed in Table 2, along with the other global parameters. For the $10^{10}M_\odot$ split point, which gives less balanced high- and low-mass bins, the ability to constrain $R_V$ well for the low-mass subset is limited by sample size, with a fairly high posterior standard deviation for $R_V$ in this run. Here, we estimate $R_V = 3.33 \pm 0.50$ for hosts with $M_* < 10^{10}M_\odot$, and $R_V = 2.67 \pm 0.21$ for those with $M_* \geq 10^{10}M_\odot$. The estimated difference $\Delta R_V \approx 0.66 \pm 0.54$ is statistically consistent with zero within $\approx 1.2\sigma$, but may be suggestive of a larger value. The considerable $R_V$ uncertainty in the low-mass subsample for this more unbalanced split choice likely arises from the dual effect of having fewer SNe generally, and from having fewer of the more extinguished SNe (see Fig. 10) that have the most leverage to constrain $R_V$. Nevertheless, for all of the Partial-Split runs, $R_V \lesssim 2.2$ tends to be strongly disfavoured on both sides of the mass split, with less than 5 per cent posterior probability in all cases.

From the Full-Split analysis, our conclusions about $\tau_A$ and $R_V$ do not change significantly, with all values above and below the various split points being consistent with their Partial-Split equivalents. The consistency of $\tau_A$ between high- and low-mass hosts remains within $\lesssim 1.2\sigma$. For all three split points, we estimate $R_V$ to be consistent within $1\sigma$ between high- and low-mass hosts, with $R_V \lesssim 2$ disfavoured with at least 95 per cent probability in all cases. The consistency of our dust inferences between the Full-Split and Partial-Split models indicates that our results are robust to how the intrinsic components are estimated.

### 4.3.2 Population Distribution of Dust Law $R_V$

In Figure 5, the blue contours indicate the finite extent of the inferred intrinsic distribution under our global dust law model. This is key to the interpretation of the apparent colour-luminosity distribution and the inference of dust laws. If one naively assumes that the apparent properties of every reddened SN should “map” back to the same intrinsic value (e.g. the mean of the distribution) under its dust correction, then one might conclude that a wide range of $R_V$ is present. However, this assumption would logically contradict the finding of a intrinsic population distribution with finite width: if the intrinsic covariance is non-zero, then not every reddened SN should simultaneously map back to the same intrinsic point (with zero variance). Instead, our statistical model simultaneously infers and accounts for the nonzero extent of the intrinsic distribution along with the distribution of dust laws, over the full griz light curves. Using this probabilistic approach, we find in this section that the data are consistent with a narrower distribution of dust laws.

Under this mode of analysis, an individual $R_V^*$ is permitted for every supernova, with these drawn from a population distribution described by a Gaussian with mean $\mu_R$, and standard deviation $\sigma_R$, truncated such that $R_V^* \geq 0.5$ (see Equation 7). This dust law distribution is included alongside our model’s usual treatment of the intrinsic colour distribution and dust extinction distribution (see Section 2.1 for a summary of the technical details), with all aspects being inferred and accounted for simultaneously. For the No-Split, and all of the Partial-Split configurations where we try this model extension, our inferences of $\mu_R$ are consistent with the global values we found previously. Our $\sigma_R$ posteriors unanimously prefer small values, with the strength at which large values are disfavoured depending on the exact model configurations. Table 3 lists parameter summaries for $\mu_R$ and $\sigma_R$ for the No-Split and Partial-Split models.

The recent work of Broust & Scolnic (2021) also investigated the possibility that dust in low- and high-mass SN Ia host galaxies has different properties. Similarly to the present work, they modelled host galaxy reddening as following separate exponential population distributions above and below their chosen mass split point of $10^{10}M_\odot$, with separate Gaussian population distributions of $R_V$ on either side of this split. For their low-redshift sample (including Foundation DR1), they estimate that supernovae in massive hosts have a higher mean level of dust reddening. Our posterior mean $\tau_A$ values for the same choice of split point are suggestive of qualitatively similar be-
Figure 7. (top row) Posterior distributions of $R_V$ and $\tau_A$ from BayesSN training runs on Foundation DR1 carried out without (left panel) and with (right panel) these parameters split by host galaxy stellar mass at the median. (bottom row) Posterior distributions of $\mu_R$, $\sigma_R$ and $\tau_A$ from the variant of our analysis where we allow a population distribution of $R_V$. Parameter summaries are posterior mean and standard deviation, except for $\sigma_R$, where 68 (95)th percentiles are quoted. These are also indicated on the relevant marginal plots as dashed lines.

haviour – something borne out more prominently when looking at the individual reddening values for the supernovae in our sample (see Section 4.5 and Figure 10) – although at $\leq 1.2\sigma$, the difference we find is not statistically significant. For $R_V$, Brout & Scolnic (2021) reported peak values of $2.75 \pm 0.35$ and $1.50 \pm 0.25$ for low- and high-mass hosts, respectively – a difference of $2.9\sigma$ – with fairly large population standard deviations of $1.3 \pm 0.2$ for both subsamples. When not splitting by host galaxy mass, they estimate a mean $R_V$ of $2.0 \pm 0.2$, and a similarly wide population distribution with $\sigma_R = 1.4 \pm 0.2$.

Under our No-Split model, we find an $R_V$ population distribution with mean parameter $\mu_R = 2.70 \pm 0.25$. For this configuration, a wide population distribution of $R_V$ is strongly disfavoured, with 68 (95) per cent of posterior probability falling at $\sigma_R < 0.37 \pm 0.61$. 
This disfavours the lower peak value of \( \mu_R = 2.0 \pm 0.2 \), and wide distribution with \( \sigma_R = 1.4 \pm 0.2 \) reported by Brout & Scolnic (2021).

The bottom left panel of Figure 7 shows the joint posterior of \( \mu_R, \sigma_R \) and \( \tau_A \) (which is insensitive to the adoption of a population distribution of \( R_V^a \)) for the No-Split model.

Under the Partial-Split results, when a split point of \( 10^{10}M_\odot \) is assumed in line with Brout & Scolnic (2021), we find \( \mu_R = 3.49 \pm 0.61 \) and \( \mu_R = 2.72 \pm 0.23 \) for supernovae in low- and high-mass hosts, respectively. For the more massive hosts, we find that \( \sigma_R < 0.32 \) (0.62) with 68 (95) per cent posterior probability. For low-mass hosts, our constraining power is weakened (especially for this choice of split), with the 68 and 95th percentiles of our posterior falling at \( \sigma_R = 0.89 \) and \( \sigma_R = 1.89 \), respectively. The joint \( \mu_R, \sigma_R \) posterior for low-mass hosts is consistent with its more tightly constrained high-mass counterpart. For host galaxies with \( M_* \geq 10^{10}M_\odot \), these results disfavour the wide population distribution of \( R_V^a \) with \( \sigma_R = 1.3 \pm 0.2 \) estimated by Brout & Scolnic (2021). For host galaxies with \( M_* < 10^{10}M_\odot \), such a diffuse distribution cannot be strongly ruled out, but it is less probable than a distribution with low \( \sigma_R \).

As in the global dust law configuration, our ability to constrain the low-mass host dust behaviour is improved when the sample is split at the median host mass. In this case, the joint posterior of \( \mu_R \) and \( \sigma_R \) for low-mass hosts is more compressed than for the \( 10^{10}M_\odot \) choice of mass split. This brings it much closer to its high-mass counterpart. Here, we find \( \mu_R = 2.97 \pm 0.37 \) for low-mass hosts, and \( \mu_R = 2.66 \pm 0.27 \) for high-mass hosts. For the low-mass subsample, 68 (95) per cent of our posterior probability lies at \( \sigma_R < 0.55 \) (1.02), whilst for the high-mass subsample we find that \( \sigma_R < 0.35 \) (0.70) with 68 (95) per cent probability. The bottom right panel of Figure 7 illustrates the joint posterior of \( \mu_R, \sigma_R \) and \( \tau_A \) for the Partial-Split model with the median host mass as split point. For a brief discussion of the prior sensitivity of \( \sigma_R \), see Appendix B.

For all choices of mass split, we obtain \( \mu_R \) values consistent with the values of \( R_V \) obtained in our global dust law analyses. The \( \mu_R \)'s across the mass split are consistent within \( \leq 1.2 \sigma \). However, the smallest \( \Delta M_0 = 0.042 \pm 0.028 \) (1.5\( \sigma \)) occurs for the \( 10^{10}M_\odot \) mass split, where we estimate \( \Delta \mu_R = 0.77 \pm 0.65 \), larger than for the median or MLE mass split. Although this estimated difference in \( \mu_R \) is consistent within the statistical uncertainties with zero, it may be suggestive of some possible trade-off between \( \Delta \mu_R \) and the size of the mass step, which is qualitatively similar to the trade-off inferred by Brout & Scolnic (2021). However, we caution that, for this mass split, our constraint on \( \Delta \mu_R \) is more uncertain than for other splits, due to the paucity of more reddened SNe in low-mass hosts. As in the global dust law analysis, \( \mu_R \leq 2.2 \) is disfavoured on both sides of the mass split in all cases. This disfavours the low \( \mu_R = 1.50 \pm 0.25 \) estimated by Brout & Scolnic (2021) for high-mass host galaxies. Our \( \tau_A \) inferences are insensitive to the inclusion of a population distribution of \( R_V^a \).

### 4.4 Distribution of SED Shape Parameters

It is interesting to investigate how the SED shape parameter, \( \theta_1 \), is distributed for the low and high mass subsamples. It is only meaningful to do this for the No-Split or Partial-Split models, where the \( W_1 \) component modulated by \( \theta_1 \) is common to both sides of the mass split. We will focus on the Partial-Split results here, although those from the No-Split model are almost identical.

For all choices of mass split, the SNe in low-mass hosts tend to have \( \theta_1 \) values which are more concentrated between \( -1.5 \) and 0 (i.e. at the bright, slow-declining end of the population), albeit with a tail extending to positive \( \theta_1 \). High-mass host galaxies seem to have a broader, flatter distribution, with fairly even numbers of SNe falling either side of \( \theta_1 = 0 \). The top panel of Figure 8 shows histograms of the posterior means of \( \theta_1 \) from a Partial-Split analysis where the sample was divided at the median host galaxy mass.

We compare this to the behaviour of the SALT2 stretch parameter, \( x_1 \), which captures similar behaviour to BayesSN’s \( \theta_1 \), with the two being shown to be highly correlated by M20. Indeed, this holds true for the Foundation sample – the top panel of Figure 9 plots SALT2’s \( x_1 \) (from Foley et al. 2018; Jones et al. 2019) against BayesSN’s \( \theta_1 \) for the 157 supernovae in our Foundation cut. We see a very similar correlation to the one observed in M20. The correlation extends to the way \( \theta_1 \) and \( x_1 \) are distributed for the low- and high-mass subsamples – the bottom panel of Figure 8 plots histograms of \( x_1 \) for the two populations. The bins are plotted from high to low stretch, and scaled so that they are approximately equivalent to the \( n_{\text{th}} \) bin of the \( x_1 \) histogram should contain supernovae with similar light curve widths and decline rates to those in the \( n_{\text{th}} \) bin of the \( \theta_1 \) histogram. Particularly for the low-mass subsample, the \( x_1 \) and \( \theta_1 \) distributions are very similar, with the \( x_1 \) values being concentrated towards the high-stretch, bright and slow-declining supernovae, with a small number of negative stretch cases. This similarity holds true for all choices of mass split point.

These results agree with expectations based on previous works (e.g.Neill et al. 2009; Sullivan et al. 2010; Childress et al. 2013; Rigault et al. 2013; Kelsey et al. 2021). It is typical to see a tighter stretch distribution in low-mass hosts, which tend to be dominated by the brighter, slow-declining, supernovae, with the dim, fast-decliners only appearing frequently in more massive galaxies. In fact, when we explicitly plot our \( \theta_1 \) values against host galaxy mass (Figure 10, top panel), we see exactly the kind of L-shaped distribution that has been identified in SALT2’s \( x_1 \) previously (Rigault et al. 2013), with the quadrant of parameter space corresponding to fast-declining supernovae in low-mass hosts being very sparsely populated.

### 4.5 Dust Reddening and Intrinsic Colour

Although we have already considered the host extinction population distribution implied by our results, we can also derive explicit values of the inferred reddening due to dust, \( E(B-V) = A_V/R_V \), for all of the supernovae in our sample. This allows a more direct comparison to the SALT2 colour parameter, \( c \). The bottom panel of Figure 9 plots SALT2 colour, \( c \) (from Foley et al. 2018; Jones et al. 2019), against \( E(B-V) \) from the Partial-Split BayesSN model. We see that the parameters correlate, albeit not as strongly as \( x_1 \) and \( \theta_1 \) (top panel). For supernovae with low reddening due to dust, we
would expect the fitted SALT2 colour (an apparent colour that rolls intrinsic and extrinsic effects into a single value – see Mandel et al. 2017, for extensive discussion of this effect) to be dominated by the intrinsic supernova colour, making the deterioration of the \( c \) vs. \( E(B - V) \) relation at \( E(B - V) \lesssim 0.1 \) unsurprising. For the more extinguished supernovae, the contribution of dust reddening to \( c \) becomes dominant, giving rise to the scattered correlation we can see in Figure 9.

It is not immediately obvious from Fig. 9 if \( E(B - V) \) correlates with host galaxy mass although we might expect it to from the behaviour of SALT2’s \( c \) parameter in previous analyses (e.g. Sullivan et al. 2010; Childress et al. 2013; Kelsey et al. 2021), which typically see the reddest supernovae only in massive hosts. If we directly plot our computed \( E(B - V) \) values against host mass (Figure 10, middle panel), some evidence for this behaviour begins to emerge, with very few supernovae with \( E(B - V) \gtrsim 0.15 \) occurring in hosts less massive than around \( 10^{10} \text{M}_\odot \). This mass range is fairly sparsely populated, however, so it is hard to say if this is a genuine difference or a statistical fluctuation in this particular sample.

Since BayesSN separately models intrinsic variation and extrinsic dust effects, we can also investigate if intrinsic supernova colour correlates with host galaxy mass – something which would be difficult to untangle from dust reddening in an analysis based on SALT2 colour. For each supernova in the sample, we compute the posterior mean and standard deviation of its intrinsic \( g - r \) colour at peak (i.e. \( \epsilon_0, \theta_1, W_1, \) and \( \epsilon_g \)). The bottom panel of Figure 10 plots these derived colours against host galaxy stellar mass. From this, there perhaps appears to be a greater diversity of intrinsic colours found in the more massive galaxies, with these seeming to host both redder and bluer supernovae than their low-mass counterparts. As with \( \theta_1 \) and \( E(B - V) \), however, there are relatively few supernovae below the mass where the behaviour appears to change most strikingly, so such conclusions should be regarded cautiously until analyses are completed on a sample which is even better populated in this mass range.

Figure 8. (top panel) Histograms showing how the shape parameter \( \theta_1 \) – controlling how strongly the principal SED component, \( W_1 \), is present for a given supernova – is distributed for the low- and high-host mass subsamples of Foundation DR1. (bottom panel) Histograms showing the distribution of the SALT2 stretch parameter \( x_1 \) (from Foley et al. 2018; Jones et al. 2019) in the same subsamples. The bins here are those in the top panel, scaled to cover equivalent ranges of parameter space.

Figure 9. Plots of SALT2 parameters against analogous BayesSN quantities, colour coded by host galaxy stellar mass. Values of SALT2 parameters from Foley et al. (2018); Jones et al. (2019). (top panel) SALT2 \( x_1 \) parameter against the BayesSN \( \theta_1 \) SED shape parameter. (bottom panel) SALT2 colour, \( c \), plotted against \( E(B - V) = A_V / R_V \) dust reddening derived from BayesSN posterior estimates of \( A_V \) and \( R_V \).
For a given training and resubstitution run, we can construct a Hubble diagram plotting photometric distances, $\mu^\text{phot}_s$, against CMB frame redshifts, $z^\text{CMB}$. During training, we employed external distance constraints (based on a fiducial $\Lambda$CDM cosmology from Riess et al. 2016) centred on $\mu^\text{phot}_0 = \mu^\text{phot}_{\Lambda CDM}(z^\text{CMB}; H_0 = 73.24, \Omega_M = 0.28, \Omega_\Lambda = 0.72)$, with uncertainties derived from peculiar velocity (assuming $\sigma_{pec} = 150\text{ km s}^{-1}$; Carrick et al. 2015) and spectroscopic redshift uncertainties. To assess the accuracy of our estimates, we compute Hubble residuals for each set of photometric distances from $\mu^\text{phot}_s - \mu^\text{ext}_s$. Figure 12 shows Hubble diagrams (upper panels) and residuals (lower panels) constructed in this way from the training and resubstitution of the Partial-Split (left column) model, assuming the median host mass as its split point.

For comparison, in the right column we plot a Hubble diagram computed by applying a linear Tripp (1998) standardisation,

$$
\mu_s = m^s_B - M_B + \alpha x^s_1 - \beta c^s - \gamma \delta s,
$$

to the SALT2 (Guy et al. 2007, 2010; Retoule et al. 2014) fit results ($m^s_B, x^s_1, c^s$: $B$-band apparent magnitude, stretch, and colour, respectively) from Foley et al. (2018); Jones et al. (2019). Here $I^s_2$ is an indicator variable which obeys $I^s_2 = 1$ if $M_{e,s} \geq 10^{10.331}M_\odot$ and $I^s_2 = 0$ otherwise. We fit the values of the Tripp parameters to the full sample and find $M_B = -19.201, \alpha = 0.125, \beta = 2.834, \gamma = 0.056$, using a simple MCMC sampling of the Tripp distance likelihood using the SALT2 fit parameters and external distance constraints\(^3\).

When fitting the low- and high-mass host subsamples independently, we find, for all choices of the mass split, the $\sigma$ and $\beta$ coefficients are consistent within the uncertainties with the full-sample values, while the differences in $M_B$ values capture the mass step. Apart from the light curve fit uncertainties and peculiar velocity uncertainties, the SALT2 Hubble diagram has a residual scatter of $\sigma_{res} = 0.134$.

For a given run, we quantify the accuracy of our photometric distance estimates using two metrics: the root mean square (RMS) of the Hubble residuals, and the level of scatter ($\sigma_{pv}$) unaccounted for by an assumed peculiar velocity uncertainty of $\sigma_{pec} = 150\text{ km s}^{-1}$ (computed as in M20, eq. 32). Table 4 summarises the values of the two scatter statistics under our different model configurations. In the latter case, the statistics are also computed separately for the portions of the sample above and below the chosen split point. The RMS and $\sigma_{pv}$ values for the SALT2 Hubble diagram are printed on the right hand panel of Figure 12.

Under training and resubstitution, all No-Split and Partial-Split BayesSN configurations tested yield a total raw RMS training error $< 0.125$ mag, with the dispersion after removal of the contribution from peculiar velocity uncertainties being $< 0.115$ mag in all cases. For the median mass split (shown in Figure 12), we find an RMS ($\sigma_{pv}$) of 0.121 (0.110) mag. These results compare favourably to the larger scatter found in the Hubble diagram derived from the SALT2 results, which shows an RMS ($\sigma_{pv}$) of 0.161 (0.151) mag. As in Mandel et al. (2020), we sample with replacement 1,000 bootstrapped sets of Hubble residuals from the BayesSN and SALT2 photometric distance results, computing the $\Delta$RMS each time. For the two Hubble diagrams shown in Figure 12, we find that $\Delta$RMS $= 0.041 \pm 0.009$ mag. As well as improving on the overall Hubble diagram scatter, by adopting a Bayesian approach, and

\(^3\)A Bayesian method correcting for regression dilution from covariate measurement error (Kelly 2007, see also March et al. 2011) obtains $\alpha = 0.13 \pm 0.01, \beta = 3.20 \pm 0.18$, with a negligible change in RMS (0.001).
Figure 11. Example photometric distance fit to two of the Foundation DR1 supernovae: ASASSN-16cs (left panels; bright, low-redshift, slower-declining) and AT2016ajl (right panels; dim, higher-redshift, faster-declining). (top panels) Fits to the griz light curves. Vertical offsets are $(0, -2, -6, -12)$ for $(g, r, i, z)$. Indicated in the lower left of the plots are posterior predictive $p$-values (see Appendix C and Figure C1) assessing the fitness of the model. An extreme $p$-value close to 0 or 1 would indicate that the model may not be a good fit to the data. (bottom panels) Joint posterior distributions over $\theta$, $\mu_{\text{phot}}$, and $A_V$. The photometric distance estimates used for constructing our Hubble diagram are obtained by marginalising over $\theta$, $\mu_{\text{phot}}$, and $A_V$(as in the second panel along the diagonal of the corner plot).

marginalising over the other latent parameters in obtaining its distance estimates, BayesN naturally yields well motivated distance uncertainties. Under the No-Split model, the distribution of scaled errors, $(\mu_{\text{phot}} - \mu_{\text{ext}})/\sigma_{\mu_{\text{phot}}}$, computed from our photometric distances, $\mu_{\text{phot}}$, and their uncertainties, $\sigma_{\mu_{\text{phot}}}$, has a standard deviation of 1.07 – close to the ideal value of 1. The results from the Partial-Split analyses are similar. Such agreement is only achieved in the SALT2 results if the fitted residual scatter term, $\sigma_{\text{res}}$, is added in quadrature to the distance uncertainties.

One interesting result that emerges from the Partial-Split analyses is the somewhat lower level of scatter in the Hubble residuals of supernovae in massive hosts. This could suggest that, despite their greater diversity of inferred properties (see Sections 4.4–4.5 and Figure 10), we are more able to effectively standardise supernovae in massive hosts and obtain accurate distances to them. Splitting at the median host mass (which gives the greatest difference of the split choices we consider here), the Hubble diagram RMS for supernovae in high-mass hosts is 0.033 mag lower than in low-mass hosts. Applying bootstrap resampling to estimate the RMS uncertainty in the two subsamples suggests that this is a roughly $2\sigma$ difference, so is not hugely significant. For the difference choices of mass split, the difference in Hubble diagram scatter between mass bins appears to align with the difference in the value of $\sigma_0$ (the level of time- and wavelength-independent residual brightness scatter; see Section 4.2) estimated during training. This is because the mode
of SED variation described by \( \sigma_0 \), contributes to the photometric distance uncertainty. In the SALT2 Hubble diagram, we estimate \( \Delta \text{RMS} = 0.024 \pm 0.018 \) mag between low- and high-mass hosts, a 1.4\( \sigma \) difference in the same direction as BayeSN.

The results from our Full-Split analyses (see Table 4) are largely similar to the Partial-Split results. The Hubble diagram dispersion for the median and MLE mass splits is fairly consistent between the Full-Split and Partial-Split runs, for both the low- and high-mass subsamples. The Full-Split configuration should be invulnerable to the inference of the intrinsic SED parameters \((W_0, W_1, \Sigma_x)\) being driven by supernovae in one mass bin at the expense of the characterisation of supernovae in the other. The consistency between our Full-Split and Partial-Split results thus disfavours this as the cause of the imbalance in Hubble diagram scatter between mass bins in the Partial-Split runs.

4.6.2 Cross-Validation

Similarly to previous supernova studies (Mandel et al. 2009, 2011, 2020; Blondin et al. 2011), we perform cross-validation (CV) to test the sensitivity of the model to the finite training set and confirm that its performance is not significantly overfit by the re-use of our training set in the resubstitution process. As in M20, we perform 10-fold cross-validation. We retrain our models ten times, each time leaving out 10 per cent of the data, which act as an unseen test set for that model. In this way, we obtain photometric distance estimates for all 157 supernovae, each one obtained from a model whose training set excluded it. The cross-validated values of our Hubble diagram scatter metrics for the No-Split and Partial-Split models are given in the bottom rows of Table 4. Under cross-validation, we find that the RMS (\( \sigma_{pv} \)) for the No-Split model rises to 0.134 (0.122), a small increase of 0.01 mag by both scatter metrics—comparable to what was seen in the M20 cross-validation analysis.

For the Partial-Split model, the total RMS increase under cross-validation is \( \lesssim 0.013 \) mag. This generally holds on both sides of the mass split, although for the \( 10^{10} M_{\odot} \) mass split the increase on the low-mass side is slightly larger at 0.017 mag. The increase in the total value of \( \sigma_{pv} \) is consistently \( \lesssim 0.014 \) mag. For the median mass split, we find an RMS (\( \sigma_{pv} \)) of 0.132 (0.122) mag under cross validation. Since this model under CV is trained on 10% less...
data than the same model under resubstitution, we would expect a somewhat less accurate model under CV. Thus, we expect the true RMS out-of-sample error rate of the model trained on the full sample would be somewhere between the two, i.e. \( \approx 0.127 \) mag.

For the BayesSN cross-validation, we repeatedly retrain a new BayesSN SED model on the training set excluding each held-out fold. Although we can not retrain the full SALT2 SED model in the same fashion, we can partially cross-validate the SALT2 Hubble diagram. We refit the Tripp (1998) formula (Eq. 9) to the same 9 of 10 training folds as in our BayesSN analysis, using the \((\alpha, \beta, M_R, \gamma)\) each time to compute the photometric distances in the held-out fold. The SALT2 RMS rises from 0.161 to 0.166 mag under this partial cross-validation. Applying the same bootstrapping procedure as in section 4.6.1, we find ARMS = 0.034 ± 0.009 between the cross-validated SALT2 and BayesSN Hubble diagrams, a 3.8σ improvement. This improvement in the accuracy of photometric distance estimates is a good indication that the BayesSN SED model is an effective statistical description of the \(gri\) SN Ia light curves of the Foundation sample.

### 4.7 Simulation-Based Validation

As a sanity check of our code, we conduct simulation-based tests. We simulate Foundation DR1-like \(gri\) light curve datasets from the BayesSN forward model. We then train new BayesSN models on these simulations to check that the original parameters are recovered.

For example, to check our \(R_V\) inferences, we generate simulations with all population hyperparameters \((W_0, W_1, \Sigma_e, \sigma_{09}, \tau_A)\), except for \(R_V\), fixed to their posterior mean values from the No-Split analysis. We generated 3 simulations, each with a different input true global host dust law \(R_V = (2.0, 3.0, 4.0)\). We then trained new BayesSN models on these simulated data. We successfully recovered posterior estimates \(\hat{R_V} = (1.88 \pm 0.11, 2.87 \pm 0.19, 4.27 \pm 0.34)\) after training on the simulated datasets.

Additionally, we generated further simulations with input \(R_V\) distributions with population parameters \((\mu_R, \sigma_R) = (2.0, 1.0)\) and \((3.0, 1.0),\) to verify that we would be able to detect a wide population distribution (la Brout & Scolnic 2021) in a Foundation-like sample. We repeated our population inference on these new simulated datasets. For the input population with \((\mu_R, \sigma_R) = (2.0, 1.0),\) our posterior estimates of the population parameters were \((\hat{\mu}_R, \hat{\sigma}_R) = (2.30 \pm 0.42, 1.59 \pm 0.47).\)

### 5 CONCLUSIONS

We have used the \(gri\) light curves of 157 Type Ia supernovae from the first Foundation Supernova Survey data release (Foley et al. 2018; Jones et al. 2019) to train a BayesSN SN Ia model, continuous over 0.35–0.95 \(\mu\)m. By training on a homogeneous SN sample – with all data in Foundation DR1 having been obtained on the Pan-STARRS-1 system – we ensure that the resulting model is affected by minimal cross-calibration systematics. This is the first statistical model for SN Ia in the rest-frame \(z\)-band to have been trained on \(z\)-band light curves. Foundation DR1 is the first large homogeneous sample of rest-frame \(z\)-band observations of SNe Ia, and was not available to previous continuous SED models (e.g. SALT2; Guy et al. 2007, 2010; Betoule et al. 2014) or models for light curves in discrete passbands (e.g. Mandel et al. 2009, 2011; and SNooPy; Burns et al. 2011). By training on rest frame \(z\)-band data, we improve on the M20 SED model at these wavelengths. Using our model, we compute photometric distance estimates for the Foundation sample, presenting the first Hubble diagram based on full rest-frame \(gri\) light curves. The value added by the \(iz\)-band data, and the advantages of our modelling approach, combine to yield an excellent total RMS of 0.121 mag. This result is robust under 10-fold cross-validation, where the total RMS only rises to 0.132 mag.

The BayesSN framework incorporates the host galaxy dust extinction law (assuming Fitzpatrick 1999) at the SED level, properly modelling its impact at all phases and wavelengths. In this way, it leverages the colour–luminosity information in the full \(gri\) light curves of the sample, enabling robust inferences about the \(R_V\) value(s) parametrizing the dust law. When fitting for a single global dust law for the full Foundation sample, we estimate \(R_V = 2.61 \pm 0.21\). This result, based on an independent sample, aligns well with many previous analyses of other low-\(z\) SN Ia samples with apparent colours con-

---

**Table 4.** Hubble diagram scatter for the 157 Foundation supernovae in our sample.

| Schemea | Model | Split Point | RMSb | σc,pvc<sup>c</sup> |
|---------|-------|-------------|------|-------------------|
|         |       |             | All  | LM    | HM    | All  | LM    | HM    |
| Resubstitution | No-Split | - | 0.124 | - | - | 0.112 | - | - |
|          | Partial-Split | 10<sup>10</sup> M<sub>0</sub> | 0.121 | 0.125 | 0.119 | 0.109 | 0.115 | 0.106 |
|          | Median | 0.121 | 0.136 | 0.103 | 0.110 | 0.125 | 0.096 | |
|          | MLE | 0.118 | 0.129 | 0.106 | 0.107 | 0.117 | 0.098 | |
|          | Full-Split | 10<sup>10</sup> M<sub>0</sub> | - | 0.166 | 0.123 | - | 0.158 | 0.111 |
|          | Median | - | 0.127 | 0.104 | - | 0.115 | 0.097 | |
|          | MLE | - | 0.128 | 0.109 | - | 0.117 | 0.101 | |
| Cross-Validation | No-Split | - | 0.134 | - | - | 0.122 | - | - |
|          | Partial-Split | 10<sup>10</sup> M<sub>0</sub> | 0.134 | 0.142 | 0.130 | 0.123 | 0.133 | 0.118 |
|          | Median | 0.132 | 0.149 | 0.113 | 0.122 | 0.138 | 0.106 | |
|          | MLE | 0.129 | 0.142 | 0.115 | 0.119 | 0.131 | 0.107 | |

* a Denotes whether photometric distances were obtained through direct training and resubstitution of all supernovae, or via 10-fold cross-validation (see section 4.6).

* b Total root mean square (RMS) of the Hubble residuals.

* c Dispersion when expected variance due to peculiar velocity uncertainty (assumed to be \(\sigma_{pec} = 150 \text{ km s}^{-1}\); Carrick et al. 2015) is taken out (as in Mandel et al. 2020, eq. 32).
sistent with the cosmological cut (peak apparent $B - V \leq 0.3$) (e.g. Riess et al. 1996; Phillips et al. 1999; Folatelli et al. 2010; Chotard et al. 2011; Foley & Kasen 2011; Mandel et al. 2011, 2017, 2020; Burns et al. 2014; Saxdell et al. 2016; Léget et al. 2020; Arima et al. 2021), some of which utilised spectroscopic or NIR data to help break the degeneracy between intrinsic variation and dust effects. The recent SN Ia “sibling” analysis of Biswas et al. (2021) estimated a similar $R_V = 2.5$, albeit for a slightly more reddened supernova (AT2019lic; $c_{\text{SALT2}} = 0.57 \pm 0.04$, $E(B - V)_{\text{SNooPy}} = 0.63 \pm 0.07$, $\log_{10}(M_*/M_\odot) = 10.46_{-0.41}^{+0.11}$) than we have considered here. Allowing a population distribution of $R_V$ values, parametrized by a mean, $\mu_R$, and standard deviation, $\sigma_R$, we find $\mu_R = 2.70 \pm 0.25$ and a strong preference for small dispersion, i.e. $\sigma_R < 0.61$ at 95% posterior probability, consistent with our global dust law results.

Beyond this, we also used the Foundation data to train a modified version of the BayeSN model which separates several key parameters by host galaxy stellar mass, including the average dust extinction, and the dust law $R_V$ (or parameters describing its population distribution). When assuming a single $R_V$ for each mass group, we estimate $R_V = 2.84 \pm 0.31$ for host galaxies less massive than the median, and $R_V = 2.58 \pm 0.23$ for those more massive. Both are consistent with our single global value of $R_V = 2.61 \pm 0.21$. For choices of mass split other than the median, we still infer values of $R_V$ between low- and high-mass hosts that are consistent within $\pm 1.2 \sigma$ (see Table 2). In all cases, we find that $R_V \leq 2.2$ is excluded from the posterior at the 95 per cent level, and that the results tend to be consistent with the global-$R_V$ result. When a population distribution of $R_V$ values is allowed either side of the median mass split, we find a mean of $\mu_R = 2.97 \pm 0.37$ for less massive hosts, and $\mu_R = 2.66 \pm 0.27$ for more massive hosts, both consistent with our mass-agnostic analysis. For all choices of mass split, our $\mu_R$ estimates are consistent to within $\sim 1.2 \sigma$ across the split. In all cases, the mean values are consistent with their counterparts from the global-$R_V$ analyses, and values of $\mu_R \leq 2.2$ tend to be strongly disfavoured on both sides of the mass split. We find that the width $\sigma_R$ of the $R_V$ population distribution is consistent with small values (see lower panels of Fig. 7). For high-mass hosts, or for the whole population when host mass is ignored, wide distributions of $R_V$ are disfavoured, with $\sigma_R \leq 0.7$ with at least 95% posterior probability in all cases. For low-mass hosts, constraining power is weakened by having smaller sample sizes, and fewer SNe with $E(B - V) \gtrsim 0.15$, so a wide $R_V$ population distribution cannot be completely ruled out. While it cannot be ruled out that some individual SNe with low-to-moderate reddening may be affected by dust laws with unusually low $R_V$ (e.g. SN 2012et with $R_V \approx 1.7^{+0.6}_{-0.3}$; c.f. caveats in Amanullah et al. 2015), we find no strong evidence in this sample that this is true for the aggregate, which can be adequately explained with $R_V = 2.5$–3. Our population mean estimates, $\mu_R$, are slightly below the average Milky Way value (e.g. Schlafly et al. 2016) by at most $2.4 \sigma$.

Looking at the extinction population distribution either side of the median host galaxy mass, we estimate an average host galaxy dust extinction of $\tau_A = 0.19 \pm 0.03$ for low masses, and $\tau_A = 0.21 \pm 0.03$ for high masses. Our estimates of $\tau_A$ are consistent to within $\lesssim 1.2 \sigma$ between low- and high-mass host galaxies for all choices of mass split, and are largely insensitive to whether a single $R_V$ or a population distribution is assumed within each subsample.

Even allowing for dust to differ between low- and high-mass galaxies, we still find evidence for some level of mass step-like behaviour. We see a magnitude offset between supernovae in low- and high-mass host galaxies which suggests that the latter group are brighter on average by a $\Delta M_0$ between 0.04 and 0.07 mag. This range is consistent with the size of mass steps found by previous analyses at a range of mass splits from $\sim 10^{10}$–$10^{10.8}$M$_\odot$ (e.g. Kelly et al. 2010; Sullivan et al. 2010; Betoule et al. 2014; Roman et al. 2018; Uddin et al. 2020; Kelsey et al. 2021; Smith et al. 2020), including the Jones et al. (2019) study of the Foundation data. For all choices of mass split point, the offset we estimate is consistent with the corresponding Hubble residual step seen in our joint, mass-agnostic, analysis.

Brout & Scolnic (2021) invoked differing $R_V$ distributions between low- and high-mass host galaxies as a possible cause of the mass step. They reported that wide $R_V$ distributions ($\sigma_R = 1.3 \pm 0.2$) centred on $\mu_R = 2.75 \pm 0.35$ in low-mass hosts and $\mu_R = 1.50 \pm 0.25$ in high-mass hosts could explain away the ~ 0.06 mag mass step they otherwise saw in their data. In our analysis of the Foundation data, we find that consistency of $R_V$ between low- and high-mass hosts is favoured, with a preference for narrow population distributions. While the estimated difference in mean $R_V$ is more uncertain when the sample is split at $10^{10}$M$_\odot$ (due to the paucity of more reddened SNe in low-mass hosts), we infer that a non-zero mass step and small difference in mean $R_V$ are preferred for all split choices. A plausible explanation for the mass step should account for its empirical observation over a range of mass splits. Consequently, our results disfavour significantly different $R_V$ distributions between low- and high-mass galaxies as an explanation of the mass step. A lack of correlation between the mass step and host galaxy dust properties has also been suggested by recent work (Ponder et al. 2020; Uddin et al. 2020) finding mass steps in the near-infrared (where sensitivity to dust properties should be minimised) comparable to those in the optical (but see also Johansson et al. 2021). This is further pointed to by the results of González-Gaitán et al. (2020), who find that a mass step persists even when separate apparent colour–luminosity relations are allowed in low- and high-mass galaxies. Complementary evidence for a non-dust explanation of SN Ia host galaxy dependence is also provided by the Hubble residual v.s. specific star formation rate (sSFR) results of Hand et al. (2021, §5.6), and by recent work favouring two progenitor populations (“prompt” and “delayed”; see e.g. Mannucci et al. 2005, 2006; Scannapieco & Bildsten 2005) as the cause of the mass/local sSFR step (Briday et al. 2021).

In this paper, we have presented analysis and evidence demonstrating that allowing for differences in dust properties in low- and high-mass host galaxies does not explain the host-mass step of SN Ia magnitudes in the Foundation DR1 sample. While this is a relatively large, homogeneous, and excellent nearby sample, owing to the well-known problem of induction, we cannot rule out the possibility that other supernova samples from other surveys may exhibit different characteristics. Future investigations into these scientific questions and the refinement of our statistical modelling and inference techniques will both be greatly enhanced by the growth of well-calibrated SN Ia datasets. In particular, future data releases from the Foundation Supernova Survey and the Young Supernova Experiment (YSE; Jones et al. 2021) will increase the already large sample of low-$z$ SNe Ia with Pan-STARRS light curves. Complementing this, the Carnegie Supernova Project-II (CSP-II Phillips et al. 2019) will augment the previous CSP-I data releases (Contreras et al. 2010; Stritzinger et al. 2011; Krisciunas et al. 2017) to provide further high quality observations at $z \lesssim 0.1$. This will include vital near-infrared photometry, which can be particularly valuable in untangling the complex puzzle of supernova-host correlation (Ponder et al. 2020; Uddin et al. 2020; Johansson et al. 2021). With larger samples, we will be able to determine more surely whether any small estimated differences between the dust properties of low- and high-mass hosts become statistically significant.

In future work, more sophisticated modelling of SN Ia correlations with host mass – e.g. functional regression to examine the time-

MNRAS 000, 1–22 (2021)
and wavelength-dependence of SED components against host mass – is a natural next step, and will be possible within the BayesSN framework. As well as this, we will be able to apply our analysis scheme to study correlations with other global or local environmental properties (e.g. star formation rate, Sullivan et al. 2006; Lampeitl et al. 2010; D’Andrea et al. 2011; Childress et al. 2013; Campbell et al. 2016; Uddin et al. 2017; Rigault et al. 2013, 2015, 2020; colour, Roman et al. 2018; metallicity, Gallagher et al. 2008; D’Andrea et al. 2011; Hayden et al. 2013; Pan et al. 2014; Campbell et al. 2016; Moreno-Raya et al. 2016a,b; age, Gallagher et al. 2008; Gupta et al. 2011; Pan et al. 2014; Campbell et al. 2016; Rose et al. 2019, 2021; or morphology Hamuy et al. 1996a, 2000; Sullivan et al. 2003; Hicken et al. 2009b; Puzhinskaya et al. 2020). Incorporation of spectroscopic indicators of intrinsic colour (e.g. ejecta velocity, Foley & Kasen 2011; Dettman et al. 2021) may provide additional insight. We are working to carefully merge the present SED model with that of M20 extending through NIR H-band, and also to extend it to the ultraviolet.

The parallel development of both the data and modelling techniques will be critical to a future understanding of SNe Ia and their correlations with their host galaxies. The proper understanding and modelling of these effects will in turn be critical to fully exploit future data from the Nancy Grace Roman Space Telescope’s supernova surveys, and Vera C. Rubin Observatory’s Legacy Survey of Space and Time.

ACKNOWLEDGEMENTS

We thank the anonymous referee for their helpful comments. We thank Dan Scolnic and Dillon Brout for useful discussions. We also thank the Foundation Supernova Survey team for their work in producing and making public the DR1 dataset.

ST was supported by the Cambridge Centre for Doctoral Training in Data-Intensive Science funded by the UK Science and Technology Facilities Council (STFC). KSM acknowledges funding from the European Research Council under the European Union’s Horizon 2020 research and innovation programme (ERC Grant Agreement No. 101002652). This project has been made possible through the ASTROSTAT-II collaboration, enabled by the Horizon 2020, EU 7th Framework Programme (Grant Agreement No. 873089). Support for this work was provided by NASA through the NASA Hubble Fellowship grant #HF2-51462.001 awarded by the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., for NASA, under contract NAS5-26555. SMW was supported by the STFC. GN was supported by the University of Illinois at Urbana-Champaign and the Center for Astrophysical Surveys at the National Center for Supercomputing Applications.

This work made use of the Illinois Campus Cluster, a computing resource that is operated by the Illinois Campus Cluster Program (ICCP) in conjunction with the National Center for Supercomputing Applications (NCSA) and which is supported by funds from the University of Illinois at Urbana-Champaign.

DATA AVAILABILITY

The data for the 180 supernovae in the Foundation DR1 cosmology sample (Foley et al. 2018; Jones et al. 2019) are publicly available at https://github.com/djones1040/Foundation_D1. Code and model files will be made available at https://github.com/bayesn.

REFERENCES

Abbott T. M. C., et al., 2019, ApJ, 872, L30
Amanullah R., et al., 2014, ApJ, 788, L21
Amanullah R., et al., 2015, MNRAS, 453, 3300
Arina N., Doi M., Morokuma T., Takenashi N., 2021, preprint, (arXiv:2101.02467)
Astier P., et al., 2006, A&A, 447, 31
Avelino A., Friedman A. S., Mandel K. S., Jones D. O., Challis P. J., Kirshner R. P., 2019, ApJ, 887, 106
Barone-Nugent R. L., et al., 2012, MNRAS, 425, 1007
Betancourt M., 2016, preprint, (arXiv:1601.00225)
Betoule M., et al., 2014, A&A, 568, A22
Biswas R., et al., 2021, preprint, (arXiv:2103.16978)
Blondin S., Mandel K. S., Kirshner R. P., 2011, A&A, 562, A81
Briday M., et al., 2021, preprint, (arXiv:2109.02456)
Brout D., Scolnic D., 2021, ApJ, 909, 26
Brout D., et al., 2019, ApJ, 874, 150
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
Butler N., et al., 2012, in McLean I. S., Ramsay S. K., Takami H., eds, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 8446, Ground-based and Airborne Instrumentation for Astronomy IV, p. 844610, doi:10.1117/12.926471
Campbell H., Fraser M., Gilmore G., 2016, MNRAS, 457, 3470
Carpenter B., et al., 2017, J. Statistical Software, 76, 1
Carrick J., Turnbull S. J., Lavaux G., Hudson M. J., 2015, MNRAS, 450, 317
Chambers K. C., et al., 2016, preprint, (arXiv:1612.05560)
Childress C., et al., 2013, ApJ, 770, 108
Chotard N., et al., 2011, A&A, 529, L4
Conley A., et al., 2011, ApJS, 192, 1
Contreras C., et al., 2010, AJ, 139, 519
Currie M., Rubin D., Aldering G., Deustua S., Fruchter A., Perlmutter S., 2020, preprint, (arXiv:2007.02458)
D’Andrea C. B., et al., 2011, ApJ, 743, 172
Dark Energy Survey Collaboration et al., 2016, MNRAS, 460, 1270
Dettman C. M., et al., 2021, preprint, (arXiv:2102.06524)
Dhawan S., Jha S. W., Leibundgut B., 2018, A&A, 609, A72
Draine B. T., 2003, ARA&A, 41, 241
Elías-Rosa N., et al., 2006, MNRAS, 369, 1880
Elías-Rosa N., et al., 2008, MNRAS, 384, 107
Filippenko A. V., Li W. D., Treffers R. R., Mazzali P. A., 2001, in Paczynski B., Filippenko A. V., 2020, preprint, (arXiv:2009.13230)
Foley R. J., et al., 2018, MNRAS, 475, 193
Foley R. J., et al., 2018, MNRAS, 475, 193
Freedman W. L., et al., 2009, ApJ, 704, 1036
Freedman W. L., et al., 2011, AJ, 141, 20
Fruchter A. S., et al., 2011, ApJ, 729, 55
Friedman A. S., et al., 2015, ApJS, 220, 9
Friedman A. S., et al., 2011, ApJ, 743, 172
Freyman J. A., et al., 2008, ApJ, 685, 752
Galbany L., et al., 2019, ApJ, 877, 106
Gallagher A., et al., 2011, ApJ, 730, 91
Gallagher J. S., Garnavich P. M., Caldwell N., Kirshner R. P., Jha S. W., Li W., Ganesalingam M., Filippenko A. V., 2008, ApJ, 685, 752
Gelman A., Meng X.-L., Stern H., 1996, Statistica Sinica, 6, 733
Gelman A., Carlin J. B., Stern H. S., Dunson D., Vehtari A., Rubin D. B., 2013, Bayesian Data Analysis, 3rd Edition. Chapman & Hall/CRC, New York
González-Gaitán S., de Jaeger T., Galbany L., Mourão A., Paulina-Afonso A., Filippenko A. V., 2020, preprint, (arXiv:2009.13230)
Gupta R. V., et al., 2011, ApJ, 740, 92
Guy J., et al., 2007, A&A, 466, 11
Guy J., et al., 2010, A&A, 523, A7
Hamuy M., Phillips M. M., Suntzeff N. B., Schommer R. A., Maza J., Aviles R., 1996a, AJ, 112, 2391
Hamuy M., et al., 1996b, AJ, 112, 2408
Hamuy M., Trager S. C., Pinto P. A., Phillips M. M., Schommer R. A., Ivanov V., Suntzeff N. B., 2000, AJ, 120, 1479

MNRAS 000, 1–22 (2021)
APPENDIX A: SEARCHING FOR A MASS STEP
As described in Section 3.3, our Partial-Split and Full-Split analyses require us to define the host galaxy stellar mass which delineates low and high masses. As well as the median, and the conventional choice of $10^{10} M_\odot$, we also try a split point motivated by the data. Specifically, this is the maximum likelihood location of a ‘mass step’ (computed subject to the constraint that the step lies within the interquartile range of host masses) for the Hubble residuals from our No-Split analysis.

For a given set of Hubble residuals, $\Delta \mu = (\Delta \mu_1, \ldots, \Delta \mu_N)$, we assume that the residuals are consistent with a step function, with mean $m_c$ for host masses of $M_* < M_{*, \text{step}}$, and mean $m_>$ for host masses of $M_* \geq M_{*, \text{step}}$. We allow for the presence of some residual scatter, $\sigma_{\text{res}}$, about the step function which covers any variance in the Hubble residuals not explained by their individual errors. The likelihood for the Hubble residual of a supernova, $s$, in a host galaxy with stellar mass $M_{*, s}$ will be

$$P(\Delta \mu_s | m_c, m_>, \sigma_{\text{res}}, \log_{10} M_{*, s})$$

which is analytic, for a given step location, $M_{*, \text{step}}$, for a sample of Hubble residuals, we compute the marginal likelihood, $P(\Delta \mu | \log_{10} M_{*, \text{step}})$, of the Hubble residuals given the step location. The joint likelihood of all the data can be reduced to a product of low and high mass likelihoods,

$$P(\Delta \mu | m_c, m_>, \sigma_{\text{res}}, \log_{10} M_{*, s}) = S_< N(m_ | \nu_<, \sigma_<^2) \times S_> N(m_ | \nu_>, \sigma_>^2),$$

where

$$\nu_< = \sigma_<^2 \left[ \sum_{M_{*, i} < M_{*, \text{step}}} \frac{\Delta \mu_i}{\sigma_i^2 + \sigma_{\text{res}}^2} \right],$$

$$\sigma_<^2 = \left[ \frac{\sum_{M_{*, i} < M_{*, \text{step}}} \frac{1}{\sigma_i^2 + \sigma_{\text{res}}^2}} {\sum_{M_{*, i} < M_{*, \text{step}}} \frac{1}{\sigma_i^2 + \sigma_{\text{res}}^2}} \right]^{-1},$$

and similarly for $\nu_>$, $\sigma_>$, $S_>$. Here, $\nu_<$ and $\nu_>$ are just the variance weighted means of the Hubble residuals either side of the step, with $\sigma_<$ and $\sigma_>$ being the standard errors on these.

The marginal likelihood we are interested in can be found by imposing priors on $m_c$, $m_>$ and $\sigma_{\text{res}}$, and integrating over these parameters. We assume conditional priors of $P(m_ | \sigma_{\text{res}}) = N(m_ | 0, 1)$ and $P(\sigma_{\text{res}} | m_c) = N(m_c | 0, 1)$, which are weakly informative over the likely range of heights either side of the step. These conjugate priors mean that the marginalisation over $m_<$ and $m_>$ is analytic, giving

$$P(\Delta \mu | \sigma_{\text{res}}, \log_{10} M_{*, \text{step}}) = S_< S_\times S_>,$$

with $S_<$ (and equivalently, $S_>$) as given by Equation (A5), and

$$S_\times = \frac{1}{2 \pi \sigma_\times^2} \exp \left[ -\frac{1}{2} \sum_i \frac{\Delta \mu_i^2}{\sigma_i^2 + \sigma_{\text{res}}^2} - \frac{\nu_<^2}{\sigma_<^2} \right],$$

and

$$S_\times = \frac{1}{2 \pi \sigma_\times^2} \exp \left[ -\frac{1}{2} \sum_i \frac{\Delta \mu_i^2}{\sigma_i^2 + \sigma_{\text{res}}^2} - \frac{\nu_>^2}{\sigma_>^2} \right] \times S_\times.$$

APPENDIX B: HYPERPRIOR CHOICE IN DUST LAW POPULATION ANALYSIS
In the versions of our analysis allowing for a population (or populations) of dust law parameters, an important choice is the prior on the $R_V$ population distribution width, $\sigma_R$. As described in Section 2.1.2, our default choice is a half-Normal prior, with a standard deviation of 2, as this penalises unreasonably large $\sigma_R$ without introducing an arbitrary hard cutoff. To test the sensitivity of our results to this choice, we repeat our analyses with two other hyperprior choices, $\sigma_R \sim \text{Half-N}(0, 1^2)$, and $\sigma_R \sim U(0, 4)$. The results of the No-Split analysis, and the high mass side of the Partial-Split analyses, are largely insensitive to the hyperprior choice. For the low mass side of the Partial-Split analysis – particularly for the more unbalanced choices of split point – where constraints on the $R_V$ population properties are weaker, prior sensitivity is increased. The lower panel of figure B1 shows the marginal posterior distribution of the low-mass value of $\sigma_R$ for the three hyperprior choices we consider (which are also plotted for comparison). These results are from the Partial-Split model, with an assumed host galaxy mass split of $10^{10} M_\odot$. This split point gives a rather unbalanced low-high-mass ratio of 48:109 (see Table 1), meaning this plot reflects one of the ‘worst-case scenarios’ for $\sigma_R$ hyperprior sensitivity. However, even in this case, we see that the posterior density is fairly insensitive between the weaker choices of a half-normal Half-N(0, 1^2) or uniform $\sigma_R \sim U(0, 4)$ hyperprior. For comparison, the upper panel of Figure B1 shows the marginal posteriors of the high-mass $\sigma_R$ when splitting at the median host mass. This case is much less prior-sensitive.

APPENDIX C: POSTERIOR PREDICTIVE CHECKS FOR PHOTOMETRIC DISTANCE FITS
When fitting a trained model to (potentially unseen) light curves for the purpose of photometric distance estimation (as in Section 4.6), a means of assessing model fitness is valuable. As well as visually comparing the posterior distribution of model light curves to the data (Figure 11), we also compute a posterior predictive $p$-value (Gelman et al. 1996; Rubin 1984; Meng 1994) as a numerical estimate of model fitness. This can be readily computed from a set of posterior samples, and provides a natural way of assessing fitness whilst averaging over the model parameters. The $p$-value is defined as the probability that a test quantity is more extreme when computed based on replicated data (simulated from the posterior predictive distribution) than it is when computed based on the data. This is well defined even when the test quantity is a function of the model parameters (as would be the case with $\chi^2$).
In our case, we choose to use

\[ \chi^2(f; \phi) = \sum_{n=1}^{N} \frac{(f_n - f_{n\text{model}}(\phi))^2}{\sigma_n^2} \]  

(C1)

as our test statistic\(^4\), where \( \phi = (\mu^\text{phot}_1, \theta^1_1, A^V_1, \delta M_1, \xi_1) \) represents our model parameters; \( f = (f_1, \ldots, f_N) \) is the vector of fluxes being tested; \( f_{n\text{model}} \) is the model flux computed at the time, and in the passband, of \( f_n \), given the parameters \( \phi \); and \( \sigma_n \) is the flux error associated with \( f_n \). For a given supernova, we have a set of \( N \) flux observations, \( f^{\text{obs}} = (f_1^{\text{obs}}, \ldots, f_N^{\text{obs}}) \), with associated errors, \( \sigma = (\sigma_1, \ldots, \sigma_N) \). Conditional on \( f^{\text{obs}} \), we will also have a set of \( M \) samples, \( (\phi_1, \ldots, \phi_M) \), from the posterior distribution

\(^4\) We have also tried the mean squared error (MSE), and found this to be a reasonable alternative.

Figure B1. Plot illustrating the effect of hyperprior choice on the inference of \( \sigma_R \) under the Partial-Split model. Solid lines show the marginal posterior distributions of \( \sigma_R \) for three choices of prior, with dashed lines plotting the corresponding priors. (top panel) Posterior distribution of \( \sigma_R \) for galaxies more massive than the population median. Prior sensitivity is negligible in this case. The orange curve corresponds to our fiducial prior choice (corresponding to the purple \( \sigma_R \) marginal distribution in Figure 7). (bottom panel) Constraints on \( \sigma_R \) for host galaxies less massive than \( 10^{10} M_\odot \), a more unbalanced mass split. This illustrates the ‘worst-case scenario’ in prior sensitivity.

Figure C1. Illustration of the posterior predictive \( p \)-value estimation for the two light curve fits in Figure 11. Each point corresponds to a single posterior sample, with the \( y \)-axes plotting posterior predictive values, \( \chi^2(f^{\text{rep}}_n(\phi_m); \phi_m) \), of our test quantity, and the \( x \)-axes showing realised values, \( \chi^2(f^{\text{obs}}_n; \phi_m) \), computed based on the data. The fraction of samples lying above the dashed 1 : 1 line gives our \( p \)-value estimate.

\[ P(\phi|f^{\text{obs}}). \] For the \( m \)th posterior sample, \( \phi_m \), we can compute a vector of model fluxes, \( f^{\text{model}}_m(\phi_m) \), and a vector of replicated observations, \( f^{\text{rep}}_m(\phi_m) \), where

\[ f^{\text{rep}}_{m,N}(\phi_m) \sim N(f^{\text{model}}_m(\phi_m), \sigma_n^2) \]  

(C2)

for \( n = 1, \ldots, N \). Then, we can numerically estimate our posterior predictive \( p \)-value,

\[ p = P(\chi^2(f^{\text{rep}}; \phi) \geq \chi^2(f^{\text{obs}}; \phi)|f^{\text{obs}}), \]  

(C3)
by computing the fraction of posterior samples for which \( \chi^2(f_{\text{rep}}(\phi_m); \phi_m) \geq \chi^2(f_{\text{obs}}; \phi_m) \). An extreme \( p \)-value (close to 0 or 1) would suggest that the model may not describe the data well.

Figure C1 illustrates this process visually for the two light curve fits featured in Figure 11. For each supernova, we have plotted the distribution of \( \chi^2(f_{\text{rep}}(\phi); \phi) \) vs. \( \chi^2(f_{\text{obs}}; \phi) \), based on 1000 posterior samples of \( \{\phi, f_{\text{rep}}(\phi)\} \). The estimated \( p \)-value is computed from the fraction of samples lying above the dashed 1:1 line.