MNRAS 000, 1–22 (2021) Preprint 29 September 2021 Compiled using MNRAS LATEX style file v3.0

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 Narayan[4,5]
[1]Institute of Astronomy and Kavli Insititute for Cosmology, Madingley Road, Cambridge, CB3 0HA, UK
[2]Statistical Laboratory, DPMMS, University of Cambridge, Wilberforce Road, Cambridge, CB3 0WB, UK
[3]Department of Astronomy and Astrophysics, University of California, Santa Cruz, CA 95064, USA
[4]University of Illinois at Urbana-Champaign, 1003 W. Green St., IL 61801, USA
[5]Center 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 *griz* 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 *griz* light curves. Our *griz* 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 $B - V \lesssim 0.3$). Splitting the population at the median host mass, we find $R_V = 2.84 \pm 0.31$ in low-mass hosts, and $R_V = 2.58 \pm 0.23$ in high-mass hosts, both consistent with the global value of $R_V = 2.61 \pm 0.21$ that we estimate for the full sample. For all choices of mass split we consider, R_V is consistent across the step within $\lesssim 1.2\sigma$. 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 $\mu(R_V) = 2.70 \pm 0.25$ with a 95% upper bound on the population $\sigma(R_V) < 0.61$. The R_V population means are consistent within $\lesssim 1.2\sigma$. 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 SN 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 $(B - V) \gg 0.3$ (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 $R_V \approx 1.5 - 2$ 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 $(B-V) \lesssim 0.3$. Still, there is ongoing debate over whether 'normal' samples of SNe Ia with $(B - V) \lesssim 0.3$ are consistent with a low $R_V \approx 1$–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 $(B - V) \lesssim 0.3$ colour cut, and limited wavelength range of conventional optical analyses. The confounding of intrinsic variation with extrinsic dust effects in the apparent colours of

★ E-mail: sjt202@cam.ac.uk
† NASA Einstein Fellow

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 parametrized 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 2.75 ± 0.35 in low-mass hosts ($M_* < 10^{10}\mathrm{M}_\odot$), and 1.50 ± 0.25 in high-mass hosts, with wide standard deviations of 1.3 ± 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 with 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; Valentini 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 ± 0.05 mag (0.10 ± 0.04 mag after an outlier cut) at a best fit location of $10^{10.44}\mathrm{M}_\odot$, with a comparably sized 0.17 ± 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 $uBgVriYJH$ light curves of every supernova, they measure comparable steps for all passbands at the median host mass of $10^{10.48}\mathrm{M}_\odot$, with step sizes of between 0.07 ± 0.03 (V-band) and 0.15 ± 0.04 mag (u-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,

$$\mu_s = m_B^s - M_B + \alpha x_1^s - \beta c^s, \quad (1)$$

expresses the distance modulus, μ_s, of a supernova, s, as a linear combination of its B-band apparent magnitude m_B^s, light curve stretch x_1^s, and apparent $B-V$ colour c^s, with stretch–luminosity and colour–luminosity coefficients α and β, 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, β, in different mass and colour bins. They find a significant relation between β and host galaxy mass. Since β 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 β 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 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-

ses in non-trivial ways. An additional complication is that the CfA (see Hicken et al. 2009a, 2012) and CSP (see Krisciunas et al. 2017) samples, which form a large fraction of the current low-z dataset, are dominated by supernovae discovered through galaxy-targeted monitoring programs such as the Lick Observatory Supernova Search (LOSS; Li et al. 2000; Filippenko et al. 2001; Leaman et al. 2011; Li et al. 2011). This creates a complex selection function, and probes a distribution of host galaxies unrepresentative of the true population.

To address these problems, considerable ongoing effort is being applied to replacing the existing low redshift dataset with a more homogeneous sample observed entirely on Pan-STARRS-1 (PS1; 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 BayeSN model (Mandel et al. 2020) is a coherent framework for optical and near-infrared (NIR) SEDs, continuous in wavelength from 0.35–1.8 μm, 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 $> 10^{10}\mathrm{M}_{\odot}$).

The BayeSN 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 BayeSN analysis of Foundation DR1. As part of this study, we repeat the mass-agnostic Mandel et al. (2020) analysis on the Foundation data, verifying BayeSN's performance, and modelling for the first time the statistical properties of SNe Ia in the rest-frame z-band. Additionally, we extend the BayeSN 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 BayeSN 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 $grizy$ 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 BayeSN 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 BAYESN 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 BayeSN 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 in the sample 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 R_V value(s) parametrizing host galaxy dust laws. The first of these (Section 2.1.1) assumes a single value of R_V for all SNe in the sample, whilst the second (Section 2.1.2) assumes a population distribution of R_V, with individual values permitted for each SN. Within the *Partial-Split* configuration, we run equivalent analyses where either a separate single R_V, or a separate population distribution of R_V, 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 ($H_0 = 73.24$, $\Omega_M = 0.28$, $\Omega_\Lambda = 0.72$; Riess et al. 2016), with external distance uncertainties derived from peculiar velocity and spectroscopic redshift uncertainties. We assume a peculiar velocity uncertainty of $\sigma_{\rm pec} = 150\,{\rm km\,s^{-1}}$ (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 BayeSN 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 BayeSN 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 $S_s(t,\lambda_r)$ of a supernova s, as a function of rest-frame phase t and wavelength λ_r, is modelled as:

$$-2.5\log_{10}[S_s(t,\lambda_r)/S_0(t,\lambda_r)] = M_0 + W_0(t,\lambda_r) + \delta M_s + \theta_1^s W_1(t,\lambda_r) + \epsilon_s(t,\lambda_r) + A_V^s \xi(\lambda_r; R_V) \quad (2)$$

(as in M20 Eq. 12). Here, $S_0(t,\lambda_r)$ is an optical-NIR SN Ia baseline spectral template (Hsiao et al. 2007; Hsiao 2009); $M_0 \equiv -19.5$ is a normalisation constant; $W_0(t,\lambda_r)$ 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 ($A_V^s = 0$). δM_s is a 'grey' (time- and wavelength-independent) brightness offset specific to supernova s; $W_1(t,\lambda_r)$ is the first functional principal component, describing the primary SED features that vary between SNe; θ_1^s is a scalar coefficient describing the extent to which $W_1(t,\lambda)$ affects the sth supernova; $\epsilon_s(t,\lambda_r)$ 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_V^s is the V-band host dust extinction affecting supernova s; and $\xi(\lambda_r; R_V)$ is the Fitzpatrick (1999) dust law, parametrized by R_V. Initially, we assume R_V 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 $W_0(t,\lambda_r)$, $W_1(t,\lambda_r)$, and $\epsilon_s(t,\lambda_r)$ functions with natural cubic splines, parametrized by matrices of knots, $\boldsymbol{W}_0$, $\boldsymbol{W}_1$, and $\boldsymbol{E}_s$. 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 $\boldsymbol{l} = (3500, 4900, 6200, 7700, 8700, 9500)$ 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 $R_{\rm MW} = 3.1$ and $A_{\rm MW}^s = R_{\rm MW} \times E(B-V)_{\rm MW}^s$ 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 likelihood (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:

$$A_V^s \sim \text{Exponential}(\tau_A), \quad (3)$$

$$\theta_1^s \sim N(0,1), \quad (4)$$

$$\delta M_s \sim N(0, \sigma_0^2), \quad (5)$$

$$\boldsymbol{e}_s \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_\epsilon), \quad (6)$$

where $\boldsymbol{e}_s = \text{vec}(\boldsymbol{E}_s)$ is the vectorised $\boldsymbol{E}_s$ matrix, and τ_A, σ_0 and $\boldsymbol{\Sigma}_\epsilon$ are hyperparameters describing the population mean host galaxy extinction, population standard deviation of the δM_s offsets, and population covariance of the residual $\boldsymbol{e}_s$ perturbations, respectively. Our weak hyperpriors are as in M20.

2.1.2 Population Distribution of Dust Law R_V

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 R_V were first introduced by Mandel et al. (2009, 2011). To enable a more direct comparison with the results of Brout & Scolnic (2021), we consider an extension of our current model in which the R_V values for individual SNe are drawn from a population distribution, rather than all having a single global value. In this model variant, the R_V^s of an individual SN s is now a latent variable, drawn from a truncated normal (Gaussian)[1] population

[1] Here, $X \sim \text{Trunc-}N(\mu, \sigma^2, a, \infty)$ 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

$$P(X|\mu,\sigma,a) = \frac{1}{\sigma}\frac{\phi(\xi)}{1-\Phi(\alpha)},$$

and is zero for $X < a$. Here, $\xi = (X-\mu)/\sigma$, $\alpha = (a-\mu)/\sigma$, and $\phi(z)$ and $\Phi(z)$ are respectively the PDF and CDF of a standard normal random variable z.

distribution,

$$R_V^s \sim \text{Trunc-}N(\mu_R, \sigma_R^2, 0.5, \infty), \tag{7}$$

with a truncation of the lower tail to force $R_V^s \geq 0.5$. A physically-motivated choice of lower bound would be $R_V^s \geq 1.2$ (the Rayleigh scattering limit, see Draine 2003), but we opt for the more liberal choice of $R_V^s \geq 0.5$ to match Brout & Scolnic (2021). The population mean and variance parameters μ_R, σ_R^2 are estimated coherently with all other parameters by augmenting the global posterior density. The simple assumption of a Gaussian population distribution for R_V^s is 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 σ_R, but to place relatively little prior probability at excessively high values of $\sigma_R > 4$. Sensitivity analysis for alternate σ_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, $M_{*,\text{split}}$, 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, μ_R and σ_R, in the case where supernovae have individual R_V^s values) parametrizing the Fitzpatrick (1999) host galaxy dust extinction law; the population mean extinction τ_A; and σ_0, the population standard deviation of the 'grey' δM_s component of the intrinsic SED. We also introduce a parameter, ΔM_0, which shifts the population mean of the δM_s for supernovae in low-mass hosts – this is functionally equivalent to allowing a constant 'mass step'. This means that equation (5) becomes

$$\delta M_s \sim \begin{cases} N(0, \sigma_{0,\text{HM}}^2) & \text{if } M_{*,s} \geq M_{*,\text{split}} \\ N(\Delta M_0, \sigma_{0,\text{LM}}^2) & \text{if } M_{*,s} < M_{*,\text{split}} \end{cases}, \tag{8}$$

where $M_{*,s}$ is the host galaxy stellar mass of SN s, and the subscript LM and HM denote the values of σ_0 below and above the mass split. The hyperpriors on the low- and high-mass values of R_V (or μ_R, σ_R), τ_A, and σ_0 are taken to be the same as in the *No-Split* model. For the low- vs. high-mass offset, ΔM_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 $M_* > 10^6 M_\odot$ 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 ($\sigma_{-\text{pv}}$) of 0.131 (0.119) mag (c.f. the first row of Table 4).

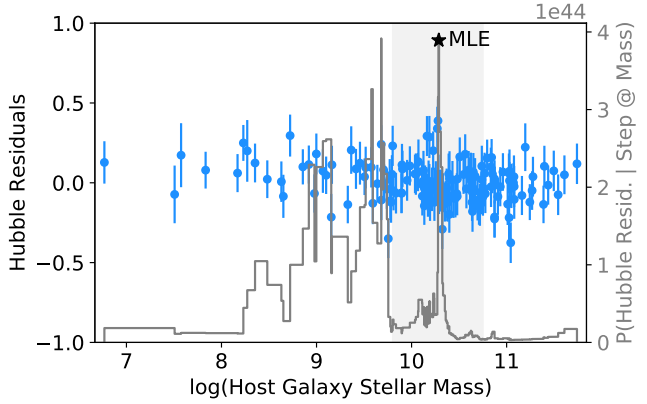

Figure 1. Hubble residuals, $\Delta\mu$ plotted along with the marginal likelihood, $P(\Delta\mu|\log_{10} M_{*,\mathrm{split}})$, of these data given a step at a particular mass (see Appendix A for computational details). Indicated with a star is the maximum likelihood step, which we use in some later analyses. The faint grey band covers the interquartile range of the host galaxy masses in the sample.

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 $M_{*,\mathrm{split}} = 10^{10}\mathrm{M}_\odot$ 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 $M_{*,\mathrm{split}}$ to the median host mass of our sample, $10^{10.331}\mathrm{M}_\odot$. Additionally, we choose a split point that is favoured as a mass step location based on Hubble residuals obtained from the *No-Split* BayeSN 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 $10^{10.286}\mathrm{M}_\odot$, 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 $10^{9.6815}\mathrm{M}_\odot$, 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 θ_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 θ_1, and with all other effects (A_V, δM, ϵ) 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 θ_1 corresponding to a faster decline. In the r band, we see a weaker version

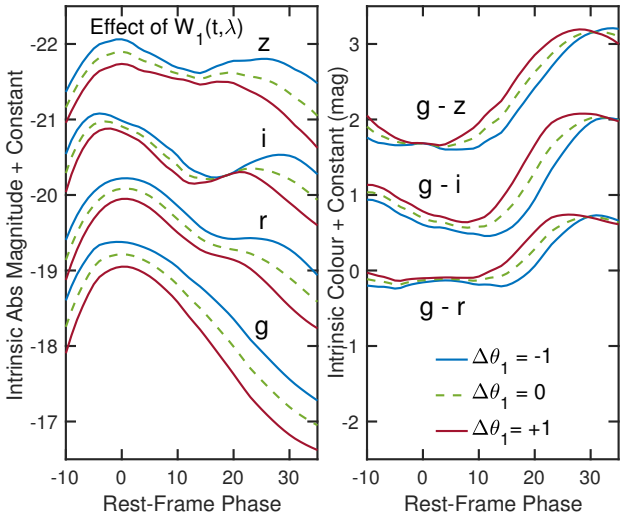

Figure 2. Synthesised rest-frame $griz$ light curves (left panel) and $g-r$, $g-i$, $g-z$, colour curves (right panel) from the BayeSN model trained on the Foundation supernovae, demonstrating the intrinsic effects of the first functional component, $W_1(t,\lambda_r)$. For each passband (or colour), the different coloured lines show the effect of varying θ_1 between -1 (blue) and 1 (red), modifying the extent to which $W_1(t,\lambda_r)$ is present. Vertical offsets are $(0,-1,-2.5,-3.5)$ for (g,r,i,z) and $(0,1.5,2.5)$ for $(g-r,g-i,g-z)$.

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 $BVriYJH$ 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 θ_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 θ_1. An interesting result is that we see very little sensitivity of the $g-r$ and $g-z$ colours around peak to varying θ_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 θ_1 increases considerably after peak, particularly around 10–25 days. Additionally, at around 30 days, each colour converges for all values of θ_1, before (most

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]	ΔM_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, ΔM_0, estimated from the *Partial-Split* posterior analysis (see Sections 2.2 and 4.2).

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' δM_s offsets, which are constant in time and wavelength; and the time- and wavelength-varying $\epsilon_s(t,\lambda_r)$ realisations, whose correlation structure is captured by a population covariance matrix, $\mathbf{\Sigma}_{\boldsymbol{\epsilon}}$.

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 $\eta_s(t,\lambda_r) = \delta M_s + \epsilon_s(t,\lambda_r)$ on rest-frame $griz$ light curves and $g-r$, $g-i$, and $g-z$ colour curves. The light curve residual variance is depicted in Figure 3 (left hand panel). The widths of the $\pm1\sigma$ envelopes, showing the residual variance within the population, are relatively constant near peak, generally showing only ~ 0.1 mag of scatter, and increasing somewhat towards later phases. The colour curve residuals are, by definition, independent of the grey residuals δM_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, σ_0^2, of the Gaussian population distribution of δM_s, and the shift, ΔM_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 σ_0, τ_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 ΔM_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 2σ greater than zero. The difference of ΔM_0 from zero is most significant (3σ) when the sample is split at the MLE mass split preferred as a step location by the Hubble residuals. For all choices of mass split, the posterior estimate of ΔM_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 ± 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* ΔM_0 estimates and *No-Split* step sizes side-by-side.

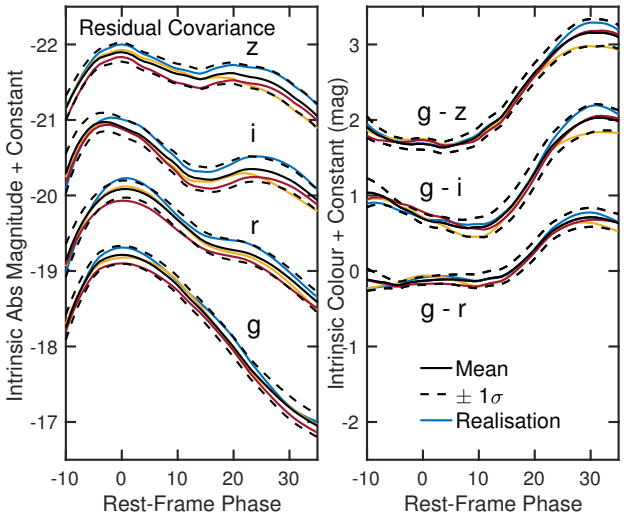

Figure 3. Same as Fig. 2, but demonstrating the effect of the intrinsic residual components, $\eta_s(t,\lambda_r) = \delta M_s + \epsilon_s(t,\lambda_r)$. The solid black curves show the mean (equivalent to the dashed green lines in Figure 2), with $\theta_1^s = \delta M_s = \epsilon_s(t,\lambda_r) = 0$. The black dashed lines show a $\pm1\,\sigma$ envelope, and the coloured lines show the effects of the residual functions of three example SNe Ia.

For all choices of split point (see Table 2, which also repeats the ΔM_0 values), our inferences of σ_0 are fairly consistent between low- and high-mass hosts. For the median and MLE split points, the posterior mean estimates of σ_0 are marginally higher in low-mass host galaxies. For the median mass split point, the joint posterior distribution of ΔM_0 and the high- and low-mass values of σ_0 is shown in Figure 4. Our inferences of ΔM_0 and σ_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 σ_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 σ_0 in low-mass hosts to be around 2.2σ higher than in high-mass hosts. For the other splits, the difference is insignificant. There is no explicit ΔM_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 2. Posterior means and standard deviations of population parameters for different choices of host mass split.

Model	Split Point	ΔM_0	σ_0		τ_A		R_V	
			LM	HM	LM	HM	LM	HM
No-Split	-	-	0.10 ± 0.01		0.19 ± 0.02		2.61 ± 0.21	
Partial-Split	$10^{10}M_\odot$	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
	Median	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	$10^{10}M_\odot$	-	0.15 ± 0.02	0.10 ± 0.01	0.16 ± 0.04	0.22 ± 0.03	3.07 ± 0.64	2.43 ± 0.22
	Median	-	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

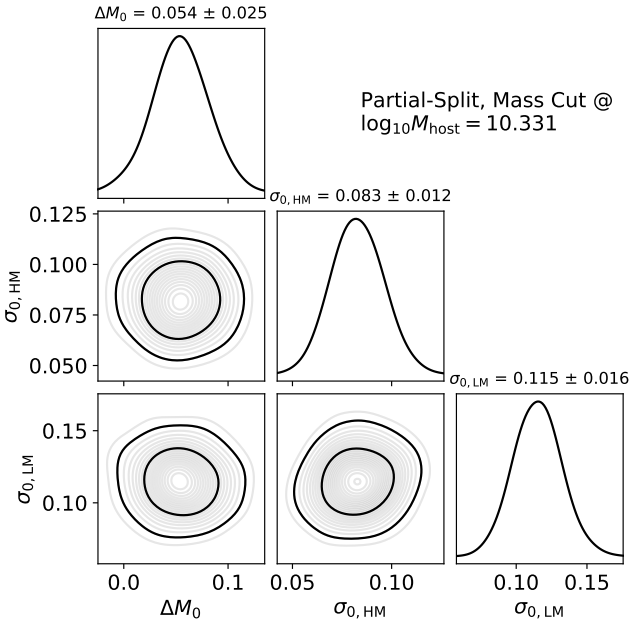

Figure 4. Posterior distribution of the low- vs. high-host-mass magnitude offset, ΔM_0, and the low- and high-host-mass values of σ_0 – the population standard deviation of the 'grey' component of the intrinsic SED. The median host mass split was used here. Darker contours indicate the regions which contain 68 and 95 per cent of the posterior probability.

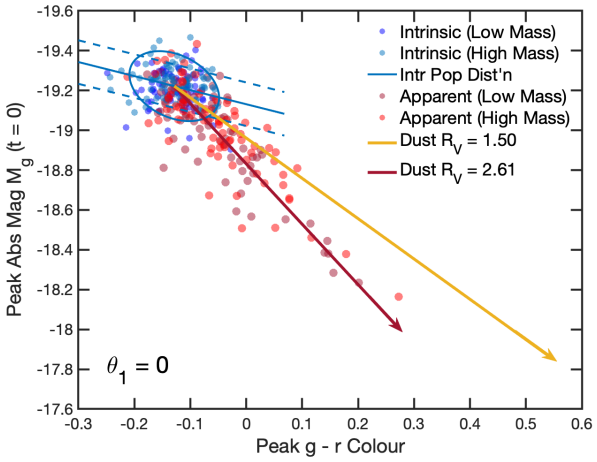

Figure 5. Posterior estimates of peak ($t = 0$) g-band absolute magnitude vs. $(g-r)$ colour for the unsplit sample, obtained from the *No-Split* model. Each SN is represented by a red and blue point. Red points indicate extinguished absolute magnitudes and apparent colours. Blue points indicate the inferred intrinsic absolute magnitudes and intrinsic colours. Bright (dark) shaded points indicate SN with high- (low-) mass hosts (where the sample is split at the median host mass of $10^{10.331}M_\odot$). The blue ellipse depicts the 1σ locus of the inferred intrinsic population distribution, while the blue solid and dashed lines indicate the inferred mean intrinsic colour–luminosity relation, and intrinsic absolute magnitude scatter around it. The magnitudes and colours have been corrected for the modelled intrinsic SED shape dependence (see Fig. 2) to $\theta_1 = 0$ using Eq. 2. The reddening–extinction vectors for $A_V = 1$ for two example R_V values ($R_V = 1.5$; and our posterior mean $R_V = 2.61$ from the full $griz$ light curves) under the Fitzpatrick (1999) law are shown by the yellow and red arrows emanating from the centre of the intrinsic distribution. Note that we do not expect that all SNe should "map back" along the dust vector to the exact centre; rather, the intrinsic properties have a distribution. The average colour–luminosity trend of the reddened SNe Ia agrees with the $R_V = 2.61$ dust law and disfavours $R_V = 1.5$ in both mass bins.

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 $R_V = 2.61 \pm 0.21$ and $\tau_A = 0.19 \pm 0.02$ – 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 $R_V = 2.89 \pm 0.20$ 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

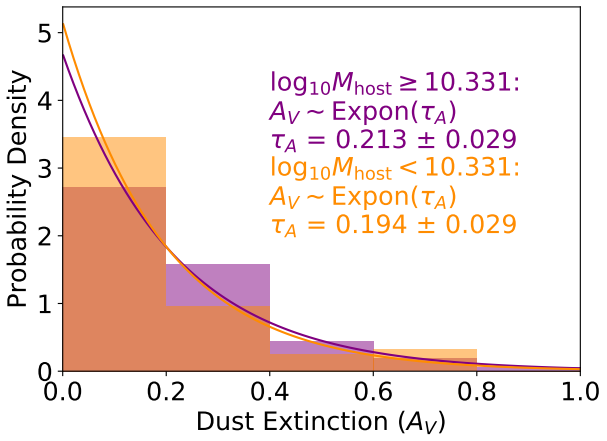

Figure 6. Population distributions of the inferred dust extinction values of the Foundation sample when split at the median host mass (under the *Partial-Split* model). The solid lines plotted over the histograms show the exponential population distributions parametrized by the inferred mean extinction values for the low- and high- host mass populations.

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 β. 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 τ_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 τ_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 τ_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 τ_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 $10^{10}\mathrm{M}_\odot$ 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 $M_* < 10^{10}\mathrm{M}_\odot$, and $R_V = 2.67 \pm 0.21$ for those with $M_* \geq 10^{10}\mathrm{M}_\odot$. 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 τ_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 τ_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 1σ 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^s is permitted for every supernova, with these drawn from a population distribution described by a Gaussian with mean μ_R, and standard deviation σ_R, truncated such that $R_V^s \geq 0.5$ (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 μ_R are consistent with the global values we found previously. Our σ_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 μ_R and σ_R for the *No-Split* and *Partial-Split* models.

The recent work of Brout & 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 $10^{10}\mathrm{M}_\odot$, 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 τ_A values for the same choice of split point are suggestive of qualitatively similar be-

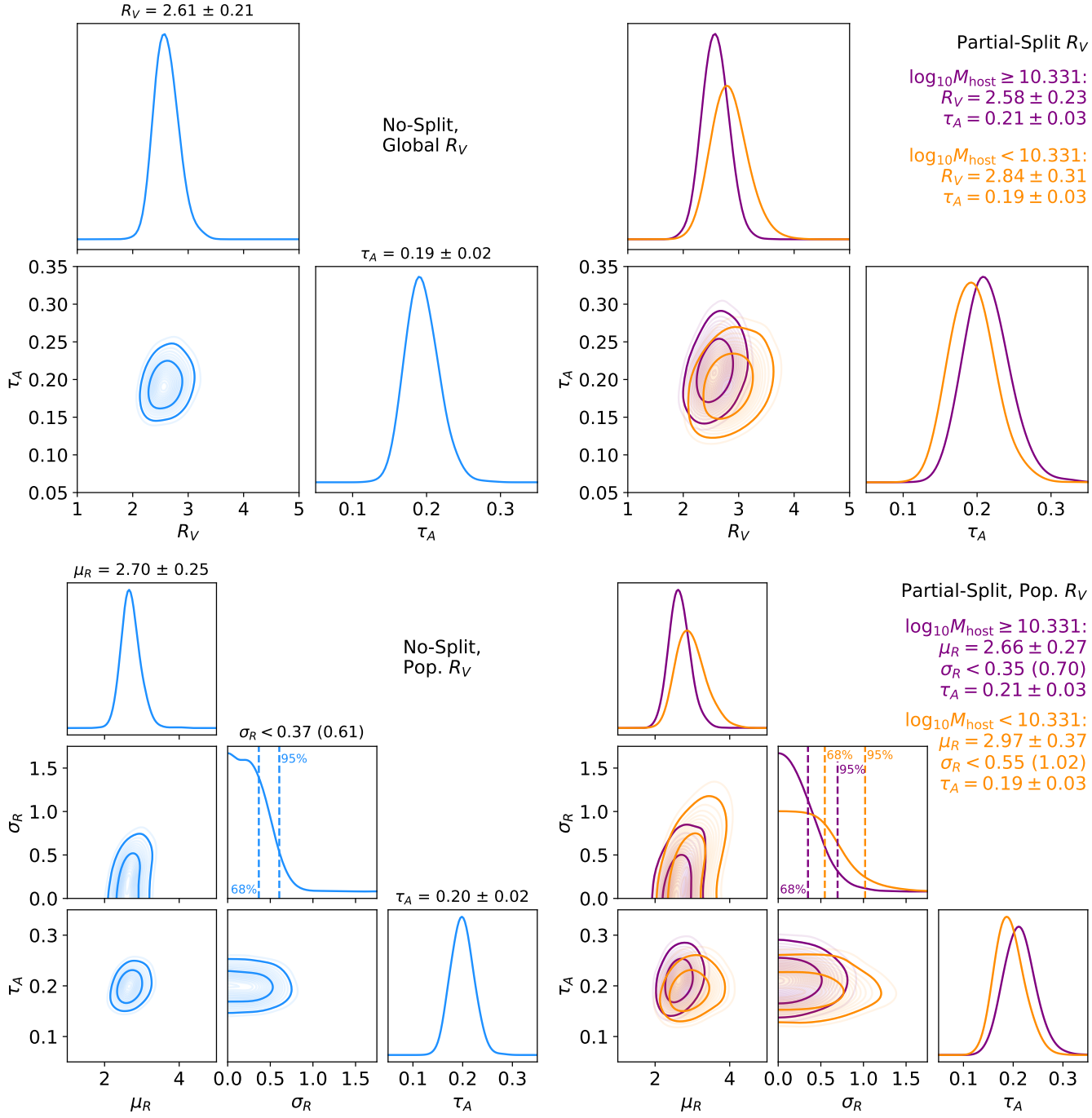

Figure 7. (top row) Posterior distributions of R_V and τ_A from BayeSN training runs on Foundation DR1 carried out without (left panel) and with (right panel) these parameters split by host galaxy stellar mass at the median. (bottom row) Posterior distributions of μ_R, σ_R and τ_A from the variant of our analysis where we allow a population distribution of R_V. Parameter summaries are posterior mean and standard deviation, except for σ_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 $\lesssim 1.2\sigma$, the difference we find is not statistically significant. For R_V, Brout & Scolnic (2021) reported peak values of 2.75 ± 0.35 and 1.50 ± 0.25 for low- and high-mass hosts, respectively – a difference of 2.9σ – with fairly large population standard deviations of 1.3 ± 0.2 for both subsamples. When not splitting by host galaxy mass, they estimate a mean R_V of 2.0 ± 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^s population distribution with mean parameter $\mu_R = 2.70 \pm 0.25$. For this configuration, a wide population distribution of R_V^s is strongly disfavoured, with 68 (95) per cent of posterior probability falling at $\sigma_R < 0.37$ (0.61).

Table 3. Inferences of R_V population parameters.

Split Point[a]	μ_R[b]		σ_R [68% (95%) u.b.][c]	
	LM	HM	LM	HM
-	2.70 ± 0.25		0.37 (0.61)	
$10^{10}M_\odot$	3.49 ± 0.61	2.72 ± 0.23	0.89 (1.87)	0.32 (0.62)
Median	2.97 ± 0.37	2.66 ± 0.27	0.55 (1.02)	0.35 (0.70)
MLE	3.03 ± 0.43	2.74 ± 0.26	0.60 (1.22)	0.34 (0.64)

[a] Host mass split point with which the *Partial-Split* model was run. The first line results are from the *No-Split* model.
[b] Posterior mean and std. dev. of the population R_V mean parameter.
[c] 68 (95)% posterior upper bounds for the population R_V std. dev. parameter.

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 μ_R, σ_R and τ_A (which is insensitive to the adoption of a population distribution of R_V^s) 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 μ_R, σ_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^s 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 σ_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 μ_R and σ_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 companion. 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 μ_R, σ_R and τ_A for the *Partial-Split* model with the median host mass as split point. For a brief discussion of the prior sensitivity of σ_R, see Appendix B.

For all choices of mass split, we obtain μ_R values consistent with the values of R_V obtained in our global dust law analyses. The μ_R's across the mass split are consistent within $\lesssim 1.2\sigma$. However, the smallest $\Delta M_0 = 0.042 \pm 0.028$ (1.5σ) 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 μ_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, $\mu_R \lesssim 2.2$ 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 τ_A inferences are insensitive to the inclusion of a population distribution of R_V^s.

4.4 Distribution of SED Shape Parameters

It is interesting to investigate how the SED shape parameter, θ_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 θ_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 θ_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 θ_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 θ_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 BayeSN's θ_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 BayeSN's θ_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 θ_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 θ_1 bins in the upper panel (i.e. the nth bin of the x_1 histogram should contain supernovae with similar light curve widths and decline rates to those in the nth bin of the θ_1 histogram). Particularly for the low-mass subsample, the x_1 and θ_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 θ_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* BayeSN model. We see that the parameters correlate, albeit not as strongly as x_1 and θ_1 (top panel). For supernovae with low reddening due to dust, we

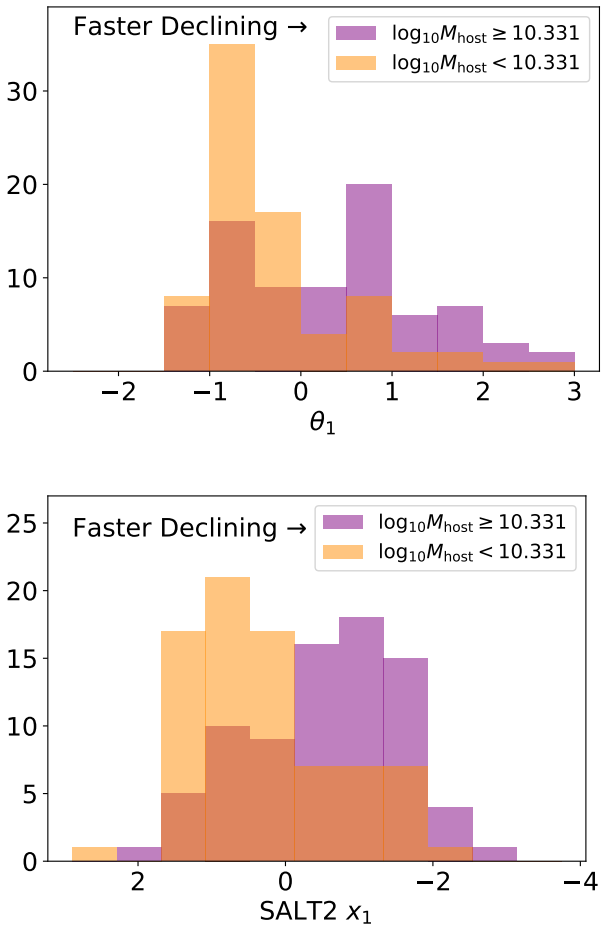

Figure 8. (top panel) Histograms showing how the shape parameter θ_1 – controlling how strongly the principal SED component, W_1, is present for a given supernova – is distributed for the low- and high- host mass subsamples of Foundation DR1. (bottom panel) Histograms showing the distribution of the 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.

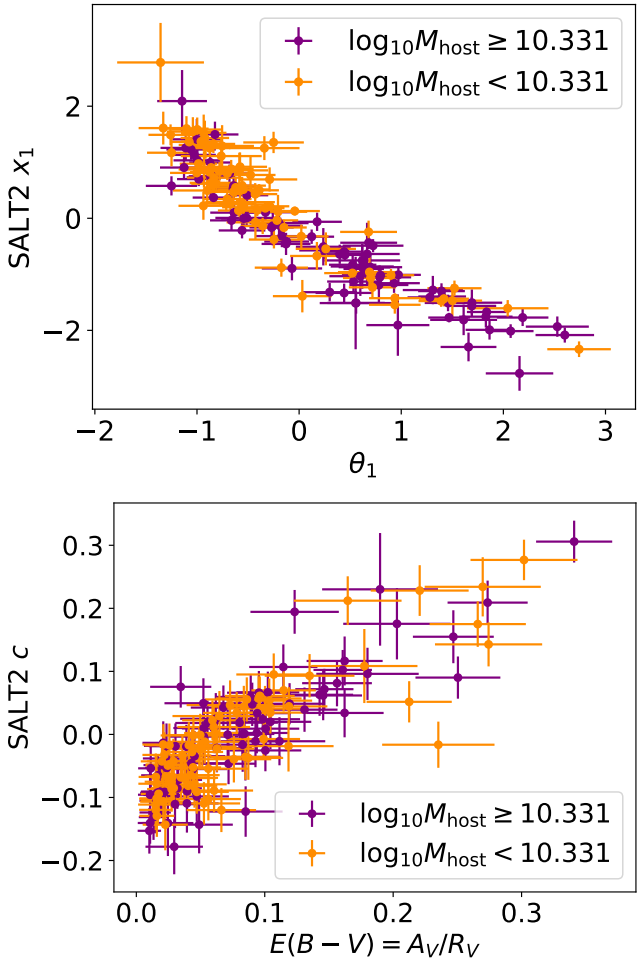

Figure 9. Plots of SALT2 parameters against analogous BayeSN quantities, colour coded by host galaxy stellar mass. Values of SALT2 parameters from Foley et al. (2018); Jones et al. (2019). (top panel) SALT2 x_1 parameter against the BayeSN θ_1 SED shape parameter. (bottom panel) SALT2 colour, c, plotted against $E(B-V) = A_V/R_V$ dust reddening derived from BayeSN posterior estimates of A_V and R_V.

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 than around $10^{10}\mathrm{M}_\odot$. 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 BayeSN 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. the colour arising purely from the effects of W_0, $\theta_1^s W_1$, and ϵ_s). 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 θ_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.

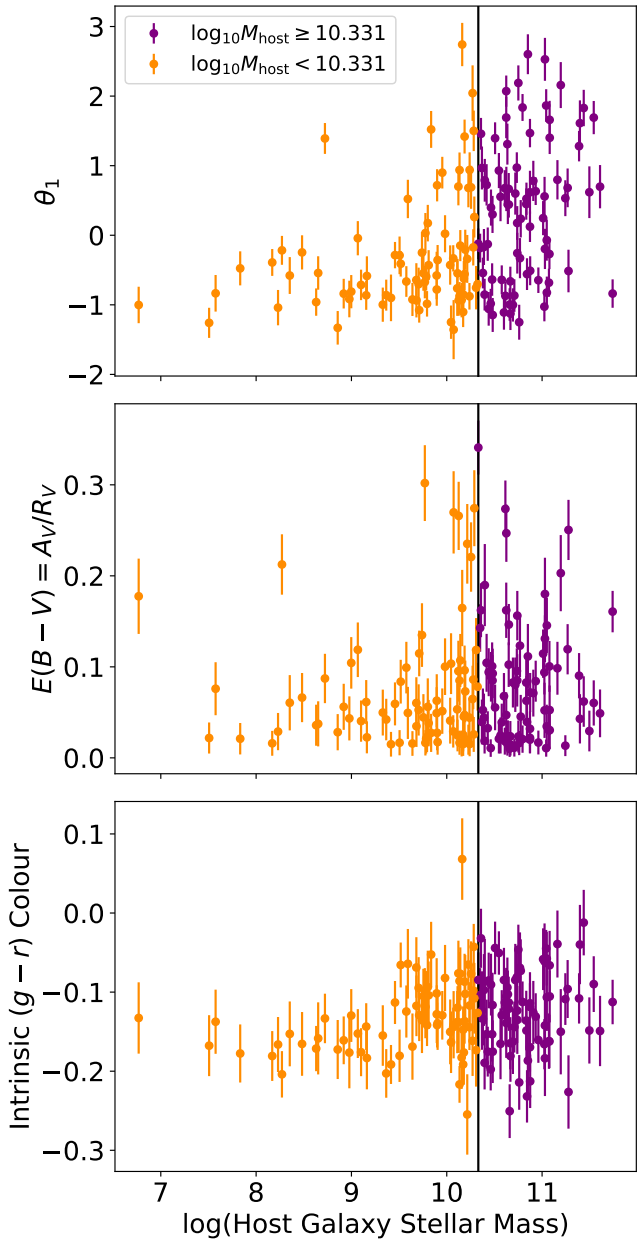

Figure 10. Plots of (top panel) BayeSN's θ_1 parameter, (middle panel) host galaxy dust reddening, $E(B-V) = A_V/R_V$, and (bottom panel) intrinsic $g-r$ colour vs. host galaxy stellar mass. These results are from the *Partial-Split* analysis, with parameters split at the median host mass (depicted as a vertical line).

4.6 Hubble Diagram Analysis

4.6.1 Training and Resubstitution

After training each of our models, we fix the global model parameters ($\boldsymbol{W}_0$, $\boldsymbol{W}_1$, $\boldsymbol{\Sigma}_\epsilon$, τ_A, R_V, σ_0, ΔM_0) to their posterior mean values. Then, we 'resubstitute' the light curves of the 157 supernovae in our training set, fitting the latent parameters (θ_1^s, A_V^s, δM_s, $\boldsymbol{e}_s$) of each supernova independently with no external distance constraints. Marginalising over the latent parameters for each supernova, we thus obtain photometric distance estimates (as in M20, section 2.8). Figure 11 shows for one SN, ASASSN-16cs, an example light curve fit and the joint posterior over θ_1^s, A_V^s, and the photometric distance, $\mu_s^{\rm phot}$.

For a given training and resubstitution run, we can construct a Hubble diagram plotting photometric distances, $\mu_s^{\rm phot}$, against CMB frame redshifts, $z_s^{\rm CMB}$. During training, we employed external distance constraints (based on a fiducial ΛCDM cosmology from Riess et al. 2016) centred on $\mu_s^{\rm ext} = \mu_{\Lambda{\rm CDM}}(z_s^{\rm CMB}; H_0 = 73.24, \Omega_M = 0.28, \Omega_\Lambda = 0.72)$, with uncertainties derived from peculiar velocity (assuming $\sigma_{\rm pec} = 150\,{\rm km\,s^{-1}}$; 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 $\mu_s^{\rm phot} - \mu_s^{\rm ext}$. 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,

$$\mu_s = m_B^s - M_B + \alpha x_1^s - \beta c^s - \gamma I_*^s, \tag{9}$$

to the SALT2 (Guy et al. 2007, 2010; Betoule et al. 2014) fit results (m_B^s, x_1^s, c^s; B-band apparent magnitude, stretch, and colour, respectively) from Foley et al. (2018); Jones et al. (2019). Here I_*^s is an indicator variable which obeys $I_*^s = 1$ if $M_{*,s} \geq 10^{10.331}{\rm M}_\odot$ and $I_*^s = 0$ 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 α and β 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 $\sigma_{\rm res} = 0.134$.

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 ($\sigma_{-\rm pv}$) unaccounted for by an assumed peculiar velocity uncertainty of $\sigma_{\rm pec} = 150\,{\rm km\,s^{-1}}$ (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 $\sigma_{-\rm pv}$ 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* BayeSN 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 ($\sigma_{-\rm pv}$) 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 ($\sigma_{-\rm pv}$) 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 BayeSN and SALT2 photometric distance results, computing the ΔRMS each time. For the two Hubble diagrams shown in Figure 12, we find that $\Delta{\rm RMS} = 0.041 \pm 0.009$ mag. As well as improving on the overall Hubble diagram scatter, by adopting a Bayesian approach,

[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).

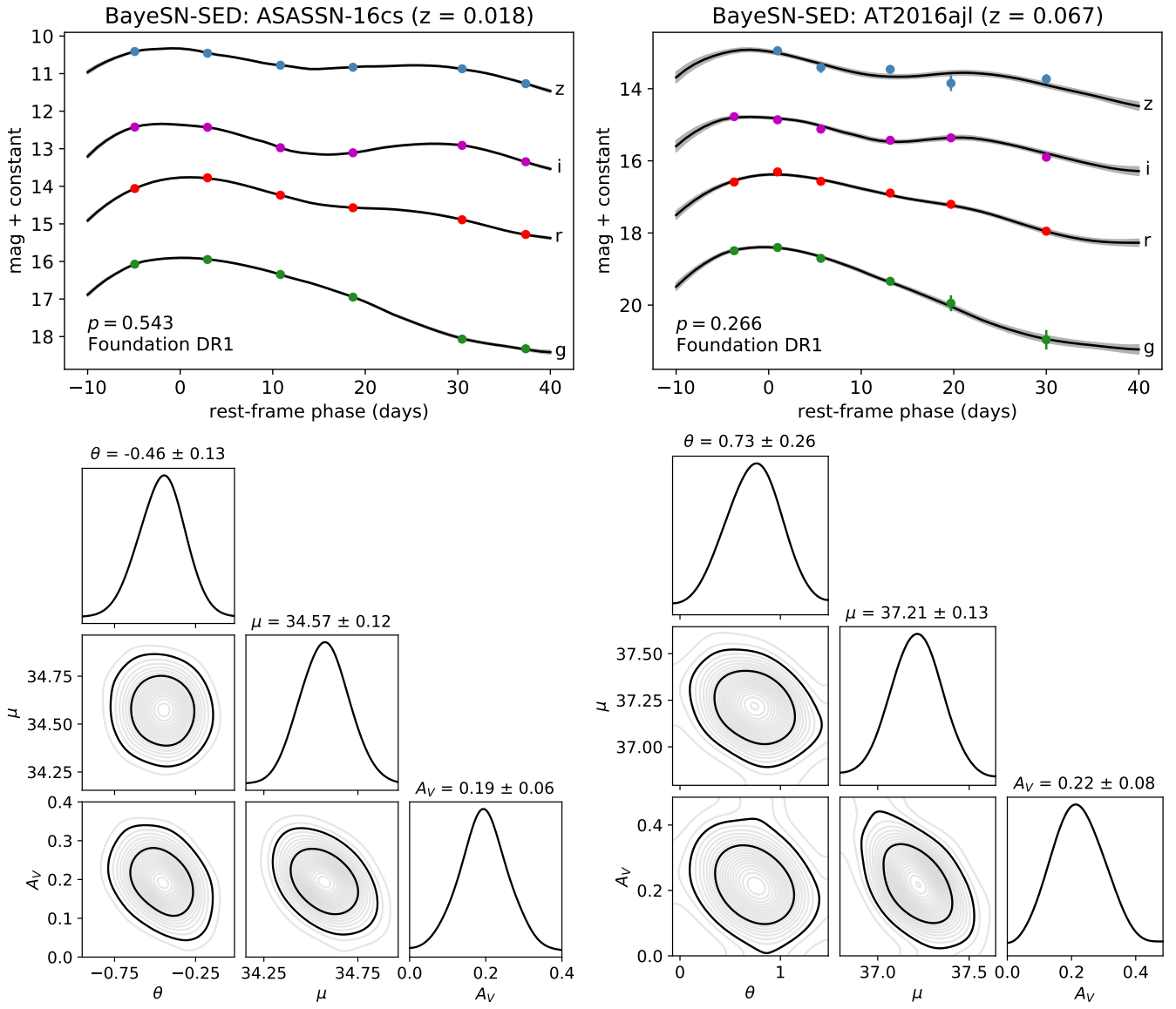

Figure 11. Example photometric distance fit to two of the Foundation DR1 supernovae: ASASSN-16cs (left panels; bright, low-redshift, slower-declining) and AT2016ajl (right panels; dim, higher-redshift, faster-declining). (top panels) Fits to the $griz$ light curves. Vertical offsets are $(0, -2, -4, -6)$ 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 θ_1^s, $\mu_s^{\rm phot}$, and A_V^S. The photometric distance estimates used for constructing our Hubble diagram are obtained by marginalising over θ_1^s and A_V^s (as in the second panel along the diagonal of the corner plot).

and 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, $(\mu_s^{\rm phot} - \mu_s^{\rm ext})/\sigma_{\mu,s}^{\rm phot}$, computed from our photometric distances, $\mu_s^{\rm phot}$, and their uncertainties, $\sigma_{\mu,s}^{\rm phot}$, 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, $\sigma_{\rm res}$, 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 2σ 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 σ_0 (the level of time- and wavelength-independent residual brightness scatter; see Section 4.2) estimated during training. This is because the mode

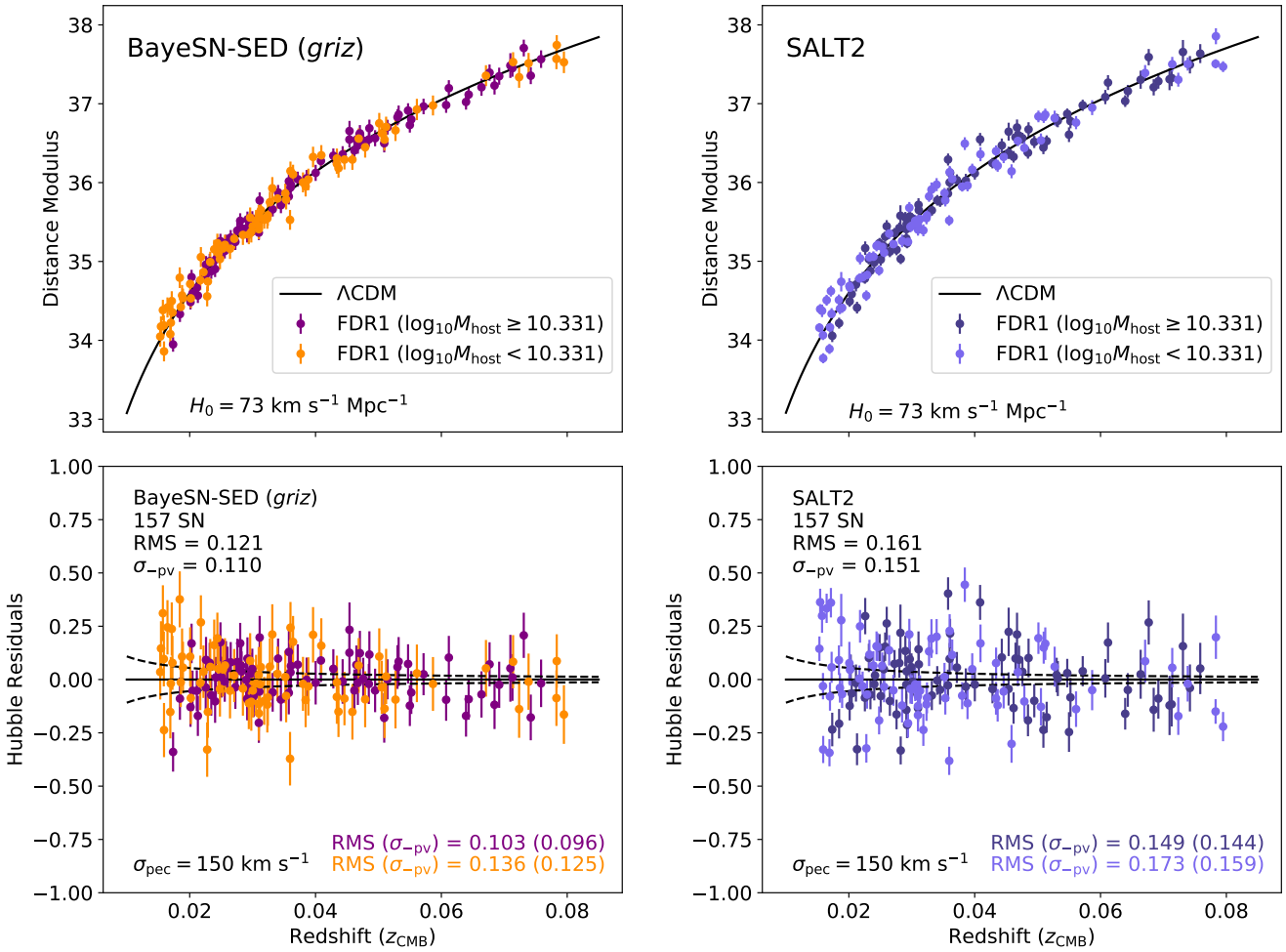

Figure 12. (left panel) Hubble diagram of photometric distances obtained by training on and resubstitution of the Foundation DR1 SNe Ia, using the *Partial-Split* 0.35–0.95 µm BayeSN SED model with key parameters split at the median mass cut. (right panel) Hubble diagram based on SALT2 fits from Foley et al. (2018); Jones et al. (2019), and the Tripp formula with the mass-step applied (Eq. 9). In all cases, dashed lines depict the level of distance uncertainty which can be expected given a peculiar velocity uncertainty of $\sigma_{\rm pec} = 150$ km s^{-1}.

of SED variation described by σ_0 contributes to the photometric distance uncertainty. In the SALT2 Hubble diagram, we estimate ΔRMS $= 0.024 \pm 0.018$ mag between low- and high-mass hosts, a 1.4σ 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 ($\boldsymbol{W}_0$, $\boldsymbol{W}_1$, $\boldsymbol{\Sigma}_\epsilon$) 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 ($\sigma_{-\rm pv}$) 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 $\sigma_{-\rm pv}$ is consistently $\lesssim 0.014$ mag. For the median mass split, we find an RMS ($\sigma_{-\rm pv}$) of 0.132 (0.122) mag under cross validation. Since this model under CV is trained on 10% less

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

Scheme[a]	Model	Split Point	RMS[b]			σ_{-pv}[c]		
			All	LM	HM	All	LM	HM
Resubstitution	*No-Split*	-	0.124	-	-	0.112	-	-
	Partial-Split	$10^{10}M_\odot$	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	$10^{10}M_\odot$	-	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	$10^{10}M_\odot$	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 $\sigma_{pec} = 150\,km\,s^{-1}$; Carrick et al. 2015) is taken out (as in Mandel et al. 2020, eq. 32).

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. ≈ 0.127 mag.

For the BayeSN cross-validation, we repeatedly retrain a new BayeSN 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 BayeSN analysis, using the $(\alpha, \beta, M_B, \gamma)$ 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 $\Delta\text{RMS} = 0.034 \pm 0.009$ between the cross-validated SALT2 and BayeSN Hubble diagrams, a 3.8σ improvement. This improvement in the accuracy of photometric distance estimates is a good indication that the BayeSN SED model is an effective statistical description of the *griz* 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 *griz* light curve datasets from the BayeSN forward model. We then train new BayeSN 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 $(\boldsymbol{W}_0, \boldsymbol{W}_1, \boldsymbol{\Sigma}_\epsilon, \sigma_0, \tau_A)$, except for R_V, fixed to their posterior mean values from the *No-Split* analysis. We generated 3 simulations, each of which has a different input true global host dust law $R_V = (2.0, 3.0, 4.0)$. We then trained new BayeSN 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 (à 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 $(\hat{\mu}_R, \hat{\sigma}_R) = (2.30 \pm 0.42, 1.59 \pm 0.47)$. For the input population with $(\mu_R, \sigma_R) = (3.0, 1.0)$, our posterior estimates were $(\hat{\mu}_R, \hat{\sigma}_R) = (3.06 \pm 0.30, 0.86 \pm 0.26)$. From the results of the two runs with $\mu_R = 2.0$ or 3.0, we would conclude with 95% posterior probability that $\sigma_R > 0.93$ or 0.50, respectively. Thus, we would detect an R_V distribution of non-zero width with high confidence.

5 CONCLUSIONS

We have used the *griz* 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 BayeSN SN Ia SED model, continuous over 0.35–0.95 μm. 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 *griz* 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 BayeSN 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 *griz* 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 consistent with the cosmological cut (peak apparent $B - V \lesssim 0.3$) (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; Sasdelli 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 $R_V \approx 2.5$, albeit for a slightly more reddened supernova (AT2019lcj; $c_{\rm SALT2} = 0.57 \pm 0.04$, $E(B-V)_{\rm SNooPy} = 0.63 \pm 0.07$, $\log_{10}(M_*/{\rm M}_\odot) = 10.48^{+0.11}_{-0.46}$) than we have considered here. Allowing a population distribution of R_V values, parametrized by a mean, μ_R, and standard deviation, σ_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 $\lesssim 1.2\sigma$ (see Table 2). In all cases, we find that $R_V \lesssim 2.2$ 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^s 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 μ_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 $\mu_R \lesssim 2.2$ tend to be strongly disfavoured on both sides of the mass split. We find that the width σ_R of the R_V^s 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^s are disfavoured, with $\sigma_R \lesssim 0.7$ 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 $R_V \approx 1.7^{+0.6}_{-0.5}$; 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 $R_V \approx 2.5$–3. Our population mean estimates, μ_R, are slightly below the average Milky Way value (e.g. Schlafly et al. 2016) by at most 2.4σ.

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 τ_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 host 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 ΔM_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 $\sim 10^{10}$–$10^{10.8}{\rm M}_\odot$ (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 $10^{10}{\rm M}_\odot$ (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- and wavelength-dependence of SED components against host mass

– is a natural next step, and will be possible within the BayeSN 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; Pruzhinskaya et al. 2020). Incorporation of spectroscopic indicators of intrinsic colour (e.g. ejecta velocity, Foley & Kasen 2011; Mandel et al. 2014; 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 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_DR1`. 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
Arima N., Doi M., Morokuma T., Takanashi N., 2021, preprint, (arXiv:2101.02407)
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, 526, 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 M., 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 K. G., 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
Elias-Rosa N., et al., 2006, MNRAS, 369, 1880
Elias-Rosa N., et al., 2008, MNRAS, 384, 107
Filippenko A. V., Li W. D., Treffers R. R., Modjaz M., 2001, in Paczynski B., Chen W.-P., Lemme C., eds, Astron. Soc. Pac. Conf. Series Vol. 246, IAU Colloq. 183: Small Telescope Astronomy on Global Scales. p. 121
Fitzpatrick E. L., 1999, PASP, 111, 63
Folatelli G., et al., 2010, AJ, 139, 120
Foley R. J., Kasen D., 2011, ApJ, 729, 55
Foley R. J., et al., 2018, MNRAS, 475, 193
Freedman W. L., et al., 2009, ApJ, 704, 1036
Friedman A. S., et al., 2015, ApJS, 220, 9
Frieman J. A., et al., 2008, AJ, 135, 338
Gallagher J. S., Garnavich P. M., Caldwell N., Kirshner R. P., Jha S. W., Li W., Ganeshalingam 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. R., 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

Hand J., Liu S., Galbany L., Perrefort D., Wood-Vasey W. M., Burns C., 2021, preprint, (arXiv:2102.08980)
Hayden B. T., Gupta R. R., Garnavich P. M., Mannucci F., Nichol R. C., Sako M., 2013, ApJ, 764, 191
Hernandez M., et al., 2000, MNRAS, 319, 223
Hicken M., et al., 2009a, ApJ, 700, 331
Hicken M., Wood-Vasey W. M., Blondin S., Challis P., Jha S., Kelly P. L., Rest A., Kirshner R. P., 2009b, ApJ, 700, 1097
Hicken M., et al., 2012, ApJS, 200, 12
Hoang T., 2017, ApJ, 836, 13
Hoffman M. D., Gelman A., 2014, J. Machine Learning Res., 15, 1593
Hounsell R., et al., 2018, ApJ, 867, 23
Hsiao E. Y., 2009, PhD thesis, Univ. Victoria
Hsiao E. Y., Conley A., Howell D. A., Sullivan M., Pritchet C. J., Carlberg R. G., Nugent P. E., Phillips M. M., 2007, ApJ, 663, 1187
Huber M., et al., 2015, The Astronomer's Telegram, 7153, 1
Ivezić Ž., et al., 2019, ApJ, 873, 111
Jha S., et al., 1999, ApJS, 125, 73
Jha S., et al., 2006, AJ, 131, 527
Johansson J., et al., 2021, preprint, (arXiv:2105.06236)
Jones D. O., et al., 2018, ApJ, 867, 108
Jones D. O., et al., 2019, ApJ, 881, 19
Jones D. O., et al., 2021, ApJ, 908, 143
Kaiser N., et al., 2010, in Stepp L. M., Gilmozzi R., Hall H. J., eds, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 7733, Ground-based and Airborne Telescopes III. p. 77330E, doi:10.1117/12.859188
Kelly B. C., 2007, ApJ, 665, 1489
Kelly P. L., Hicken M., Burke D. L., Mandel K. S., Kirshner R. P., 2010, ApJ, 715, 743
Kelsey L., et al., 2021, MNRAS, 501, 4861
Krisciunas K., Hastings N. C., Loomis K., McMillan R., Rest A., Riess A. G., Stubbs C., 2000, ApJ, 539, 658
Krisciunas K., et al., 2003, AJ, 125, 166
Krisciunas K., et al., 2004a, AJ, 127, 1664
Krisciunas K., et al., 2004b, AJ, 128, 3034
Krisciunas K., et al., 2007, AJ, 133, 58
Krisciunas K., et al., 2017, AJ, 154, 211
Lampeitl H., et al., 2010, ApJ, 722, 566
Leaman J., Li W., Chornock R., Filippenko A. V., 2011, MNRAS, 412, 1419
Léget P. F., et al., 2020, A&A, 636, A46
Leloudas G., et al., 2009, A&A, 505, 265
Li W. D., et al., 2000, in Holt S. S., Zhang W. W., eds, American Inst. Phys. Conf. Series Vol. 522, Cosmic Explosions: Tenth Astrophysics Conference. pp 103–106 (arXiv:astro-ph/9912336), doi:10.1063/1.1291702
Li W., et al., 2011, MNRAS, 412, 1441
Magnier E. A., et al., 2013, ApJS, 205, 20
Magnier E. A., et al., 2020, ApJS, 251, 6
Mandel K. S., Wood-Vasey W. M., Friedman A. S., Kirshner R. P., 2009, ApJ, 704, 629
Mandel K. S., Narayan G., Kirshner R. P., 2011, ApJ, 731, 120
Mandel K. S., Foley R. J., Kirshner R. P., 2014, ApJ, 797, 75
Mandel K. S., Scolnic D. M., Shariff H., Foley R. J., Kirshner R. P., 2017, ApJ, 842, 93
Mandel K. S., Thorp S., Narayan G., Friedman A. S., Avelino A., 2020, preprint, (arXiv:2008.07538)
Mannucci F., Della Valle M., Panagia N., Cappellaro E., Cresci G., Maiolino R., Petrosian A., Turatto M., 2005, A&A, 433, 807
Mannucci F., Della Valle M., Panagia N., 2006, MNRAS, 370, 773
March M. C., Trotta R., Berkes P., Starkman G. D., Vaudrevange P. M., 2011, MNRAS, 418, 2308
Meng X.-L., 1994, Ann. Statistics, 22, 1142
Moreno-Raya M. E., López-Sánchez Á. R., Mollá M., Galbany L., Vílchez J. M., Carnero A., 2016a, MNRAS, 462, 1281
Moreno-Raya M. E., Mollá M., López-Sánchez Á. R., Galbany L., Vílchez J. M., Carnero Rosell A., Domínguez I., 2016b, ApJ, 818, L19
Neill J. D., et al., 2009, ApJ, 707, 1449
Nobili S., Goobar A., 2008, A&A, 487, 19
Pan Y. C., et al., 2014, MNRAS, 438, 1391
Pastorello A., et al., 2007a, MNRAS, 376, 1301
Pastorello A., et al., 2007b, MNRAS, 377, 1531
Perlmutter S., et al., 1999, ApJ, 517, 565
Phillips M. M., 1993, ApJ, 413, L105
Phillips M. M., Lira P., Suntzeff N. B., Schommer R. A., Hamuy M., Maza J., 1999, AJ, 118, 1766
Phillips M. M., et al., 2006, AJ, 131, 2615
Phillips M. M., et al., 2019, PASP, 131, 014001
Pignata G., et al., 2008, MNRAS, 388, 971
Ponder K. A., Wood-Vasey W. M., Weyant A., Barton N. T., Galbany L., Garnavich P., Matheson T., 2020, preprint, (arXiv:2006.13803)
Popovic B., Brout D., Kessler R., Scolnic D., Lu L., 2021, ApJ, 913, 49
Pruzhinskaya M. V., Novinskaya A. K., Pauna N., Rosnet P., 2020, MNRAS, 499, 5121
Rau A., et al., 2009, PASP, 121, 1334
Rest A., et al., 2014, ApJ, 795, 44
Riess A. G., Press W. H., Kirshner R. P., 1996, ApJ, 473, 588
Riess A. G., et al., 1998, AJ, 116, 1009
Riess A. G., et al., 1999, AJ, 117, 707
Riess A. G., et al., 2016, ApJ, 826, 56
Riess A. G., Casertano S., Yuan W., Macri L. M., Scolnic D., 2019, ApJ, 876, 85
Rigault M., et al., 2013, A&A, 560, A66
Rigault M., et al., 2015, ApJ, 802, 20
Rigault M., et al., 2020, A&A, 644, A176
Roman M., et al., 2018, A&A, 615, A68
Rose B. M., Garnavich P. M., Berg M. A., 2019, ApJ, 874, 32
Rose B. M., Rubin D., Strolger L., Garnavich P. M., 2021, ApJ, 909, 28
Rubin D. B., 1984, Ann. Statistics, 12, 1151
Sako M., et al., 2011, ApJ, 738, 162
Sako M., et al., 2018, PASP, 130, 064002
Sasdelli M., Ishida E. E. O., Hillebrandt W., Ashall C., Mazzali P. A., Prentice S. J., 2016, MNRAS, 460, 373
Scannapieco E., Bildsten L., 2005, ApJ, 629, L85
Schlafly E. F., Finkbeiner D. P., 2011, ApJ, 737, 103
Schlafly E. F., et al., 2012, ApJ, 756, 158
Schlafly E. F., et al., 2016, ApJ, 821, 78
Scolnic D. M., Riess A. G., Foley R. J., Rest A., Rodney S. A., Brout D. J., Jones D. O., 2014a, ApJ, 780, 37
Scolnic D., et al., 2014b, ApJ, 795, 45
Scolnic D., et al., 2015, ApJ, 815, 117
Scolnic D. M., et al., 2018, ApJ, 859, 101
Shappee B. J., et al., 2014, ApJ, 788, 48
Smith M., et al., 2020, MNRAS, 494, 4426
Spergel D., et al., 2015, preprint, (arXiv:1503.03757)
Stan Development Team 2020, Stan Modelling Language Users Guide and Reference Manual v.2.25. https://mc-stan.org
Stanishev V., et al., 2007, A&A, 469, 645
Stanishev V., et al., 2018, A&A, 615, A45
Stritzinger M. D., et al., 2011, AJ, 142, 156
Stubbs C. W., Doherty P., Cramer C., Narayan G., Brown Y. J., Lykke K. R., Woodward J. T., Tonry J. L., 2010, ApJS, 191, 376
Sullivan M., et al., 2003, MNRAS, 340, 1057
Sullivan M., et al., 2006, ApJ, 648, 868
Sullivan M., et al., 2010, MNRAS, 406, 782
Tripp R., 1998, A&A, 331, 815
Uddin S. A., Mould J., Lidman C., Ruhlmann-Kleider V., Zhang B. R., 2017, ApJ, 848, 56
Uddin S. A., et al., 2020, ApJ, 901, 143
Valentini G., et al., 2003, ApJ, 595, 779
Villar V. A., et al., 2020, ApJ, 905, 94
Wang X., et al., 2008, ApJ, 675, 626
Weyant A., et al., 2018, AJ, 155, 201
Wood-Vasey W. M., et al., 2008, ApJ, 689, 377

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 $10^{10}\mathrm{M}_\odot$, 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, $\boldsymbol{\Delta\mu} = (\Delta\mu_1, \ldots, \Delta\mu_N)$, we assume that the residuals are consistent with a step function, with mean $m_<$ for host masses of $M_* < M_{*,\mathrm{step}}$, and mean $m_>$ for host masses of $M_* \geq M_{*,\mathrm{step}}$. We allow for the presence of some residual scatter, σ_{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

$$P(\Delta\mu_s|m_<, m_>, \sigma_{\mathrm{res}}, \log_{10} M_{*,\mathrm{step}}) = \begin{cases} N(\Delta\mu_s|m_<, \sigma_s^2 + \sigma_{\mathrm{res}}^2) & \text{if } M_{*,s} < M_{*,\mathrm{step}} \\ N(\Delta\mu_s|m_>, \sigma_s^2 + \sigma_{\mathrm{res}}^2) & \text{if } M_{*,s} \geq M_{*,\mathrm{step}}. \end{cases} \quad \text{(A1)}$$

To identify the optimal step location, $M_{*,\mathrm{step}}$, for a sample of Hubble residuals, we compute the marginal likelihood, $P(\boldsymbol{\Delta\mu}|\log_{10} M_{*,\mathrm{step}})$, 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,

$$P(\boldsymbol{\Delta\mu}|m_<, m_>, \sigma_{\mathrm{res}}, \log_{10} M_{*,\mathrm{step}}) = S_< N(m_<|\nu_<, \sigma_<^2) \times S_> N(m_>|\nu_>, \sigma_>^2), \quad \text{(A2)}$$

where

$$\nu_< = \sigma_<^2 \left[\sum_{\substack{i \\ M_{*,i} < M_{*,\mathrm{step}}}} \frac{\Delta\mu_i}{\sigma_i^2 + \sigma_{\mathrm{res}}^2} \right], \quad \text{(A3)}$$

$$\sigma_<^2 = \left[\sum_{\substack{i \\ M_{*,i} < M_{*,\mathrm{step}}}} \frac{1}{\sigma_i^2 + \sigma_{\mathrm{res}}^2} \right]^{-1}, \quad \text{(A4)}$$

and

$$S_< = \sqrt{\frac{2\pi\sigma_<^2}{\prod_i 2\pi(\sigma_i^2 + \sigma_{\mathrm{res}}^2)}} \exp\left[-\frac{1}{2} \left(\sum_i \frac{\Delta\mu_i^2}{\sigma_i^2 + \sigma_{\mathrm{res}}^2} - \frac{\nu_<^2}{\sigma_<^2} \right) \right], \quad \text{(A5)}$$

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 $m_<$, $m_>$ and σ_{res}, and integrating over these parameters. We assume conditional priors of $P(m_<|\sigma_{\mathrm{res}}) = N(m_<|0, 1)$ and $P(m_>|\sigma_{\mathrm{res}}) = N(m_>|0, 1)$, 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

$$P(\boldsymbol{\Delta\mu}|\sigma_{\mathrm{res}}, \log_{10} M_{*,\mathrm{step}}) = S_<\tilde{S}_< \times S_>\tilde{S}_>, \quad \text{(A6)}$$

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

$$\tilde{S}_< = \sqrt{\frac{1}{2\pi(\sigma_<^2 + 1)}} \exp\left[-\frac{\nu_<^2}{2(\sigma_<^2 + 1)} \right], \quad \text{(A7)}$$

with $\tilde{S}_>$ being computed similarly. Introducing a uniform prior, $P(\sigma_{\mathrm{res}}) = U(\sigma_{\mathrm{res}}|0, 0.2)$, on the residual scatter parameter, we can numerically marginalise over this to compute the marginal likelihood of interest,

$$P(\boldsymbol{\Delta\mu}|\log_{10} M_{*,\mathrm{step}}) = \int_{-\infty}^{\infty} P(\boldsymbol{\Delta\mu}|\sigma_{\mathrm{res}}, \log_{10} M_{*,\mathrm{step}}) P(\sigma_{\mathrm{res}})\, \mathrm{d}\sigma_{\mathrm{res}} = \frac{1}{0.2} \int_0^{0.2} S_<\tilde{S}_< \times S_>\tilde{S}_>\, \mathrm{d}\sigma_{\mathrm{res}}, \quad \text{(A8)}$$

for a given step location, $\log_{10} M_{*,\mathrm{step}}$.

For the *No-Split* Hubble residuals, we evaluate Equation (A8) for a range of host masses to find the step position(s) for which those Hubble residuals are most probable.

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^s population distribution width, σ_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 σ_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^s 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 σ_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 $10^{10}\mathrm{M}_\odot$. 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 σ_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 $\text{Half-}N(0, 2^2)$ 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 σ_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 χ^2).

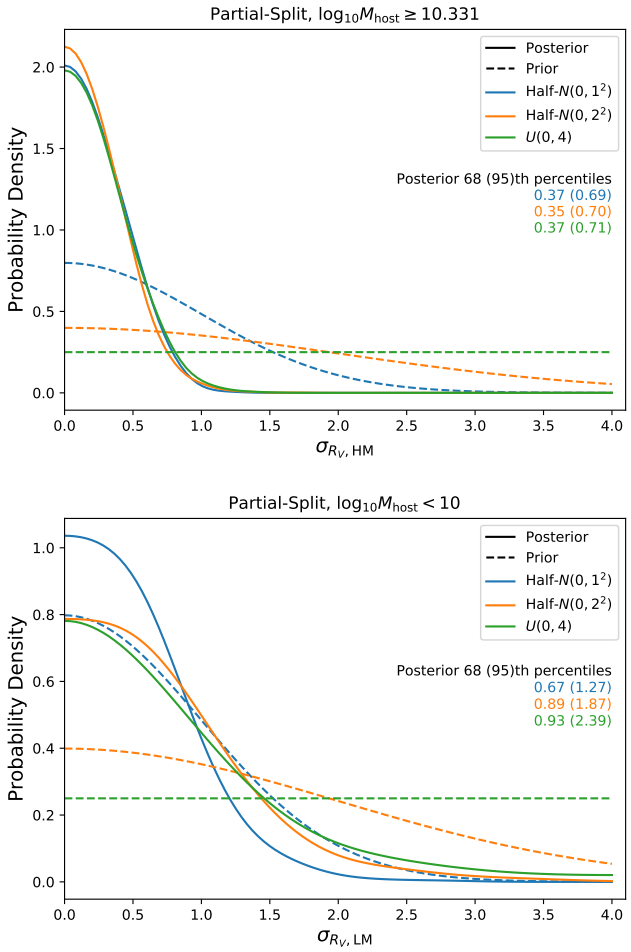

Figure B1. Plot illustrating the effect of hyperprior choice on the inference of σ_R under the *Partial-Split* model. Solid lines show the marginal posterior distributions of σ_R for three choices of prior, with dashed lines plotting the corresponding priors. (top panel) Posterior distribution of σ_R 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 σ_R marginal distribution in Figure 7). (bottom panel) Constraints on σ_R for host galaxies less massive than $10^{10}\mathrm{M}_\odot$, a more unbalanced mass split. This illustrates the 'worst-case scenario' in prior sensitivity.

In our case, we choose to use

$$\chi^2(\boldsymbol{f};\boldsymbol{\phi}) = \sum_{n=1}^{N} \frac{[f_n - f_n^{\rm model}(\boldsymbol{\phi})]^2}{\sigma_n^2} \tag{C1}$$

as our test statistic[4], where $\boldsymbol{\phi} = (\mu_s^{\rm phot}, \theta_1^s, A_V^s, \delta M_s, \boldsymbol{e}_s)$ represents our model parameters; $\boldsymbol{f} = (f_1, \ldots, f_N)$ is the vector of fluxes being tested; $f_n^{\rm model}$ is the model flux computed at the time, and in the passband, of f_n, given the parameters $\boldsymbol{\phi}$; and σ_n is the flux error associated with f_n. For a given supernova, we have a set of N flux observations, $\boldsymbol{f}^{\rm obs} = (f_1^{\rm obs}, \ldots, f_N^{\rm obs})$, with associated errors, $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_N)$. Conditional on $\boldsymbol{f}^{\rm obs}$, we will also have a set of M samples, $(\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_M)$, from the posterior distribution

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

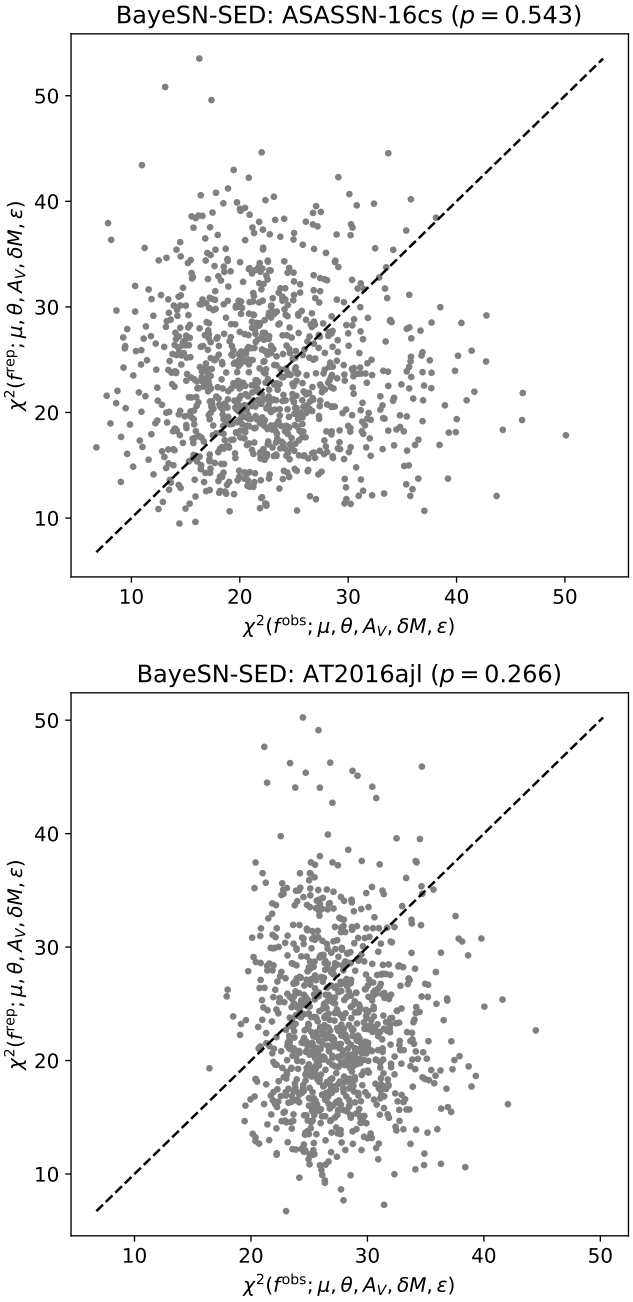

Figure C1. Illustration of the posterior predictive p-value estimation for the two light curve fits in Figure 11. Each point corresponds to a single posterior sample, with the y-axes plotting posterior predictive values, $\chi^2(\boldsymbol{f}_m^{\rm rep}(\boldsymbol{\phi}_m); \boldsymbol{\phi}_m)$,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

$P(\boldsymbol{\phi}|\boldsymbol{f}^{\rm obs})$. For the mth posterior sample, $\boldsymbol{\phi}_m$, we can compute a vector of model fluxes, $\boldsymbol{f}_m^{\rm model}(\boldsymbol{\phi}_m)$, and a vector of replicated observations, $\boldsymbol{f}_m^{\rm rep}(\boldsymbol{\phi}_m)$, where

$$f_{m,n}^{\rm rep}(\boldsymbol{\phi}_m) \sim N(f_{m,n}^{\rm model}(\boldsymbol{\phi}_m), \sigma_n^2) \tag{C2}$$

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

$$p = P[\chi^2(\boldsymbol{f}^{\rm rep}(\boldsymbol{\phi}); \boldsymbol{\phi}) \geq \chi^2(\boldsymbol{f}^{\rm obs}; \boldsymbol{\phi})|\boldsymbol{f}^{\rm obs}], \tag{C3}$$

by computing the fraction of posterior samples for which $\chi^2(\boldsymbol{f}_m^{\rm rep}(\boldsymbol{\phi}_m);\boldsymbol{\phi}_m) \geq \chi^2(\boldsymbol{f}^{\rm obs};\boldsymbol{\phi}_m)$. 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 $\chi^2(\boldsymbol{f}^{\rm rep}(\boldsymbol{\phi});\boldsymbol{\phi})$ vs. $\chi^2(\boldsymbol{f}^{\rm obs};\boldsymbol{\phi})$, based on 1000 posterior samples of $[\boldsymbol{\phi}, \boldsymbol{f}^{\rm rep}(\boldsymbol{\phi})]$. The estimated p-value is computed from the fraction of samples lying above the dashed 1 : 1 line.

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