Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS 510, 3939–3966 (2022)
Advance Access publication 2021 December 2

https://doi.org/10.1093/mnras/stab3496

A hierarchical Bayesian SED model for Type Ia supernovae in the optical to near-infrared

Kaisey S. Mandel[1,2,3]★ Stephen Thorp,[1] Gautham Narayan[4,5] Andrew S. Friedman[6] and Arturo Avelino[7]

[1]Institute of Astronomy and Kavli Institute for Cosmology, Madingley Road, Cambridge CB3 0HA, UK
[2]Statistical Laboratory, DPMMS, University of Cambridge, Wilberforce Road, Cambridge CB3 0WB, UK
[3]The Alan Turing Institute, Euston Road, London NW1 2DB, UK
[4]Department of Astronomy, 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
[6]Center for Astrophysics and Space Sciences, University of California, San Diego, CA 92093, USA
[7]Center for Astrophysics, Harvard and Smithsonian, Cambridge, MA 02138, USA

Accepted 2021 November 28. Received 2021 November 28; in original form 2020 August 26

ABSTRACT

While conventional Type Ia supernova (SN Ia) cosmology analyses rely primarily on rest-frame optical light curves to determine distances, SNe Ia are excellent standard candles in near-infrared (NIR) light, which is significantly less sensitive to dust extinction. An SN Ia spectral energy distribution (SED) model capable of fitting rest-frame NIR observations is necessary to fully leverage current and future SN Ia data sets from ground- and space-based telescopes including *HST*, LSST, *JWST*, and *RST*. We construct a hierarchical Bayesian model for SN Ia SEDs, continuous over time and wavelength, from the optical to NIR (B through H, or $0.35-1.8$ µm). We model the SED as a combination of physically distinct host galaxy dust and intrinsic spectral components. The distribution of intrinsic SEDs over time and wavelength is modelled with probabilistic functional principal components and the covariance of residual functions. We train the model on a nearby sample of 79 SNe Ia with joint optical and NIR light curves by sampling the global posterior distribution over dust and intrinsic latent variables, SED components and population hyperparameters. Photometric distances of SNe Ia with NIR data near maximum obtain a total RMS error of 0.10 mag with our BayeSN model, compared to 0.13–0.14 mag with SALT2 and SNooPy for the same sample. Jointly fitting the optical and NIR data of the full sample up to moderate reddening (host $E(B-V) < 0.4$) for a global host dust law, we find $R_V = 2.9 \pm 0.2$, consistent with the Milky Way average.

Key words: methods: statistical – transients: supernovae – distance scale.

1 INTRODUCTION

Type Ia supernovae (SNe Ia) are effective cosmological probes as ‘standardizable candles’: their peak luminosities can be inferred from their optical light-curve shapes and colours, so their distances can be estimated from their apparent brightnesses. Precise and accurate SN Ia distances with small systematic errors are essential to accurate constraints on the cosmic expansion history, including local measurements of the Hubble constant (Burns et al. 2018; Riess et al. 2019), the late-time cosmic acceleration (Riess et al. 1998; Perlmutter et al. 1999), and the properties of the dark energy driving it, in particular, its equation-of-state parameter w (e.g. Scolnic et al. 2018; Abbott et al. 2019). Currently, there is a significant 4.4σ tension between the value of H_0 locally inferred from SNe Ia via the distance ladder (74.03 ± 1.42 km s^{-1} Mpc^{-1}; Riess et al. 2019) and the value derived from Planck CMB analysis assuming the ΛCDM cosmological model (67.4 ± 0.5 km s^{-1} Mpc^{-1}; Planck Collaboration 2020). Since this tension could potentially be a sign of new physics, it is imperative to test for systematic errors with empirical cross-checks (e.g. Dhawan, Jha & Leibundgut 2018). With increasing sample sizes, nearby SNe Ia will be able to constrain the growth of structure as probes of the peculiar velocity field (e.g. Howlett et al. 2017; Huterer et al. 2017; Graziani et al. 2020). Further cosmological constraints can also be derived from strong or weak lensing of SNe Ia (e.g. Goldstein et al. 2018; Dhawan et al. 2020; Macaulay et al. 2020). In this paper, we present a new data-driven statistical model, BayeSN, for SN Ia spectral energy distributions (SEDs) to analyse light curves and extract more precise and accurate distances from current and future surveys by exploiting the advantageous properties of SNe Ia in the near-infrared (NIR).

The current global sample used for cosmology, derived from the SDSS-II, SNLS, Pan-STARRS (PS1), low-z and *HST* surveys, has grown to over a thousand SNe Ia (Pantheon; Scolnic et al. 2018). Future surveys, such as the Legacy Survey of Space and Time (LSST, Ivezić et al. 2019) provided by the Vera Rubin Observatory, will boost that number by orders of magnitude. The constraints on dark energy with the current sample will soon be limited, not by

★ E-mail: kmandel@ast.cam.ac.uk

© The Author(s) 2021.
Published by Oxford University Press on behalf of Royal Astronomical Society. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Downloaded from https://academic.oup.com/mnras/article/510/3/3939/6448478 by University of Cambridge user on 14 December 2022

statistical uncertainties from the numbers of SNe, but by systematic errors. In recent analyses, photometric calibration and SN model uncertainties dominate the systematic error budget (JLA: Betoule et al. 2014; PS1: Scolnic et al. 2018; DES: Brout et al. 2019). The calibration systematics are now being tamed by improved instrumental calibration (e.g. Regnault et al. 2015), better cross-calibration between surveys (Scolnic et al. 2015; Currie et al. 2020), better networks of photometric standards (Narayan et al. 2016, 2019), and by ongoing efforts to replace the heterogeneous low-redshift sample with a large, unbiased, homogeneous sample obtained on a precisely calibrated photometric system (PS1, Foundation Survey; Foley et al. 2018b; Jones et al. 2019). LSST will increase the cosmologically useful SN Ia sample to $\sim 10^5$ over its 10 yr duration. It will further diminish cross-survey calibration systematics by replacing previous high-redshift SN Ia surveys with a single, homogeneous, and large SN Ia sample taken on a single system. However, systematic errors due to the statistical models and methods used to analyse SN Ia light-curve data will remain.

Observations probing the rest-frame near-infrared (NIR, particularly $\lambda \gtrsim 1\,\mu$m, e.g. *YJH* bands) are a route to more precise and accurate distances. NIR observations of SNe Ia significantly improve their cosmological utility. Unlike in the optical, where they must be standardized via correlations of optical luminosity with light-curve shape and colour, SNe Ia are excellent, nearly standard candles in the NIR, showing little intrinsic luminosity variation ($\sim$0.1 mag) at peak (e.g. Krisciunas, Phillips & Suntzeff 2004c; Wood-Vasey et al. 2008; Mandel et al. 2009; Contreras et al. 2010; Barone-Nugent et al. 2012; Kattner et al. 2012; Phillips 2012; Burns et al. 2018; Stanishev et al. 2018; Avelino et al. 2019). The NIR also has significantly reduced sensitivity to dust extinction relative to the optical (by factors of 4–8, comparing NIR *YJH* to optical *B*). Dhawan et al. (2018) showed how a small set of SNe Ia, used as NIR standard candles to measure H_0, can replace a much larger optical sample, while still providing a 4.3 per cent measurement (consistent with Riess et al. 2019), without any light-curve shape or colour corrections as are required in the optical. We recently compiled a sample of 89 nearby SNe Ia with optical and NIR light curves passing standard quality cuts (Avelino et al. 2019). Using 56 SNe Ia with NIR data near peak brightness, where the luminosity dispersion is minimal, we found a 35 per cent reduction in Hubble Diagram scatter (i.e. more precise distances) when using SNe Ia as NIR standard candles, relative to conventional optical-only fits to the same SNe.

The combination of optical and NIR data better constrains the host galaxy dust extinction and the shape of the dust law as a function of λ (parametrized by R_V) (Krisciunas et al. 2007; Burns et al. 2014), and significantly improves the accuracy and precision of SN Ia distances (Mandel, Narayan & Kirshner 2011). The nature of the dust in SN Ia host galaxies is fundamental to the largest 'correction' in standardizing SNe Ia, that due to colour. Incorrect modelling interpretation of the SN Ia colour–magnitude relation is therefore a major source of systematic error in SN distances. However, the correct values(s) of the R_V parametrizing the dust extinction law has long been a matter of confusion, and its proper estimation is fraught with statistical subtleties.

Very early analyses that found unphysically low values $R_V \lesssim 1$ (Branch & Tammann 1992) did not account for correlations between the luminosity, colour, and light-curve shape (later modelled by e.g. Phillips 1993; Riess, Press & Kirshner 1996a; Phillips et al. 1999). Riess, Press & Kirshner (1996b) noted that confusing intrinsic colour–luminosity variation with dust effects would lead to mistakenly lower estimated R_V values. Simple linear regression analyses of SN extinguished absolute magnitudes against apparent colours and light-curve shapes have led to apparent colour–magnitude slopes (e.g. β in the Tripp formula) that have sometimes been interpreted as low dust R_V values (Tripp 1998; Tripp & Branch 1999; Guy et al. 2005; Astier et al. 2006; Conley et al. 2007; Freedman et al. 2009). Kessler et al. (2009b) and Scolnic et al. (2014) highlighted the relevance of colour dispersion to estimating β, the latter finding a Milky Way dust-like colour–magnitude slope. Mandel et al. (2017) showed that statistical confounding of the intrinsic colour–luminosity correlation and dispersion with the extrinsic effects of dust leads to estimates of β that are biased low relative to the true dust R_V, and a probabilistic generative model with explicit parameters for these physically distinct effects led to a Bayesian estimate of $R_V = 2.8 \pm 0.3$, consistent with the Milky Way average.

Anomalously low $R_V \approx 1.5-1.8$ values have been estimated for a few very highly reddened SNe Ia ($E(B-V) > 1$) (e.g. Elias-Rosa et al. 2006, 2008; Wang et al. 2008; Amanullah et al. 2014).[1] While the origin of these low R_V estimates is still under investigation (Wang 2005; Goobar 2008; Amanullah & Goobar 2011; Phillips et al. 2013; Amanullah et al. 2015; Johansson et al. 2017; Bulla et al. 2018a; Bulla, Goobar & Dhawan 2018b), these very red SNe are not present in the cosmological sample, due to the standard cut on peak apparent SN colour ($B - V \lesssim 0.3$). When only low- to moderately reddened normal SNe Ia with apparent colours consistent with the cosmological sample are considered, values of $R_V \approx 2.5-3$ have generally been estimated in nearby samples, often by utilizing spectroscopic or NIR data to break the degeneracy between intrinsic colours and dust in the optical (Folatelli et al. 2010; Chotard et al. 2011; Foley & Kasen 2011; Mandel et al. 2011; Phillips 2012; Burns et al. 2014; Mandel et al. 2017; Léget et al. 2020).

The excellent properties of the NIR have not been fully integrated into and leveraged by the statistical models routinely used for SN Ia cosmology. We have constructed a new, hierarchical Bayesian model, BAYESN, for time-dependent SN Ia spectral energy distributions (SEDs) from the optical to NIR wavelengths. With NIR coverage, our model leverages the low luminosity dispersion in the NIR, while its wide optical-to-NIR wavelength range enables it to more stringently constrain the host galaxy dust, and the dust law, affecting the SNe Ia. These two advantages enable us to more accurately improve our model of the intrinsic SED coherently across all wavelengths. While it produces the best distance estimates when fitting complete light curves across the full wavelength range, as a Bayesian model, it also makes the most effective use of the observations available in any partial data set, e.g. optical-only, NIR-only, while marginalizing over the unobserved parts of the SED.

BAYESN is an important tool, not only for properly analysing current data sets, but also extracting optimal distances and robust cosmological constraints from future optical and NIR SNe Ia observations. Beyond the data sets analysed in this work, the ability to effectively leverage joint optical and NIR observations is crucial for fully exploiting a number of recent and current surveys and forthcoming data sets, including the Carnegie Supernova Project-II (CSP-II; Phillips et al. 2019), the Foundation Supernova Survey (Foley et al. 2018b), and Young Supernova Experiment (YSE; Jones et al. 2021) with Pan-STARRS, RAISIN (GO-13046, GO-14216) and SIRAH (GO-15889) with the *Hubble Space Telescope* (*HST*), the ESO VISTA Extragalactic Infrared Legacy Survey (VEILS), and the DEHVILS Survey using UKIRT. This is also important for LSST,

[1] However, not all highly reddened SNe Ia have low R_V, e.g. SN2012cu with $E(B-V) \approx 1$ and $R_V \approx 3$, compatible with the Milky Way average (Huang et al. 2017).

Downloaded from https://academic.oup.com/mnras/article/510/3/3939/6448478 by University of Cambridge user on 14 December 2022

which will observe SNe Ia in *ugrizy*, and will therefore probe rest-frame z or y to redshifts $z \lesssim 0.3$. The Nancy Grace *Roman Space Telescope* (RST, formerly WFIRST) will have a dedicated SN survey and its wide NIR filters will overlap with rest-frame *YJH* out to redshifts $z \lesssim 1$, 0.7, 0.4 respectively.

1.1 Comparison to existing models

The models used to analyse SN Ia light curves and estimate distances are entirely empirical and are learned from the data. The conventional approach has a number of shortcomings that need to be addressed to exploit fully the data and to control astrophysical and modelling systematics. The model most commonly used for fitting optical SN Ia light curves is SALT2 (Guy et al. 2007, 2010; Betoule et al. 2014). It models the SN Ia SED in phase (rest-frame time since peak luminosity) and wavelength, as a function of optical light-curve shape (x_1) and apparent colour (c) at peak. SN Ia light-curve fits estimate these parameters and the optical peak apparent magnitude m_B. Photometric distances are obtained from a fitted linear dependence of SN Ia absolute magnitude on light-curve shape and colour (Tripp 1998):

$$\mu_s = m_{B,s} - M_B + \alpha\, x_{1,s} - \beta\, c_s, \tag{1}$$

where μ_s is the distance modulus of an individual SN s, $(m_{B,s}, x_{1,s}, c_s)$ are parameters obtained from the SALT2 fit of the individual SN s, and (α, β, M_B) are global (or population) parameters describing the luminosity trends with light-curve shape and colour, and the absolute magnitude intercept at $x_1 = c = 0$, respectively.

Major shortcomings of the conventional approach are:

(i) Residual (‘Intrinsic’)[2] scatter systematic error: Spectral variations of SN Ia light-curve data around the best-fitting SED model in excess of measurement error are accounted for by an error model that contributes additional covariance to the fitted light-curve parameters. Even accounting for this, a Hubble residual scatter with $\sigma \approx 0.13$ mag around equation (1) still remains. Its wavelength-dependence is accounted for in simulations with an ‘intrinsic scatter model’ (Kessler et al. 2013; Mosher et al. 2014). It is not well constrained, and currently there are two options: one with 30 per cent chromatic variation and 70 per cent achromatic variation (Guy et al. 2010), and the other, based on Chotard et al. (2011), with a 75 per cent: 25 per cent split. Scolnic et al. (2014) showed that both models are consistent with the cosmological SN Ia data, therefore the current optical data alone cannot discriminate between the two. However, the impact of changing the assumed model for the residual scatter in a cosmological analysis results in a shift $\Delta w \sim 0.04$, and thus is a dominant systematic error.

Employing the correct residual covariances across phase and wavelength is crucial to the proper quantification of uncertainties and weighting of the SN data. Our BayeSN SED model coherently estimates the intrinsic residual covariance across phase and wavelength simultaneously with the training of the entire hierarchical model, and this covariance is employed when fitting SN light curves to estimate dust and distance, while marginalizing over the SED residuals.

(ii) Degeneracy between intrinsic versus dust colour–luminosity variations: The largest ‘correction’ in equation (1) is due to colour, but the conventional analysis treats it in a simplistic way. Fundamentally, intrinsic variation and dust have physically distinct effects on the SN Ia SED. However, the SALT2 model assumes that all colour variation can be described by the peak apparent $B - V$ colour parameter c and a single, effective colour law, $CL(\lambda)$. The conventional approach of fitting a single linear function for extinguished absolute magnitude versus apparent colour confounds the two effects (Mandel et al. 2017). In contrast, our BayeSN SED model allows for a probabilistic, physically motivated combination of different spectral effects from intrinsic SN variation and dust across time and wavelength.

(iii) Lack of NIR coverage: The most widely used SALT2.4 model is only specified over rest-frame wavelengths of 0.2−0.9 µm, though the colour law for $\lambda > 0.7$ µm is an extrapolation. Although optical surveys, such as Foundation (Foley et al. 2018b), routinely obtain z-band data, they cannot be fit by SALT2 for nearby SNe Ia. SALT2.4 is incapable of leveraging the useful properties of SNe Ia in the rest-frame NIR at $\lambda \gtrsim 1$ µm.

In contrast, our BayeSN SED model is trained on data covering optical to NIR wavelengths extending from B through H-band (0.35−1.8 µm) and uses Bayesian inference to combine information over the full phase and wavelength range for optimal estimates of dust and distance.

(iv) Lack of SED modelling of astrophysical correlations: an apparent correlation between Hubble residuals and the host galaxy stellar mass (Kelly et al. 2010; Sullivan et al. 2010) is conventionally addressed somewhat simplistically by splitting the scalar absolute magnitude term M_B in equation (1) by host mass. However, in principle, the astrophysical correlation of SN Ia luminosity with host mass should be accounted for at the fundamental level of the SED. While we do not address this in this paper, Thorp et al. (2021) recently demonstrated how BayeSN can be extended to introduce host-mass dependence.

Although SALT2 is the most common SN model used in cosmology, there are alternatives. SNooPy is an optical-NIR model for SN Ia light curves defined in discrete rest-frame *uBVgriYJH* passbands (Burns et al. 2011). It is not a model for the continuous SED; rather, for each discrete rest-frame passband it has a template light curve that varies as a function of a shape parameter (either $\Delta m_{15}(B)$ or s_{BV}). It requires the calculation of K-corrections of the photometry from each observer-frame passband into a corresponding rest-frame model passband as a preprocessing step. The template light-curve model is then fit to the K-corrected data in the rest-frame bands. This 1-to-1 mapping is not ideal, as there are redshifts at which, for example, wide *HST* WFC3 NIR filters significantly cover two rest-frame model passbands, so the observed light curves are actually sensitive to the statistical properties of the underlying SED in both rest-frame bands. Furthermore, the K-correction calculation employs an ad-hoc ‘mangling’ procedure to match a spectral template to observed colours independently at each epoch. This is prone to overfitting, its uncertainties are difficult to propagate, and is not viable for the noisy, sparse data typical of high-z light curves, in which the light curves in different passbands may be irregularly and asynchronously sampled. We compare the results from BayeSN to those from applying SALT2 and SNooPy to the same SNe Ia in Section 5.

SNEMO (Saunders et al. 2018; Rose et al. 2020) and SUGAR (Léget et al. 2020) are recent empirical models built from optical spectrophotometric time series. Whereas SNEMO is a principal components model for the optical SED, SUGAR models the spectral dependence on factors composed of spectral line characteristics. However, they only cover rest-frame $0.33 < \lambda < 0.86$ µm, and so they cannot leverage the valuable NIR at $\lambda \gtrsim 1$ µm.

[2]The terminology of ‘intrinsic scatter’ here is a confusing misnomer. In the conventional SALT2 framework that is agnostic about the distinction between intrinsic and dust effects, there is no reason to attribute all of its residual scatter to variation intrinsic to the supernovae, even if the model were true.

Downloaded from https://academic.oup.com/mnras/article/510/3/3939/6448478 by University of Cambridge user on 14 December 2022

1.2 Outline of paper

The outline of this paper is as follows. In Section 2, we describe our new hierarchical Bayesian model for SN Ia SEDs in the optical to NIR. In Section 3, we describe the compilation of optical and NIR SN Ia light-curve data that we analyse. In Section 4, we describe our computational implementation for training the BAYESN model and fitting SN Ia light curves. In Section 5, we present our results, including a Hubble diagram showing the improvement in distances (to 0.10 mag total RMS error) obtained from BAYESN fits to optical and NIR data compared to current methods applied to the same sample. We also describe our inferences about host galaxy dust, for which we constrain a global value of $R_V = 2.9 \pm 0.2$ for our sample with $E(B - V)_{\rm host} \lesssim 0.4$. In Section 6, we conclude.

In Appendix Section A, we provide some background on Bayesian functional principal component analysis, and in Appendix Section B, we describe the extension of our model to a second functional component.

2 THE STATISTICAL MODEL

To construct and train our SN Ia SED model, we employ a hierarchical Bayesian approach. Hierarchical Bayes provides a principled, coherent framework for modelling multiple uncertain and random effects underlying the data described via a probabilistic generative model. It is a natural strategy for probabilistic modelling and inference of populations as well as their constituent individuals (Loredo & Hendry 2010, 2019; Gelman et al. 2013). In a hierarchical model, parameters describing an individual (e.g. the dust extinction A_V^s for a particular SN) are called *latent variables*, and are modelled as probabilistic draws from a population distributions (e.g. the distribution of A_V^s values across the SN sample), which are in turn described by *hyperparameters* (e.g. the mean parameter τ_A of the population distribution). The prior probability densities placed on the hyperparameters are called *hyperpriors*. A fully Bayesian treatment coherently infers the latent variables of all individuals in the sample along with the population hyperparameters, conditional on the observed data and the model assumptions, through the joint posterior probability density.

The first applications of hierarchical Bayes to supernova analyses were demonstrated by Mandel et al. (2009, 2011), who developed probabilistic models for SN Ia optical and NIR light curves in discrete passbands. Mandel, Foley & Kirshner (2014) constructed a hierarchical Bayesian model to disentangle dust reddening from intrinsic colours in the optical by leveraging the velocity–colour relation (VCR; Foley & Kasen 2011). Other hierarchical Bayesian models for SN Ia analysis have focused exclusively on analysing the 3-parameter output from SALT2 fits to SN Ia light curves (March et al. 2011; Rubin et al. 2015; Shariff et al. 2016b; Mandel et al. 2017; Hinton et al. 2019), rather than the observed data itself. Since they do not attempt to directly model the irregularly and asynchronously sampled multivariate, multiband light curve (time series) data, they are dependent on the internal shortcomings of SALT2 described in Section 1.1.

In contrast, our BAYESN SED model combines the hierarchical Bayesian strategy with techniques from functional data analysis (e.g. Ramsay & Silverman 2005) to deal with the full complexity of observed photometric time series, and to perform probabilistic inference on the multiple time- and wavelength-dependent latent functions underlying the observed data. In particular, we model the modes of variation of the intrinsic SED in terms of a Bayesian formulation of functional principal components. While principal components analysis (PCA) is a standard tool for dimensionality reduction of multivariate data, in its conventional use, however, it lacks a probabilistic framework. Probabilistic and Bayesian formulations of PCA for multivariate vectorial data were described by Roweis (1998), Tipping & Bishop (1999), Bishop (1999, 2006). In particular, Tipping & Bishop (1999) constructed a probabilistic PCA as a special case of a Gaussian latent variable model for factor analysis, with an associated likelihood function mapping between a low-dimensional latent space and the high-dimensional data space, and a prior distribution over the latent variables. Bishop (1999) further developed Bayesian PCA by introducing priors on the principal components and residual variance. These useful probabilistic formulations enable us to embed a principal components model within our hierarchical Bayesian framework while simultaneously accounting for multiple random effects and sources of uncertainty, such as dust, distance, and measurement error. Thus, we can determine the intrinsic principal components while marginalizing over the other uncertainties in the global inference problem.

A primary goal of BAYESN is to model populations of latent SED functions over time and wavelength, so we extend these concepts to functional data, by incorporating continuity and smoothness constraints on the functional principal components, and by modelling the time- and wavelength-dependent covariance of the residual functions. In this paper, we deal mainly with photometric flux data, which are essentially *integral constraints* (under the passband throughput and with measurement errors) on, or *functionals* of, the latent SED component functions. Embedding the functional inference within a hierarchical Bayesian structure enables us to solve the inverse problem by finding a low-dimensional latent function space for parsimoniously modelling intrinsic variations of the SN Ia SED distribution, while simultaneously deconvolving it from the SED effects of the dust distribution, and coherently accounting for the uncertainties in both. See Appendix Section A for a further explanation of Bayesian FPCA.

A schematic depiction of the probabilistic forward model of the SED for a single supernova's light-curve data is shown in Fig. 1. We construct a log intrinsic SN SED across time and optical to NIR wavelengths by modifying a mean intrinsic SED function with functional principal components scaled by latent SED shape parameters. This is further modified by the dust extinction law as a function of wavelength, scaled by the dust extinction parameter. A random function described by a covariance matrix models the SED residuals, as a function of time and wavelength, that are not captured by the previous main modes of variation. The combination of these effects yields the latent host-reddened SED in the SN rest-frame. Finally, the effects of distance, redshifting, and time dilation, integration of the flux under the observer's filter functions, the observational cadence of the survey, and photometric measurement error yields the observed multiband optical and NIR time series (light curves) of a SN Ia.

2.1 Flux data model

Suppose supernova s with spectroscopic redshift[3] z_s has a distance modulus μ_s. The ith photometric observation of SN s is taken at observer-frame Modified Julian Date (MJD) T_s^i through a filter with an effective transmission function $\mathbb{T}_{s,i}(\lambda_o)$ as a function of observed wavelength λ_o. The calibration standard has an SED $F_{\rm std}(\lambda_o)$, which defines the reference magnitude in the passband. The calibrated flux ('FluxCal' in SNANA; Kessler et al. 2009a) is the ratio of the SN flux at the observer through the passband to the flux of the standard

[3]In Section 2.1, z_s refers to the observer-frame, heliocentric redshift.

Downloaded from https://academic.oup.com/mnras/article/510/3/3939/6448478 by University of Cambridge user on 14 December 2022

star through the same passband:

$$f_{s,i} = 10^{0.4\times Z_{s,i}} \times \frac{\int F_{\rm obs}^{s,i}(\lambda_o)\, \mathbb{T}_{s,i}(\lambda_o)\lambda_o\, \mathrm{d}\lambda_o}{\int F_{\rm std}(\lambda_o)\, \mathbb{T}_{s,i}(\lambda_o)\lambda_o\, \mathrm{d}\lambda_o} = 10^{0.4\times Z_{s,i}} \int F_{\rm obs}^{s,i}(\lambda_o)\, \mathbb{B}_{s,i}(\lambda_o)\lambda_o\, \mathrm{d}\lambda_o. \quad (2)$$

The passbands used in this analysis are described in Section 3.2. We define the *normalized* transmission function as:

$$\mathbb{B}_{s,i}(\lambda_o) \equiv \frac{\mathbb{T}_{s,i}(\lambda_o)}{\int F_{\rm std}(\lambda'_o)\, \mathbb{T}_{s,i}(\lambda'_o)\lambda'_o\, \mathrm{d}\lambda'_o} = \frac{\mathbb{T}_{s,i}(\lambda_o)}{\int \mathbb{T}_{s,i}(\lambda'_o)\lambda'_o\, \mathrm{d}\lambda'_o} \times \frac{\int \mathbb{T}_{s,i}(\lambda'_o)\lambda'_o\, \mathrm{d}\lambda'_o}{\int F_{\rm std}(\lambda'_o)\, \mathbb{T}_{s,i}(\lambda'_o)\lambda'_o\, \mathrm{d}\lambda'_o}. \quad (3)$$

$Z_{s,i}$ is the zeropoint for this observation.[4] The model flux value can be converted to an apparent magnitude, on the system of the standard, like so:

$$m_{s,i} = -2.5\log_{10}\left(f_{s,i}\right) + Z_{s,i} = -2.5\log_{10} \int F_{\rm obs}^{s,i}(\lambda_o)\, \mathbb{B}_{s,i}(\lambda_o)\lambda_o\, \mathrm{d}\lambda_o. \quad (4)$$

Now we model the observable flux density $F_{\rm obs}^{s,i}(\lambda_0)$ (per unit wavelength) for observation i of SN s. If the MJD date of B-band maximum is $T_s^{\rm max}$, then we define the rest-frame phase of this observation as $t_s^i \equiv (T_s^i - T_s^{\rm max})/(1+z_s)$. We denote the effective SED in the SN rest-frame, extinguished by host galaxy dust, as $S_s(t, \lambda_r)$. The flux density of the light from SN s at observed wavelength λ_o and at time T_s^i at the Earth is:

$$F_{\rm obs}^{s,i}(\lambda_o) = (1+z_s)^{-1}\, 10^{-0.4\,\mu_s} \times S_s\left(t_s^i, \lambda_r = \frac{\lambda_o}{1+z_s}\right) \times 10^{-0.4\, A_{\rm MW}^s\, \xi(\lambda_o;R_{\rm MW})}. \quad (5)$$

The last term is the attenuation of flux by dust along the line of sight within the Milky Way Galaxy. The V-band Milky Way extinction is obtained from the reddening map (Schlafly & Finkbeiner 2011), $A_{\rm MW}^s = E(B-V)_{\rm MW}^s \times R_{\rm MW}$, and we adopt $R_{\rm MW} = 3.1$ and the Fitzpatrick (1999) extinction law for $\xi(\lambda_o; R_{\rm MW})$.

The range in observed wavelength λ_o over which the transmission is effectively non-zero is denoted as $[\lambda_o^{\rm min}, \lambda_o^{\rm max}]$. The effective rest-wavelength range is then $[\lambda_r^{\rm min} = \lambda_o^{\rm min}/(1+z_s),\ \lambda_r^{\rm max} = \lambda_o^{\rm max}/(1+z_s)]$. Combining equation (2) with equation (5), we can rewrite the model flux for the ith observation of SN s as an integral over the rest-frame wavelength $\lambda_r = \lambda_o/(1+z_s)$:

$$f_{s,i} = (1+z_s)\, 10^{-0.4\,\mu_s} \times 10^{+0.4\times Z_{s,i}} \times \int_{\lambda_r^{\rm min}}^{\lambda_r^{\rm max}} S_s(t_s^i, \lambda_r) \times 10^{-0.4\, A_{\rm MW}^s\, \xi(\lambda_r[1+z_s];R_{\rm MW})} \times \mathbb{B}_{s,i}(\lambda_r[1+z_s])\, \lambda_r\, \mathrm{d}\lambda_r. \quad (6)$$

This calibrated flux is measured with some photometric noise with a given variance $\sigma_{s,i}^2$, and we assume a Gaussian sampling distribution for the measured flux $\hat{f}_{s,i}$:

$$P(\hat{f}_{s,i}|\, f_{s,i}) = N\left(\hat{f}_{s,i}|\, f_{s,i}, \sigma_{s,i}^2\right). \quad (7)$$

For all the observations i (across all observation times and filters) of SN s, the measurement likelihood is

$$P(\hat{\boldsymbol{f}}_s|\, \boldsymbol{f}_s) = \prod_i P\left(\hat{f}_s^i|\, f_s^i\right), \quad (8)$$

assuming independence of the flux measurement errors.

2.2 Dust and intrinsic supernova SED model

The host-dust-extinguished SED is obtained from the intrinsic SED $S_s^{\rm int}(t, \lambda_r)$ in the SN rest-frame via

$$S_s(t, \lambda_r) = S_s^{\rm int}(t, \lambda_r) \times 10^{-0.4\, A_V^s\, \xi(\lambda_r;R_V)}, \quad (9)$$

where A_V^s is the host galaxy dust extinction and $\xi(\lambda_r; R_V)$ is the extinction law with parameter R_V. We adopt the extinction law of Fitzpatrick (1999).

Our model intrinsic SN spectral energy distribution is a function of rest-frame phase t and λ_r. We decompose it into a *global* spectral template modified by *individual* effects that vary per SN s.

$$S_s^{\rm int}(t, \lambda_r) = S_0(t, \lambda_r) \times 10^{-0.4 M_0} \times 10^{-0.4\, W_0(t,\lambda_r)} \times 10^{-0.4\,\delta M_s} \times 10^{-0.4\,\delta W_s(t,\lambda_r)}, \quad (10)$$

where $M_0 \equiv -19.5$ is fixed normalization factor,[5] and the fixed function $S_0(t, \lambda_r)$ is the updated spectral template of Hsiao (2009). This template spans 0.1 to 2.5 μm from −20d to +85d past B-band maximum, and was constructed from over 1000 spectra, including NIR spectra from Marion et al. (2009), using the procedure described in Hsiao et al. (2007). It is arbitrarily normalized to have a B-band magnitude of zero at peak phase $t = 0$.

The terms on the top line altogether describe the *global* spectral template. They model the baseline mean intrinsic SED as the Hsiao (2009) spectral template, normalized and smoothly warped by $M_0 + W_0(t, \lambda_r)$ to match the inferred mean intrinsic absolute magnitudes and intrinsic colours of the training sample (c.f. Section 2.4 for details).

The terms on the bottom line describe the *individual* effects, the modifications to the global SED that are specific to each supernova s. The δM_s term corresponds to an overall shift of the log SED that is independent of phase and wavelength. The function $\delta W_s(t, \lambda_r)$ corresponds to phase- and wavelength-dependent effects. We further decompose this function as:

$$\delta W_s(t, \lambda_r) = \left[\sum_{k=1}^{K} \theta_k^s\, W_k(t, \lambda_r)\right] + \epsilon_s(t, \lambda_r). \quad (11)$$

[4] $Z_{s,i} = 27.5 + m_{s,i}^{\rm std}$, where $m_{s,i}^{\rm std}$ is the reference magnitude of the reference standard with SED $F_{\rm std}(\lambda_o)$ in the passband, and is typically adopted to be zero for both AB and Vega-based magnitude systems. Each SN survey reports magnitudes with respect to this standard magnitude, though the reference standard itself is seldom observed directly. Rather, the surveys calibrate their photometry using the reported magnitudes of a network of standard stars (e.g. CALSPEC or Landolt), or local stars, in the same frame as the SN, which have themselves been calibrated with respect to the reference standard. Formally then, the comparison of the measured magnitude $\hat{m}_{s,i}$ (in equation 15) and the model magnitude $m_{s,\,i}$ (equation 4) generated from synthetic photometry of the model SED surfaces, effectively involves two zeropoints that may be subtly different (zeropoint error). The 27.5 is a conventional scaling applied in SNANA files, i.e the flux ratios are multiplied by $10^{0.4\times 27.5} = 10^{11}$ for convenience.

[5] This value is chosen for convenience, but is somewhat arbitrary. Any additional global magnitude normalization is absorbed into $W_0(t, \lambda)$ during training. In particular, a global shift in all distance moduli (preserving all distance ratios) due to a change in the assumed H_0 would trivially result in a constant shift of $5\Delta\log_{10}H_0$ in $W_0(t, \lambda)$.

Downloaded from https://academic.oup.com/mnras/article/510/3/3939/6448478 by University of Cambridge user on 14 December 2022

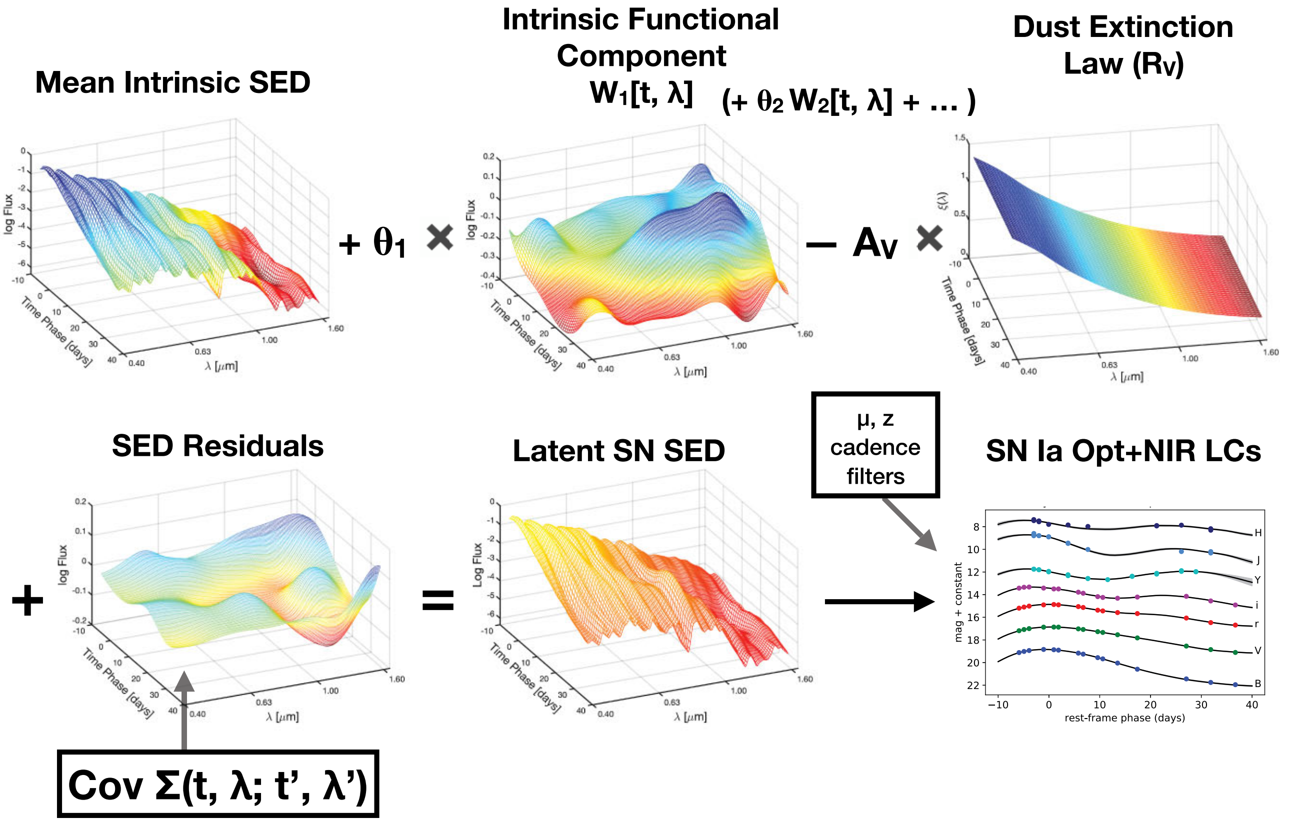

Figure 1. Schematic of the BAYESN forward generative model for the optical and NIR light-curve (time series) data of a single SN. The log SN SED across time and wavelength comprises a mean intrinsic SED function modified by intrinsic functional principal components scaled by latent SED shape parameters, and extinguished and reddened by the host galaxy dust law, parametrized by the optical slope R_V and scaled by the visual extinction A_V. Variations not captured by these major modes are modelled by residual SED functions whose statistical properties across time and wavelength are captured by a covariance function. The resulting latent host-dust-reddened rest-frame SED undergoes the effects of distance, redshifting and time dilation, integration of the flux under the observer's filter functions, the survey cadence, and measurement error to yield the observed multiband optical and NIR time series (light curves) of an SN Ia.

The $W_k(t, \lambda_r)$ functions are the *functional principal components* (FPCs) describing the major modes of (t, λ_r) variation in the log SED underlying the light curves of individual SN s. The θ_k^s coefficients are scores describing the degree of component $W_k(t, \lambda_r)$ present in SN s. The functions $\epsilon_s(t, \lambda_r)$ describe the phase- and wavelength-dependent SED *residuals* that are not captured by the other effects. They source the remaining time-dependent intrinsic colour variations. The total residual SED function of an SN Ia is $\eta_s(t, \lambda_r) = \delta M_s + \epsilon_s(t, \lambda_r)$.

In this work, we mainly focus on training a model with $K = 1$ intrinsic functional principal component. In Appendix B, we describe the $W_2(t, \lambda_r)$ inferred for the $K = 2$ model. For $K > 1$, under training, the $W_k(t, \lambda_r)$ are learned such that the coefficients θ_k^s are uncorrelated in their population distribution (Section 2.5), i.e. $\mathrm{Cov}[\theta_k^s, \theta_{k'}^s] = \delta_{kk'}$. When we train the $K = 2$ model to learn the second FPC, we are effectively extracting it from the covariance of the residual functions $\epsilon_s(t, \lambda)$ under the $K = 1$ model.

The above equations express a linear model for the logarithm of the host-dust-extinguished SN SED:

$$-2.5\log_{10}[S_s(t, \lambda_r)/S_0(t, \lambda_r)] = M_0 + W_0(t, \lambda_r) + \delta M_s + \left[\sum_{k=1}^{K} \theta_k^s\, W_k(t, \lambda_r)\right] + \epsilon_s(t, \lambda_r) + A_V^s\, \xi(\lambda_r; R_V). \quad (12)$$

Note that M_0, $W_0(t, \lambda_r)$, δM_s, and $\delta W_s(t, \lambda_r)$ are in units of magnitude, like μ_s and A_V^s. The advantage of modelling the logarithm of the SED is that we can easily preserve positive flux at all phases and wavelengths while specifying priors on the functional components $W_k(t, \lambda_r)$ and the latent principal component scores that span positive and negative reals.

2.3 Magnitude approximation

For the vast majority of the nearby training set used in this work, the flux data have high signal to noise. Therefore, it is a good approximation to convert these data to magnitudes. The magnitude measurement and the variance of its measurement error are $\hat{m}_{s,i} = -2.5\log_{10}(\hat{f}_{s,i}) + Z_{s,i}$ and

$$\sigma^2_{m,s,i} = \left(\frac{2.5}{\ln 10}\frac{\sigma_{s,i}}{\hat{f}_{s,i}}\right)^2. \quad (13)$$

Transformation of the model flux (equation 6) to the model magnitude $m_{s,i}$ yields

$$m_{s,i} = \mu_s + M_0 + \delta M_s - 2.5\log_{10}\left[(1+z_s)\int_{\lambda_r^{\min}}^{\lambda_r^{\max}} S_0\left(t_s^i, \lambda_r\right) \times 10^{-0.4\,[W_0(t_s^i,\lambda_r)+\delta W_s(t_s^i,\lambda_r)+A_V^s\xi(\lambda_r;R_V)]} \times 10^{-0.4\,A^s_{\mathrm{MW}}\,\xi(\lambda_o;R_{\mathrm{MW}})} \times \mathbb{B}_{s,i}(\lambda_o)\,\lambda_r\,\mathrm{d}\lambda_r\right]. \quad (14)$$

Using this, we can change the measurement-likelihood function,

Downloaded from https://academic.oup.com/mnras/article/510/3/3939/6448478 by University of Cambridge user on 14 December 2022

equation (7), to:

$$P(\hat{m}_{s,i}\,|\,m_{s,i}) = N\left(\hat{m}_{s,i}\,|\,m_{s,i}, \sigma^2_{m,s,i}\right). \tag{15}$$

This form is useful since the model magnitude inside the likelihood is linear in some of the parameters. However, the full flux model (equation 7) allows us to use low signal to noise, or even negative, flux measurements which cannot be reliably converted into magnitudes with Gaussian errors. Hence, we can use the flux model to fit the flux data of high-redshift SNe Ia with typically lower signal to noise.

2.4 2D SED surface models

We model the unknown functions $\{W_k(t, \lambda_r)$: $k = 0, \ldots, K\}$, and $\{\epsilon_s(t, \lambda_r)$: $s = 1, \ldots, N_{\rm SN}\}$ in a flexible, data-driven manner. Each function is represented as a surface defined by a 2D grid of knots. We specify a 2D grid as the Cartesian product of a 1D grid in rest-frame phase, $\boldsymbol{\tau}$, and a 1D grid in rest-frame wavelength $\boldsymbol{l}$. Each 1D grid can be irregularly spaced. The essential idea is that a generic, smooth surface $g(t, \lambda_r)$ at any point (t, λ_r) in the 2D domain of the SED can be modelled as $g(t, \lambda_r) = \boldsymbol{s}(\lambda_r; \boldsymbol{l})^T\, \boldsymbol{G}\, \boldsymbol{s}(t; \boldsymbol{\tau})$, where $\boldsymbol{s}(x; \boldsymbol{\xi})$ denotes the 1D natural cubic spline smoother (column) vector for knots $\boldsymbol{\xi}$ at evaluated at point x. The knots matrix $\boldsymbol{G}$ has elements $G_{ij} = g(t = \tau_j, \lambda_r = l_i)$, which define the values the surface must pass through at the knot locations, and are parameters for inference. The surface $g(t, \lambda_r)$ is linear in the knots matrix $\boldsymbol{G}$.

Using this, we model the functions of phase and wavelength in terms of knot matrices $\{\boldsymbol{W}_k : k = 0, \ldots, K\}$ and $\{\boldsymbol{E}_s : s = 1, \ldots, N_{\rm SN}\}$, like so. For the global correction to the mean template:

$$W_0(t, \lambda_r) = \boldsymbol{s}(\lambda_r; \boldsymbol{l})^T\, \boldsymbol{W}_0\, \boldsymbol{s}(t; \boldsymbol{\tau}). \tag{16}$$

For the functional components ($k = 1, \ldots, K$),

$$W_k(t, \lambda_r) = \boldsymbol{s}(\lambda_r; \boldsymbol{l})^T\, \boldsymbol{W}_k\, \boldsymbol{s}(t; \boldsymbol{\tau}). \tag{17}$$

For the residual SED functions of each SN s,

$$\epsilon_s(t, \lambda_r) = \boldsymbol{s}(\lambda_r; \boldsymbol{l})^T\, \boldsymbol{E}_s\, \boldsymbol{s}(t; \boldsymbol{\tau}). \tag{18}$$

These latent functions are determined by the unknown matrices $\{\boldsymbol{W}_k : k = 0, \ldots, K\}$, and $\{\boldsymbol{E}_s : s = 1, \ldots, N_{\rm SN}\}$, which are inferred as hyperparameters and latent variables.

We specify a set of knots on a grid in rest-frame phase and wavelength. The phase coordinates are $\boldsymbol{\tau} = (-10, 0, 10, 20, 30, 40)$ d. The phase spacing is chosen as inspection of the light-curve data indicates they vary smoothly on ∼10 d time-scales. We found empirically that this spacing works well in practice for our light-curve fits, and strikes a balance between temporal resolution/regularization and statistical/computational feasibility. The wavelength coordinates are $\boldsymbol{l}$. We place a knot at the central wavelengths of the filters *BVriYJH* plus two outer knots bracketting these: $\boldsymbol{l} = (0.3, 0.43, 0.54, 0.62, 0.77, 1.04, 1.24, 1.65, 1.85)$ μm. The purpose of the first and last knots in wavelength is to ensure that our spline surfaces are defined throughout the entire first (*B*) and last (*H*) broad-band filters. To avoid degeneracies, we 'tie down' the residual knot matrices at the first and last wavelength knots for every phase knot: $E_{s,ij} = 0$ if $i = 1$ or $i = \dim(\boldsymbol{l})$, $\forall\, j$.

2.5 Population distributions and hyperpriors

We specify the population distributions on the latent parameters of individual supernovae.

For the latent functional SED effects, following the probabilistic PCA formulation (Tipping & Bishop 1999), we adopt an independent standard Gaussian prior $\theta^s_k \sim N(0, 1)$ for the individual score of each SN s in each component $k = 1, \ldots, K$, i.e.

$$\boldsymbol{\theta}^s \sim N(0, \boldsymbol{I}_{K\times K}). \tag{19}$$

Thus, the resulting functions $W_k(t, \lambda_r)$ ($k \geq 1$) are not scaled to have unit norm, as they would be in standard PCA. Rather, because the latent scores θ^s_k are normalized to have a population variance of one, $W_k(t, \lambda_r)$ absorbs a factor of the population standard deviation in that component. A '1σ' effect of the k-th component on the SED is thus computed from $\theta_k\, W_k(t, \lambda_r)$ by varying $\Delta\theta_k \pm 1$ around the mean.

For the elements of the $\boldsymbol{W}_0$ matrix, we adopt improper flat priors $W_{0,ij} \sim U(-\infty, \infty)$, as this ensures the global joint probability density is mathematically invariant under a global shift in all distance moduli (e.g. a change in $5\log_{10}H_0$), that preserves all relative SN Ia distance ratios. With this prior, $W_0(t, \lambda)$ would simply absorb the constant.

For the $\boldsymbol{W}_k$ matrices that parametrize our functional components ($k \geq 1$), we use an independent standard normal hyperprior on the value of each knot: $W_{k,ij} \sim N(0, 1)$. This is a weak constraint, since we have scaled the problem to expect these variations to be of the order of a few tenths of a magnitude.

For the residual SED perturbations, we assume a multivariate Gaussian distribution on the column-wise vectorization of each residual matrix $\boldsymbol{E}_s$:

$$\boldsymbol{e}_s = \mathrm{vec}[\boldsymbol{E}_s] \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_\epsilon). \tag{20}$$

A matrix $\boldsymbol{\Gamma}(t, \lambda_r; \boldsymbol{\tau}, \boldsymbol{l})$ can be constructed so that equation (18) can be written equivalently as,

$$\epsilon_s(t, \lambda_r) = \boldsymbol{\Gamma}(t, \lambda_r; \boldsymbol{\tau}, \boldsymbol{l})\, \boldsymbol{e}_s. \tag{21}$$

While equation (18) and equation (21) are equivalent, equation (18) is the more compact representation, since $\boldsymbol{\Gamma}(t, \lambda_r; \boldsymbol{\tau}, \boldsymbol{l})$ tends to a very large (but sparse) matrix. However, equation (21) is useful, because, together with the residual distribution equation (20), it implies that the residual functions $\epsilon_s(t, \lambda_r)$ are realizations of a Gaussian process (GP; Rasmussen & Williams 2005):

$$\epsilon_s(t, \lambda_r) \sim \mathcal{GP}[\boldsymbol{0}, k_\epsilon(t, \lambda_r; t', \lambda'_r)] \tag{22}$$

with a zero prior mean and a non-stationary kernel for the covariance of the residuals at any two coordinates:

$$\begin{aligned} k_\epsilon(t, \lambda_r; t', \lambda'_r) &\equiv \mathrm{Cov}[\epsilon_s(t, \lambda_r), \epsilon_s(t', \lambda'_r)] \\ &= \boldsymbol{\Gamma}(t, \lambda_r; \boldsymbol{\tau}, \boldsymbol{l})\, \boldsymbol{\Sigma}_\epsilon\, \boldsymbol{\Gamma}(t', \lambda'_r; \boldsymbol{\tau}, \boldsymbol{l})^T. \end{aligned} \tag{23}$$

We adopt this non-stationary covariance structure rather than the more popular stationary kernels, such as squared exponential, since we do not expect the complex physical mechanisms of SN Ia explosions to generate statistical properties that are invariant to phase or wavelength shifts.

The covariance matrix $\boldsymbol{\Sigma}_\epsilon$ encodes the variances and correlation structures of the residual functions: $\boldsymbol{\Sigma}_\epsilon = \mathrm{diag}(\boldsymbol{\sigma}_\epsilon)\, \boldsymbol{R}_\epsilon\, \mathrm{diag}(\boldsymbol{\sigma}_\epsilon)$. Following the separation strategy proposed by Barnard, McCulloch & Meng (2000), we specify separate priors on the standard deviation parameters $\boldsymbol{\sigma}_\epsilon$ and the correlation matrix $\boldsymbol{R}_\epsilon$. For each q-th element $\sigma_{\epsilon,q} \geq 0$, we adopt a weakly informative half-Cauchy hyperprior with unit scale (Gelman 2006; Polson & Scott 2012), i.e. $P(\sigma_{\epsilon,q}) = HC(\sigma_{\epsilon,q}\,|\,a = 1)$, with probability density

$$HC(x\,|\,a) \propto (a^2 + x^2)^{-1} \tag{24}$$

for $x \geq 0$, and zero otherwise. This hyperprior is proper, and relatively flat for small x. It is sensible because we have scaled the problem to expect $\sigma_{\epsilon,q}$ to be less than a magnitude. For the correlation matrix, we adopt the LKJ hyperprior as implemented in STAN and derived

Downloaded from https://academic.oup.com/mnras/article/510/3/3939/6448478 by University of Cambridge user on 14 December 2022

from Lewandowski, Kurowicka & Joe (2009),

$$P(\boldsymbol{R}_\epsilon) \propto |\boldsymbol{R}_\epsilon|^{\eta-1} \quad (25)$$

with $\eta = 1$. This places a uniform prior on positive semidefinite correlation matrices.

The δM_s terms model a phase- and wavelength-independent shift of the SED in overall log luminosity. Since these shifts are indistinguishable from the effect of distance on the apparent light curves, this propagates into an uncertainty floor on photometric distance estimates. We model the population of these independent shifts as $\delta M_s \sim N(0, \sigma_0^2)$ and estimate their variance σ_0^2 as a hyperparameter. We use a weak half-Cauchy prior (equation 24) on σ_0 with scale $a = 0.1$, since we expect this to be of the order of a tenth of magnitude. However, we have checked that our posterior estimate of σ_0 is insensitive to the hyperprior scale over the range $a = [0.1, 0.5]$.

We assume that host galaxy extinction A_V^s is drawn from an independent exponential distribution with mean extinction hyperparameter τ_A:

$$P\left(A_V^s|\tau_A\right) = \tau_A^{-1} \exp\left(-A_V^s/\tau_A\right), \quad (26)$$

for $A_V \geq 0$ and zero otherwise. This is a sensible choice, since the true A_V^s must be non-negative, and we expect the most lines of sight through the host galaxies to pass through little dust, with the probability density decreasing with increasing column density. This model distribution has been used before by, e.g. Jha, Riess & Kirshner (2007) and Mandel et al. (2009). The hyperprior we adopt for τ_A is also a unit half-Cauchy, $P(\tau_A) = HC(\tau_A, 1)$, reflecting our expectations that the typical τ_A is on the order of tenths of a magnitude. For the unknown R_V, we assume a single global value with a uniform hyperprior $R_V \sim U(1, 5)$ reflecting a wide range of possible values. Thorp et al. (2021) expands our framework to allow per-SN variation in R_V^s by modelling and inferring their population distribution, as was done previously by Mandel et al. (2011).

2.6 External distance constraints

In the training phase, we use estimates $\hat{\mu}_{\mathrm{ext},s}$ of the SN distance moduli that are external to the photometric SN data, as described in Avelino et al. (2019). We assume they have Gaussian errors around the true distance modulus.

For the vast majority of the training set, we utilize the redshift as an indicator of distance conditional on the fiducial cosmological model $\hat{\mu}_{\mathrm{ext},s} = \mu_{\Lambda CDM}(z_s)$ with $\Omega_M = 0.28$ and $\Omega_\Lambda = 0.72$. However, at these redshifts $z < 0.04$, these distances are relatively insensitive to the cosmological parameters, other than H_0 which only sets an overall scale for all absolute magnitudes, and for which we adopt 73.24 kms^{-1} Mpc^{-1} (Riess et al. 2016). These redshifts are corrected to the CMB frame and corrected for bulk flows. The distance modulus uncertainty, due to errors in observed redshift z_s as estimates for the cosmological redshift z_s^c, from redshift and peculiar velocity uncertainties is

$$\hat{\sigma}_{\mathrm{ext},s}^2 \approx \left(\frac{5}{z_s \ln 10}\right)^2 \left[\sigma_{\mathrm{pec}}^2/c^2 + \sigma_{z,s}^2\right], \quad (27)$$

where we have adopted $\sigma_{\mathrm{pec}} = 150$ km s $^{-1}$ (Carrick et al. 2015). The external distance constraint can be expressed as $P(\mu_s|\, z_s) \propto N(\hat{\mu}_{\mathrm{ext},s}|\, \mu_s, \hat{\sigma}_{\mathrm{ext},s}^2)$ after marginalizing out the unknown z_s^c.

For eight SNe Ia in our training set at $z < 0.01$, we use external distance estimates $\hat{\mu}_{\mathrm{ext},s}$ from available redshift-independent measures (e.g. Cepheids), and their uncertainties $\hat{\sigma}_{\mathrm{ext},s}^2$, as listed in table 4 of Avelino et al. (2019). These external distance constraints can be expressed as $P(\mu_s|\, \hat{\mu}_{\mathrm{ext},s}) \propto N(\hat{\mu}_{\mathrm{ext},s}|\mu_s, \hat{\sigma}_{\mathrm{ext},s}^2)$.

2.7 The global joint posterior distribution

For an individual SN s, the joint probability density of its flux light curve data $\hat{\boldsymbol{f}}_s$ and its latent parameters $\boldsymbol{\phi}_s \equiv (\boldsymbol{\theta}_s, \boldsymbol{e}_s, \delta M_s, A_V^s)$ and distance modulus μ_s conditional on the population hyperparameters $\boldsymbol{H} \equiv (\boldsymbol{W}_{0:K}, \boldsymbol{\Sigma}_\epsilon, \sigma_0, \tau_A, R_V)$ and redshift is

$$P(\hat{\boldsymbol{f}}_s, \boldsymbol{\phi}_s, \mu_s|\, \boldsymbol{H}; z_s) = P(\hat{\boldsymbol{f}}_s|\, \boldsymbol{\phi}_s, \mu_s; \boldsymbol{W}_{0:K}, R_V) \times P(\boldsymbol{\theta}_s) P(\boldsymbol{e}_s|\, \boldsymbol{\Sigma}_\epsilon) P(\delta M_s|\, \sigma_0) P(A_V^s|\, \tau_A) P(\mu_s|\, z_s), \quad (28)$$

where $\boldsymbol{W}_{0:K} \equiv \{\boldsymbol{W}_0, \boldsymbol{W}_1, \ldots, \boldsymbol{W}_K\}$ is the collection of matrices describing the intrinsic mean and K functional components of the SED, and $\boldsymbol{\theta}_s \equiv (\theta_1^s, \ldots, \theta_K^s)^T$ are the intrinsic coefficients of SN s. The first factor on the right-hand side is the data likelihood defined by equations (6), (8), and (12). For the eight SN with redshift-independent distance measurements, we replace $P(\mu_s|\, z_s)$ with $P(\mu_s|\, \hat{\mu}_{\mathrm{ext},s})$. During training, the dates of optical maxima $T_s^{\max}$ are fixed to their pre-fitted values, which are very accurate for this training set of well-sampled light curves.

The global posterior distribution of all the latent variables of individual supernovae and the population hyperparameters given the data, external distance constraints, and redshifts is

$$P(\{\boldsymbol{\phi}_s, \mu_s\}; \boldsymbol{H}\, |\, \{\hat{\boldsymbol{f}}_s; z_s\}) \propto \left[\prod_{s=1}^{N_{\mathrm{SN}}} P(\hat{\boldsymbol{f}}_s, \boldsymbol{\phi}_s, \mu_s, |\, \boldsymbol{H}; z_s)\right] \times P(\boldsymbol{W}_{0:K})\, P(\boldsymbol{\sigma}_\epsilon)\, P(\boldsymbol{R}_\epsilon)\, P(\sigma_0)\, P(\tau_A)\, P(R_V). \quad (29)$$

This global posterior distribution is the objective function for training our model to learn the population hyperparameters, covariance structure, and SED components while marginalizing over the latent variables of individual SNe Ia. It provides a coherent, probabilistic quantification of uncertainty of over all parameters and hyperparameters.

2.8 Photometric distance estimation

The training process gives us posterior estimates of the hyperparameters $\hat{\boldsymbol{H}} \equiv (\hat{\boldsymbol{W}}_{0:K}, \hat{\boldsymbol{\Sigma}}_\epsilon, \hat{\sigma}_0, \hat{\tau}_A, \hat{R}_V)$ marginalized over all latent variables in the sample. For simplicity, we take the posterior means of these hyperparameters as point estimates. Under distance-fitting mode, we condition on the hyperparameters, and the posterior density of the latent parameters $\boldsymbol{\phi}_s$ and distance modulus μ_s of SN s is

$$P(\boldsymbol{\phi}_s, \mu_s|\, \hat{\boldsymbol{f}}_s; \hat{\boldsymbol{H}}) \propto P(\hat{\boldsymbol{f}}_s|\, \boldsymbol{\phi}_s, \mu_s; \hat{\boldsymbol{W}}_{0:K}, \hat{R}_V) \times P(\boldsymbol{\theta}_s) \times P(\boldsymbol{e}_s|\, \hat{\boldsymbol{\Sigma}}_\epsilon) \times P(\delta M_s|\, \hat{\sigma}_0) \times P(A_V^s|\, \hat{\tau}_A), \quad (30)$$

where we omit any external distance constraint. When fitting individual SN with the trained model, the dates of optical maxima $T_s^{\max}$ are included in $\boldsymbol{\phi}_s$ and fit. By sampling this joint posterior, we can approximate the marginal posterior density of the photometric distance modulus,

$$P(\mu_s|\, \hat{\boldsymbol{f}}_s; \hat{\boldsymbol{H}}) = \int P(\boldsymbol{\phi}_s, \mu_s|\, \hat{\boldsymbol{f}}_s; \hat{\boldsymbol{H}})\, \mathrm{d}\boldsymbol{\phi}_s, \quad (31)$$

as well as its posterior summaries such as the mean and variance via marginalization. However, this distribution is not necessarily Gaussian, nor is it required to be, since dust effects introduce some asymmetry.

In principle, instead of fixing the hyperparameters to their posterior means from training, we could use the samples of the joint posterior over the hyperparameters (equation 29) to incorporate their uncertainties into the photometric distance estimates. However, this is a more computationally burdensome process, and we have found the posterior means to be sufficient for our current goal of evaluating the

Downloaded from https://academic.oup.com/mnras/article/510/3/3939/6448478 by University of Cambridge user on 14 December 2022

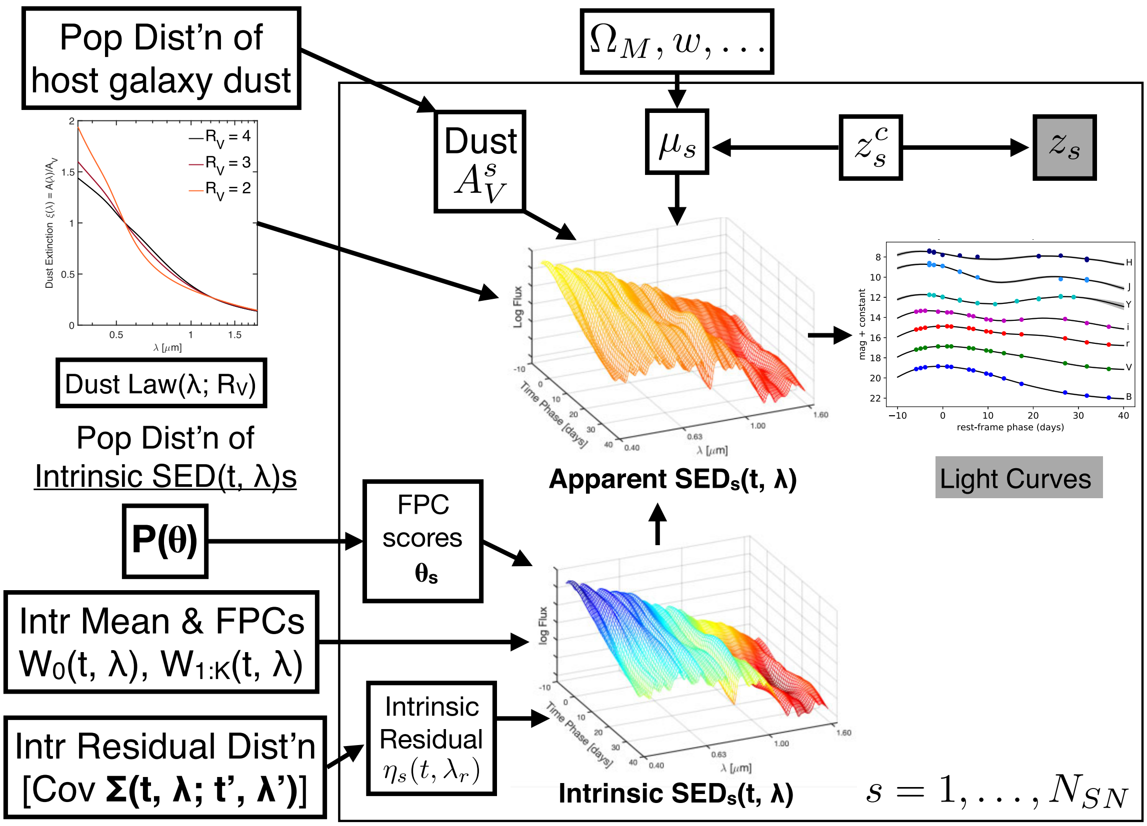

Figure 2. Probabilistic graphical model depicting the hierarchical BAYESN model for optical-NIR SN Ia light-curve data. Each open box presents a set of unknown parameters or hyperparameters, each grey-shaded box represents observed data, and the arrows indicate relations of conditional probability. Parameters within the plate, labelled $s = 1, \ldots, N_{\rm SN}$, are latent variables or functions sampled for every SN s, whereas parameters outside the plate represent global or population hyperparameters. This graph is further discussed as a probabilistic generative model in Section 2.9. The hierarchical global posterior density (equation 29) estimates the unknown latent variables and hyperparameters conditional on the observed data of the entire SN Ia sample.

precision of photometric distances on the Hubble diagram. Nevertheless, propagating the joint hyperparameter uncertainties will be relevant for assessing systematic errors in a full cosmological analysis.

2.9 Probabilistic graphical model

Our hierarchical Bayesian model can be depicted with a type of probabilistic graphical model called a directed acyclic graph, shown in Fig. 2. The graphical model depicts a probabilistic process for generating the SN Ia data, via links between the priors, global or population hyperparameters, latent variables and functions of individual SNe Ia, and their observed light-curve data. The intrinsic SED of a single SN Ia s is constructed from the mean SED and functional principal components $W_{0:K}(t, \lambda_r)$, a draw of the FPC scores $\boldsymbol{\theta}_s$ from its population distribution, and a draw of an intrinsic residual SED function from its population distribution described by a covariance function over time and wavelength. The host galaxy dust extinction A_V^s of a SN s is drawn from a population distribution of extinction values, whereas the unknown R_V, parametrizing the dust law, is given a wide prior. The effects of dust and distance modulus μ_s on the intrinsic SED combine (with appropriate redshifting and time dilation) to yield the apparent SN SED. This is observed with some cadence and noise through the observer's filter functions to yield the optical and NIR light-curve data. During training, the distance is constrained externally to the light curve by the cosmological redshift and the fiducial cosmological parameters (fixed in this low-z analysis). The redshift is observed with some uncertainty due to local peculiar velocities. Bayesian inference with the hierarchical model solves the inverse problem through the computation of the posterior probability density (equation 29) of the unknown latent variables and hyperparameters conditional on the observed data of the entire SN Ia sample.

3 DATA

3.1 Optical and NIR light curve data

We use the compilation of low-z SNe Ia with joint optical and NIR light curves described in Avelino et al. (2019). For our purposes, we define the optical as the *BVRI* filters and the NIR as the *YJH*. In our various analyses below, we fit either the available optical (*BVRI*) or optical + NIR (*BVRIYJH*) for a given SNe Ia. The selection criteria and cuts were described in Avelino et al. (2019) and detailed information on the specific SNe is listed in their tables 2 and 3. In particular, a colour excess cut $E(B-V)_{\rm host} \leq 0.4$ was applied for consistency with the cosmological sample. For each SN Ia, we use published optical and NIR data only from the same survey; we do not mix data sources within a single SN. Consequently, SN 2005bo, SNF20080514-002, SN 2010iw, SN 2010kg, SN 2011ao, SN 2011B, SN 2011by, SN 2011df were removed because they had NIR data, but no published optical data, from the CfA. SN 2006bt was removed because it is a known peculiar supernova (Foley et al. 2010). We failed to fit the light curves of SN 2000E (Valentini et al. 2003)

Downloaded from https://academic.oup.com/mnras/article/510/3/3939/6448478 by University of Cambridge user on 14 December 2022

with our current model (although it was not an outlier in the Hubble Diagram), so we have omitted it to avoid biasing the training.

The resulting sample comprises 79 SNe Ia with joint optical and NIR light curves. Avelino et al. (2019) further defined a subset with NIR data near maximum light (see their table 13). To enter into this subset, an SN Ia was required to have at least one NIR observation at least 2.5 d before maximum light. We have a total of 48 SNe Ia in this cut, which we refer to as 'NIR@max'. All SNe Ia in the full sample have some NIR data available regardless of the phase of the first NIR observation. These SNe Ia and cuts are listed in Table 1. Additional information can be found in Avelino et al. (2019). The full data set consists of 22 SN from the CfA Supernova Program (CfA; Jha et al. 1999; Wood-Vasey et al. 2008; Hicken et al. 2009, 2012; Friedman et al. 2015), 44 from the Carnegie Supernova Project (CSP; Krisciunas et al. 2017), eight from the Las Campanas Observatory (K04a,b: Krisciunas et al. 2004a,b), as well as five others from individual papers in the literature (K03: Krisciunas et al. 2003; P08: Pignata et al. 2008; St07: Stanishev et al. 2007; L09: Leloudas et al. 2009; K07: Krisciunas et al. 2007).

The statistical characteristics of observations in this sample are as follows. For each observed (B, Y, J, H) band light curve, there are on average (3.8, 3.6, 3.2, 2.7) data points at pre-maximum phases ($t < 0$ d), (8.2, 8.2, 7.1, 6.3) data points during the main post-maximum decline ($0 < t < 20$ d), and (5.8, 4.9, 4.7, 4.4) data points at later phases $t > 20$ d, respectively. The filter with the least data is the NIR Y-band, which is covered by only 46 SNe Ia in our training set (mainly from CSP). The median photometric errors ($\sigma_{s,i}$) across all observations in (B, Y, J, H) are (0.016, 0.020, 0.038, 0.053) mag, respectively. The median apparent $B - V$ colour error among observations near peak is 0.016 mag.

The size of our training set reflects the recent progress of ground-based surveys in accumulating quality joint optical and NIR SN Ia light-curve data (Friedman et al. 2015; Krisciunas et al. 2017). The number of SNe Ia in our current compilation more than doubles those used to train previous NIR-capable light-curve models. The training set for the first BAYESN models included 37 SNe Ia with both optical and NIR coverage (Mandel et al. 2009, 2011), and the training set for SNooPy comprised $\lesssim$30 SNe Ia (Burns et al. 2011). Further increases in the training set will soon be possible with forthcoming data from CSP-II (Phillips et al. 2019) and the *Supernovae in the Infrared avec Hubble* (SIRAH) program (*HST* GO-15889, P.I. S. Jha).

3.2 Passband throughput

For each observation in the data compilation, we specify a model for the effective passband throughput. A model passband throughput is needed to forward model the observed flux, regardless of whether that flux is reported in the natural system of the telescope or transformed on to a 'standard' system such as SDSS *ugriz* (Fukugita et al. 1996). For a measurement reported on the natural system of a telescope, the total passband throughput must include all terrestrial elements of the measurement chain – site atmospheric transmission, mirror reflectivity, filter transmission, transmission of camera optics, and detector quantum efficiency. For measurements reported on a standard system, the passband throughput must reflect the original measurement chain used to observe the standard stars that were employed in calibrating the SN flux, in addition to the measurement chain of the facility used to observe the SN itself.

For the Carnegie Supernova Project and related objects observed at Las Campanas Observatory (Krisciunas et al. 2004a,b, 2017), we use the total natural system passband throughputs[6] as defined in the implementation of the SNooPy (Burns et al. 2011). We take care to include any changes in the CSP passband throughputs when a filter was replaced. For the NIR observations by the CfA using the 1.3m PAIRITEL telescope at Mt. Hopkins (Wood-Vasey et al. 2008; Friedman et al. 2015), we use the natural system passband throughputs measured by the 2MASS project,[7] which used the same facility. Finally, for objects observed by the CfA Supernova Program (Jha et al. 1999; Hicken et al. 2009, 2012) and remaining literature objects (K03, K07, St07, P08, and L09) we use the published *standard* system photometry and model the passband throughput using the shifted Bessell filters described in Stritzinger et al. (2005). While the CfA SN program published both natural and standard system photometry, and the former is generally preferred as it avoids some potential systematic errors in transforming the flux, using the natural system photometry relies on having a good description of the passband throughput of the natural system. Unfortunately, there are no determinations of all the elements in the measurement chain for objects observed by the CfA SN survey, and the current model for passband throughput included in the SNDATA repository[8] does not include any model for the site atmosphere at all. The CfA Supernova measurements were the result of an extensive effort over almost two decades with four separate cameras, through a variety of filters, using a telescope that underwent numerous mirror coatings, and the provenance of each measurement cannot easily be determined retrospectively. By contrast, the standard system photometry for CfA objects is known to be consistent with standard system photometry measured by the CSP and LOSS (Ganeshalingam et al. 2010). Thus, we prefer to use the standard system photometry over the natural system photometry in this work.

Ultimately, we plan on training a version of BAYESN (Thorp et al. 2021) exclusively on SNe Ia observed by the Foundation Survey and the Young Supernova Experiment, which have well-determined measurements of the PS1 natural system passband throughput.

4 IMPLEMENTATION

4.1 BayeSN

We have implemented our Bayesian model in the STAN probabilistic programming language (Carpenter et al. 2017; Stan Development Team 2021) to specify and sample the global posterior density over all latent variables and hyperparameters conditional on the training set data. STAN implements a variant of dynamic Hamiltonian Monte Carlo (HMC; Neal 2011; Betancourt 2017), originally based on the No-U-Turn Sampler (NUTS) (Hoffman & Gelman 2014). STAN utilizes automatic differentiation to compute gradients of the log posterior (equations 29, 30) and guide efficient exploration and convergence to the target density in high-dimensional parameter spaces. We typically run four chains in parallel, each initialized with random jitter to start at a different point in parameter space. We follow standard procedures to assess convergence and mixing of the chains (Gelman & Rubin 1992; Gelman et al. 2013). The first half of the iterations, which are used for adaptation of the HMC algorithm and burn-in, are discarded. The algorithm adapts the integration time to yield samples that are nearly serially uncorrelated, and we run it long enough so that the effective sample size is approximately 1000.

[6] https://csp.obs.carnegiescience.edu/data/filters

[7] https://old.ipac.caltech.edu/2mass/releases/second/doc/sec3_1b1.html#s18

[8] http://snana.uchicago.edu/downloads/SNDATA_ROOT.tar.gz

Downloaded from https://academic.oup.com/mnras/article/510/3/3939/6448478 by University of Cambridge user on 14 December 2022

Table 1. Table of supernovae.

SN	Source	Cut	Filters	z^a_{CMB}	$\hat{\mu}^b_{\mathrm{ext}}$	$\hat{\mu}_{\mathrm{phot}}$ (resub)^c	$\hat{\mu}_{\mathrm{phot}}$ (CV)^d
SN1998bu	CfA	NIR@max	*BVRIJH*	0.003	30.07 ± 0.20	29.99 ± 0.10	29.96 ± 0.09
SN1999ee	K04a	NIR@max	*BVRIJH*	0.011	33.33 ± 0.10	33.25 ± 0.10	33.21 ± 0.09
SN1999ek	K04b	NIR@max	*BVRIJH*	0.018	34.34 ± 0.06	34.18 ± 0.10	34.18 ± 0.10
SN2000bh	K04a	–	*BVRIYJH*	0.024	35.00 ± 0.05	34.94 ± 0.10	34.93 ± 0.08
SN2000ca	K04a	NIR@max	*BVRIJH*	0.024	34.99 ± 0.05	35.00 ± 0.10	34.99 ± 0.10
SN2001ba	K04a	NIR@max	*BVIJH*	0.030	35.51 ± 0.04	35.66 ± 0.09	35.66 ± 0.09
SN2001bt	K04b	NIR@max	*BVRIJH*	0.014	33.85 ± 0.08	33.79 ± 0.10	33.80 ± 0.09
SN2001cn	K04b	–	*BVRIJH*	0.015	34.03 ± 0.07	33.94 ± 0.10	33.96 ± 0.10
SN2001cz	K04b	NIR@max	*BVRIJH*	0.017	34.25 ± 0.06	33.96 ± 0.10	33.95 ± 0.10
SN2001el	K03	NIR@max	*BVRIJH*	0.004	31.31 ± 0.05	31.28 ± 0.10	31.17 ± 0.09
SN2002dj	P08	NIR@max	*BVRIJH*	0.008	32.65 ± 0.40	32.95 ± 0.10	32.95 ± 0.10
SN2003du	St07	–	*BVRIJH*	0.009	32.92 ± 0.06	32.87 ± 0.10	32.86 ± 0.09
SN2003hv	L09	–	*BVRIYJH*	0.005	31.15 ± 0.25	31.30 ± 0.09	31.34 ± 0.10
SN2004S	K07	–	*BVRIJH*	0.011	33.23 ± 0.10	33.27 ± 0.10	33.24 ± 0.10
SN2004ef	CSP	–	*BVriYJH*	0.030	35.50 ± 0.04	35.52 ± 0.09	35.50 ± 0.08
SN2004eo	CSP	NIR@max	*BVriYJH*	0.015	34.00 ± 0.07	33.82 ± 0.10	33.88 ± 0.09
SN2004ey	CSP	NIR@max	*BVriYJH*	0.015	34.02 ± 0.07	34.12 ± 0.10	34.11 ± 0.08
SN2004gs	CSP	–	*BVriYJH*	0.029	35.39 ± 0.04	35.38 ± 0.10	35.39 ± 0.09
SN2005cf	CfA	NIR@max	$BVr'i'JH$	0.007	32.26 ± 0.10	32.30 ± 0.09	32.31 ± 0.10
SN2005el	CSP	NIR@max	*BVriYJH*	0.015	34.00 ± 0.07	33.98 ± 0.09	34.02 ± 0.09
SN2005iq	CSP	NIR@max	*BVriYJH*	0.034	35.74 ± 0.03	35.88 ± 0.09	35.90 ± 0.09
SN2005kc	CSP	NIR@max	*BVriYJH*	0.015	33.89 ± 0.07	33.75 ± 0.10	33.74 ± 0.09
SN2005ki	CSP	NIR@max	*BVriYJH*	0.020	34.63 ± 0.05	34.62 ± 0.10	34.62 ± 0.09
SN2005lu	CSP	–	*BVriY*	0.032	35.62 ± 0.03	35.71 ± 0.12	35.72 ± 0.11
SN2005na	CfA	–	$BVr'i'JH$	0.027	35.28 ± 0.04	35.23 ± 0.11	35.24 ± 0.11
SN2006D	CfA	NIR@max	$BVr'i'JH$	0.009	32.84 ± 0.12	32.91 ± 0.09	32.89 ± 0.09
SN2006N	CfA	–	$BVr'i'JH$	0.015	33.89 ± 0.08	33.82 ± 0.10	33.78 ± 0.10
SN2006ac	CfA	–	$BVr'i'JH$	0.024	34.98 ± 0.05	35.08 ± 0.10	35.06 ± 0.10
SN2006ax	CSP	NIR@max	*BVriYJH*	0.018	34.36 ± 0.06	34.31 ± 0.09	34.30 ± 0.10
SN2006bh	CSP	NIR@max	*BVriYJH*	0.011	33.24 ± 0.10	33.34 ± 0.09	33.33 ± 0.10
SN2006cp	CfA	–	$BVr'i'JH$	0.022	34.84 ± 0.05	34.97 ± 0.10	34.98 ± 0.12
SN2006ej	CSP	–	*BVriYJH*	0.021	34.66 ± 0.05	34.67 ± 0.10	34.67 ± 0.10
SN2006kf	CSP	NIR@max	*BVriYJH*	0.019	34.53 ± 0.06	34.71 ± 0.10	34.72 ± 0.08
SN2006lf	CfA	NIR@max	$BVr'i'JH$	0.012	33.49 ± 0.09	33.52 ± 0.10	33.53 ± 0.09
SN2007A	CSP	NIR@max	*BVriYJH*	0.017	34.27 ± 0.06	34.25 ± 0.10	34.24 ± 0.10
SN2007af	CSP	NIR@max	*BVriYJH*	0.006	31.79 ± 0.05	31.94 ± 0.09	31.98 ± 0.09
SN2007ai	CSP	NIR@max	*BVriYJH*	0.033	35.69 ± 0.03	35.52 ± 0.09	35.52 ± 0.09
SN2007as	CSP	NIR@max	*BVriYJH*	0.018	34.41 ± 0.08	34.42 ± 0.09	34.43 ± 0.09
SN2007bc	CSP	NIR@max	*BVriYJH*	0.021	34.72 ± 0.05	34.74 ± 0.10	34.73 ± 0.09
SN2007bd	CSP	NIR@max	*BVriYJH*	0.031	35.57 ± 0.04	35.60 ± 0.10	35.57 ± 0.10
SN2007ca	CSP	NIR@max	*BVriYJH*	0.015	33.89 ± 0.08	34.04 ± 0.10	34.04 ± 0.08
SN2007co	CfA	–	$BVr'i'JH$	0.027	35.30 ± 0.04	35.43 ± 0.10	35.42 ± 0.10
SN2007cq	CfA	–	$BVr'i'JH$	0.025	35.11 ± 0.04	34.87 ± 0.10	34.86 ± 0.11
SN2007jg	CSP	NIR@max	*BVriYJH*	0.038	36.02 ± 0.03	36.14 ± 0.10	36.16 ± 0.09
SN2007le	CSP	NIR@max	*BVriYJH*	0.006	32.13 ± 0.17	32.20 ± 0.09	32.20 ± 0.10
SN2007qe	CfA	–	$BVr'i'JH$	0.024	34.96 ± 0.05	35.18 ± 0.10	35.19 ± 0.09
SN2007sr	CSP	–	*BVriYJH*	0.004	31.29 ± 0.11	31.62 ± 0.09	31.63 ± 0.09
SN2007st	CSP	–	*BVriYJH*	0.021	34.72 ± 0.05	34.42 ± 0.10	34.40 ± 0.09
SN2008C	CSP	–	*BVriYJH*	0.018	34.31 ± 0.06	34.37 ± 0.10	34.39 ± 0.09
SN2008af	CfA	–	$BVr'i'JH$	0.034	35.78 ± 0.03	35.66 ± 0.12	35.63 ± 0.11
SN2008ar	CSP	NIR@max	*BVriYJH*	0.029	35.42 ± 0.04	35.30 ± 0.10	35.29 ± 0.10
SN2008bc	CSP	NIR@max	*BVriYJH*	0.016	34.05 ± 0.07	34.14 ± 0.09	34.12 ± 0.08
SN2008bf	CSP	NIR@max	*BVriYJH*	0.025	35.13 ± 0.05	35.13 ± 0.09	35.10 ± 0.09
SN2008fl	CSP	–	*BVriYJH*	0.020	34.59 ± 0.06	34.49 ± 0.09	34.50 ± 0.09
SN2008fr	CSP	–	*BVriYJH*	0.038	36.04 ± 0.12	36.10 ± 0.09	36.08 ± 0.10
SN2008fw	CSP	–	*BVriYJH*	0.009	32.76 ± 0.13	33.05 ± 0.10	33.05 ± 0.09
SN2008gb	CfA	NIR@max	$BVr'i'JH$	0.038	36.03 ± 0.03	35.94 ± 0.10	35.87 ± 0.11
SN2008gg	CSP	–	*BVriYJH*	0.031	35.58 ± 0.03	35.66 ± 0.10	35.65 ± 0.09
SN2008gl	CSP	–	*BVriYJH*	0.033	35.72 ± 0.03	35.79 ± 0.10	35.84 ± 0.09
SN2008gp	CSP	NIR@max	*BVriYJH*	0.034	35.74 ± 0.03	35.71 ± 0.09	35.70 ± 0.09
SN2008hj	CSP	NIR@max	*BVriYJH*	0.037	35.97 ± 0.03	36.01 ± 0.10	36.00 ± 0.08
SN2008hm	CfA	–	$BVr'i'JH$	0.021	34.70 ± 0.05	34.76 ± 0.10	34.75 ± 0.10

Downloaded from https://academic.oup.com/mnras/article/510/3/3939/6448478 by University of Cambridge user on 14 December 2022

Table 1 – *continued*

SN	Source	Cut	Filters	$z^{a}_{\rm CMB}$	$\hat{\mu}^{b}_{\rm ext}$	$\hat{\mu}_{\rm phot}$ (resub)^{c}	$\hat{\mu}_{\rm phot}$ (CV)^{d}
SN2008hs	CfA	NIR@max	$BVr'i'JH$	0.019	34.47 ± 0.06	34.70 ± 0.10	34.80 ± 0.10
SN2008hv	CSP	NIR@max	*BVriYJH*	0.014	33.81 ± 0.08	33.85 ± 0.10	33.85 ± 0.09
SN2008ia	CSP	–	*BVriYJH*	0.022	34.86 ± 0.05	34.84 ± 0.10	34.82 ± 0.09
SN2009D	CSP	NIR@max	*BVriYJH*	0.024	35.03 ± 0.04	35.03 ± 0.09	35.00 ± 0.09
SN2009Y	CSP	NIR@max	*BVriYJH*	0.009	32.95 ± 0.12	33.01 ± 0.09	32.95 ± 0.09
SN2009aa	CSP	NIR@max	*BVriYJH*	0.029	35.40 ± 0.04	35.27 ± 0.10	35.27 ± 0.09
SN2009ab	CSP	–	*BVriYJH*	0.010	33.14 ± 0.11	33.47 ± 0.10	33.49 ± 0.08
SN2009ad	CSP	NIR@max	*BVriYJH*	0.029	35.40 ± 0.04	35.33 ± 0.10	35.31 ± 0.10
SN2009ag	CSP	NIR@max	*BVriYJH*	0.010	33.12 ± 0.11	33.09 ± 0.09	33.07 ± 0.10
SN2009al	CfA	NIR@max	$BVr'i'JH$	0.023	34.94 ± 0.05	34.84 ± 0.09	34.83 ± 0.09
SN2009an	CfA	NIR@max	$BVr'i'JH$	0.011	33.23 ± 0.10	33.32 ± 0.09	33.31 ± 0.09
SN2009bv	CfA	NIR@max	$BVr'i'JH$	0.038	36.05 ± 0.03	36.13 ± 0.10	36.13 ± 0.10
SN2009cz	CSP	NIR@max	*BVriYJH*	0.022	34.79 ± 0.05	34.79 ± 0.10	34.78 ± 0.09
SN2009kk	CfA	–	$BVr'i'JH$	0.012	33.51 ± 0.09	33.96 ± 0.10	33.97 ± 0.09
SN2009kq	CfA	–	$BVr'i'JH$	0.013	33.58 ± 0.09	33.72 ± 0.10	33.75 ± 0.10
SN2010ai	CfA	NIR@max	$BVr'i'JH$	0.024	34.99 ± 0.05	34.96 ± 0.10	34.95 ± 0.10
SN2010dw	CfA	–	$BVr'i'JH$	0.039	36.09 ± 0.03	35.99 ± 0.10	35.95 ± 0.09

Notes. [a]Redshift with corrections for local flows and CMB as described in Avelino et al. (2019). For eight nearby SN with available redshift-independent distance estimates from Cepheids, Tully–Fisher, or surface brightness fluctuations (SNs 1998bu, 2001el, 2002dj, 2003du, 2003hv, 2005cf, 2007af, 2007sr), this is an effective redshift as described in Avelino et al. (2019).
[b]External distance estimate and standard deviation, either from redshift-independent distance estimate or from redshift and assumed $H_0 = 73.24$ km s $^{-1}$ Mpc^{-1}. See Avelino et al. (2019) tables 2 and 4.
[c]Optical + NIR BAYESN photometric distance estimate obtained by resubstitution (c.f. Section 5.3).
[d]Optical + NIR BAYESN photometric distance estimate obtained by cross-validation (c.f. Section 5.3).

We discretize the integrals over wavelength (equation 6) as numerical Riemann sums with resolution $\Delta\lambda_r = 20$ Å. This provides sufficient precision for evaluating the model fluxes (with discretization error < 0.2 per cent and therefore much smaller than typical photometric error – across all passbands, the median photometric error is 0.018 mag).

We validated our training code via simulations. We set values of the hyperparameters, similar to those found from training on the real data, and used the forward model to generate simulated SN Ia light curves with characteristics similar to our real training set. The training code was then run on the simulated data set and we recovered all true values of the SED components and hyperparameters within the posterior uncertainties.

We can employ the model and Bayesian inference code in two modes. In *training* mode, we condition on the external distance estimates and their uncertainties, along with the SN Ia light curves and redshifts, to sample the joint posterior of all hyperparameters and latent variables. Trying to find the single optimal point of the global posterior in the high-dimensional parameter space is vulnerable to overfitting. Instead, we use the Bayesian approach to sample the global joint posterior equation (29), which allows us to marginalize over the posterior uncertainties in the latent variables when estimating the hyperparameters, including the SED components. In *distance-fitting* mode, we use posterior estimates of the hyperparameters of the already-trained model, and we remove the external distance constraint. Redshifts are only used to shift the SED between the rest-frame and observer-frame and to account for time-dilation. We then compute posterior inference on the latent parameters of individual SNe, and marginalize to obtain the posterior the photometric distance from the SN Ia light curve (equation 31).

For BAYESN and SNOOPY we fit the *BVRIYJH* bands, where *RI* includes *ri* and $r'i'$ filters, where applicable. The version of BAYESN described here has not been trained on *U*-band data; preliminary analysis with a BAYESN prototype including the *U*-band does not show a significant improvement in results on this sample. We apply our current model either to the available *BVRI* (optical) or *BVRIYJH* (optical + NIR) data.

4.2 SALT2 and SNooPy fitting

We used the `SALT2.4` model of Betoule et al. (2014) with the error model covariance to fit the optical light curves. The specific implementation of SALT2 used is available in the `sncosmo` package (Barbary et al. 2016). For each object, we initially adopt the Avelino et al. (2019) estimates of the time of *B*-band maximum to select observations between -10d and + 40d in phase with $S/N > 3$. These estimates were originally obtained from SNOOPY fits to these well-sampled light curves, and are very precise. This ensures that the same observations are used by both SALT2 and BAYESN. SALT2 has a range of 2000–9000 Å and therefore can fit the *UBVRI* bands, but as with BAYESN, we do not fit the *U*-band and restrict the comparison to *BVRI*. We compared the SALT2 results with or without *U*-band, and found that the *U*-band data did not improve the results for our sample. The limited template range of SALT2 also prevents us from comparison with *BVRIYJH* fits. Pierel et al. (2018) created a NIR extension to the SALT2 model that is suitable for simulations, but that did not use the same training procedure as that used to create the SALT2.4 model templates. Therefore, it is not suitable for fitting real light curves and does not yield calibrated distances.

For each object, we begin with an initial guess for the parameters, which we refine with `Minuit` (James & Roos 1975). We use the result from `Minuit` to set the initial positions of 32 walkers used to sample the posterior distribution with the `emcee` Markov Chain Monte Carlo package (Foreman-Mackey et al. 2013). We generate 2000 samples per walker after discarding the first 500 steps as burn-in. We visually inspect the parameter chains and 2D marginalized posterior distributions. We compute the median value of the samples as the 'best-fit' estimate and use the 16th and 84th percentiles of the

Downloaded from https://academic.oup.com/mnras/article/510/3/3939/6448478 by University of Cambridge user on 14 December 2022

samples as a credible interval. As these are well-sampled high-S/N light curves, the parameters are well-constrained. To obtain distances, we fit the parameters of the Tripp formula (equation 1) using the full sample, obtaining: $\hat{M}_B = -19.01$, $\hat{\alpha} = 0.117$, $\hat{\beta} = 2.939$. For consistent comparisons, these SALT2 distance estimates are on a scale of $H_0 = 73.24$ km s $^{-1}$ Mpc^{-1}.

We use the SNooPy `EBV_model2`[9] to fit the observations using templates parametrized by the light-curve stretch, s_{BV}. The `EBV_model2` uses the same algorithm as Prieto, Rest & Suntzeff (2006) to build the templates together with the updated calibration of 24 CSP supernovae presented in Burns et al. (2011). The resulting `EBV_model2` rest-frame light curves templates cover *uBgVriYJH*. To be consistent with our comparison to SALT2, which is restricted to modelling only the optical observations, we fit *BVRI* (optical-only) as well as *BVRIYJH* (optical + NIR) data with SNooPy. We use the same initial guesses for the SNooPy fit parameters as used for the SALT2 fits. SNooPy uses a non-linear least-squares Levenberg-Marquadt algorithm to minimize the variance weighted residuals to the model. As with SALT2, we report the statistical uncertainties on the fit parameters derived from inverting the Hessian matrix at the best-fitting parameters, and we have adjusted the SNooPy distance estimates to a scale of $H_0 = 73.24$ km s $^{-1}$ Mpc^{-1}. Our low-redshift SNe have well-sampled light curves with high S/N and thus the likelihood and posterior are highly Gaussian and peaked around the best-fitting values. We do not find any significant differences between the Levenberg-Marquadt results and those using MCMC sampling. The SNooPy light-curve fitting procedure weights the light-curve fit only by the photometric errors; there is no residual covariance model.

5 RESULTS AND DISCUSSION

In the following sections (Sections 5.1 and 5.2), we describe results obtained from our model trained on the optical and NIR light curves of the full sample of 79 SNe Ia.

5.1 Light-curve inference for individual SNe Ia

As an example, Fig. 3 demonstrates a BayeSN light-curve fit to optical and NIR observations of SN 2005iq (CSP, $z = 0.034$). It also shows the posterior distribution of the latent parameters (θ_1, A_V, μ) obtained under distance-fitting mode. To obtain the marginal distribution of the photometric distance modulus $\mu_s^{\rm phot}$, the other latent parameters of the SN (including the residuals $\boldsymbol{e}_s$) are integrated over. The photometric distance modulus is well constrained to ±0.09 mag using the joint optical and NIR data at all phases.

In Fig. 4, we show a visual comparison between the BayeSN and SALT2 parameter estimates. In the top panel, we plot the SED shape parameter θ_1, which is the score of the first functional component, against the SALT2 x_1 ‘stretch’ parameter for the same SNe Ia. The sign of θ_1 has been chosen to be in the same sense as the decline rate $\Delta m_{15}(B)$ of Phillips (1993), which is the magnitude change in *B*-band between *B*-band peak and 15 d afterwards. Larger values of θ_1 correspond to faster (larger) post-maximum optical decline rates.

[9]SNooPy also has a `max_model` mode that allows one to fit (*K*-corrected) light curve data to a template light curve model in a single rest-frame filter to find a single magnitude at maximum. We do not compare against this mode, since the purpose of BayeSN is to fit the SED over the entire phase and wavelength range covered by the available data in multiple passbands simultaneously, without using *K*-corrections to compute a 1-to-1 map between photometry in observer-frame and rest-frame filters.

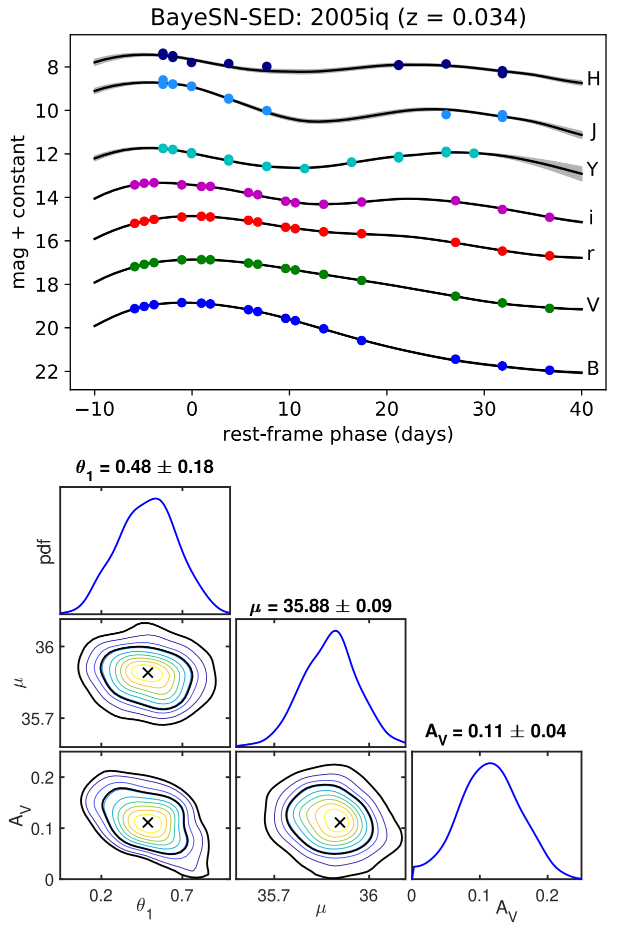

Figure 3. (top) Example BayeSN light-curve fit of optical and NIR *BVriYJH* CSP observations of the Type Ia SN 2005iq. (bottom) Posterior distribution of latent parameters of light-curve fit to CSP observations of SN 2005iq. In the 2D contour plots, the black contours contain 68 per cent and 95 per cent of the marginal posterior probability, and the mode is indicated. The 1D marginal plots depict a kernel density estimate applied to the MCMC samples for each parameter. The SED shape parameter θ_1 and host galaxy dust extinction A_V are marginalized over to obtained the posterior distribution of the photometric distance modulus μ_s.

Larger x_1 values correspond to broader optical light curves, which have slower (smaller) optical decline rates. There is a fairly tight, slightly non-linear correlation between θ_1 and x_1, suggesting that they are capturing the same underlying major mode of variation.

In the bottom panel of Fig. 4, we compare the SALT2 colour parameter *c* and the BayeSN fitted value of the apparent $B - V$ colour at peak $t = 0$. The latter is determined by evaluating the rest-frame SED model (at redshift $z_s = 0$) with the fit parameters (θ_1^s, A_V^s, $\boldsymbol{e}_s$) for each SN, and integrating it under reference *B* and *V* bandpasses, which we take to be those of the CSP. There is a strong but not exactly 1-to-1 correlation between the two. The BayeSN model is able to leverage the optical and NIR data of the full light curve to probabilistically decompose the apparent colour into an intrinsic $B - V$ colour and dust reddening $E(B - V)$. The former is computed from the light-curve fit by evaluating the rest-frame SED with the light-curve fit parameters (θ_1^s, $\boldsymbol{e}_s$) and setting $A_V^s = 0$, and integrating it under the reference passbands, and the latter is determined by $E(B - V)_s = A_V^s / R_V$. Our model finds that the apparent colours are the sum of two different effects and captures these two different sources of

Downloaded from https://academic.oup.com/mnras/article/510/3/3939/6448478 by University of Cambridge user on 14 December 2022

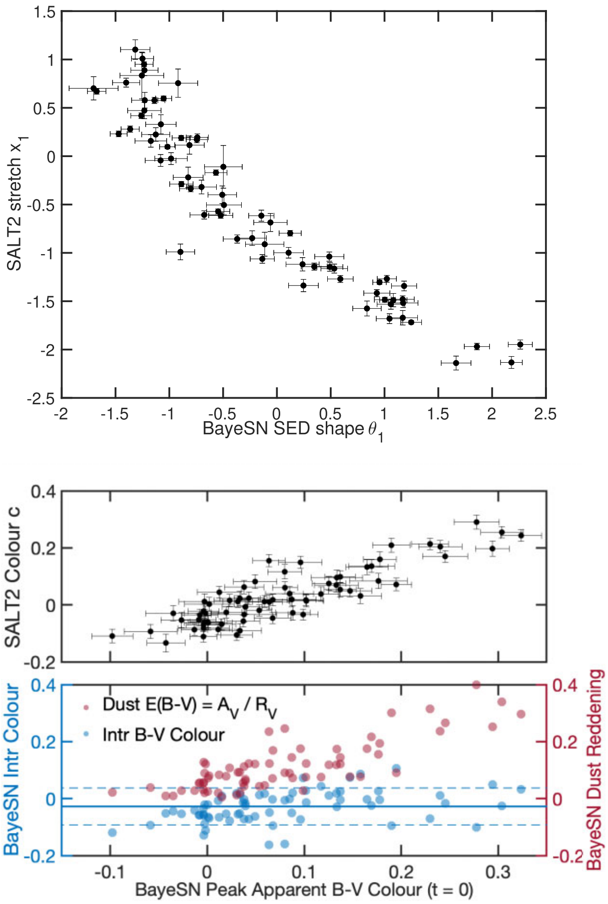

Figure 4. Comparison of BayeSN and SALT2 parameters. (upper panel) Strong correlation between the θ_1 coefficient of the first principal SED component and the SALT2 light-curve shape 'stretch' parameter x_1. (lower panel, top) correlation between the peak ($t = 0$) $B - V$ apparent colour from BAYESN light-curve fit and the SALT2 colour parameter c. (lower panel, bottom) BAYESN models the apparent colour as the sum of two latent components: the intrinsic colour (blue) and the positive reddening due to dust, $E(B - V) = A_V/R_V$ (red). The inferred population mean (blue solid) and standard deviation (blue dashed) of the intrinsic $B - V$ colour distribution are indicated. We plot the SNe with B and V measurements within ±5 d of B maximum light.

variation, which are each correlated with the rest of the SED (and thus luminosity as a function of wavelength) in different ways.

The population standard deviation of the peak intrinsic $B - V$ colour is estimated to be 0.065 ± 0.005 (blue dashed), consistent with the previous estimate of 0.067 ± 0.009 of Mandel et al. (2017). Of course, the attribution of this residual colour scatter to variation intrinsic to the SNe depends on the model assumptions being true. Possible misspecifications that might contribute some non-intrinsic scatter to this variance include colour calibration error and population variation of the dust law R_V. Burns et al. (2018) found the typical zeropoint calibration error ≈0.02 mag for CSP (the majority of our data set), and Scolnic et al. (2015) found that zeropoint errors are typically uncorrelated between bands, so we expect this contamination to be small. For a population variance σ_R^2 in dust laws around the single R_V of the model, a very rough calculation shows that the error in dust reddening that may leak into intrinsic colour scatter, $\sigma_{E(B-V)} \approx (\tau_A/R_V)(\sigma_R/R_V) \lesssim 0.02$ mag even with moderate variation $\sigma_R \lesssim 0.6$ (Thorp et al. 2021) for this range of low-to-moderate reddening, and thus is also expected to be subdominant.

5.2 Population inference

The statistical properties of the latent SED, captured by the intrinsic FPC, residual covariance, and dust distribution, are learned during the BAYESN model training phase by sampling the global posterior density, equation (29).

5.2.1 Intrinsic SED components

The baseline intrinsic SED depicted in Fig. 1 is obtained with $\theta_1^s = A_V^s = \boldsymbol{e}_s = \delta M_s = 0$, and is equal to $S_0(t, \lambda_r)10^{-0.4[M_0+W_0(t,\lambda_r)]}$. The first functional principal component (FPC) $W_1(t, \lambda)$ is also shown in Fig. 1. The top panel of Fig. 5 shows the effect of our first functional component $W_1(t, \lambda)$ on the baseline intrinsic SED at phases $t = 0$ and $t = 20$ as one changes the coefficient θ_1. In the bottom panel, for comparison, we show the effect of the dust extinction on the SED. An interesting difference between the two is the sign flip of the effect of θ_1 in the NIR at phase $t = 20$. Under this effect, SNe Ia that are dimmer in the optical are actually brighter in the NIR *YJ* bands at this later phase. This is an indication of the correlation of dimmer SNe Ia having earlier rises to the secondary NIR maximum. In contrast, the effect of dust is to make SNe Ia dimmer at all phases. This sign-flip distinction may help break the degeneracy between intrinsic SN and extrinsic dust effects.

The figure also shows a reference set of filter passband functions (with arbitrary scaling for visualization purposes). We visualize the effect of our functional components on rest-frame photometric light curves by integrating our SED model with various parameter values under this reference set. The reference set we choose for illustration are the CSP *BVriYJH* passbands and the *z*-band filter from Pan-STARRS1 (PS1). The rest-frame *z*-band (at ≈ 0.9 µm, between *i* and *Y*) region of SN Ia SEDs is regularly probed by low-*z* surveys such as Foundation and YSE, but is not modelled by either SALT2 or SNooPy. In Section 5.4, we demonstrate an example of BAYESN fitting of a rest-frame *z*-band SN Ia light curve from Foundation DR1 (Foley et al. 2018b).

By integrating the SED model under these reference optical and near-infrared passbands, we show in Fig. 6 the effect of the 1st FPC $W_1(t, \lambda_r)$ on the intrinsic optical and NIR light curves. We see that this intrinsic component captures the optical width–luminosity relation (Phillips 1993): intrinsically brighter supernovae have more slowly declining (or broader) light curves, whereas dimmer ones decline faster. This effect is seen most clearly in the *B* and *V* bands. In the redder optical bands (*r* and *i*) and into the NIR *zYJH* bands, we see that this same effect is also correlated with the timing of the second peak at $t = 20$–30 d: brighter supernovae tend to have later secondary NIR peaks, while dimmer SNe Ia have earlier ones, which is a further reflection of the trend seen in Fig. 5. In *iYJH* bands, the effect also correlates to more pronounced second peaks. The empirical relation we capture correlates strongly with the theoretical models of Kasen (2006), who found that brighter SNe Ia should have more pronounced NIR secondary maxima at later phases due to role of the ionization evolution of iron group elements in the SN ejecta in redistributing energy from the optical to the NIR. Similar trends have been seen by Dhawan et al. (2015), and Shariff et al. (2016a) explored the use of the phase of the secondary NIR maximum for standardizing SN Ia optical magnitudes.

The first NIR peak typically occurs a few days before the optical (*B*) peak ($t = 0$). Estimation of the 1st FPC at early pre-maximum phases in the NIR is somewhat limited by the relative scarcity of quality NIR observations there in the current data set (particularly in the *H*-band). Hence, the apparent sensitivity of the early

Downloaded from https://academic.oup.com/mnras/article/510/3/3939/6448478 by University of Cambridge user on 14 December 2022

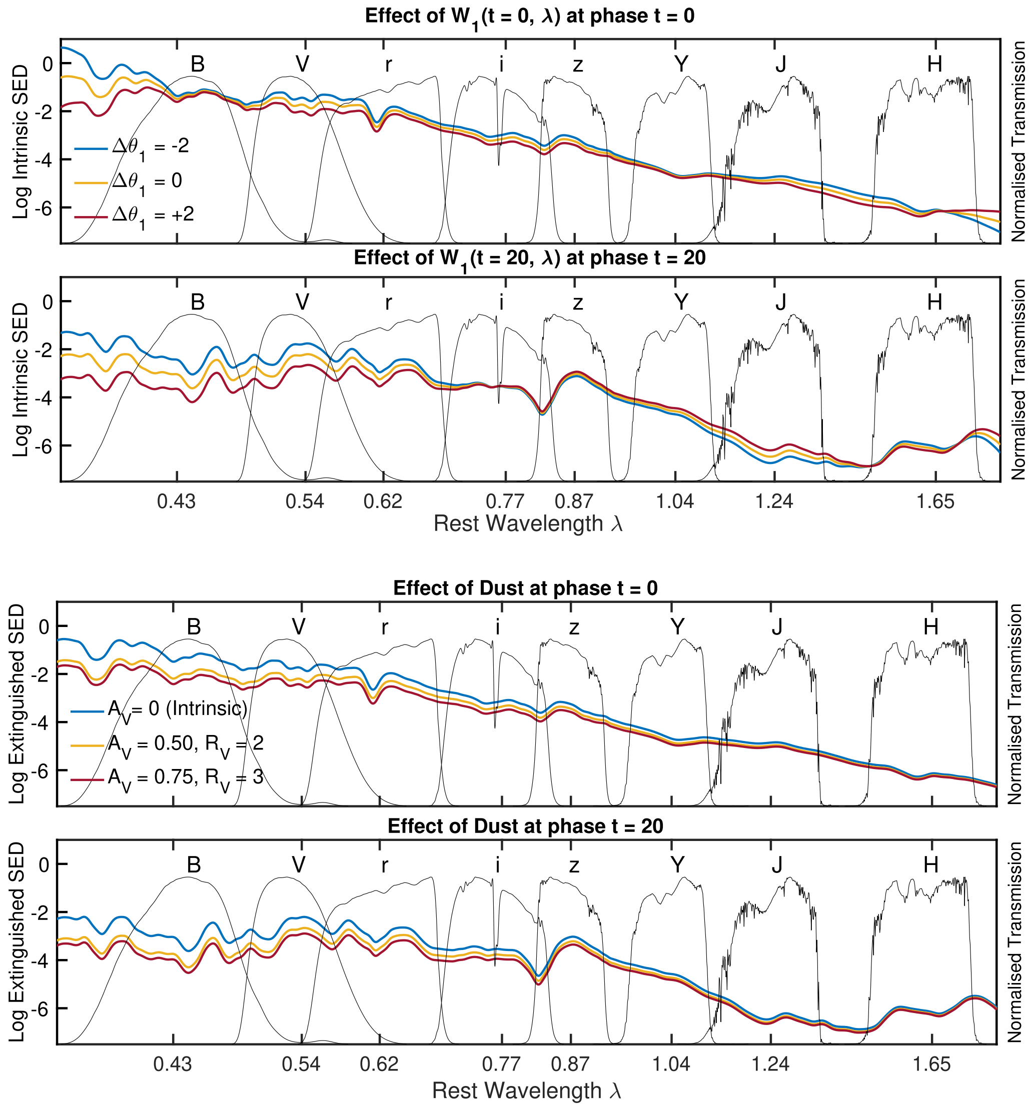

Figure 5. (top) Variation in the optical and NIR intrinsic SED captured by the first functional component $W_1(t, \lambda)$ at $t = 0$ and 20 d. We vary the value of θ_1 by $\bar{\theta}_1 \pm 2\sigma$, holding all other SN parameters to zero. (bottom) The effect of dust extinction on the optical and NIR SED. We apply dust extinction to the baseline mean intrinsic SED with different combinations of A_V, R_V that produce the same optical colour excess $E(B - V) = A_V/R_V$.

($t < -5$) H-band light curve to $\theta_1 W_1(t, \lambda)$ may be spurious. Future data releases with greater NIR coverage at early phases will help us improve the model.

In Fig. 7, we illustrate the dependence of optical and NIR absolute magnitudes on the SED shape parameter θ_1 of the FPC. The extinguished absolute magnitudes $M_s^{\rm ext}$ of an SN s are obtained by evaluating the model SED with its fitted parameters (θ_1^s, $\boldsymbol{e}_s$, δM_s, A_V^s), setting $\mu_s = 0$, and integrating it under the reference passbands in the SN rest-frame. The intrinsic absolute magnitudes $M_s^{\rm int}$ are obtained in the same way but by setting $A_V^s = 0$. In the optical B-band the average dust extinction correction is a 0.40 mag shift in absolute magnitude for the sample. In the NIR Y, H-bands, the mean shift is 0.10 and 0.06 mag, respectively, which reflects the much diminished effect of dust extinction in the NIR compared to the optical (Fig. 1). The relatively steep mean dependence of the B intrinsic absolute magnitude on θ_1 captures the optical width–luminosity relation (Phillips 1993). In the NIR, the slopes of the dependence of Y and H intrinsic absolute magnitudes with θ_1 are consistent with zero, after marginalizing over the posterior uncertainties. The scatter about the mean intrinsic relation due to the SED residual functions is approximately 0.10 mag. We note that the scatter around the mean intrinsic relation is not necessarily identical to the photometric distance uncertainty nor the expected scatter in the Hubble diagram. This is because the SED shape θ_1 and the dust extinction A_V factors must themselves be estimated from the data, and their uncertainties are themselves influenced by the intrinsic residual covariance. Instead, proper inference of the photometric distance uncertainty comes from the marginalization in equation (31). However, the diminished effect of dust A_V and the insensitivity to θ_1 in the NIR do significantly reduce their contributions to the derived photometric distance uncertainties.

Colour curves, derived from flux ratios or magnitude differences between different filters, provide a useful window for understanding SNe Ia, since they are independent of the distance estimate and its errors. In the top panel of Fig. 8, we illustrate the effect of the $W_1(t, \lambda_r)$ on the intrinsic optical-NIR colour curves by varying θ_1. At each epoch t, these are obtained by integrating the resulting rest-frame SED under each passband taking the difference with respect to the V-band magnitude. The general trend is that the intrinsically

Downloaded from https://academic.oup.com/mnras/article/510/3/3939/6448478 by University of Cambridge user on 14 December 2022

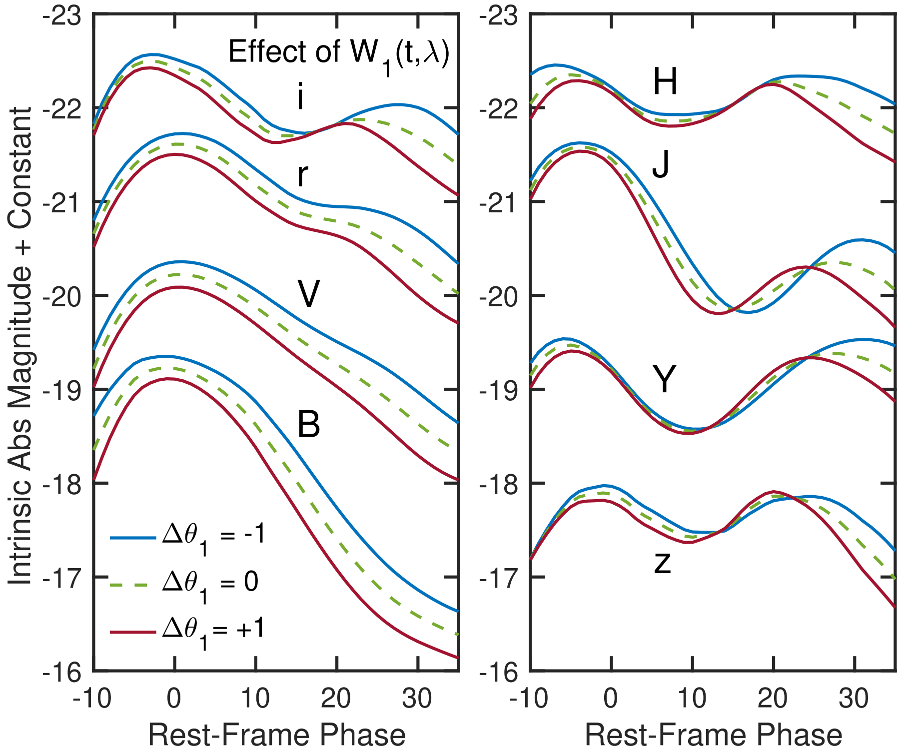

Figure 6. Intrinsic variation in optical and NIR intrinsic absolute light curves captured by the first functional component $W_1(t, \lambda)$. We vary the value of θ_1 by $\bar{\theta}_1 \pm 1\sigma$, while fixing A_V and other SN parameters to zero. Variation in $\theta_1 W_1(t, \lambda)$ captures the width–luminosity relation in the optical (Phillips 1993). Variation in this component simultaneously modulates the amplitude and timing of the second peak in the near-infrared. For visual clarity, the absolute light curves have been shifted vertically by arbitrary constants (B: 0, V: −1, r: −2.5, i: −4, z: 0.5, Y: −1, J: −3, H: −4).

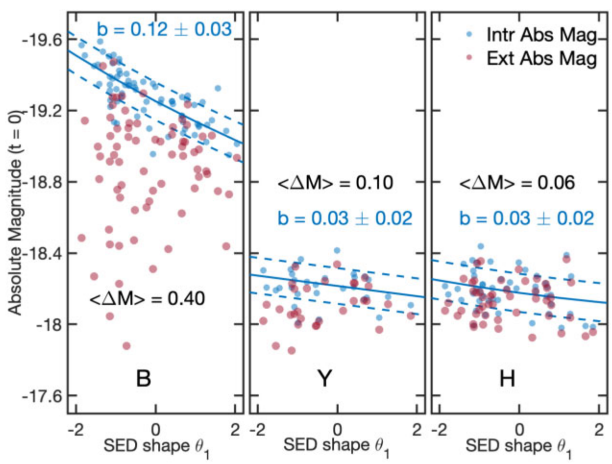

Figure 7. Intrinsic variation and host galaxy dust effects on peak absolute magnitudes at $T_{B,\max}$ (phase $t = 0$) in the rest-frame optical B and NIR Y, H bands. Each point is a posterior realization of the intrinsic absolute magnitude $M_s^{\rm int}$ (blue) or host dust-extinguished absolute magnitude $M_s^{\rm ext}$ (red) of each SN. In each panel, we plot the SNe with data in a given filter. The solid line indicates the mean effect of the intrinsic $W_1(t, \lambda)$ model component on the intrinsic absolute magnitude through the coefficient θ_1. The slope of this line is indicated as b. The dashed lines indicate ±1 standard deviation captured by the intrinsic residual covariance. The mean effect of host galaxy dust extinction in each band, quantified by the sample average difference between each SN's extinguished and intrinsic absolute magnitude, is shown.

brighter, and more slowly declining, SNe Ia (more negative θ_1) tend to have bluer (more negative) colour curves in each of the colours shown. The first FPC $W_1(t, \lambda_r)$ modulates the colour curves in a time-dependent fashion. While there are fixed points in phase when particular intrinsic colours are fairly insensitive to θ_1, at phases $10 < t < 20$ d, there is significant intrinsic colour variation in all optical-NIR colours relative to V-band.

In the bottom panel of Fig. 8, we compare this with the impact of host galaxy dust reddening on the optical-NIR colour curves. In contrast to the intrinsic FPC, the effect of dust on colour curves is relatively constant in phase,[10] and the main effect is across different colours. We show the mean intrinsic colour curves with no dust $A_V = 0$ (thin blue), as well as two combinations of the dust parameters [(A_V, R_V) = (0.75, 3) or (0.50, 2)] that result in the same colour excess $E(B-V) = A_V/R_V = 0.25$. The plot demonstrates that, with apparent colour information in optical BVr data alone, it is very difficult to distinguish between the two possibilities. In contrast, the optical-NIR $V - YJH$ colour information helps us to break the degeneracy and distinguish between the two values of the dust law R_V.

5.2.2 Intrinsic SED residual distribution

The model captures the population distribution of residual SED variations that are unexplained by the intrinsic FPC, the host

[10] In principle, it is not exactly time-independent: since the intrinsic SN SED is time-evolving, even if the amount of dust extinction A_V is truly constant, the reddening effect on each magnitude has some time-dependence (e.g. Phillips et al. 1999; Jha et al. 2007). However, this effect is too small to be seen on the plot.

Downloaded from https://academic.oup.com/mnras/article/510/3/3939/6444478 by University of Cambridge user on 14 December 2022

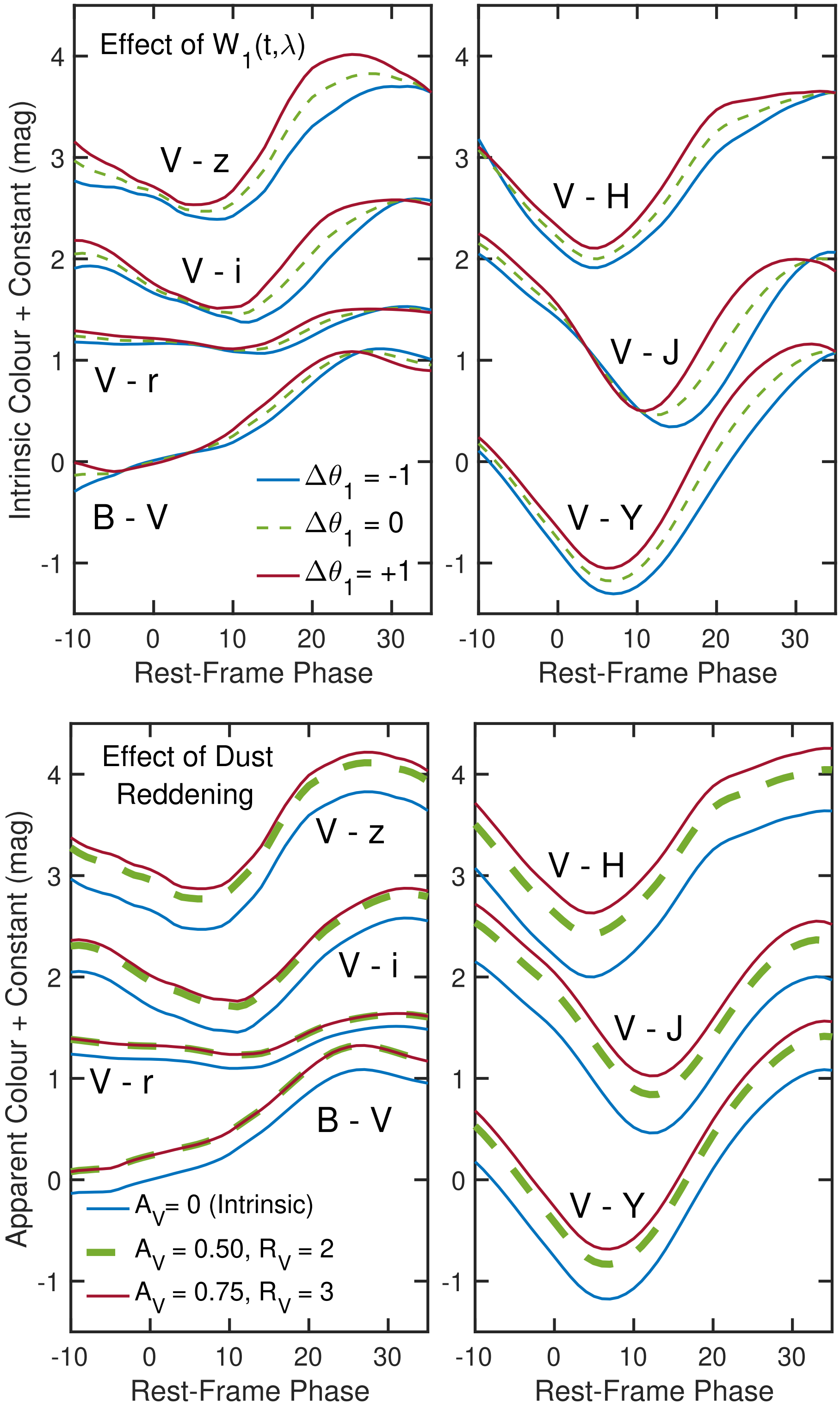

Figure 8. (top) Intrinsic variation in optical and NIR colour curves captured by the first functional component $W_1(t, \lambda)$. We vary the value of θ_1 by $\bar{\theta}_1 \pm 1\sigma$, while fixing A_V and other SN parameters to zero. (bottom) Effect of host galaxy dust extinction on optical and NIR colour curves. We show unreddened, intrinsic colour curves (blue), and two apparent colour curves with the same amount of optical $E(B-V)$ colour excess due to dust, but two different values of the dust law $R_V = 2$ or 3. We fix θ_1 and other SN parameters to zero. The phase-dependence of the $W_1(t, \lambda)$ component on intrinsic colour curves makes it distinguishable from dust. The effect of dust reddening on colour curves is approximately constant with phase. (bottom left-hand panel) With optical data only, it is difficult to distinguish between two different combinations of host dust A_V, R_V that produce the same colour excess $E(B-V) = A_V/R_V$. (bottom right-hand panel). Since the dust extinction in the NIR is smaller and less dependent on R_V, the optical-NIR colour curves help to break this degeneracy. For visual clarity, the colour curves have been shifted vertically by arbitrary constants ($B-V$: 0, $V-r$: 1.3, $V-i$: 2.5, $V-z$: 3.5, $V-Y$: 0.25, $V-J$: 2.25, $V-H$: 3.25).

Downloaded from https://academic.oup.com/mnras/article/510/3/3939/6448478 by University of Cambridge user on 14 December 2022

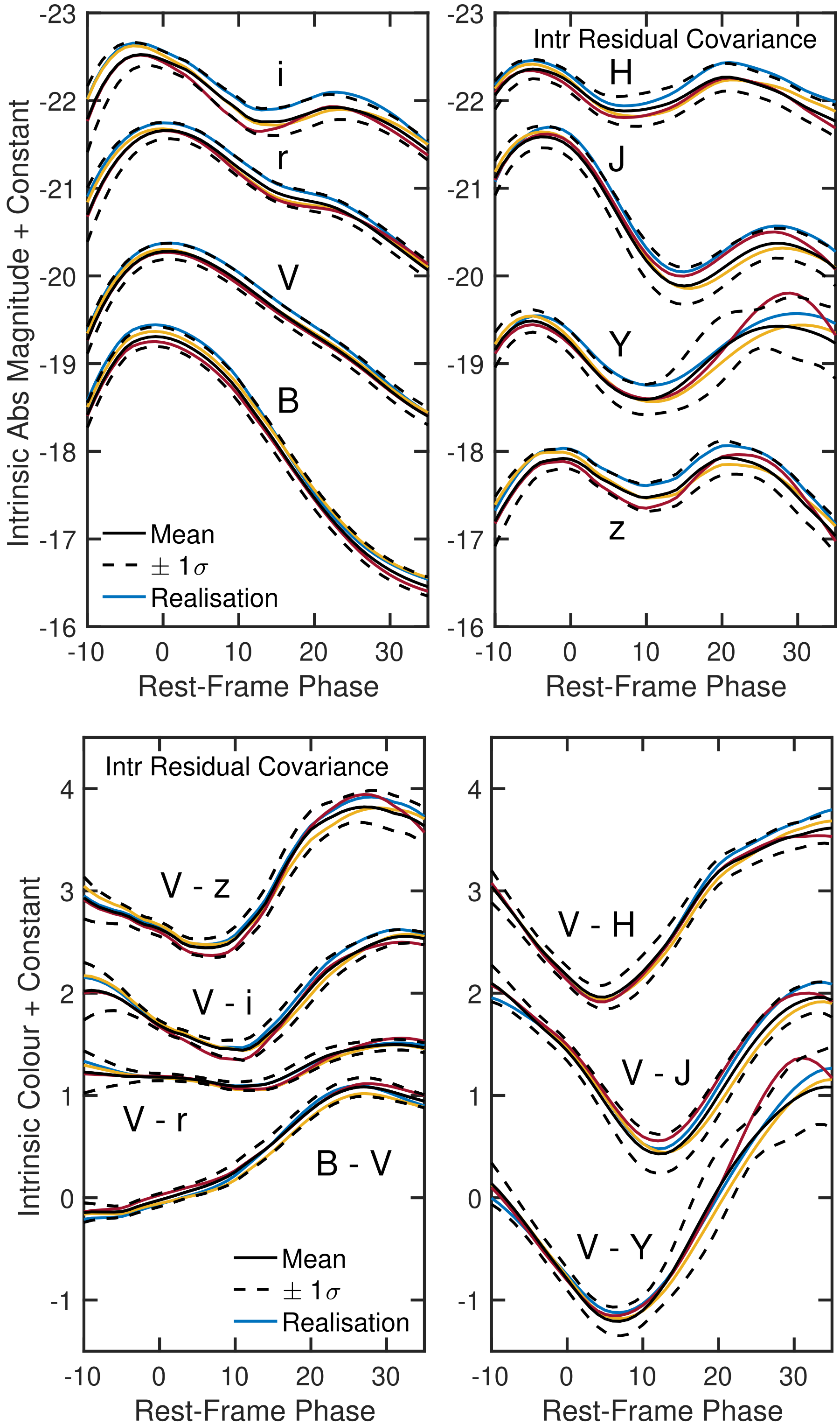

Figure 9. Effects of the covariance of phase- and wavelength-dependent intrinsic SED residuals on optical and NIR light curves (top) and colour curves (bottom). We fix the main effects $\theta_1 = A_V = 0$. The black solid lines represent the light curves or colour curves generated from the mean intrinsic SED model. The dashed lines correspond to ± 1 population standard deviation around the mean curves captured by the intrinsic SED residual covariance. The light curves or colour curves corresponding to the effects of the inferred intrinsic SED residual functions $\eta_s(t, \lambda)$ of three example SNe in the training set are shown as blue, yellow, or red curves. For example, the red curves in all the panels correspond to the effect of the intrinsic SED residual function of a single SN.

galaxy dust extinction, peculiar velocities, or other external distance uncertainties, or measurement error, through the residual covariance. The total residual SED function of an SN Ia s is $\eta_s(t, \lambda_r) = \delta M_s + \epsilon_s(t, \lambda_r)$. An example of an SED residual function is shown in Fig. 1.

Fig. 9 shows the effect of intrinsic SED residuals on rest-frame optical and NIR light and colour curves. We hold $\theta_1 = A_V = 0$, and we compute the impact of the distribution of SED residuals on the light curves and colour curves by integrating through the

Downloaded from https://academic.oup.com/mnras/article/510/3/3939/6448478 by University of Cambridge user on 14 December 2022

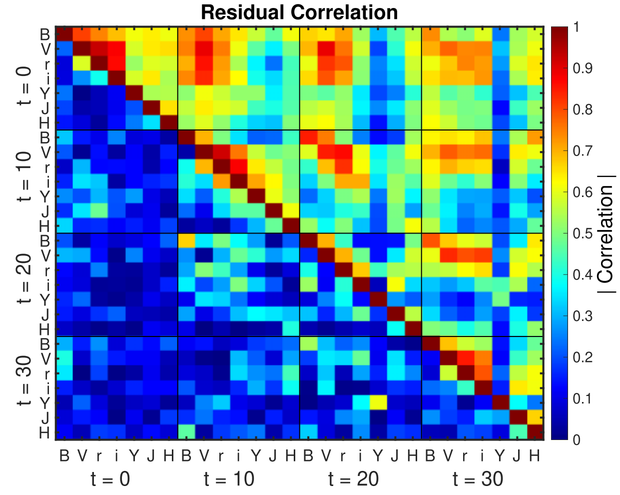

Figure 10. Correlation matrix of optical-NIR light-curve residuals over phase for the $K = 1$ model. (top right-hand panel) The correlation matrix of magnitude residuals in optical-NIR filters attributed to the total SED residual functions $\eta(t, \lambda_r) = \delta M_s + \epsilon_s(t, \lambda_r)$. (bottom left-hand panel) The correlation matrix of magnitude results in optical-NIR filters attributed to the time- and wavelength-dependent residual functions $\epsilon_s(t, \lambda_r)$. For visualization purposes, the absolute value of the correlation coefficient is plotted.

reference passbands. We compute the $\pm 1\sigma$ range at each epoch t. We do not unrealistically assume the residuals are statistically independent at each phase or in each filter; rather the residuals manifest as continuous perturbations around the main effects. The model captures continuous residual SED functions correlated across phase and wavelength. To illustrate this, we show the effect of three realizations of the intrinsic residual functions on the light curves. The residual variance is generally narrow at phases around the first peak. In later phases, particularly in the NIR, there is more intrinsic residual variation because the 1st FPC does not capture the full range of variation of the second peak.

Fig. 10 depicts the correlations over phase and wavelength of the impact of the total residual functions $\eta_s(t, \lambda_r)$ and the time- and wavelength-dependent part $\epsilon_s(t, \lambda_r)$ on the rest-frame optical-NIR light-curve magnitudes. The total residual map (top right-hand panel) indicates moderate-to-strong correlations between the optical bands, but weaker cross-correlations between optical and NIR bands. While this is clearest at peak ($t = 0$), a similar pattern persists at later phases, as well as in cross-phase correlations. At late phases ($t =$ 20, 30), the Y-band residual appears to have low correlation with other bands; this is likely due to significant variations in the second peak at these phases, as seen in Fig. 9. When the inferred grey time-constant scatter is removed, the correlations due to $\epsilon_s(t, \lambda_r)$ (bottom left-hand panel) are reduced, but there is still interesting structure. In particular, there are still moderate correlations between bands in the post-decline phases ($t = 10, 20$) as well as intertemporal correlations (e.g. between $t = 10, 20, 30$). In Section B, we show that some of the additional structure there may be captured with higher order FPCs.

5.2.3 *Host galaxy dust population*

Fig. 11 shows the distribution of posterior mean estimates of the individual dust extinction A_V values. It is well described by an exponential distribution with an average value of $\tau_A = 0.329 \pm 0.045$ mag, consistent with previous estimates in the range of 0.3–0.4 mag found for similar samples (Jha et al. 2007; Mandel et al. 2011,

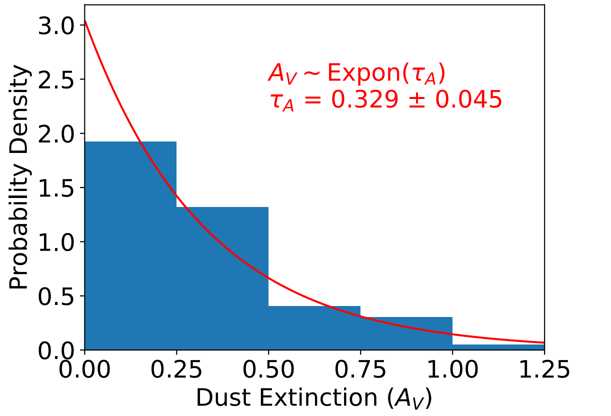

Figure 11. Distribution of posterior mean estimates of A_V^s for the SNe Ia sample. The model exponential distribution with the inferred scale (average) τ_A is shown.

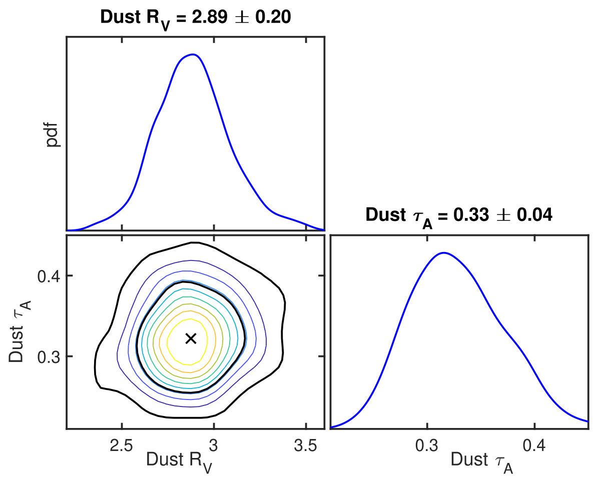

Figure 12. Posterior distribution of the inferred global R_V of the host galaxy dust law and τ_A, the population mean A_V. The black contours of the 2D contour plot contain 68 per cent and 95 per cent of the posterior probability, and the mode is marked. The 1D marginals are depicted by kernel density estimates of the MCMC samples.

2014). Fig. 12 shows posterior inferences of the average τ_A and the global value of the dust law slope R_V. The posterior constraints are determined during the training phase, and thus are obtained by marginalizing over all other components and hyperparameters of the hierarchical model. These posterior estimates are well constrained fairly independently. In particular, for this sample with colour excess $E(B - V)_{\rm host} \lesssim 0.4$, the estimated global $R_V = 2.89 \pm 0.20$ is consistent with the average for normal Milky Way dust. This is in good agreement with previous analyses of nearby samples, which have found, at these relatively low-to-moderate values of reddening (which are similar to those found in the cosmological sample), average values of the host dust $R_V \approx 2.5$–3 (Chotard et al. 2011; Foley & Kasen 2011; Mandel et al. 2011; Burns et al. 2014; Mandel et al. 2017; Léget et al. 2020).

The hierarchical model constrains R_V by analysing and weighing the entire distribution of SEDs over phase and optical to NIR wavelengths using the entire training set of SNe Ia. For visualization purposes, however, it is useful to inspect a 'slice' of this inference by examining a low-dimensional summary. Multidimensional colour information is useful as it provides constraints on the dust distribution

Downloaded from https://academic.oup.com/mnras/article/510/3/3939/6448478 by University of Cambridge user on 14 December 2022

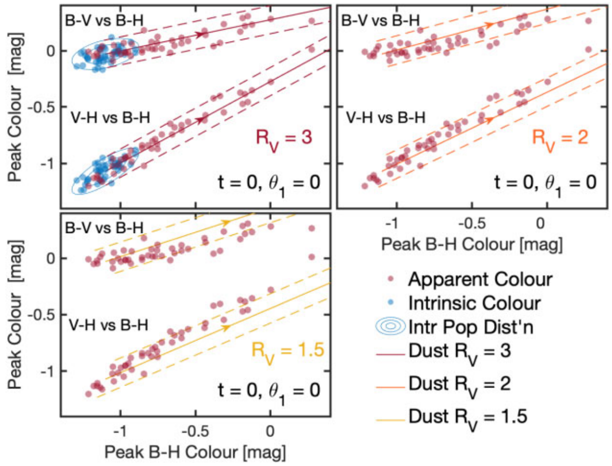

Figure 13. Constraints on the host galaxy dust R_V from the optical and NIR colour–colour diagram of SNe Ia observed in B, V, and H near peak ($t = 0$). (top left-hand panel) Each point is a posterior realization of the peak apparent colours (red) or intrinsic colours (blue) of an SN, corrected for the inferred intrinsic colour-shape relation to $\theta_1 = 0$. The blue ellipses are (68 per cent, 95 per cent) contours of the intrinsic colour population distribution inferred during the training phase, which estimated a global dust law parameter $R_V = 2.89 \pm 0.20$. For comparison, the red solid (dashed) lines have the slope of the reddening vector for $R_V = 3$ in these colours, and intercept the mean (are tangent to the 95 per cent contour) of the intrinsic distribution. Nearly all of the SNe Ia apparent colours should lie within the dashed lines under the correct dust reddening law. (top right-hand panel) Comparison of the apparent colour distribution with the inconsistent dust reddening vector for $R_V = 2$. (bottom lefthand panel) Comparison of the apparent colour distribution with the inconsistent dust reddening vector for $R_V = 1.5$.

while being insensitive to the distance estimate (and its errors). We exploit the fact that the optical and NIR data allows us to constrain the dust effects over a much larger wavelength range than is possible conventionally with the optical data alone. The plot of the dust law in Fig. 2 shows that the extinction at NIR H-band ($\approx 1.6\,\mu$m, cf. Fig. 5) is only 16 per cent of that in optical V-band ($\approx 0.54\,\mu$m), and very insensitive to R_V. Thus, the differential extinction (the colour excess) between V- and H-bands probes a large net dust effect ($\approx 0.83\, A_V$), while itself being insensitive to R_V. Meanwhile, the colour excess between B ($\approx 0.43\,\mu$m) and H-bands similarly covers a large wavelength range and therefore a large dust effect, but because of the high sensitivity of A_B to R_V (for a given A_V), this colour excess is very sensitive to R_V. The complementary optical $B - V$ colours cover only a narrow range in the optical, and therefore captures a smaller differential effect of dust, but is also highly sensitive to R_V. The advantage of measurements spanning optical to NIR is that we can leverage the joint colour information in these SNe Ia to constrain and break the degeneracy between A_V and R_V in the optical-only colours (Fig. 8).

Fig. 13 shows a 'slice' of the constraints on R_V in these colours from training the BayeSN model (equation 29). The top left-hand panel shows the distribution of peak ($t = 0$) apparent colours and intrinsic colours inferred by the model, and corrected for the inferred intrinsic colour-shape relation (Fig. 8) to $\theta_1 = 0$. The inferred intrinsic colour distribution is anchored by the SNe Ia with the least inferred amount of dust. The red arrow indicates the dust reddening vector for each colour pair for the dust law $R_V = 3$ and illustrates a shift corresponding to $A_V = 0.57$ mag from the centre of the intrinsic colour distribution. The colour distributions are consistent with an $R_V = 3$ dust law (and the posterior estimate $R_V = 2.89 \pm 0.20$).

The other panels show that the apparent colour distribution in these colour pairs are inconsistent with the dust reddening vectors for $R_V = 2$ or $R_V = 1.5$. That is, given the apparent colours of the low-reddening set (low $B - H$), assuming a low R_V would predict bluer (more negative) average $V - H$ apparent colours for a given $B - H$ for more reddened SNe Ia (e.g. $B - H > 0.5$) than is observed. Conversely, the apparent colours of the high-reddening set (high $B - H$) would imply that the apparent $V - H$ colours of the low-reddening set ought to be redder (more positive) than is observed, when assuming a low R_V. For $B - V$ colours, the same inconsistencies persist but in the opposite sense. The high- and low-reddening ends of the apparent colour distribution are most consistent with each other for $R_V \approx 3$.

Because the estimation of R_V hinges on the comparison of the colours of high-reddening SNe to those of low-reddening SNe, the most highly reddened SNe have the most leverage. In our sample, SN 1998bu has the largest extinction estimate ($A_V = 1.15 \pm 0.08$). To test that our R_V estimate is not entirely driven by this SN, we retrained the full hierarchical model omitting SN 1998bu. We found $R_V = 2.83 \pm 0.19$, indicating that our estimate is robust to the reddest SN.

5.2.4 *Covariance structure of optical and NIR peak absolute magnitudes*

During the training phase, we estimate the population covariance structure of SN Ia SEDs. The covariance structure is implied by the model equation (12) and the population distribution of the latent parameters. The total population covariance of the log latent SED at two different rest-frame coordinates (t, λ_r) and (t', λ_r') is captured by the model as

$$\begin{aligned}\mathrm{Cov}[\log S(t, \lambda_r), \log S(t', \lambda_r')] = {} & \mathrm{Var}[A_V]\, \xi(\lambda_r; R_V)\, \xi(\lambda_r'; R_V) \\ & + \left[\sum_{i=1}^{K} W_i(t, \lambda_r) W_i(t', \lambda_r') \right] \\ & + \sigma_0^2 + k_\epsilon(t, \lambda_r; t', \lambda_r'), \end{aligned} \quad (32)$$

where $k_\epsilon(t, \lambda_r; t', \lambda_r')$ is given by equation (23), and we invoke the statistical properties of the latent variables: e.g. $\mathrm{Cov}[\theta^i, \theta^j] = \delta_{ij}$, $\mathrm{Cov}[\theta^i, \delta M] = 0$ (consequences of the independent prior and population distributions specified in Section 2.5).[11] On the right-hand side, the top line describes the covariance across rest-frame wavelength induced by the dust extinction and the dust law $\xi(\lambda)$, which depends on R_V. The second line describes covariance across both phase and wavelength induced by the K intrinsic functional principal components of the SED. The third line describes the covariance of the intrinsic residual terms $\eta_s(t, \lambda_r) = \delta M_s + \epsilon_s(t, \lambda_r)$. Because the absolute magnitude in any one passband at some phase t is obtained by exponentiating equation (12) and then performing an integral of the SED under the transmission function, the covariance between any pair of absolute magnitudes in different filters at different phases is not analytic and must be computed numerically.

The population variance of the time- and wavelength-independent 'grey' magnitude offsets δM_s is captured in σ_0^2. Since this mode is indistinguishable from distance in the light-curve data, this term sets an uncertainty floor for the photometric distances. At low-z, the

[11] In practice, after training the $K = 1$ model, we estimate a sample covariance $\widehat{\mathrm{Cov}}[\theta_1, \delta M] = 0.0013$, corresponding to a sample correlation 0.016, both consistent with zero. For the $K = 2$ model (Appendix B), we estimate a sample covariance $\widehat{\mathrm{Cov}}[\theta_1, \theta_2] = 0.01$ corresponding to a sample correlation of 0.01, also consistent with zero.

Downloaded from https://academic.oup.com/mnras/article/510/3/3939/6448478 by University of Cambridge user on 14 December 2022

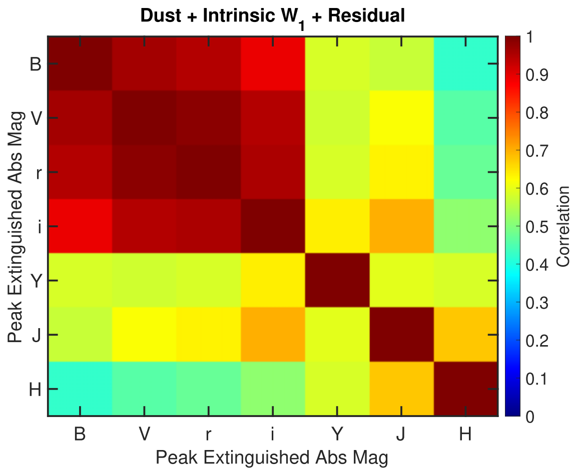

Figure 14. Map of population correlations between peak ($t = 0$) extinguished absolute magnitudes in optical and NIR passbands. These include all modelled sources of latent SED variation, including dust extinction, the intrinsic FPC $\theta_1 W_1(t, \lambda_r)$, and the residual SED covariance. Dust effects induce significant wavelength-dependent correlations in the optical, but have significantly diminished effect in the NIR. While the optical magnitudes are significantly correlated with themselves, they are less so with the NIR magnitudes, with optical-NIR cross-correlations as low as ≈ 40 per cent. This indicates there is additional information in the NIR that helps improve distance estimates.

estimated value of σ_0 is sensitive to the assumed value of $\sigma_{\rm pec}$, since the latter determines how much of the Hubble residual scatter can be attributed to peculiar velocity uncertainty. On training with our assumed value of $\sigma_{\rm pec} = 150$ km s^{-1} (Carrick et al. 2015), our model estimates $\sigma_0 = 0.09 \pm 0.02$ mag.[12] However, with a higher value of $\sigma_{\rm pec} = 250$ km s^{-1} (e.g. Scolnic et al. 2018), our model estimates $\sigma_0 = 0.06 \pm 0.02$ mag. In the future, a higher redshift optical and NIR sample would help determine σ_0 more robustly against peculiar velocities.

The full covariance structure over rest-frame phase and wavelength learned by the model is complex, and we defer a detailed discussion to future work. Here, we distill some of its key aspects. Fig. 14 depicts the population cross-correlation structure between peak (at $t = 0$) optical and NIR absolute magnitudes. The variation in absolute magnitudes is generated by the combination of the various latent component effects on the SED, and is obtained by integrating the SED through the reference filters. The map shows the correlation of the peak extinguished absolute magnitudes across optical and NIR passbands, inclusive of dust, intrinsic θ_1 SED variation, and residual covariance. The peak absolute magnitudes in the optical have a very strong total correlation, whereas the cross-correlation between optical and NIR peak absolute magnitudes is as low as ≈ 40 per cent. This is caused in part by the strong, coherent wavelength-dependence of the host galaxy dust extinction. However, the dust extinction is significantly diminished in the NIR. This reduced cross-correlation indicates there is additional information in the NIR magnitudes that helps us to improve distance estimates.

[12]There is a minor mathematical degeneracy between δM_s and the time- and wavelength-averaged mean $\langle \epsilon_s(t, \lambda_r)\rangle$, so that the full grey time-constant scatter is the sum of the two. However, this is largely broken via the priors during training. After training, if we reassign the $\langle \epsilon_s(t, \lambda_r)\rangle$ to δM_s, this merely increases the σ_0 by 0.004 mag, five times smaller than the posterior uncertainty.

5.3 Hubble diagram analysis

After training the model by sampling equation (29), we obtain posterior estimates of the FPC and population hyperparameter $\hat{\boldsymbol{H}} \equiv (\hat{\boldsymbol{W}}_{0:K}, \hat{\boldsymbol{\Sigma}}_\epsilon, \hat{\sigma}_0^2, \hat{\tau}_A, \hat{R}_V)$. We then use these to evaluate the photometric distances, derived from the light curves alone, using equation (31). We take the posterior mean and standard deviation of the posterior probability density of the photometric distances. Table 1 lists the redshifts, external distance estimates, and BAYESN photometric distance moduli for the sample.

We assess the accuracy and precision of our photometric distance estimate by comparison to the external distance estimates, via the Hubble residuals, $\hat{\mu}_s^{\rm phot} - \hat{\mu}_s^{\rm ext}$. We compare them using two summary statistics, listed in Table 2. First, we report the simple total RMS the differences between our posterior mean estimate photometric distance modulus $\hat{\mu}_s^{\rm phot}$ and the external distance estimate $\hat{\mu}_s^{\rm ext}$. Secondly, we report a statistic we denote $\hat{\sigma}_{\text{-pv}}$, obtained by minimizing

$$\hat{\sigma}_{\text{-pv}} = \underset{\sigma_{\text{-pv}}}{\arg\max} \log \left[\prod_s N\left(\hat{\mu}_s^{\rm phot} |\, \hat{\mu}_s^{\rm ext}, \sigma^2_{{\rm ext},s} + \sigma^2_{\text{-pv}}\right) \right]. \quad (33)$$

This is a maximum-likelihood estimate of the amount of dispersion in the Hubble residuals not accounted for by the uncertainties in the external distance estimate, which is dominated by the peculiar velocity uncertainty $\sigma_{\rm pec} = 150$ km s^{-1} for the vast majority of this low-z sample.

It is conventional in the SALT2 analysis to compute an 'intrinsic dispersion'[13] of the Hubble residuals, by estimating the amount of scatter in the Hubble residuals in excess of the expected contributions of 'measurement error' (which is really the estimated uncertainty on the fit parameters m_B, x_1, c), and the peculiar velocity uncertainties. This is necessary because only the light-curve fitting uncertainties on the SALT2 parameters are propagated through the Tripp formula, equation (1), to compute the distance modulus uncertainties, and the results are typically much smaller than the total RMS in the Hubble diagram. Similarly, SNooPy only uses the photometric measurement uncertainties in the light curve fit. In contrast, BAYESN produces distance uncertainties via Bayesian marginalization of the SED fit to the light-curve data, coherently incorporating θ_1 and A_V uncertainties and the residual covariance over phase and wavelength (equation 31). Since each method has a different way of reporting the distance errors, we do not 'subtract' the reported distance errors from the total RMS. Instead, to ensure consistent comparisons across methods, we use $\hat{\sigma}_{\text{-pv}}$ to remove from the total RMS only the expected contribution from external distance errors (e.g. peculiar velocities), which are the same for each method applied to the same set of SNe Ia.

Table 2 lists these Hubble residual dispersion measures for different subsets of the SN Ia sample. The vast majority comes from two large surveys with homogeneously reduced data, the CfA (Hicken et al. 2009, 2012; Friedman et al. 2015) and CSP-I (Krisciunas et al. 2017). We label this set 'CfA + CSP'. Including the minority of other SNe Ia drawn from the more heterogeneous data sources in the literature results in the 'All' sample. Furthermore, a subset of the full 'AnyNIR' sample with NIR observations near maximum light is labelled 'NIR@max.' We run BAYESN and SNOOPY on either optical-only (*BVRI*) or optical + NIR (*BVRIYJH*) light-curve data, while SALT2 is only run on optical *BVRI* data.

[13]But see footnote 2.

Downloaded from https://academic.oup.com/mnras/article/510/3/3939/6448478 by University of Cambridge user on 14 December 2022

Table 2. Summary of Hubble residuals.

SN source[a]	NIR cut[b]	$N_{\rm SN}$	λ^c	Model[d]	Total rms[e]	$\sigma_{\text{-pv}}(150)^{\rm f}$
CfA + CSP	NIR@max	40	*BVRIYJH*	BayeSN-tr	0.096	0.083
CfA + CSP	NIR@max	40	*BVRIYJH*	BayeSN-cv	0.108	0.099
CfA + CSP	NIR@max	40	*BVRIYJH*	SNooPy	0.141	0.110
CfA + CSP	NIR@max	40	*BVRI*	SALT2	0.129	0.112
All	NIR@max	48	*BVRIYJH*	BayeSN-tr	0.113	0.091
All	NIR@max	48	*BVRIYJH*	BayeSN-cv	0.123	0.107
All	NIR@max	48	*BVRIYJH*	SNooPy	0.148	0.110
All	NIR@max	48	*BVRI*	SALT2	0.131	0.114
CfA + CSP	AnyNIR	66	*BVRIYJH*	BayeSN-tr	0.135	0.109
CfA + CSP	AnyNIR	66	*BVRIYJH*	BayeSN-cv	0.145	0.124
CfA + CSP	AnyNIR	66	*BVRIYJH*	SNooPy	0.157	0.128
CfA + CSP	AnyNIR	66	*BVRI*	SALT2	0.147	0.125
All	AnyNIR	79	*BVRIYJH*	BayeSN-tr	0.137	0.109
All	AnyNIR	79	*BVRIYJH*	BayeSN-cv	0.147	0.123
All	AnyNIR	79	*BVRIYJH*	SNooPy	0.161	0.125
All	AnyNIR	79	*BVRI*	SALT2	0.148	0.122
CfA + CSP	AnyNIR	66	*BVRI*	BayeSN-tr	0.149	0.128
CfA + CSP	AnyNIR	66	*BVRI*	BayeSN-cv	0.156	0.140
CfA + CSP	AnyNIR	66	*BVRI*	SNooPy	0.158	0.142
CfA + CSP	AnyNIR	66	*BVRI*	SALT2	0.147	0.125
All	AnyNIR	79	*BVRI*	BayeSN-tr	0.150	0.126
All	AnyNIR	79	*BVRI*	BayeSN-cv	0.157	0.137
All	AnyNIR	79	*BVRI*	SNooPy	0.158	0.140
All	AnyNIR	79	*BVRI*	SALT2	0.148	0.122

Note. [a]Data Source. 'All' = CfA+CSP + Others [b]The 'NIR@max' cut requires NIR data near maximum light. 'AnyNIR' does not. [c]In optical + NIR fitting, all available data in *BVRIYJH* is used. In optical-only fitting, only available data in *BVRI* is used, where *R* and *I* can also include r, $r^{'}$, and i, $i^{'}$. In either case, the model was trained on optical + NIR data. [d]'Bayesn-tr' refers to the error of photometric distances from resubstitution of the whole training set. 'BayeSN-cv' refers to the error of photometric distances from 10-fold cross-validation. We cannot do equivalent cross-validation with SALT2 or SNooPy. [e]Simple total RMS of the Hubble residuals. [f]Dispersion estimate after removing expected variance due to peculiar velocity uncertainties, assuming $\sigma_{\rm pec} = 150\ {\rm km\,s^{-1}}$.

5.3.1 *Resubstitution or training error*

The resubstitution, or training error, is obtained by training the model on the optical+NIR data of the full sample, and then applying it to fit the optical + NIR or optical-only light curves of SNe Ia in the training set to determine their photometric distances. In Table 2, these estimates are labelled 'BayeSN-tr'. Fig. 15 shows the Hubble diagram obtained with BAYESN fits of optical and NIR data of the CfA + CSP NIR@max sample. With joint optical and NIR data, BAYESN achieves a low total RMS = 0.096 mag on this set. Removing the expected contribution from external distance error and peculiar velocities, we obtain $\hat{\sigma}_{\text{-pv}} = 0.083$ mag. Meanwhile, on the same set of SNe Ia, SNOOPY, and SALT2 have larger RMS ≈ 0.13–0.14 mag, with $\hat{\sigma}_{\text{-pv}} \approx 0.11$ mag. Notably, the photometric distance modulus uncertainties of individual SNe Ia from SNOOPY or SALT2 with the standard procedure are small in comparison to the total RMS, because they only propagate the photometric light curve uncertainties (and in SALT2, the error model covariance). In Fig. 15, we show these error bars, and, in grey, we have added to those in quadrature the residual variance needed to make the total reduced χ^2 of each Hubble Diagram equal to one. In contrast, the BAYESN photometric distance uncertainties are obtained in a principled manner by marginalization of the latent components including the residual covariance (equation 31). The individual photometric distance uncertainties from BAYESN listed in Table 1 already reflect the scatter in the Hubble diagram.

We assess the significance of the difference between the RMS Hubble residual of distance from our model compared to those from SALT2 using bootstrap. From the full training set SNe Ia, we construct a bootstrapped set by sampling with replacement. For each method, we compute the Hubble residual RMS of the SNe Ia within the bootstrapped set. We compute the difference in RMS between the two methods within the bootstrapped set. We repeat this 1000 times and then compute the variance of the differences in RMS across the bootstraps. This procedure accounts for the fact that each method is analysing the same set of SNe Ia, and therefore the joint sampling distribution of both methods' RMS over bootstraps is correlated. For the CfA + CSP NIR@max subset, we compare SALT2 using optical (which has the lowest RMS of the alternate methods) versus BAYESN using optical + NIR, and we find a $\Delta{\rm RMS} = 0.033 \pm 0.012\,(2.7\sigma)$.

Avelino et al. (2019) recently obtained RMS scatter of 0.11–0.12 mag for the same SNe in the NIR@max subsets using only NIR $YJHK_s$ light-curve data, and without any host-galaxy dust correction. Our optical + NIR results are a slight improvement over that. Since this sample has already been restricted to low-to-moderate reddening $E(B-V)_{\rm host} < 0.4$, we expect the NIR extinction corrections to be small. With the fitted exponential dust distribution and dust law, the H-band extinction A_H has a population standard deviation ≈0.06 mag, a subdominant component of the total variance. However, when fitting more highly reddened SNe Ia, or light curves with a

Downloaded from https://academic.oup.com/mnras/article/510/3/3939/6448478 by University of Cambridge user on 14 December 2022

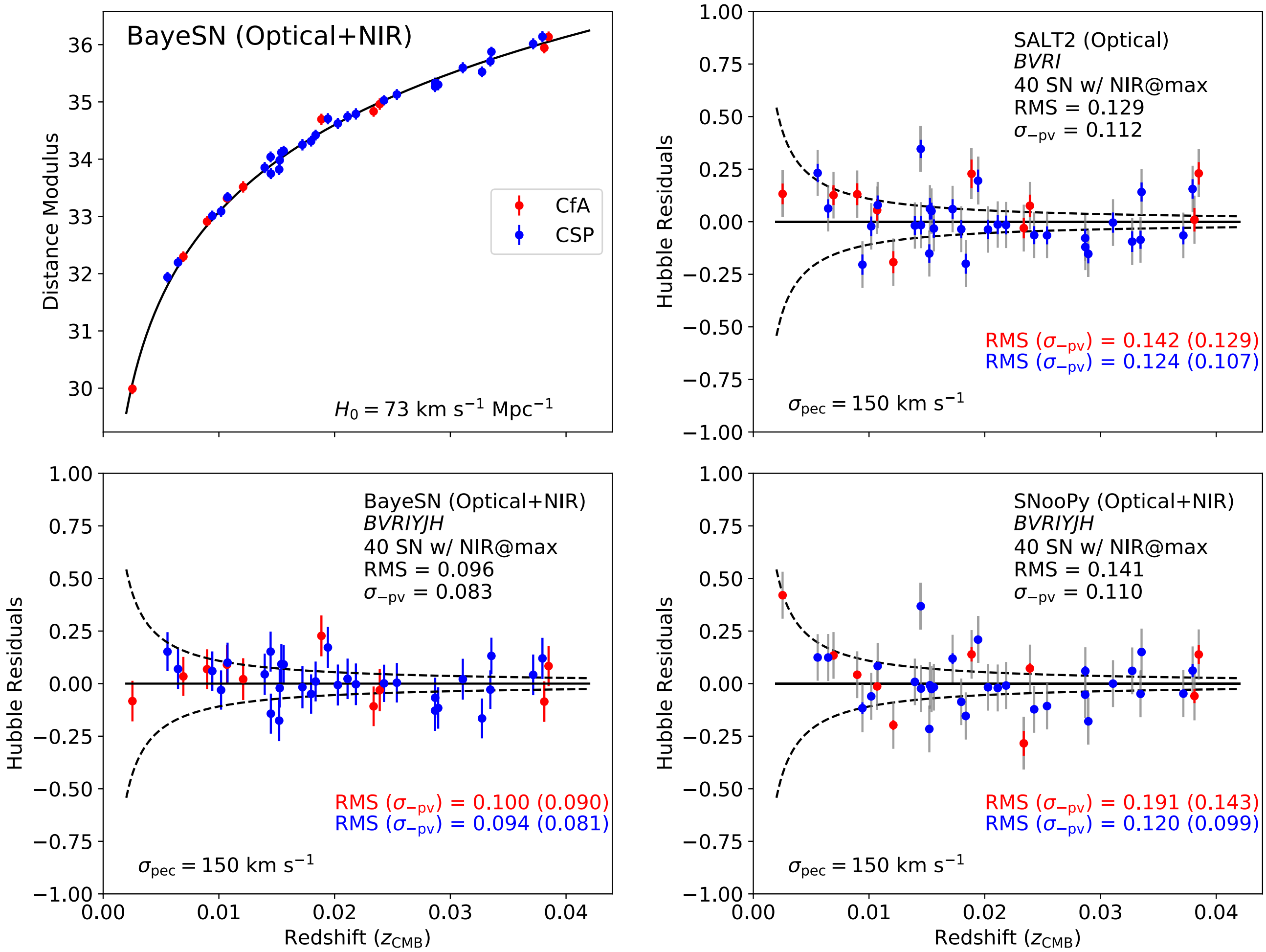

Figure 15. Comparison of Hubble diagrams and Hubble residuals from BAYESN, SNOOPY, and SALT2, applied to the same set of CfA and CSP SNe Ia with NIR data near maximum light. (top left-hand panel) Hubble Diagram of photometric distances obtained by fitting the optical and NIR light curves with BAYESN, compared to the local distance-redshift relation under standard cosmological parameters. (bottom left-hand panel) Hubble residuals for BAYESN. The simple total RMS is 0.096 mag. After removing the expected variance due to peculiar velocity uncertainty (dashed, $\sigma_{\mathrm{pec}} = 150$ km s^{-1}), the remaining dispersion is $\hat{\sigma}_{\text{-pv}} = 0.083$ mag. The distance uncertainties are determined via marginalization accounting for the residual covariance (equation 31). (top right-hand panel) Hubble residuals from SALT2 applied to the optical-only data (*BVRI*) of the same sample. (bottom right-hand panel) Hubble residuals from SNooPy applied to the optical and NIR data of the same sample. For SALT2 and SNooPy, we show two error bars for each SN: one obtained from the light-curve fit uncertainties, and, in grey, those augmented in quadrature with the residual variance needed to make the total $\chi^2_\nu = 1$ for each Hubble diagram. *Y*-band data is only available for CSP objects (c.f. Table 1).

shorter NIR wavelength range, we expect even the NIR extinction corrections, determined from joint fits with the optical, to become more important. Current NIR SN Ia light curves are usually obtained as follow-up observations to optical discoveries and BAYESN is able to also leverage the complementary optical data already obtained to analyse the full wavelength range in an automated, consistent way.

Table 2 summarizes of Hubble diagram dispersions of the other subsets of the SN Ia sample. We find that the addition of the literature sample to the CfA + CSP sample (to constitute All) increases the dispersion slightly in nearly all cases, which is to be expected since these SNe Ia come from more heterogeneous data sources. BAYESN optical + NIR distances are still more precise than SNooPy and SALT2 in the AnyNIR sample, when we do not require NIR measurements near maximum light, but the advantage is reduced. This highlights the importance of obtaining the NIR data near maximum light. On optical-only data (*BVRI*), all three methods perform similarly, with total RMS ≈0.15−0.16 mag.

5.3.2 *Cross-validation*

Cross-validation techniques to test the sensitivity of SN Ia models and their distance estimates to the finite training set have been previously employed by Mandel et al. (2009, 2011) and Blondin, Mandel & Kirshner (2011). These procedures address the double use of the data inherent in resubstitution. We performed 10-fold cross-validation to assess the out-of-training sample distance error. We equally divided the full training set into 10 folds, each with a roughly similar redshift distribution. First, we hold out one fold, and train a new BAYESN SED model on the optical+NIR data of the SNe Ia in the other 9 folds. Then we used the new trained model to estimate the photometric distances of the SNe Ia in the held-out fold, by fitting either their optical + NIR or optical-only light curves. We repeated this procedure 10 times, each time holding out a different fold, training a new model on the complement, and using it to evaluate the photometric distances of the held-out SNe. The Hubble residual summaries of the cross-validation out-of-training

Downloaded from https://academic.oup.com/mnras/article/510/3/3939/6448478 by University of Cambridge user on 14 December 2022

sample photometric distances thus obtained are listed in Table 2 as 'BayeSN-cv'.

In the best case, for the CfA + CSP NIR@max subset, the total RMS of the photometric distances relative to the external distances is 0.108 mag. As expected, this is slightly higher than the RMS training error (0.096 mag) because the cross-validated distance of each SN is obtained using a model trained on a set that excludes that SN. This is an overestimate of the true error of the fully-trained model, since each model under CV is trained on a 10 per cent smaller training set than the full sample. We expect the difference between the training and cross-validation error to narrow as more training data becomes available. Still, the difference between the two numbers is already small (0.012 mag), so it is reasonable to conclude that the typical distance error for similar optical and NIR light curves with peak NIR data is ≈0.10 mag.

A large fraction of our training set SNe Ia were also used in the training sets for both SNooPy (Burns et al. 2011) and SALT2 (Guy et al. 2010). To our knowledge, there has been no equivalent cross-validation analysis, including hold-out and iterative retraining, for these other models. Since we are unable to retrain these other models on partitions or resampled subsets, it is difficult to make equivalent, direct comparisons of these models to our cross-validation results.

Our cross-validation runs demonstrate the capability of our training code to straightforwardly and repeatedly train new models on different SN Ia data sets automatically without human intervention. We will be able to use this modularity to train and compare new BayeSN SED models based on data sets partitioned by survey or by astrophysical classes (e.g. SN Ia host galaxy properties or spectroscopic subclasses) to investigate the statistical and physical differences in the learned SED components and latent variables.

5.4 Application to Foundation SN Ia light curves

The optical and NIR light curves in our training set listed in Table 1 are mainly from the Carnegie Supernova Project and CfA Supernova Program, which typically measured high-quality light curves with relatively frequent time sampling (c.f. Fig. 3). However, most SN Ia light curves used for cosmology are not sampled as well in phase or wavelength. To test our model on SN Ia light curves outside of our training set with more typical sampling, we have fit *griz* light curves obtained by the Foundation Supernova Survey using the Pan-STARRS1 (PS1) telescope (Foley et al. 2018b).

Fig. 16 demonstrates a BayeSN fit to Foundation observations of the Type Ia SN 2016gou / ATLAS16cxr. It shows the well-constrained joint posterior distribution of the parameters obtained from the MCMC fit: the θ_1 coefficient of the 1st FPC, the dust extinction A_V, and the photometric distance μ. Because BayeSN is a model for the continuous SED spanning 0.35 to 1.8 µm, we are able to integrate the model SED under the *griz* PS1 passbands to fit this data, even though these exact passbands were not used in the training phase. Our SED model does not require K-corrections to be computed as pre-processing step to map observer-frame to rest-frame passbands. Notably, the SALT2.4 model cannot properly fit rest-frame z-band due to the wavelength limits of its SED template, and SNooPy lacks a rest-frame z-band light-curve template. However, proper modelling of the rest-frame z-band is important for fully utilizing *griz* data from low-z surveys such as Foundation and the Young Supernova Experiment. In a companion paper, we present a full analysis of the Foundation DR1 data set using our new BayeSN model (Thorp et al. 2021).

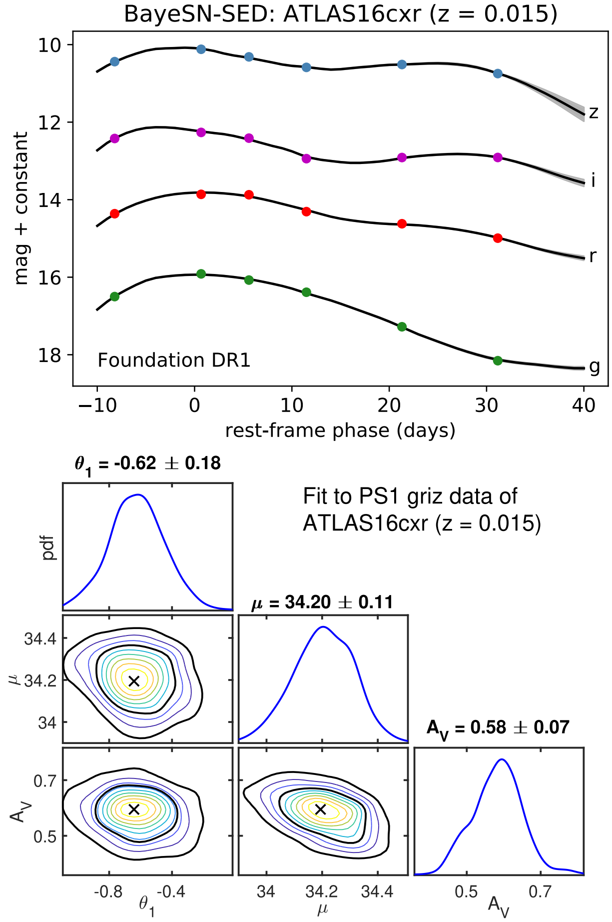

Figure 16. (top) BayeSN light-curve fit of Foundation DR1 *griz* observations of ATLAS16cxr. (bottom) Posterior distribution of BayeSN parameters from the light-curve fit.

6 CONCLUSION

6.1 Improvements over current models

We have constructed a new hierarchical Bayesian model, BayeSN, for SN Ia SEDs from the optical through NIR. This is the first statistical model for continuous SN Ia SEDs designed for fitting observed optical and NIR light curve data, and is crucial for properly analysing NIR observations from current and future SN Ia surveys. Our model is capable of statistically leveraging the powerful properties of SN Ia in the NIR, in particular the narrow dispersion in NIR luminosities at peak, and the much diminished effect of dust in the SN Ia host galaxies. BayeSN jointly leverages the optical and NIR data to constrain the dust extinction A_V and the reddening law R_V more stringently, thereby controlling systematic errors due to the dust correction. BayeSN coherently estimates the covariance of the residual SED functions across time and wavelength, and incorporates them into the dust and distance estimates in a principled, probabilistic manner.

By generalizing the previous hierarchical Bayesian framework of Mandel et al. (2009, 2011) from modelling light curves in fixed discrete rest-frame filters to modelling a continuous SED function in phase and wavelength, we obviate the need for ad-hoc K-correction pre-processing procedures to compute 1-to-1 mappings between observer-frame and rest-frame filters, which is required by SNooPy. Instead, observed data are compared directly against the model fluxes

Downloaded from https://academic.oup.com/mnras/article/510/3/3939/6448478 by University of Cambridge user on 14 December 2022

implied by the redshifted SED model integrated under the observer's passbands. Redshifting effects are thereby incorporated directly into the statistical model.

Furthermore, BAYESN has a number of advantages over the SALT2.4 model conventionally used in cosmological analyses. Whereas the SALT2.4 spectral template has coverage only up to rest-frame 0.9 μm (inclusive of rest-frame *i*-band), our BAYESN SED model extends to 1.8 μm (i.e. through rest-frame *H*-band). The SALT2 model does not internally discern distinct SED components separately describing the effects of SN Ia intrinsic variation versus host galaxy dust extinction. Instead, it uses a single colour law to fit a single apparent colour parameter, effectively confounding the two physically distinct sources of spectral variation.

In contrast, our BAYESN SED model internally encodes the continuous wavelength-dependent host galaxy dust reddening and extinction at the SN Ia SED level, as effects physically distinct from the time-dependent intrinsic components of SED variation. Our model leverages the photometric constraints on the entire continuous SED to determine the dust properties, fit for the intrinsic modes of variation, and coherently weigh the uncertainties and combine information from across phase and wavelength to compute the probability distribution of the photometric distance modulus. With the low- z compilation analysed here, BAYESN can determine the distance moduli for SNe Ia with optical and NIR coverage near maximum light to ≈0.10 mag precision (total RMS), compared to 0.13–0.14 mag using SALT2 or SNooPy on the same SNe Ia. Combining optical and NIR data across the entire phase and wavelength range, we used BAYESN to derive tight constraints on the host galaxy dust law. For this sample with colour excess $E(B-V)_{\rm host} \lesssim 0.4$, we found $R_V = 2.9 \pm 0.2$, consistent with the Milky Way average.

6.2 Applications to current and future data sets

Beyond the data compilation analysed here, our BAYESN SED model will be broadly applicable for analysing optical and NIR SN Ia light-curve data from more recent and current surveys. Forthcoming data from the Carnegie Supernova Project-II (Phillips et al. 2019) will enable us to expand our nearby training set with high-quality optical and NIR light curves of SNe Ia further into the Hubble flow (limiting the impact of peculiar velocity uncertainties). We are using Foundation DR1 *griz* light curves obtained with the well-calibrated Pan-STARRS telescope for training and analysis with BAYESN (Thorp et al. 2021).

BAYESN is important for fully analysing data from recent and ongoing programs that use the *Hubble Space Telescope* to observe SNe Ia in the rest-frame NIR at high-z (RAISIN) and low-z (SIRAH), in conjunction with optical data from ground-based surveys. The ESO VISTA Extragalactic Infrared Legacy Survey (VEILS)[14] recently concluded a time-domain survey that observed SNe Ia in the observer-frame *J*-band up to $z \approx 0.6$, in conjunction with the Dark Energy Survey and the ESO VOILETTE survey in *griz*.

LSST's observer-frame y filter will probe the rest-frame NIR z or y bands to redshifts $z \lesssim 0.3$. The Nancy Grace *Roman Space Telescope* (*RST*)'s wide imaging filters will extend to 2.0 μm (e.g. Hounsell et al. 2018), and thus will overlap with rest-frame H to $z \lesssim 0.4$, J to $z \lesssim 0.7$, and Y to $z \lesssim 1$. The addition of a K-band filter will extend the NIR coverage further (Rubin 2020). BAYESN will be crucial for properly leveraging the full wavelength range of these surveys both to constrain the host galaxy dust properties and to produce optimal distance estimates. It will also be important for fully analysing any potential simultaneous observations of SNe Ia by LSST and RST (e.g. Foley et al. 2018a) or *Euclid* (Rhodes et al. 2017).

6.3 Future analyses and model extensions

Our hierarchical Bayesian SED modelling and inference framework is modular and flexible and will enable us to expand upon the SED model presented here to explore in greater depth various aspects of SNe Ia. In Thorp et al. 2021, we investigate dust distributions by allowing R_V^s to vary for each SN Ia within a population governed by hyperparameters to be inferred, as was done previously by Mandel et al. (2011). We will also be able to test alternative forms of the dust extinction law (e.g. Goobar 2008; Amanullah et al. 2015). We will further probe the statistical properties of the intrinsic SED residuals over phase and wavelength, through the modelling and assessment of additional $K > 2$ functional components and improved estimation of residual covariance.

A further shortcoming of current SN Ia models is the lack of incorporation of astrophysical correlations at the fundamental level of the SED. A 'host mass step' captures an apparent correlation between host galaxy stellar masses and SN Ia optical luminosities controlling for light-curve shape and colour (Kelly et al. 2010; Sullivan et al. 2010; Smith et al. 2020). While the astrophysical nature of this correlation is still under active investigation (Jones et al. 2018; Rigault et al. 2020; Brout & Scolnic 2021; González-Gaitán et al. 2021; Thorp et al. 2021) it is typically addressed simplistically by correcting derived distances, or equivalently splitting the scalar absolute magnitude constant in equation (1), according to the host mass. The correlation of SN Ia NIR absolute magnitudes with host mass has been investigated recently by Burns et al. (2018), Ponder et al. (2020), Uddin et al. (2020), and Johansson et al. (2021). Our current low-z training set has roughly an average log host mass $\log_{10}(M_*/M_\odot) \approx 10.3$ and approximately 80 per cent lie in the 'high-mass' category $\log_{10}(M_*/M_\odot) > 10$. In future work, we will apply BAYESN to a broader set of SNe Ia to conduct a Bayesian statistical analysis of this effect.

Similarly, SN Ia ejecta velocities, measured from spectral lines, are correlated with SN Ia intrinsic colour, and can be used to gain leverage on dust estimation and improve the accuracy of distances (Foley & Kasen 2011; Foley 2012; Mandel et al. 2014). Recently, Siebert et al. (2020) found correlations between ejecta velocity and SALT2 Hubble residuals. However, these astrophysical correlations should be accounted for at the fundamental physical level of the SN Ia SED functions, rather than by correcting derived distances. In future work, we will expand our BAYESN framework to explore, estimate, and incorporate the impact of these astrophysical effects on the full SED function $S(t, \lambda_r)$ in a coherent statistical model. We will do this by adding functional regression terms proportional to $f(M_*)\, W_{M_*}(t, \lambda_r)$ or $f(v)\, W_v(t, \lambda_r)$ to our SED model (equation 12), and by modelling potential correlations with host dust population parameters.

In this work, we have leveraged joint optical and NIR broad-band photometry of SNe Ia to learn the statistical properties of the latent intrinsic and dust components of SN Ia SEDs, while using the Hsiao (2009) template as a baseline 'skeleton' to model spectral features at finer resolutions than the typical passband. Some of the residual SED covariance and scatter in the Hubble residuals indeed may be caused by per-SN variation in spectral features on wavelength scales much smaller than the typical filter. In future development, we will

[14]https://people.ast.cam.ac.uk/~mbanerji/VEILS/index.html

Downloaded from https://academic.oup.com/mnras/article/510/3/3939/6448478 by University of Cambridge user on 14 December 2022

increase the wavelength resolution of our model, so that we can train simultaneously on spectroscopic sequences and photometric light curves of SNe Ia to improve the latent SED model. We will be able to leverage databases of optical spectra (Blondin et al. 2012; Silverman et al. 2012; Folatelli et al. 2013; Siebert et al. 2019), as well as forthcoming ground-based NIR spectra from the Magellan FIRE instrument obtained by the CSP-II and CfA Supernova Group (Hsiao et al. 2019), and space-based NIR spectra from the ongoing *Hubble Space Telescope* SIRAH program (GO-15889).

In future work, our probabilistic inference framework can be extended to coherently estimate SN Ia SED components and cosmological parameters together while accounting for population drift of intrinsic and dust parameters and survey selection effects. Our BayeSN SED model will serve as the centrepiece of a fully hierarchical Bayesian statistical framework for principled supernova cosmology analysis.

ACKNOWLEDGEMENTS

We dedicate this paper to our friend and colleague, Dr. Andrew S. Friedman (1979–2020), who passed away during preparation of the manuscript. This work builds upon ASF's observations and research on SNe Ia in the NIR. KSM, GN, and AA had the privilege to work with ASF as members of the CfA Supernova Group. Andy's absence is keenly felt.

We thank the anonymous referee for a thorough and thoughtful review of the manuscript. We thank David Jones, Robert Kirshner, and members of the Nancy Grace Roman Space Telescope Science Investigation Team for supernova cosmology led by Ryan Foley for useful discussions. 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. This work was performed in part at Aspen Center for Physics, which is supported by National Science Foundation grant PHY-1607611. The participation of KSM at the Aspen Center for Physics was supported by the Simons Foundation. 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 published data sets used are available from the original sources listed in Table 1. 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., Goobar A., 2011, ApJ, 735, 20
Amanullah R. et al., 2014, ApJ, 788, L21
Amanullah R. et al., 2015, MNRAS, 453, 3300
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
Barbary K. et al., 2016, SNCosmo: Python Library for Supernova Cosmology, record ascl:1611.017
Barnard J., McCulloch R., Meng X.-L., 2000, Stat. Sinica, 10, 1281
Barone-Nugent R. L. et al., 2012, MNRAS, 425, 1007
Betancourt M., 2017, preprint (arXiv:1701.02434)
Betoule M. et al., 2014, A&A, 568, A22
Bishop C. M., 2006, Pattern Recognition and Machine Learning. Springer, New York
Bishop C., 1999, in Larochelle H., Ranzato M., Hadsell R., Balcan M.F., Lin H., eds, Advances in Neural Information Processing Systems. MIT Press, p. 382
Blondin S., Mandel K. S., Kirshner R. P., 2011, A&A, 526, A81
Blondin S. et al., 2012, AJ, 143, 126
Branch D., Tammann G. A., 1992, ARA&A, 30, 359
Brout D., Scolnic D., 2021, ApJ, 909, 26
Brout D. et al., 2019, ApJ, 874, 150
Bulla M., Goobar A., Amanullah R., Feindt U., Ferretti R., 2018a, MNRAS, 473, 1918
Bulla M., Goobar A., Dhawan S., 2018b, MNRAS, 479, 3663
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
Carpenter B. et al., 2017, J. Stat. Softw., 76, 1
Carrick J., Turnbull S. J., Lavaux G., Hudson M. J., 2015, MNRAS, 450, 317
Chotard N. et al., 2011, A&A, 529, L4
Conley A., Carlberg R. G., Guy J., Howell D. A., Jha S., Riess A. G., Sullivan M., 2007, ApJ, 664, L13
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)
Dhawan S., Leibundgut B., Spyromilio J., Maguire K., 2015, MNRAS, 448, 1345
Dhawan S., Jha S. W., Leibundgut B., 2018, A&A, 609, A72
Dhawan S. et al., 2020, MNRAS, 491, 2639
Elias-Rosa N. et al., 2006, MNRAS, 369, 1880
Elias-Rosa N. et al., 2008, MNRAS, 384, 107
Fitzpatrick E. L., 1999, PASP, 111, 63
Folatelli G. et al., 2010, AJ, 139, 120
Folatelli G. et al., 2013, ApJ, 773, 53
Foley R. J., 2012, ApJ, 748, 127
Foley R. J., Kasen D., 2011, ApJ, 729, 55
Foley R. J., Narayan G., Challis P. J., Filippenko A. V., Kirshner R. P., Silverman J. M., Steele T. N., 2010, ApJ, 708, 1748
Foley R. J. et al., 2018a, preprint (arXiv:1812.00514)
Foley R. J. et al., 2018b, MNRAS, 475, 193
Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, PASP, 125, 306
Freedman W. L. et al., 2009, ApJ, 704, 1036
Friedman A. S. et al., 2015, ApJS, 220, 9
Fukugita M., Ichikawa T., Gunn J. E., Doi M., Shimasaku K., Schneider D. P., 1996, AJ, 111, 1748
Ganeshalingam M. et al., 2010, ApJS, 190, 418
Gelman A., 2006, Bayesian Anal., 1, 515
Gelman A., Rubin D. B., 1992, Stat. Sci., 7, 457
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
Goldstein D. A., Nugent P. E., Kasen D. N., Collett T. E., 2018, ApJ, 855, 22
González-Gaitán S., de Jaeger T., Galbany L., Mourão A., Paulina-Afonso A., Filippenko A. V., 2021, MNRAS, 508, 4656
Goobar A., 2008, ApJ, 686, L103
Graziani R. et al., 2020, preprint (arXiv:2001.09095)
Guy J., Astier P., Nobili S., Regnault N., Pain R., 2005, A&A, 443, 781
Guy J. et al., 2007, A&A, 466, 11
Guy J. et al., 2010, A&A, 523, A7
Hicken M. et al., 2009, ApJ, 700, 331

Downloaded from https://academic.oup.com/mnras/article/510/3/3939/6448478 by University of Cambridge user on 14 December 2022

Hicken M. et al., 2012, ApJS, 200, 12
Hinton S. R. et al., 2019, ApJ, 876, 15
Hoffman M. D., Gelman A., 2014, J. Mach. Learn. Res., 15, 1593
Hounsell R. et al., 2018, ApJ, 867, 23
Howlett C., Robotham A. S. G., Lagos C. D. P., Kim A. G., 2017, ApJ, 847, 128
Hsiao E. Y., 2009, Ph.D. 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
Hsiao E. Y. et al., 2019, PASP, 131, 014002
Huang X. et al., 2017, ApJ, 836, 157
Huterer D., Shafer D. L., Scolnic D. M., Schmidt F., 2017, J. Cosmol. Astropart. Phys., 2017, 015
Ivezić Ž. et al., 2019, ApJ, 873, 111
James F., Roos M., 1975, Comput. Phys. Commun., 10, 343
Jha S. et al., 1999, ApJS, 125, 73
Jha S., Riess A. G., Kirshner R. P., 2007, ApJ, 659, 122
Johansson J. et al., 2017, MNRAS, 466, 3442
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
Kasen D., 2006, ApJ, 649, 939
Kattner S. et al., 2012, PASP, 124, 114
Kelly P. L., Hicken M., Burke D. L., Mandel K. S., Kirshner R. P., 2010, ApJ, 715, 743
Kessler R. et al., 2009a, PASP, 121, 1028
Kessler R. et al., 2009b, ApJS, 185, 32
Kessler R. et al., 2013, ApJ, 764, 48
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., Phillips M. M., Suntzeff N. B., 2004c, ApJ, 602, L81
Krisciunas K. et al., 2007, AJ, 133, 58
Krisciunas K. et al., 2017, AJ, 154, 211
Léget P. F. et al., 2020, A&A, 636, A46
Leloudas G. et al., 2009, A&A, 505, 265
Lewandowski D., Kurowicka D., Joe H., 2009, J. Multivariate Anal., 100, 1989
Loredo T. J., Hendry M. A., 2010, in Hobson M. et al., ed., Bayesian Methods in Cosmology. Cambridge Univ. Press, Cambridge, p. 245
Loredo T. J., Hendry M. A., 2019, preprint (arXiv:1911.12337)
Macaulay E. et al., 2020, MNRAS, 496, 4051
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
March M. C., Trotta R., Berkes P., Starkman G. D., Vaudrevange P. M., 2011, MNRAS, 418, 2308
Marion G. H., Höflich P., Gerardy C. L., Vacca W. D., Wheeler J. C., Robinson E. L., 2009, AJ, 138, 727
Mosher J. et al., 2014, ApJ, 793, 16
Murphy K. P., 2022, Probabilistic Machine Learning: An introduction. MIT Press, Cambridge, MA. Available at: probml.ai
Narayan G. et al., 2016, ApJ, 822, 67
Narayan G. et al., 2019, ApJS, 241, 20
Neal R. M., 2011, Brooks S., Gelman A., Jones G., Meng X.-L., eds, Handbook of Markov Chain Monte Carlo, vol. 54. CRC Press, Boca Raton, FL, p. 113
Nolan T. H., Goldsmith J., Ruppert D., 2021, preprint (arXiv:2104.00645)
Perlmutter S. et al., 1999, ApJ, 517, 565
Phillips M. M., 1993, ApJ, 413, L105
Phillips M. M., 2012, Publ. Astron. Soc. Aust., 29, 434
Phillips M. M., Lira P., Suntzeff N. B., Schommer R. A., Hamuy M., Maza J., 1999, AJ, 118, 1766
Phillips M. M. et al., 2013, ApJ, 779, 38
Phillips M. M. et al., 2019, PASP, 131, 014001
Pierel J. D. R. et al., 2018, PASP, 130, 114504
Pignata G. et al., 2008, MNRAS, 388, 971
Planck Collaboration, 2020, A&A, 641, A6
Polson N. G., Scott J. G., 2012, Bayesian Anal., 7, 887
Ponder K. A., Wood-Vasey W. M., Weyant A., Barton N. T., Galbany L., Garnavich P., Matheson T., 2020, preprint (arXiv:2006.13803)
Prieto J. L., Rest A., Suntzeff N. B., 2006, ApJ, 647, 501
Ramsay J., Silverman B., 2005, Functional Data Analysis, Springer Series in Statistics. Springer, Berlin
Rasmussen C. E., Williams C. K. I., 2005, Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning). The MIT Press, Cambridge, MA
Regnault N. et al., 2015, A&A, 581, A45
Rhodes J. et al., 2017, ApJS, 233, 21
Riess A. G., Press W. H., Kirshner R. P., 1996a, ApJ, 473, 88
Riess A. G., Press W. H., Kirshner R. P., 1996b, ApJ, 473, 588
Riess A. G. et al., 1998, AJ, 116, 1009
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., 2020, A&A, 644, A176
Rose B. M. et al., 2020, ApJ, 890, 60
Roweis S., 1998, in Jordan M. I., Kearns M. J., Solla S. A., eds, Advances in Neural Information Processing Systems 10. MIT Press, Cambridge, MA, p. 626
Rubin D., 2020, preprint (arXiv:2010.15112)
Rubin D. et al., 2015, ApJ, 813, 137
Saunders C. et al., 2018, ApJ, 869, 167
Schlafly E. F., Finkbeiner D. P., 2011, ApJ, 737, 103
Scolnic D. M., Riess A. G., Foley R. J., Rest A., Rodney S. A., Brout D. J., Jones D. O., 2014, ApJ, 780, 37
Scolnic D. et al., 2015, ApJ, 815, 117
Scolnic D. M. et al., 2018, ApJ, 859, 101
Shariff H., Dhawan S., Jiao X., Leibundgut B., Trotta R., van Dyk D. A., 2016a, MNRAS, 463, 4311
Shariff H., Jiao X., Trotta R., van Dyk D. A., 2016b, ApJ, 827, 1
Siebert M. R. et al., 2019, MNRAS, 486, 5785
Siebert M. R., Foley R. J., Jones D. O., Davis K. W., 2020, MNRAS, 493, 5713
Silverman J. M. et al., 2012, MNRAS, 425, 1789
Smith M. et al., 2020, MNRAS, 494, 4426
Stan Development Team, 2021, Stan Modelling Language Users Guide and Reference Manual v.2.27. Available at: https://mc-stan.org
Stanishev V. et al., 2007, A&A, 469, 645
Stanishev V. et al., 2018, A&A, 615, A45
Stritzinger M., Suntzeff N. B., Hamuy M., Challis P., Demarco R., Germany L., Soderberg A. M., 2005, PASP, 117, 810
Suarez A. J., Ghosal S., 2017, Bayesian Anal., 12, 311
Sullivan M. et al., 2010, MNRAS, 406, 782
Thorp S., Mandel K. S., Jones D. O., Ward S. M., Narayan G., 2021, MNRAS, 508, 4310
Tipping M. E., Bishop C. M., 1999, J. R. Stat. Soc. B, 61, 611
Tripp R., 1998, A&A, 331, 815
Tripp R., Branch D., 1999, ApJ, 525, 209
Uddin S. A. et al., 2020, ApJ, 901, 143
Valentini G. et al., 2003, ApJ, 595, 779
van der Linde A., 2008, Comput. Stat. Data Anal., 53, 517
Wang L., 2005, ApJ, 635, L33
Wang X. et al., 2008, ApJ, 675, 626
Wood-Vasey W. M. et al., 2008, ApJ, 689, 377

APPENDIX A: BAYESIAN FUNCTIONAL PCA

Classical PCA is an oft-used procedure for dimensionality reduction of multivariate data $\{\boldsymbol{x}_i \in \mathbb{R}^D\}_{i=1}^N$ by finding the leading L eigenvectors of the sample covariance matrix that explain the most variance

Downloaded from https://academic.oup.com/mnras/article/510/3/3939/6448478 by University of Cambridge user on 14 December 2022

(eigenvalues) in the data. However, it does not itself explicitly model measurement error in the data nor quantify uncertainties. Generalizing PCA with a probabilistic formulation, Tipping & Bishop (1999) developed probabilistic PCA as a linear Gaussian latent variable model with an explicit likelihood function and residual (noise) term. For each individual i,

$$\boldsymbol{\theta}_i \sim N(\mathbf{0}, \boldsymbol{I}) \tag{A1}$$

$$\boldsymbol{\epsilon}_i \sim N(\mathbf{0}, \sigma^2 \boldsymbol{I}) \tag{A2}$$

$$\boldsymbol{x}_i = \boldsymbol{W}\boldsymbol{\theta}_i + \boldsymbol{\epsilon}_i, \tag{A3}$$

where $\boldsymbol{\theta}_i \in \mathbb{R}^L$ is the score vector, $\boldsymbol{\epsilon}_i \sim N(0, \sigma^2 \boldsymbol{I})$ is a Gaussian noise vector, and $\boldsymbol{W} \in \mathbb{R}^{D\times L}$ contains L orthonormal column vectors, and σ^2 models an isotropic residual variance. This probabilistic model has a marginal likelihood $\boldsymbol{x}_i \sim N(\mathbf{0}, \boldsymbol{W}\boldsymbol{W}^T + \sigma^2\boldsymbol{I})$. They proved that the columns of PPCA's maximum-likelihood solution $\hat{\boldsymbol{W}}_{\rm mle}$ recovers the classical principal component vectors in the limit of $\sigma \rightarrow 0$. However, a non-zero σ^2 enables us to explicitly model the residual variation. (See also Bishop 2006; Murphy 2022).

Bishop (1999) generalized this further as Bayesian PCA by introducing priors over the components $\boldsymbol{W}$ and σ^2 to define the posterior over these unknowns. In BAYESN, we embed Bayesian PCA within the overall hierarchical framework. The analogous equations to equations (A1)–(A3) are equations (19), (20), and (11) in our model. We generalize the isotropic noise covariance from $\sigma^2\boldsymbol{I}$ to $\boldsymbol{\Sigma}_\epsilon$ to capture correlated residual structure (Section 5.2.2) and introduce hyperpriors on the covariance matrix through equations (24) and (25). Furthermore, since we use this structure to describe the joint distribution of spline knots defining continuous SED surfaces (Section 2.4), it also models the functional principal components and the residual functions over time and wavelength. Hence, embedding this Bayesian FPCA in a hierarchical Bayesian framework enables us to model the distribution of SN SEDs in terms of functional principal components and scores while simultaneously accounting for other physical factors (e.g. dust) and their uncertainties, while expressing the light-curve data as noisy estimates of functionals of the latent SED sampled irregularly in time. By sampling the global posterior equation (29), we can then coherently estimate all these components and quantify their joint uncertainties. Bayesian formulations of FPCA have also been described by, e.g. van der Linde (2008), Suarez & Ghosal (2017), Nolan, Goldsmith & Ruppert (2021).

APPENDIX B: THE W_2 SED COMPONENT

We describe the second intrinsic functional principal component $W_2(t, \lambda)$ that is learned when we train the model with $K = 2$. This component can be viewed as the first functional PC of the intrinsic covariance of the residual functions $\epsilon_s(t, \lambda)$ under the $K = 1$ model. In the $K = 2$ model, we pull out this secondary mode of variation and parametrize its effect through the coefficient θ_2.

In Fig. B1, we show the effect of $W_2(t, \lambda)$ on the intrinsic ($A_V = 0$) absolute light curves obtained via integration of the SED model with θ_2 varying between the mean value and $\pm 1\sigma$. This component captures some overall luminosity variation in the optical B and V bands, while modulating the relative amplitudes of the first peak, trough, and second peak in the NIR bands. Unlike $W_1(t, \lambda)$, the second FPC does not significantly change the timing of the second NIR peak, except slightly in Y-band. Fig. B2 illustrates the effect of $W_2(t, \lambda)$ on the intrinsic optical and optical-NIR colours curves. This

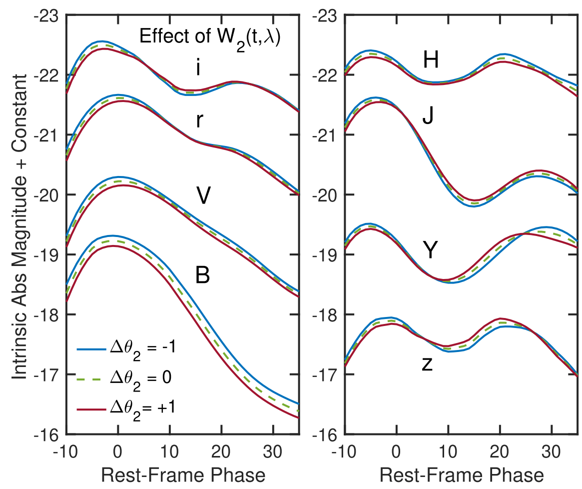

Figure B1. Intrinsic variation in optical and NIR light curves captured by the second functional component $W_2(t, \lambda)$. We fix $\theta_1 = A_V = 0$ and vary the value of θ_2 by $\bar{\theta}_2 \pm 1\sigma$. This component captures luminosity variation in the optical that appears to be correlated with the relative amplitudes of the NIR trough and second peak.

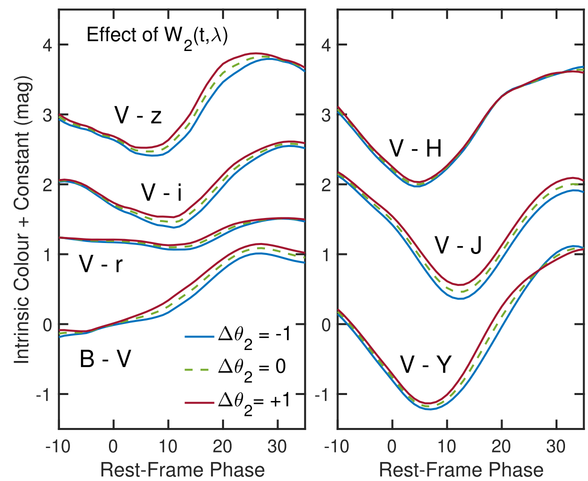

Figure B2. Variation in optical and NIR intrinsic colour curves captured by the second functional component $W_2(t, \lambda)$. We fix $\theta_1 = A_V = 0$ and vary the value of θ_2 by $\bar{\theta}_2 \pm 1\sigma$. This component captures intrinsic colour variation in the post-maximum phases at $t \approx 10$ to 30 d.

component captures variation in the post-peak colours from 5 to 25 rest-frame days in phase.

We did not use the $K = 2$ model including the $W_2(t, \lambda)$ component in the main analysis of the paper, because it did not significantly improve the precision of distances in the Hubble diagram with the current data set, compared to the $K = 1$ model. In future work with larger data sets, we will further investigate higher order functional principal components.

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

Downloaded from https://academic.oup.com/mnras/article/510/3/3939/6448478 by University of Cambridge user on 14 December 2022