Type Ia Supernova Light Curve Inference: Hierarchical Models in the Optical and Near-infrared

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. Submit a story.

Accessibility

TYPE Ia SUPERNOVA LIGHT CURVE INFERENCE: HIERARCHICAL MODELS IN THE OPTICAL AND NEAR-INFRARED

Kaisey S. Mandel, Gautham Narayan, and Robert P. Kirshner
Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138, USA; kmandel@cfa.harvard.edu

ABSTRACT

We have constructed a comprehensive statistical model for Type Ia supernova (SN Ia) light curves spanning optical through near-infrared (NIR) data. A hierarchical framework coherently models multiple random and uncertain effects, including intrinsic supernova (SN) light curve covariances, dust extinction and reddening, and distances. An improved BayeSN Markov Chain Monte Carlo code computes probabilistic inferences for the hierarchical model by sampling the global probability density of parameters describing individual SNe and the population. We have applied this hierarchical model to optical and NIR data of 127 SNe Ia from PAIRITEL, CfA3, Carnegie Supernova Project, and the literature. We find an apparent population correlation between the host galaxy extinction A_V and the ratio of total-to-selective dust absorption R_V. For SNe with low dust extinction, $A_V \lesssim 0.4$, we find $R_V \approx 2.5$–2.9, while at high extinctions, $A_V \gtrsim 1$, low values of $R_V < 2$ are favored. The NIR luminosities are excellent standard candles and are less sensitive to dust extinction. They exhibit low correlation with optical peak luminosities, and thus provide independent information on distances. The combination of NIR and optical data constrains the dust extinction and improves the predictive precision of individual SN Ia distances by about 60%. Using cross-validation, we estimate an rms distance modulus prediction error of 0.11 mag for SNe with optical and NIR data versus 0.15 mag for SNe with optical data alone. Continued study of SNe Ia in the NIR is important for improving their utility as precise and accurate cosmological distance indicators.

Key words: distance scale – methods: statistical – supernovae: general

Online-only material: color figures, machine-readable tables

1. INTRODUCTION

Type Ia supernova (SN Ia) rest-frame optical light curves have been of great utility for measuring fundamental quantities of the universe. As standardizable candles, they were critical to the detection of cosmic acceleration (Riess et al. 1998; Perlmutter et al. 1999). SNe Ia have been used to constrain the equation-of-state parameter w of dark energy (Garnavich et al. 1998), and recent efforts have measured w to ~10%, (Wood-Vasey et al. 2007; Astier et al. 2006; Kowalski et al. 2008; Hicken et al. 2009b; Kessler et al. 2009; Freedman et al. 2009; Amanullah et al. 2010). SNe Ia have also been used to establish the extragalactic distance scale and measure the Hubble constant (Freedman et al. 2001; Jha et al. 1999; Riess et al. 2005, 2009a, 2009b).

SN Ia distance indicators exploit empirical relations between peak optical luminosities of SNe Ia and distance-independent measures such as light curve shape observed in the sample of nearby low-z SNe Ia (Hamuy et al. 1996a; Riess et al. 1999; Jha et al. 2006; Hicken et al. 2009a). Methods include $\Delta m_{15}(B)$ (Phillips 1993; Hamuy et al. 1996b; Phillips et al. 1999; Prieto et al. 2006), multicolor light curve shapes (MLCS; Riess et al. 1996a, 1998; Jha et al. 2007), "stretch" (Goldhaber et al. 2001), CMAGIC (Wang et al. 2003), SALT (Guy et al. 2005, 2007), and SiFTO (Conley et al. 2008). One of the largest systematic uncertainties limiting the precision of distance estimates from rest-frame optical light curves is dust extinction in the host galaxy and the confounding of dust reddening with the intrinsic color variations of SNe Ia (Conley et al. 2007). Current approaches differ conceptually and practically on how apparent colors, intrinsic colors, and dust effects are modeled. While most methods make use of the optical-luminosity–light-curve-width correlation, some methods, such as MLCS (Riess

et al. 1996a, 1998; Jha et al. 2007), attempt to separately model the intrinsic colors of the SN Ia and host galaxy dust reddening and extinction, whereas others model both effects with a single factor (e.g., SALT2; Guy et al. 2007).

Early observations of SNe Ia in the infrared were made by Kirshner et al. (1973), Elias et al. (1981), Elias et al. (1985), Frogel et al. (1987), and Graham et al. (1988). Observations of nearby SNe Ia in the NIR revealed that the peak near-infrared luminosities of SNe Ia have a dispersion smaller than 0.20 mag (Elias et al. 1985; Meikle 2000; Krisciunas et al. 2004a, 2004c). Wood-Vasey et al. (2008, hereafter WV08) compiled a sample of NIR SN Ia observations taken with the Peters Automated InfraRed Imaging TELescope (PAIRITEL; Bloom et al. 2006). They found that the H-band peak absolute magnitude had small scatter $\sigma(M_H) \approx 0.15$ mag and could provide distance estimates competitive with those derived from optical light curve shapes. The effect of dust extinction is significantly diminished at NIR wavelengths, relative to the optical. The combination of optical and NIR observations of SN Ia light curves could lead to even better estimates of SN Ia distances (Krisciunas et al. 2007).

A significant source of puzzlement in the analysis of SN Ia light curves is the nature of the apparent color and brightness variations among supernovae (SNe), which are comprised of color and luminosity variations intrinsic to the SN Ia population, and also random reddening and extinction by dust in the host galaxies of SNe Ia. The function of dust absorption over wavelength is typically parameterized by the ratio of total to selective extinction, $R_V = A_V/(A_B - A_V)$. This ratio has an average value of 3.1 for interstellar dust in the Milky Way Galaxy, although it can vary between 2.1 and 5.8 (Draine 2003). Studies of external galaxies have found similar extinction curves; for example, Finkelman et al. (2008, 2010) found average values of $R_V \approx 2.8$. Early studies of SNe Ia found

values of $R_V < 1$ (c.f. Branch & Tammann 1992 for a review), although these analyses did not take into account relationships between the luminosity, color, and light curve shape of the events. Using the first version of the MLCS (Riess et al. 1996a) to model these relationships for SN Ia optical light curves, Riess et al. (1996b) analyzed the colors of 20 nearby SNe Ia and found $R_V = 2.6 \pm 0.3$, consistent with the Milky Way average. Tripp (1998) and Tripp & Branch (1999) found $R_V \approx 1$, but they modeled intrinsic color and dust reddening as a single factor. More recently, Conley et al. (2007) found that the relation between SN Ia optical luminosity and apparent color, controlling for light curve shape, required a low value of $R_V \approx 1$–1.7, if the total color variation is interpreted as interstellar dust in the host galaxy. An analysis of the color curves of SNe Ia by Nobili & Goobar (2008) found $R_V \approx 1$–1.7. Hicken et al. (2009b) found that a dust absorption profile with $R_V = 1.7$ was favored using the MLCS2k2 model for the CfA3 sample (Hicken et al. 2009a). These studies assumed that a universal color or dust absorption profile applied to all SNe Ia. Recently, Wang et al. (2009a) separately fit $R_V \approx 1.6$ for a subset of SNe Ia with high ejecta velocities, and $R_V \approx 2.4$ for a subset with normal ejecta velocities. However, Foley & Kasen (2011) find $R_V \approx 2.5$ for both subsets if the reddest SNe are excluded. For individual, highly extinguished SNe with multi-wavelength coverage, R_V can be fit precisely and values of 1.5–1.8 have been reported (Krisciunas et al. 2007; Elias-Rosa et al. 2006, 2008; Wang et al. 2008). It has been suggested that low R_V values could result from scattering by circumstellar dust in the local environment of the SN (Wang 2005; Goobar 2008).

Contreras et al. (2010) have recently presented the initial sample of nearby SN Ia light curves observed by the Carnegie Supernova Project (CSP), a subset of which were observed in the optical and near-infrared. Folatelli et al. (2010) compared the apparent colors of the SN to a subset suspected of having no contamination by dust, and found a dust law slope $R_V \approx 1.7$ best fit the SN sample, assuming a global value of R_V. However, when the two reddest SNe were removed from the sample, this global value changed to $R_V \approx 3.2$. When minimizing dispersion in the Hubble diagram, they found $R_V \approx 1$–2, both with or without the reddest SN. The different and unusual values of R_V in the literature are problematic for the proper analysis and interpretation of SN Ia observables and for cosmological applications.

Mandel et al. (2009) presented a hierarchical Bayesian approach to constructing probabilistic models for SN Ia light curves. This strategy was applied to the modeling of the extant near-infrared (NIR) light curves of SNe Ia from PAIRITEL (WV08) and the literature in the JHK_s passbands. Using a Markov Chain Monte Carlo (MCMC) algorithm (BayeSN) designed specifically for hierarchical SN Ia light curve models, they computed coherent probabilistic inferences for individual SNe and the population, taking into account multiple sources of randomness and uncertainty. It was found that the variances of the peak absolute magnitudes were small, particularly in the H band: $\sigma(M_H) \approx 0.11 \pm 0.03$ mag. Since observations in the NIR passbands are insensitive to dust extinction, the estimation of host galaxy dust extinction was omitted from the NIR-only analysis in that paper.

In this paper, we expand upon the hierarchical modeling approach for SNe Ia first described by Mandel et al. (2009), and apply it to statistical modeling of SN Ia light curves in both the optical and near-infrared, including the effects of host galaxy dust. We describe a general mathematical representation of SN Ia light curves in terms of multiple decline rates over phase in different passbands. This differential decline rates representation is employed within a hierarchical model incorporating multiple random effects: measurement error, peculiar velocities, dust extinction, and intrinsic variation. In particular, the intrinsic correlation structure of the light curves over phase and over wavelengths spanning optical to near-infrared is explicitly modeled and estimated. We also model the joint distribution of dust extinction and R_V. To estimate the parameters of individual SNe and the characteristics of the host galaxy dust distribution and intrinsic SN Ia populations, we have implemented a new BayeSN MCMC algorithm. This new algorithm incorporates enhancements to improve efficiency and convergence in the global parameter space.

We apply the hierarchical model to optical (*BVRI*) and NIR (*JH*) data of nearby SN Ia light curves from the PAIRITEL, CfA3, CSP samples, and the literature. We present inferences about the host galaxy dust population and the correlation structure of the intrinsic SN Ia light curve population. To check the fit of the hierarchical model to the sample of data, we compare posterior predictive replications from the model to the observed parameter distributions of colors, magnitudes, and light curve shapes. We quantify the utility of including NIR observations of SNe Ia for improving estimates of host galaxy dust properties and distance predictions. Analyzing optical and NIR light curves within the same model, we demonstrate that distance moduli to SNe Ia observed with optical and near-infrared light curve data can be predicted more accurately and precisely (rms = 0.11 mag) than with optical data alone (rms = 0.15 mag).

This paper is organized as follows. In Section 2, we describe our hierarchical Bayesian approach to constructing statistical models for SN Ia light curves. In Section 3, we outline a new version of the BayeSN algorithm for computing probabilistic inferences with the hierarchical model using the SN data. In Section 4, the application of the hierarchical model to nearby SN Ia data in the optical and near-infrared is described, and posterior inferences about the dust and SN Ia light curve populations are summarized in Section 5. In Section 6, we describe checks on our model inferences. In Section 7, we employ cross-validation (CV) to construct a Hubble diagram of predicted distances to SNe Ia, and demonstrate the advantages of including the NIR data for making more precise inferences about dust extinction and luminosity distances. We conclude in Section 8.

In Appendix A, we describe a non-parametric representation for individual light curves used by our model. In Appendix B, we describe the calculation of *K*-correction and Milky Way extinction effects based on spectral templates. Hyperprior distributions are stated in Appendix C. Mathematical details of the new BayeSN algorithm are given in Appendix D.

2. STATISTICAL MODELS FOR SN Ia LIGHT CURVES

Inferences in SN cosmology are based on statistical models built from empirical data. The application of statistical models for SNe Ia to constraining the cosmological parameters has focused on using information in the apparent, optical light curves of SNe Ia to infer their peak luminosities and to estimate the luminosity distances. In particular, these models capture empirical light curve width and color correlations with luminosity that allow SNe Ia to be used as "standardizable candles" (Phillips 1993; Hamuy et al. 1996b; Phillips et al. 1999; Riess et al. 1996a, 1998; Goldhaber et al. 2001; Prieto et al. 2006;

Jha et al. 2007; Guy et al. 2005, 2007). Recent work has explored the utility of using spectral ratios or characteristics of spectral lines correlated with the luminosity to predict distances (Bailey et al. 2009; Blondin et al. 2011). These methods show promise, but need to be validated on larger spectroscopic samples to be competitive with light curve methods.

The observed SN Ia light curve data are the result of the combination of multiple random effects. Different "normal" SNe Ia can have intrinsically distinct absolute light curves, peak luminosities, and colors. Each event can be extinguished and reddened by a different, and random, amount of host galaxy dust along the line of sight, and this dust may have different extinction laws as a function of wavelength for each event. Before the light curve is recorded by an astronomer, it is subject to redshift effects, absorption due to Milky Way dust, and measurement error. The measured redshift of each SN host galaxy is different from the cosmological redshift by a random peculiar velocity. Sensible statistical models for SN Ia light curves must account for these multiple random effects in the data.

In the absence of the other effects, the apparent colors (e.g., $B - V$) of SNe Ia at any phase are the sum of random intrinsic colors (e.g., $(B - V)_{int} \equiv M_B - M_V$) and random amounts of reddening by host galaxy dust ($A_B - A_V \equiv E(B - V)_{dust}$). Hence, the joint distribution of the apparent colors over different wavelength ranges and at different phases is the convolution of the intrinsic color distribution and the dust reddening distribution. Similarly, when the SN distances are known, the extinguished absolute magnitudes (e.g., $V(t) - \mu$) at different wavelengths and phases are the sum of random intrinsic absolute light curves (e.g., $M_V(t)$) and random amounts of dust extinction (e.g., A_V) over wavelength. The joint distribution of extinguished absolute light curves is the convolution of the intrinsic absolute light curve distribution and the dust extinction distribution. Since dust extinction only makes objects appear dimmer and redder, the convolution with the dust distribution distorts the intrinsic distribution into the apparent distribution in the following ways. Clearly, the apparent distribution will have a dimmer and redder average light curve. The distribution of extinguished absolute magnitudes will be wider than the intrinsic distribution at any phase or wavelength. The dust will also induce or increase apparent positive correlations between absolute magnitude and color (in the redder–dimmer sense), between two colors (redder–redder), and between absolute magnitudes (dimmer–dimmer) at different wavelengths and phases. If we want to use SN Ia light curves to understand the statistical intrinsic properties of these physical events, or those of dust in distant galaxies, it is necessary to deconvolve these two effects in the observed data.

Selecting a subsample of the observed SN, for example, the apparent blue end of the full sample, as representative of the "intrinsic" distribution, does not necessarily alleviate these distortions. Some previous studies have selected an "unreddened" subsample based on auxiliary data, such as elliptical host galaxies or large physical separation of the SN from the center of the galaxy, which may suggest lack of dust extinction. However, unless these auxiliary criteria can guarantee negligible dust effects, the resulting subsamples may still be distorted if there is a chance for some dust extinction. Hicken et al. (2009b) showed with the large CfA3 sample (Hicken et al. 2009a) that SNe with moderate estimated dust extinction ($A_V \approx 0.4$ mag) are found in elliptical host galaxies or at large projected galactocentric distances between the host galaxy and the SN. Furthermore, selecting an "intrinsic" subsample based on auxiliary data might

distort inferences if the auxiliary properties are correlated with intrinsic properties, the distribution of which one is trying to identify.

Statistical errors in the estimates of the random effects can distort inferences of the intrinsic distribution. For example, if distances to nearby SNe Ia are estimated via recession velocities and the Hubble law to infer absolute magnitudes, the effects of random peculiar velocities on distance errors can distort the inferred intrinsic distributions of SN Ia absolute light curves. Even in the absence of dust, a histogram of simple point estimates of peak absolute magnitudes for each SN (e.g., $V_0 - \mu(z)$) will appear broader than the true, intrinsic distribution, $P(M_V)$. Similarly, random peculiar velocities can apparently induce or strengthen a positive correlation between the absolute magnitudes at two wavelengths, and distort correlations of absolute magnitudes with other observables if these random effects are not properly modeled. Measurement errors and errors in dust extinction corrections will also tend to distort joint distributions of inferred variables. Since extinction in the NIR is greatly diminished relative to the optical, inferences on the distribution of intrinsic SN Ia light curve properties in the NIR are much less vulnerable to distortions by dust, but are still affected by the other sources of error. SNe far enough into the Hubble flow, so that peculiar velocity effects are negligible, will still be vulnerable to dust effects, especially at optical wavelengths.

Hierarchical Bayes provides a framework for the probabilistic modeling of multiple sources of randomness and uncertainty. Its application to statistical modeling of SNe Ia was first presented by Mandel et al. (2009) who constructed hierarchical models for SN Ia light curves in the near-infrared. The hierarchical framework provides a unified method of inference for populations and individuals of those populations. It includes a population distribution that models intrinsic variations and correlations of SN Ia light curves, a population distribution for the host galaxy dust to each SN, and models individual light curves, dust extinction, distances, and redshifts. Using Bayes' theorem, probabilistic estimates for the unknown parameters of individual SNe, as well as the hyperparameters of the populations, can be computed coherently and consistently.

Statistical inference with hierarchical models provides a principled method of probabilistic deconvolution of physically distinct and random effects that are combined in the observed data. Probabilistic inference allows for not only the estimation of each separate effect, but also the exploration of the joint uncertainties and tradeoffs between the multiple effects. It enables the estimation of the statistical characteristics of an underlying intrinsic population distribution while accounting for the distortions in the observed distribution caused by measurement error or other random effects. Similar issues regarding inferring the intrinsic distributions of inferred quantities in the presence of random error have been discussed and specific Bayesian techniques have been applied by Kelly (2007), Kelly & Bechtold (2007), Hogg et al. (2010), Loredo & Hendry (2010), among others, in other astrophysical contexts.

Statistical modeling of SN Ia light curves is inherently a multi-dimensional problem. Light curve observations are essentially noisy, usually irregular time series in multiple filters at different wavelengths. However, the absolute light curves exhibit regularities; for example, the fast declining light curves tend to be intrinsically dimmer. Existing models for optical light curves, e.g., MLCS (Riess et al. 1996a, 1998; Jha et al. 2007), SALT (Guy et al. 2005, 2007), and $\Delta m_{15}(B)$ templates (Hamuy et al. 1996b; Prieto et al. 2006) attempt to capture

regularities by assuming strong functional forms, governing the "global" behavior of light curves over a wide range in phase and wavelength, and controlled by one or two parameters (e.g., Δ; x_1, c; $\Delta m_{15}(B)$, respectively). Although these formulations can be useful as a form of dimensionality reduction to project gross variations in high-dimensional data onto a small-dimensional latent parameter space, it is not clear that this reduction can be done cleanly without loss of statistical information contained in the light curves. Detailed studies of well-sampled light curves reveal that, for example, SNe with the same $\Delta m_{15}(B)$ measurement (the magnitude change in the B-band light curve after 15 days from the peak) can display significant differences in their multi-band light curves over a range in phase (Folatelli et al. 2010; Höflich et al. 2010), signifying that a single light curve shape parameter does not capture the full variety of light curve signals that are generated by the underlying explosion physics. Furthermore, these global parameters lack direct interpretability: even with a continuously well-sampled SN light curve, it is impossible to estimate the Δ parameter without knowing the particular templates that attempt to project it onto the latent parameter space.

In this paper, we take a different approach to modeling the light curves. Instead of adopting a strong parameterization of the global behavior of the absolute light curves, we take a non-parametric approach that models the shape of the light curves "locally." This does not mean that there are no parameters; rather, we use local parameters, describing the variations in signal in each neighborhood of phase and wavelength, to build up a model for the light curve over the full range of phase and wavelength. We develop a differential decline rates model (Section 2.1, Appendix A) to represent the light curves in filters at multiple wavelengths using the decline rates over intervals in phase in each passband. Regularities underlying the population of light curves are then captured, not by one or two global parameters, but by inferring the correlations between the local parameters in the training set of well-sampled SNe. From this perspective, the light curves are modeled as stochastic processes with covariance structure over phase and wavelength that must be estimated. The correlation structure in the light curves is modeled in the intrinsic population distribution for SN Ia light curves. Although there may be many local parameters, they are not each statistically independent once the correlation structure of the population is learned. Indeed, the intrinsic dimensionality of the light curves (i.e., the effective number of "global" degrees of freedom) is implicit in the estimated covariance structure, and does not need to be fixed a priori. By incorporating this non-parametric light curve model in the hierarchical Bayes framework, we can coherently estimate the joint uncertainties in the correlation structure and incorporate them into distance predictions for SNe Ia.

The probabilistic hierarchical approach also provides a principled framework for dealing with missing data. The observations are typically not obtained at an exactly regular cadence: observation times often can be random or clustered, with gaps in temporal coverage due to weather or instrumentation. The SNe in a given sample may not all be observed in the full set of passbands; in this paper, the SNe are observed in the optical filters, but only a subset is observed in the NIR. The Bayesian approach deals with this by marginalizing over the unobserved light curves in the posterior distribution, thus incorporating this lack of information into inferences without omitting good incomplete data, which would be necessary if the analysis required the entire data set to be complete. In the absence of complete
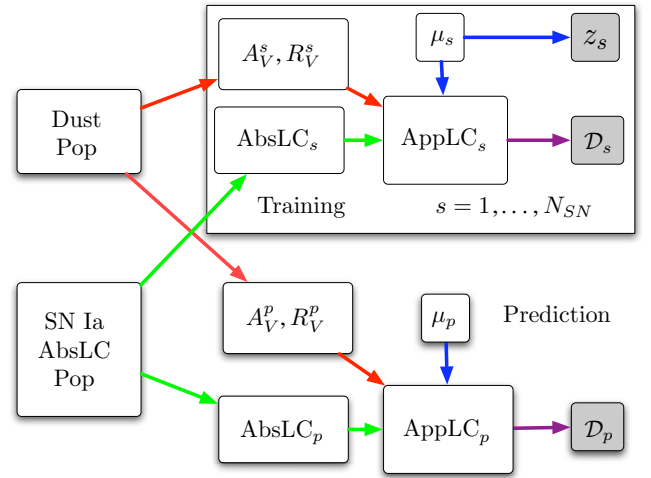

Figure 1. Hierarchical framework for statistical inference with SN Ia light curves. The global posterior density of the hierarchical model parameters given the full SN data set is represented formally with a directed acyclic graph. Unknown parameters are represented by open nodes. Observed data (redshifts z and measured light curves $\mathcal{D}$) are represented by shaded nodes. Each arrow or link describes a relationship of conditional probability. The hierarchical model coherently incorporates randomness and uncertainties due to measurement error (purple), intrinsic SN variations (green), dust extinction and reddening (red), and peculiar velocities and distances (blue) into inferences about individual SNe and the population. The graph can be understood as a generative model for the data. "SN Ia AbsLC Pop" represents parameters describing the population of SN Ia light curves, including intrinsic variations and correlations in shape, color, and luminosity across multiple wavelengths. From this population, each SN randomly draws a set of multi-wavelength light curves "AbsLC." The box "Dust Pop" represents parameters governing the population distribution of host galaxy dust values. Each SN randomly draws dust parameters A_V, R_V from this distribution. These dust parameters combine with the individual absolute light curves and distance modulus to generate an apparent light curve "AppLC," which is sampled with noise to produce the observed multi-wavelength light curve data $\mathcal{D}$. In the nearby universe, the distance modulus is related to the observed recession velocity or redshift through the Hubble law plus a noise term representing random peculiar velocities of host galaxies. This random generative process is conceptually repeated for each SN in the data set. The difference between "training" and distance prediction is that the latter does not condition on the redshift–distance likelihood information of the SN (bottom).

(A color version of this figure is available in the online journal.)

data, the model makes the best estimates and predictions given the available observed data.

We have built upon the basic framework described by Mandel et al. (2009). The overall structure of the hierarchical Bayesian model is depicted by Figure 1. We describe each component of the model in turn.

2.1. Representation of Apparent Light Curves

An apparent light curve model at phase t in rest-frame filter F with parameters F_0 and $\boldsymbol{\theta}^F = (\boldsymbol{\theta}_L^F, \boldsymbol{\theta}_{NL}^F)$ is generally described by

$$\mathrm{LC}^F(t; F_0, \boldsymbol{\theta}^F) = F_0 + l^F(t, \boldsymbol{\theta}^F)$$

$$= F_0 + l_0^F\left(t; \boldsymbol{\theta}_{NL}^F\right) + \boldsymbol{l}_1^F\left(t; \boldsymbol{\theta}_{NL}^F\right) \cdot \boldsymbol{\theta}_L^F, \quad (1)$$

where F_0 is the apparent magnitude at $t = 0$ in rest-frame filter F and $l^F(t; \boldsymbol{\theta}^F)$ is the normalized light curve in filter F, so that $l^F(0) = 0$. The vector of linear light curve shape parameters is $\boldsymbol{\theta}_L^F$, and $\boldsymbol{\theta}_{NL}^F$ is a vector of nonlinear light curve shape parameters. The phase is defined in the rest frame of the SN, with $t = 0$ corresponding to the time of maximum light in

B, T_0: $t = (T - T_0)/(1 + z)$, where z is the measured redshift and T is the time of observation. For a multi-wavelength model using light curve observations corresponding to rest-frame filters $F^1, \ldots, F^N$, the apparent light curves of an SN are described by a vector of apparent light curve parameters

$$\boldsymbol{\phi} = \left(F_0^1, \ldots, F_0^N, \boldsymbol{\theta}^{F_1}, \ldots, \boldsymbol{\theta}^{F_N}\right) \qquad (2)$$

and the time of maximum light in B, T_0.

The models employed in this paper do not use nonlinear shape parameters, so $\boldsymbol{\theta}^F = \boldsymbol{\theta}_L^F$. To specify the light curve functions $l_0^F(t)$ and $\boldsymbol{l}_1^F(t)$, we construct a representation in terms of differential decline rates, as described in Appendix A. In this representation, the light curve shape parameters $\boldsymbol{\theta}^F$ for each filter are simply the changes in magnitude over disjoint intervals in phase. Let $\boldsymbol{d}^F$ be a vector of decline rates of a light curve in filter F on a grid in phase, set to $\boldsymbol{\tau} = (-12, -8, -4, -2, 0, 2, 4, 6, 8, 10, 12, 15, 18, 23, 30, 37.5, 45)$ days. The decline rates are positive after maximum light and negative before peak in each filter. For a given set of decline rates $\boldsymbol{\theta}^F = \boldsymbol{d}^F$, the normalized light curve for each filter at arbitrary phase is constructed with a smooth curve defined by non-parametric regression cubic spline. In this representation, the light curve function $l_0^F(t) = 0$, and $\boldsymbol{l}_1(t)$ is determined by a linear smoothing spline: $LC^F(t; F_0, \boldsymbol{\theta}^F) = F_0 + \boldsymbol{l}_1^F(t) \cdot \boldsymbol{\theta}^F$.

2.2. Likelihood Function for Apparent Light Curves

In this section, we describe the likelihood function for the apparent light curve model parameters, conditional on the observed light curve data. The likelihood function explicitly accounts for K-corrections from the rest-frame filter to the observer-frame filter, Milky Way extinction, and photometric measurement error.

A light curve measurement in the observer frame O at time T is m^O. This observation differs from the apparent light curve model in the rest frame through K-corrections (to account for the redshifting of the SN spectrum), Milky Way extinction, and measurement error. At each redshift, we construct a unique mapping between each observer-frame filter O and a rest-frame filter F (in this paper, $O, F \in \{B, V, R, I, J, H\}$):

$$m^O = kc^{OF}(t; z, \boldsymbol{\phi}) + gx^{OF}(t; z, \boldsymbol{\phi}, E_{MW})$$
$$+ LC^F(t; F_0, \boldsymbol{\theta}^F) + \epsilon. \qquad (3)$$

The K-correction for the redshift z SN magnitude at phase t from rest-frame filter F to observer-frame filter O is $kc^{OF}(t; z, \boldsymbol{\phi})$. The K-correction has a dependence on the SN spectral energy distribution (SED), and this is modeled as a function of apparent color. Thus, it depends on the apparent light curve parameters $\boldsymbol{\phi}$ only through the apparent model colors at the same phase. For example, at low redshifts $z < 0.05$, if the observer-frame filter is $O = B$, then the rest-frame filter is $F = B$, and the dependence of $kc^{OF}(t; z, \boldsymbol{\phi})$ on the model light curve parameters $\boldsymbol{\phi}$ is through the apparent color $B - V$ at phase t. The effective Milky Way extinction, $gx^{OF}(t; z, \boldsymbol{\phi}, E_{MW})$, also depends on the SN SED through the colors, and is also a function of the estimated color excess due to Milky Way dust, $E_{MW} \equiv E(B - V)_{MW}$, which is obtained from the Schlegel et al. (1998) dust maps. Details regarding the calculation of K-corrections and Milky Way extinction for SNe Ia are presented in Appendix B. The variance of the random error term ϵ includes photometric error, and estimated uncertainties in K-corrections and Milky Way extinction.

For each measurement in observer-frame filter O at observed time T, we can write down Equation (3) relating the measurement to the rest-frame light curve model. Let the observations in filter O be arranged into a time-ordered vector $\boldsymbol{m}^O$, with each observation listed from earliest to latest. The corresponding equations can also be time-ordered. If the SN light curve is observed in multiple filters, $O_1, \ldots, O_N$, we can arrange the full data in time-filter ordering, so that $\boldsymbol{m} = (\boldsymbol{m}^{O_1}, \ldots, \boldsymbol{m}^{O_N})$, with the observer-frame filters ranked from the shortest to the longest central wavelength. With this arrangement, a time-filter-ordered vector equation can be written for each SN:

$$\boldsymbol{m} = \mathbf{KC}(T_0; z, \boldsymbol{\phi}) + \mathbf{GX}(T_0; z, \boldsymbol{\phi}, E_{MW}) + \boldsymbol{L}_2(T_0, z)\boldsymbol{\phi} + \boldsymbol{\epsilon}. \quad (4)$$

We suppress the explicit dependence on the known observation times T. Each of the terms depends on the time of maximum T_0 and the time-dilating redshift through the phase $t = (T - T_0)/(1 + z)$. Each element of the vectors $\mathbf{KC}$ and $\mathbf{GX}$ corresponds to the K-correction or Milky Way extinction scalar in Equation (3). Here $\boldsymbol{L}_2$ is the unique matrix that, when multiplied with the apparent light curve parameter vector $\boldsymbol{\phi}$, computes the rest-frame apparent light curve model corresponding to each time-filter-ordered observation in $\boldsymbol{m}$. Its rows are constructed from the individual vectors $\boldsymbol{l}_1(t)$.

If the random errors $\boldsymbol{\epsilon}$ are normally distributed, the likelihood function of the unknowns T_0, $\boldsymbol{\phi}$ for the full data set for a single SN is

$$P(\boldsymbol{m} \,|\, T_0, \boldsymbol{\phi}, z) = N[\boldsymbol{m} \,|\, \mathbf{KC}(T_0; z, \boldsymbol{\phi}) + \mathbf{GX}(T_0; z, \boldsymbol{\phi}, E_{MW})$$
$$+ \boldsymbol{L}_2(T_0, z)\boldsymbol{\phi}, \boldsymbol{W}], \qquad (5)$$

where $\boldsymbol{W}$ is the error covariance matrix and $N(\boldsymbol{x} \,|\, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ is a multivariate Gaussian probability density with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. In general, $\boldsymbol{W}$ can be a full positive definite symmetric matrix, if the errors due to photometric calibration, K-corrections, and Milky Way extinction are correlated across the observer filters and observation times. Our algorithms allow for this to be a full matrix, but for this paper, we take the simple approach of assuming it is diagonal, using only the measurement error variances. We symbolize the light curve data, the observed magnitudes and times, and the error covariance for each SN, as $\mathcal{D}_s = \{\boldsymbol{m}, \boldsymbol{W}, \boldsymbol{T}\}$.

2.3. Redshift–Distance Likelihood Function

The theoretical relation between the cosmological redshift z_c and the luminosity distance d_L to an SN in a smooth cosmological model depends on the cosmological parameters $\Omega_M, \Omega_\Lambda, w$, and the Hubble constant $h = H_o/100 \text{ km s}^{-1}$. At low redshifts, distances are insensitive to the cosmological model, and if we are concerned only with ratios of distances (or differences of distance moduli), then it is sufficient to fix h. For this paper, we are not concerned with constraining cosmological parameters, but on the statistical properties of SN Ia light curves, so we fix $\Omega_M = 0.73$, $\Omega_\Lambda = 0.27$, $w = -1$, and $h = 0.72$. The cosmological redshift may differ from the measured redshift z through measurement error and random peculiar velocities. The expected value of the distance modulus at redshift z is $f(z) = 25 + 5 \log_{10}[d_L(z) \text{ Mpc}^{-1}]$. As described in Mandel et al. (2009), the likelihood function of the distance modulus given the measured redshift

$$P(\mu \,|\, z) = N\big[\mu \,|\, f(z), \sigma_\mu^2 = [f'(z)]^2\big(\sigma_z^2 + \sigma_{\text{pec}}^2/c^2\big)\big]. \qquad (6)$$

In the low-z regime, where $d_L(z)$ is linear in z (the Hubble law), the variance is

$$\sigma_\mu^2 = \left(\frac{5}{z \ln 10}\right)^2 \left[\sigma_z^2 + \frac{\sigma_{pec}^2}{c^2}\right], \tag{7}$$

where σ_z^2 is the redshift measurement variance and σ_{pec}^2 is the expected variance due to random peculiar velocities. In this paper, we have alternately taken $\sigma_{pec} = 150$ (Radburn-Smith et al. 2004) and 300 km s^{-1}. Our results are consistent between the two values. We used the measured redshifts corrected to the cosmic microwave background frame and the local infall flow model of Mould et al. (2000).

In this paper, we are only concerned with evaluating distance predictions for low-z SNe Ia, so the cosmological parameters are fixed to their concordance values. The dependence of the redshift–distance relation on the cosmological parameters could be made explicit by writing $P(\mu \,|\, z; \Omega_M, \Omega_\Lambda, w)$ and allowing them to be free parameters that appear in the global posterior density (Equation (19)) of a cosmological sample of SNe Ia.

2.4. Latent Variable Model and Host Galaxy Dust

The vector ϕ in Equation (2) encodes the information needed to construct the apparent light curve model in the rest-frame filters. Using the differential decline rates representation, this vector encodes the peak apparent magnitudes and the decline rates of the apparent light curve in multiple filters over intervals in phase. The latent parameter vector ψ encodes the information for constructing the absolute light curve model in the rest-frame filters: the peak absolute magnitudes in rest-frame filters and the decline rates of the absolute light curves. The two sets of parameters are related by host galaxy dust extinction and distance:

$$\phi = \psi + A + v\mu. \tag{8}$$

The vector $A = A_V(\alpha + \beta/R_V)$ represents the effect of extinction on the absolute light curve parameters, and is a function of the host galaxy extinction, A_V, and the slope of the extinction law, R_V, using the dust extinction law of Cardelli et al. (1989). We model the effect of host galaxy extinction as described in Jha et al. (2007). For a given (A_V, R_V), the effective dust extinction in filter F at phase t is

$$A_F(t) = A_V \zeta_F(t)(a_F + b_F/R_V). \tag{9}$$

The coefficients a_F and b_F model the effect of dust extinction on the SN SED within each passband F at the time of maximum light. The functions $\zeta_F(t)$ model the change of this effect with phase due to the evolving SN SED. The constant vector α is constructed with components

$$\alpha_j = \begin{cases} a_F, & \text{if } \phi_j \text{ is a peak magnitude, } F_0 \\ \Delta\zeta_F^k a_F, & \text{if } \phi_j \text{ is a decline rate in filter } F \\ & \text{between phases } \tau_k \text{ and } \tau_{k+1}, \end{cases} \tag{10}$$

where $\Delta\zeta_F^k \equiv [\zeta_F(\tau_{k+1}) - \zeta_F(\tau_k)]$. The constant vector β is defined analogously, in terms of ζ_F and b_F. This accounts for the effect of dust extinction on the apparent magnitudes and light curve shape through the evolving SN Ia SED with phase.

Since distance only changes the magnitude, but not the shape of the light curve (after accounting for time dilation), the constant vector v is defined with components

$$v_j = \begin{cases} 1, & \text{if } \phi_j \text{ is a peak magnitude, } F_0 \\ 0, & \text{otherwise.} \end{cases} \tag{11}$$

With these constructions, we use Equation (8) to relate the apparent light curve parameters ϕ to the latent variables of extinction A_V, R_V, the distance modulus μ, and absolute (intrinsic) light curve parameters

$$\psi = (M^{F_1}, \ldots, M^{F_N}, \tilde{\theta}^{F_1}, \ldots, \tilde{\theta}^{F_N}), \tag{12}$$

where $\tilde{\theta}^F$ contain the decline rates of the absolute model light curves in each rest-frame filter. Since the model for the extinction in each passband and phase, Equation (9), is linear in the inverse of R_V, we find it useful to define $r_V = R_V^{-1}$ to simplify the notation.

2.5. Population Distribution Model for Intrinsic Absolute Light Curves

Even normal SNe Ia do not all have the same luminosities, intrinsic colors, or light curve shapes. The heterogeneities of these properties in the population of SNe Ia—which cannot be explained by dust or distance—are called intrinsic variations. Estimation of the covariances in the population of SN Ia light curves is crucial to the utility of SNe Ia as standardizable candles for distance estimation. For example, the well-known width–luminosity correlation of optical light curves (Phillips 1993; Hamuy et al. 1996b; Phillips et al. 1999) allows us to estimate the distance modulus to an SN to ∼0.2 mag.

To model and capture intrinsic correlations between the absolute magnitudes at multiple wavelengths and the shapes of their light curves, we need to specify a general correlation structure for the population distribution of ψ. Ideally, the population distribution $P(\psi)$ would be specified by reliable astrophysical theory. However, current explosion models for SNe Ia do not provide such detailed guidance regarding the expected distribution of absolute light curve properties. Thus, we seek to model the population distribution of light curves generally and infer the statistical properties of the intrinsic variations from the data. We capture the intrinsic variations and correlations of SN Ia absolute light curves by modeling the distribution of the intrinsic parameters ψ as a multivariate Gaussian:

$$\psi_s \sim N(\mu_\psi, \Sigma_\psi) \tag{13}$$

with mean vector μ_ψ and intrinsic covariance matrix Σ_ψ. The intrinsic covariance matrix models population correlations between the peak absolute magnitudes at different wavelengths, correlations between the light curve decline rates at different phases and wavelengths, and correlations between peak absolute magnitudes and light curve decline rates at different phases and wavelengths. By capturing the population correlations of the absolute light curves in multiple filters at different wavelengths, we implicitly also model the correlation structure of the intrinsic colors. The intrinsic covariance matrix can be readily decomposed into a matrix of intrinsic correlations, R_ψ, and a vector of intrinsic standard deviations, σ_ψ, one for each component ψ^i: $\Sigma_\psi = \text{diag}(\sigma_\psi) R_\psi \text{diag}(\sigma_\psi)$. Each element of R_ψ is a correlation coefficient between -1 and 1. A valid Σ_ψ must be positive definite and symmetric. Since ψ directly describes the multi-band absolute light curves, Equation (13) models them as a stochastic (Gaussian) process.

Suppose the intrinsic parameters vector $\psi = (\psi^u, \psi^o)$ of a particular SN can be partitioned into parameters ψ^o tightly constrained by its observations and parameters ψ^u that are not directly observed. For example, ψ^o may be the decline rates describing the shapes of this SN's light curves, and ψ^u may

be the peak absolute magnitudes of this SN under distance prediction. With this population model for intrinsic light curve variations, it is simple to estimate $\boldsymbol{\psi}^u$ given $\boldsymbol{\psi}^o$. The full hyperparameters can be partitioned in the same way:

$$\boldsymbol{\mu}_\psi = \begin{pmatrix} \boldsymbol{\mu}_\psi^u \\ \boldsymbol{\mu}_\psi^o \end{pmatrix}, \quad \boldsymbol{\Sigma}_\psi = \begin{pmatrix} \boldsymbol{\Sigma}_\psi^{uu} & \boldsymbol{\Sigma}_\psi^{uo} \\ \boldsymbol{\Sigma}_\psi^{ou} & \boldsymbol{\Sigma}_\psi^{oo} \end{pmatrix}. \quad (14)$$

Using standard theorems for the multivariate normal distribution, the expected value of $\boldsymbol{\psi}^u$, conditional on $\boldsymbol{\psi}^o$, is $\mathbb{E}[\boldsymbol{\psi}^u|\boldsymbol{\psi}^o] = \boldsymbol{\mu}_\psi^u + \boldsymbol{\Sigma}_\psi^{uo}(\boldsymbol{\Sigma}_\psi^{oo})^{-1}(\boldsymbol{\psi}^o - \boldsymbol{\mu}_\psi^o)$, and its conditional covariance (uncertainty) is $\mathrm{Cov}[\boldsymbol{\psi}^u|\boldsymbol{\psi}^o] = \boldsymbol{\Sigma}_\psi^{uu} - \boldsymbol{\Sigma}_\psi^{uo}(\boldsymbol{\Sigma}_\psi^{oo})^{-1}\boldsymbol{\Sigma}_\psi^{ou}$. This example demonstrates how this model uses the correlation structure of the absolute light curves to relate inferred variables to observable quantities and vice versa.

In the absence of host galaxy dust ($A_V = 0$) and measurement error, an estimator for the predicted distance can be derived straightforwardly for well-sampled light curves. Suppose that apparent light curve parameters $\boldsymbol{\phi}$ could be measured perfectly for well-sampled light curves with vanishing measurement error. If the intrinsic mean $\boldsymbol{\mu}_\psi$ and covariance $\boldsymbol{\Sigma}_\psi$ were known, then the posterior prediction of the distance modulus has mean $\hat{\mu} = \hat{V}_\mu \boldsymbol{v}^T \boldsymbol{\Sigma}_\psi^{-1}(\boldsymbol{\phi} - \boldsymbol{\mu}_\psi)$ and variance $\hat{V}_\mu = (\boldsymbol{v}^T \boldsymbol{\Sigma}_\psi^{-1} \boldsymbol{v})^{-1}$. In fact, $\hat{\mu}$ is the minimum variance unbiased linear estimator of the distance modulus, a result that does not depend on the Gaussianity of the intrinsic distribution of $\boldsymbol{\psi}$, the absolute light curve parameters. This can be shown by noting that $\hat{\mu}$ is the generalized least-squares solution of Equation (8) and invoking the Gauss–Markov theorem. However, the presence of a random amount of host galaxy dust for each SN, and finite sampling and measurement error of the light curves necessitates modeling these other aspects of the hierarchical structure.

Modeling intrinsic variations of $\boldsymbol{\psi}$ in the population using the covariance structure of a multivariate Gaussian is the simplest choice. However, if non-Gaussianities become important then it will be possible to replace this simple assumption with more complex distributions. For example, a Gaussian mixture model could be used to describe a multi-modal population, Student-t distributions can be employed to model fat-tailed populations, and nonlinear regression could capture nonlinear correlation structure. Alternatively, one might seek a representation or parameterization (Section 2.1) for $\boldsymbol{\psi}$ that makes its population distribution more amenable to modeling with a simple form. For the application in this paper, we did not find these more advanced approaches to be necessary, so we postpone their discussion for future work.

Observables that are not derived from light curve data, such as host galaxy masses (e.g., Kelly et al. 2010; Sullivan et al. 2010) or spectroscopic measurements (e.g., Bailey et al. 2009; Blondin et al. 2011; Foley & Kasen 2011), can be correlated with the intrinsic absolute light curves. They can be included in this framework by augmenting $\boldsymbol{\psi}$ with an auxiliary parameter and specifying a likelihood function describing the uncertainty in the new observable. The joint distribution $P(\boldsymbol{\psi})$ would model the covariance structure of the intrinsic light curves along with the auxiliary property, which can be leveraged to compute distance predictions using the extra information.

2.6. Population Models for Host Galaxy Dust

We also adopt models for the population distribution of host galaxy dust parameters A_V and r_V for each SN. Their joint population distribution can be factored as $P(A_V, r_V) =$

$P(r_V | A_V)P(A_V)$. The extinction A_V values are assumed to be drawn from an exponential distribution describing dust along lines of sight from SN host galaxies (Jha et al. 2007): $A_V \sim \mathrm{Expon}(\tau_A)$. The probability density is

$$P(A_V | \tau_A) = \begin{cases} \tau_A^{-1} e^{-A_V/\tau_A}, & A_V > 0 \\ 0 & A_V \leqslant 0 \end{cases} \quad (15)$$

with an unknown hyperparameter, the exponential scale length τ_A, which is inferred from the hierarchical posterior probability density conditional on the data.

Even along lines of sight within the Milky Way, interstellar dust can cause varying amounts of reddening for a given amount of absorption or extinction. This ratio is captured by the parameter $R_V \equiv A_V/E(B - V)$. Although the average value within the Milky Way is 3.1, it can range from 2.1 to 5.8, and depends on the nature of the dust grains (Cardelli et al. 1989; Draine 2003; and references therein).

Previous studies have focused on treating R_V as a constant for all SNe, either set to the Milky Way value, or left as a fit parameter. However, this is a rather strong assumption, so here we consider the possibility that the R_V of dust within the distant host galaxies of SNe Ia may vary within a common distribution. We also wished to explore whether R_V could be systematically different for SNe with different A_V dust extinctions. To test for and capture potential population correlations between A_V and r_V, we consider several models for the conditional population distribution $r_V \equiv R_V^{-1}$ given A_V. We consider six cases with the following assumptions.

1. $\boldsymbol{R_V = 3.1}$. The host galaxy dust law slope is fixed to the Milky Way interstellar average for all SNe.
2. **CP**: ($\boldsymbol{R_V = \mathrm{const}}$). Complete pooling; R_V has the same value for each SN, but this value is unknown and inferred from the posterior density.
3. **NP**. No pooling. $\{r_V^s\}$ for each SN are completely independent with a flat prior $P(r_V^s) \propto 1$ on each.
4. **PP: $m = 0$**. Partial pooling. $\{r_V^s\}$ are conditionally independent draws from a common Gaussian population distribution independent from the magnitude of extinction A_V:

$$r_V^s \sim N(\mu_r, \sigma_r^2). \quad (16)$$

The mean μ_r and variance σ_r^2 of this population are unknown and inferred from the posterior density.

5. **PP: $m = 1$**. Partial pooling. $\{r_V^s\}$ are conditionally independent draws from a common population distribution with a mean linearly dependent on A_V:

$$r_V^s | A_V \sim N(\beta_0 + \beta_1 A_V^s, \sigma_r^2). \quad (17)$$

The regression coefficients $\boldsymbol{\beta}$ and residual variance σ_r^2 of this population are unknown and inferred from the posterior density. The intercept β_0 represents the population mean r_V value at vanishing extinction and β_1 captures a potential trend of r_V with increasing A_V.

6. **PP: Steps**. Partial pooling with the step function distribution. The range in A_V is divided up into two or four intervals (c.f. Tables 2 and 3). Within each interval, r_V^s for each SN is a conditionally independent draw from a common Gaussian distribution with mean μ_r and variance σ_r^2 (Equation (16)). These hyperparameters are estimated from the posterior density.

In all cases, we limited R_V to an allowed range: $0.18 < r_V < 0.7$ ($1.4 < R_V < 5.6$).

Cases 4, 5, and 6 perform partial pooling, or shrinkage estimation, of the r_V parameters, which is characteristic of hierarchical Bayes models. In the complete pooling case, it is assumed that the dust in the host galaxies of SNe all have the same values of R_V, and thus all the information in the sample of SNe is used ("pooled") to infer the single R_V value. In the case of no pooling, it is assumed that the host galaxy dust for different SNe can have different, independent values of r_V, and that only the information from each SN is used to infer the R_V value for that SN. These two cases are limits of partial pooling, in which the information for each SN is combined with that of the population to produce the individual R_V estimates. The appropriate weight between the individual SN information and that of the population is negotiated by the inferred population variance, σ_r^2, and the precision with which R_V can be estimated independently for each SN. As $\sigma_r^2 \to 0$, we obtain complete pooling, and for $\sigma_r^2 \to \infty$, we have effectively no pooling. At intermediate values, partial pooling finds a middle ground between the noisy and possibly unstable estimates of no pooling and the possibly unrealistic strong assumptions of complete pooling. The hyperparameter σ_r^2 can be understood as the residual variance of r_V in the sample after accounting for the other sources of error for each individual SN. Shrinkage accounts for the fact that a histogram or scatter plot of individual, simple, point estimates of derived quantities will be wider than the true, intrinsic distribution of those quantities, if those point estimates have significant uncertainties, and it reduces the mean squared error of each parameter estimate. From a non-Bayesian perspective, shrinkage can be regarded as an adaptive regularization that determines from the data how much to allow an individual estimate to deviate from the population average or trend. Shrinkage estimation with multi-level models has been discussed recently by Loredo & Hendry (2010).

We can test the hypothesis that the dust law parameter r_V has no dependence on host galaxy extinction A_V by comparing the results from fitting Cases 4 and 5. Since Case 4 is a nested case of Case 5 with $\beta_1 = 0$, we can check to see whether or not the marginal posterior density of β_1 is consistent with zero when fitting Case 5. Similarly, with Case 6, we can check to see whether the population means μ_r in each interval in A_V are consistent with each other across the range of A_V, or if there are significant differences.

For brevity and specificity, in subsequent sections describing statistical computation, we adopt the assumptions of Case 5; the hyperparameters governing the r_V population are $\boldsymbol{\beta}$ and σ_r^2. For other cases, the hyperparameters are $\boldsymbol{\mu}_r$ and σ_r^2 and the algorithms are appropriately modified to account for the different models.

2.7. Specifying the Hyperpriors

Diffuse, or non-informative, hyperpriors are adopted on the highest-level hyperparameters of the hierarchical model: $\boldsymbol{\mu}_\psi$, $\boldsymbol{\Sigma}_\psi$ for the SN light curve population and τ_A, $\boldsymbol{\mu}_r$, or $\boldsymbol{\beta}$, and σ_r^2 for the host galaxy dust population. As the number of well-observed SNe, N_{SN}, becomes larger, the influence of the hyperpriors on the posterior estimates of the hyperparameters diminishes. Thus, so long as we include a sufficient number of SNe in the hierarchical model, we can choose the particular, analytic forms of diffuse hyperpriors for computational convenience. We discuss some mathematical details in Appendix C.

2.8. Global Posterior Probability Density

Mandel et al. (2009) derived the global posterior probability density over all the SNe in the training set as a product of the light curve and redshift–distance likelihood functions for individual SNe, population distributions for SN Ia light curves and host galaxy dust, and the hyperpriors on the hyperparameters of the population distributions. We construct the global posterior using the new component probability models described above. Let the time-filter-ordered light curve data in the observer frame for SN s be $\mathcal{D}_s$, with measured redshift z_s. The unknown parameters for an individual SN are the apparent light curve parameters $\boldsymbol{\phi}_s$, the distance modulus μ_s, the extinction A_V^s, the slope of the dust law r_V^s, and the time of maximum T_0^s. For a given set of population hyperparameters, $\boldsymbol{\mu}_\psi$, $\boldsymbol{\Sigma}_\psi$, τ_A, $\boldsymbol{\beta}$, σ_r^2, the conditional posterior density for an individual SN is

$$P\left(\boldsymbol{\phi}_s, T_0^s, \mu_s, A_V^s, r_V^s \,\middle|\, \mathcal{D}_s, z_s; \boldsymbol{\mu}_\psi, \boldsymbol{\Sigma}_\psi, \tau_A, \boldsymbol{\beta}, \sigma_r^2\right)$$
$$\propto P\left(\boldsymbol{m}_s \,\middle|\, T_0^s, \boldsymbol{\phi}_s, z_s\right) \times P(\mu_s | z_s)$$
$$\times P(\boldsymbol{\psi}_s = \boldsymbol{\phi}_s - \boldsymbol{v}\mu_s - \boldsymbol{A}_s \,|\, \boldsymbol{\mu}_\psi, \boldsymbol{\Sigma}_\psi)$$
$$\times P\left(A_V^s, r_V^s \,\middle|\, \tau_A, \boldsymbol{\beta}, \sigma_r^2\right). \tag{18}$$

The training set data are comprised of the light curve data for all the SNe in the training set $\mathcal{D} = \{\mathcal{D}_s\}$ and their redshifts $\mathcal{Z} = \{z_s\}$. The unknown hyperparameters of the populations are the mean and covariance of the absolute light curve parameters $\boldsymbol{\mu}_\psi$, $\boldsymbol{\Sigma}_\psi$, the exponential scale of the extinction distribution τ_A, and the hyperparameters describing the R_V^{-1} distribution. The global joint posterior density of all SN observables $\{\boldsymbol{\phi}_s\}$, distance moduli $\{\mu_s\}$, dust parameters $\{A_V^s, r_V^s\}$, and the population hyperparameters conditioned on the database $\mathcal{D}, \mathcal{Z}$ is proportional to the product of N_{SN} individual densities multiplied by the hyperpriors:

$$P\left(\{\boldsymbol{\phi}_s, T_0^s, \mu_s, A_V^s, r_V^s\}; \boldsymbol{\mu}_\psi, \boldsymbol{\Sigma}_\psi, \tau_A, \boldsymbol{\beta}, \sigma_r^2 \,\middle|\, \mathcal{D}, \mathcal{Z}\right)$$
$$\propto \left[\prod_{s=1}^{N_{SN}} P\left(\boldsymbol{m}_s \,\middle|\, T_0^s, \boldsymbol{\phi}_s, z_s\right) \times P(\mu_s | z_s)\right.$$
$$\times P(\boldsymbol{\psi}_s = \boldsymbol{\phi}_s - \boldsymbol{v}\mu_s - \boldsymbol{A}_s \,|\, \boldsymbol{\mu}_\psi, \boldsymbol{\Sigma}_\psi)$$
$$\left.\times P\left(A_V^s, r_V^s \,\middle|\, \tau_A, \boldsymbol{\beta}, \sigma_r^2\right)\right] \times P(\boldsymbol{\mu}_\psi, \boldsymbol{\Sigma}_\psi) \times P\left(\tau_A, \boldsymbol{\beta}, \sigma_r^2\right).$$
$$\tag{19}$$

To predict the distance of an SN using its light curve data, one sets the redshift–distance likelihood $P(\tilde{\mu}_s | \tilde{z}_s) \propto 1$ in Equation (18), where we use tilde on parameters and data for prediction set SNe. The tilde redshift $\tilde{z}_s$ is used for time-dilation and K-correction in the light curve likelihood function, but not to constrain the distance modulus in the redshift–distance likelihood. The marginal posterior predictive density is $P(\tilde{\mu}_s | \tilde{D}_s, \tilde{z}_s; \mathcal{D}, \mathcal{Z})$, obtained by integrating Equation (19).

Equation (19) is an explicit statement of the objective function for statistical inference for both training and prediction with the hierarchical model. By computing Equation (19), we solve for the most likely distributions of host galaxy dust and intrinsic luminosities, colors, and light curve shapes that best account for the apparent distributions, conditional on the model assumptions. A directed acyclic graph (DAG) representing the hierarchical model and the global posterior density is shown in Figure 1. The graph depicts a generative probabilistic model

linking together the populations and individuals, and parameters and hyperparameters to the data (Mandel et al. 2009). For simplicity, we have not shown the dependence of the light curve likelihood function on redshift through time-dilation and K-correction, and only show the redshift–distance dependence, which is the key difference between training and prediction.

3. IMPROVED MCMC WITH BAYESN

It is important to distinguish between the tasks of statistical inference and statistical computation. The former entails deriving estimators for unknown quantities, conditional on data, and the assumptions of the statistical model. We have done that in the previous section by deriving the global posterior probability density, Equation (19). The task of statistical computation is accomplished by specifying, constructing, and implementing algorithms for computing the numerical values of these estimators for the observed values of the data, under the assumptions of the model. In this section, we describe our strategy for statistical computation of the posterior estimates by stochastic sampling from the global posterior probability density, Equation (19).

We perform a fully Bayesian inference of the hierarchical model using MCMC sampling methods such as the Metropolis–Hastings algorithm (Metropolis et al. 1953; Hastings 1970) and Gibbs sampling (Geman & Geman 1984). Mandel et al. (2009) presented the first MCMC algorithm (BAYESN) for hierarchical Bayesian inference with SN light curves. We have made many modifications to the original BAYESN algorithm to incorporate the modeling of host galaxy dust, K-corrections from the rest-frame model light curves to the observer-frame measured magnitudes, and to improve the computational speed and convergence rate of the chains. The new algorithm is largely comprised of Gibbs sampling and Metropolis–Hastings jumps that do not require the user to "tune" jump sizes for each SN, as would be required for ordinary Metropolis steps. Instead, the algorithm only requires the specification of a jump size for the scalar time of maximum for each SN, T_0^s, which is a relatively easy task: an rms jump proposal of ~ 0.5 day was generically successful for rapid convergence for all SNe. Movement along the other dimensions of parameter space is accomplished by Gibbs sampling or more general Metropolis–Hastings proposals that exploit the conditional structure of the posterior distribution to adaptively propose more global moves. By minimizing the amount of manual tuning required before running the MCMC, we have increased the ease-of-use and practical applicability of MCMC for fitting comprehensive hierarchical models for SNe Ia. A basic introduction to MCMC algorithms for SN Ia probabilistic inference was given by Mandel et al. (2009); statistical references include Liu (2002), Gelman et al. (2003), and Robert & Casella (2005).

We use this new BAYESN code to sample from the global posterior probability distribution of all individual parameters and the population hyperparameters of SN Ia light curves and host galaxy dust, conditioned on the SN light curve data and redshifts. Here, we sketch the improved BAYESN Gibbs sampling algorithm. Further mathematical details can be found in Appendix D. The state of the chain is determined by the current values of all the parameters and hyperparameters:

$$\mathcal{S} = \left(\{\boldsymbol{\phi}_s, T_0^s, \mu_s, A_V^s, r_V^s\}, \boldsymbol{\mu}_\psi, \boldsymbol{\Sigma}_\psi, \tau_A, \boldsymbol{\beta}, \sigma_r^2\right). \qquad (20)$$

We generate random samples from the global posterior distribution $P(\mathcal{S}|\mathcal{D}, \mathcal{Z})$ using a Markov chain, conditioned on the photometric light curve data for all SNe $\mathcal{D} = \{\mathcal{D}_s\}$, and

their redshifts $\mathcal{Z} = \{z_s\}$. We begin with crude, randomized guesses of the individual SN parameters $\{\boldsymbol{\phi}_s, T_0^s, \mu_s, A_V^s, r_V^s\}$ for all SNe s in the data set. In each step, we update a subset, or block, of parameters from their conditional posterior density with the complement set of parameters (and the data) fixed to their current values. The choice of parameter blocks exploits the conditional independence structure of the DAG of the inference; see Figure 1.

1. Compute the absolute light curve parameters $\{\boldsymbol{\psi}_s\}$ using Equation (8). The conditional posterior density of the light curve hyperparameters is $P(\boldsymbol{\mu}_\psi, \boldsymbol{\Sigma}_\psi | \cdot, \mathcal{D}, \mathcal{Z}) = P(\boldsymbol{\mu}_\psi | \boldsymbol{\Sigma}_\psi; \{\boldsymbol{\psi}_s\})P(\boldsymbol{\Sigma}_\psi | \{\boldsymbol{\psi}_s\})$. We update $\boldsymbol{\Sigma}_\psi$ from the second factor, and then update $\boldsymbol{\mu}_\psi$ given $\boldsymbol{\Sigma}_\psi$ from the first (see Appendix D for details). (We use the notation $(\cdot)$ to indicate all the other parameters and data that have not been explicitly indicated.)

2. Draw a new extinction scale hyperparameter τ_A from the conditional distribution $P(\tau_A | \cdot, \mathcal{D}, \mathcal{Z}) = P(\tau_A | \{A_V^s\})$, by drawing a random number from an inverse gamma distribution.

3. Obtain new values of $\boldsymbol{\beta}$ and σ_r^2 from $P(\boldsymbol{\beta}, \sigma_r^2 | \cdot, \mathcal{D}, \mathcal{Z}) = P(\boldsymbol{\beta} | \sigma_r^2, \{r_V^s, A_V^s\})P(\sigma_r^2 | \{r_V^s, A_V^s\})$. First draw a new σ_r^2 from an inverse gamma distribution. Conditional on that value, draw a new $\boldsymbol{\beta}$ from a Gaussian.

4. Repeat the following steps for each SN s.

   (a) Propose a new time of maximum $T_{0,s}^* \sim N(T_{0,s}, s_T^2)$ according to a random walk from the previous estimate. It is usually sufficient to use $s_T \approx 0.5$ day. Given $T_{0,s}^*$, propose a new set of apparent light curve parameters for all filters, $\boldsymbol{\phi}^*$, from the distribution in Appendix D. Compute the Metropolis–Hastings ratio r to decide whether to accept the joint proposal $(T_0^*, \boldsymbol{\phi}^*)$ or to stay at the current values $(T_0, \boldsymbol{\phi})$.

   (b) Update the new distance modulus μ_s from the conditional probability density $P(\mu_s | \cdot, \mathcal{D}_s, z_s) = P(\mu_s | \phi_s, A_V^s, r_V^s; \boldsymbol{\mu}_\psi, \boldsymbol{\Sigma}_\psi; z_s)$ by sampling from a Gaussian distribution.

   (c) Draw a new host galaxy extinction A_V^s from the conditional posterior probability density $P(A_V^s | \cdot, \mathcal{D}_s, z_s) = P(A_V^s | \boldsymbol{\phi}_s, \mu_s, r_V^s; \boldsymbol{\mu}_\psi, \boldsymbol{\Sigma}_\psi, \tau_A, \boldsymbol{\beta}, \sigma_r^2)$. This can be shown to be a truncated Gaussian distribution in $A_V^s > 0$.

   (d) Draw a new host galaxy dust law slope r_V^s from the conditional posterior $P(r_V^s | \cdot, \mathcal{D}_s, z_s) = P(r_V^s | \boldsymbol{\phi}_s, \mu_s, A_V^s; \boldsymbol{\mu}_\psi, \boldsymbol{\Sigma}_\psi, \boldsymbol{\beta}, \sigma_r^2)$ by evaluating the probability density function (pdf) on a fine grid on $0.18 < r_V < 0.7$ and using griddy Gibbs sampling (Ritter & Tanner 1992).

   (e) (optional) Perform a translation in the space of distance versus extinction: $(A_V^s, \mu_s) \to (A_V^s, \mu_s) + \gamma(1, -x)$. Here, x determines a random direction along the tradeoff. We select a random γ using generalized conditional sampling (Liu 2002), and then translate to update (A_V^s, μ_s).

5. Finally, record the current state of the chain $\mathcal{S}$. Repeat all steps until convergence.

This algorithm generates an irreducible and ergodic Markov chain that will converge to a stationary distribution equal to the global posterior density, Equation (19), according to standard theorems (e.g., Robert & Casella 2005). It converges without Step 4e, but this step speeds the exploration of the tradeoff

between extinction and distance for each SN. We found that including this step reduces the auto-correlation scale for the slowest converging A_V^s by a factor of ~3.

4. APPLICATION

4.1. Data Sets

In this section, we describe the application of the hierarchical framework to inference with nearby SNe with optical and NIR light curve observations (*BVRIJH*).

Mandel et al. (2009) analyzed a sample of SNe with near-infrared JHK_s light curve observations compiled from the PAIRITEL sample (WV08), and available published light curves from the literature (Jha et al. 1999; Hernandez et al. 2000; Krisciunas et al. 2000, 2001; Di Paola et al. 2002; Valentini et al. 2003; Krisciunas et al. 2003, 2004b, 2004c, 2007; Elias-Rosa et al. 2006; Pastorello et al. 2007; Stanishev et al. 2007; Pignata et al. 2008). Nearly all of the SNe in the PAIRITEL sample were also observed in optical filters *UBVRI* or *UBVr'i'* by Hicken et al. (2009a), and nearly all of the SNe from the literature were also observed at optical wavelengths. We selected the same set of SNe with NIR light curves used in Mandel et al. (2009) with the following exceptions: SN 2005ao was omitted for lack of quality data points and SN 2006lf was omitted because its position was near the Galactic plane, so its Milky Way dust reddening estimate was very large and unreliable. Extensive studies of two of the well-sampled SNe, SN 2005cf and SN 2006X, were presented by Wang et al. (2008, 2009b).

Since the number of SNe Ia with NIR light curves is still small, we supplemented this set with SNe from the recent CfA3 sample of nearby SNe Ia (Hicken et al. 2009a) that had well-sampled optical light curves. The additional light curves increase the statistical strength in the optical bands and stabilize estimation of the full hierarchical model. They also provide a set of optical-only light curves to compare against the set of SNe with optical and NIR light curves (Section 7). We included SNe with high-quality light curve data at $0.01 < z < 0.065$, with more than five observations in *B* band, and with the first observation in *B* occurring less than 10 days past maximum. Since these light curves lack NIR data, the model marginalizes over the missing light curves for all inferences. Some of the SNe in the CfA3 sample were observed in the *RI* passbands, while some were observed in the *r'i'* passbands. In both cases, we map these observer-frame passbands to rest-frame *RI* passbands, and in the latter case, the *K*-correction takes into account the cross-filter transformation. The *K*-corrections for observations in the *B* and *V* filters mapped them to rest-frame *B* and *V* filters. The *J* and *H* observations were mapped to rest-frame *J* and *H* filters defined by the Two Micron All Sky Survey (2MASS).

We use only normal SNe Ia with *B* decline rates $0.75 < \Delta m_{15}(B) < 1.6$. As discussed by Hicken et al. (2009a), fast decliners and SN 1991bg-like objects have different luminosity–light-curve-shape relations than normal SNe Ia, and should be modeled separately, so we do not include them in our analysis. We have also excluded the peculiar SN 2006bt from the sample (Foley et al. 2010). The full "CfA+literature" sample consists of 110 SNe, 37 of which have NIR observations.

We modeled the rest-frame *BVRIJH* light curves of this set of SNe listed in Table 4. For each SN, if there were data in an observer-frame passband mapping to a given rest-frame filter, we list the fitted peak apparent magnitude in the rest-frame filter. If there were no observations for a given rest-frame passband, then the estimated peak magnitude is not listed. We have omitted

ultraviolet data for the analysis in this paper. Kessler et al. (2009) found that differences in the *U*-band model between MLCS2k2 and SALT2 lead to large ($\Delta w \sim 0.2$) differences in the cosmology using high-redshift samples. As this work focuses on low-redshift data, we have omitted the *U* data to avoid calibration and standardization problems, selection effects, and *S*-corrections between the *u* and *U* bands. Future work, applying our framework to high-redshift SN Ia samples for cosmological analysis, will carefully incorporate *U*-band data.

Contreras et al. (2010) have recently published light curve observations for a set of nearby SNe as part of the CSP, and an analysis of these data was discussed by Folatelli et al. (2010). These are high-quality well-sampled light curves with optical coverage, and a subset included contemporaneous NIR observations as well. Comparison of the CSP and PAIRITEL data reduction and calibration for SNe observed in both samples is ongoing. For joint analysis of the current published photometry of these sets, in Section 7, we augment our sample with 27 SNe with CSP data passing the criteria and examine the distance predictions for each set as a consistency check. Twenty of these SNe have joint optical and NIR light curve observations.

4.2. Statistical Computation

We ran the new BAYESN code (Section 3) to coherently fit the apparent light curves with the differential decline rates model (Section 2.1), to estimate host galaxy dust extinction (Section 2.4) and distance moduli (Section 2.3), and to infer intrinsic variations and correlations (Section 2.5) and the host galaxy dust population (Section 2.6). The code samples the global posterior density (Equation (19)) over all unknowns given the data.

We seeded each chain with random, initial values of the SN and dust parameters $\{\phi_s, A_V^s, \mu_s\}$. In many cases, the *B* light curve was sufficiently well sampled that the time of maximum could be unambiguously determined. In these cases, we fixed the T_0 to that value and did not re-estimate it when running the sampler (by setting $s_T = 0$ in step 4a). For SNe with more uncertain T_0 we re-estimated it during the fitting ($s_T = 0.5$d). Under Case 1 (Section 2.6), the initial values of R_V were all fixed to 3.1. For all other r_V cases, the initial $\{r_V^s\}$ were randomized.

The initial dust configuration of the SN set is random, and we do not make an initial estimation of the A_V^s value before running BAYESN. Thus, it is possible that this randomized initialization of dust values may assign a high A_V^s to an apparently blue SN, and a low A_V^s to an apparently red object. If we see that multiple chains starting with different and random initial dust configurations eventually converge to the same posterior estimate, we can be reassured that our final inferences are independent of the initial assignments of A_V, r_V values, and that the probabilistic inference has sorted out the probable dust values over the set of SNe. In Figure 2, we show that four independent and parallel chains training the hierarchical model over the set of SNe, each starting with a different initial value for the host galaxy extinction, converge to the same final estimates in the long run of the MCMC. This shows that our final inferences for the trained model are robust to the initial values of host galaxy dust, and indicates the self-consistency of estimates and convergence of the algorithm to a unique joint solution over the full set of SNe.

To perform training and prediction, the BAYESN code generated four parallel, independent chains of 2×10^4 complete cycles (Steps 1–5). The initial values for each chain were generated by using different random numbers for each independent chain. We
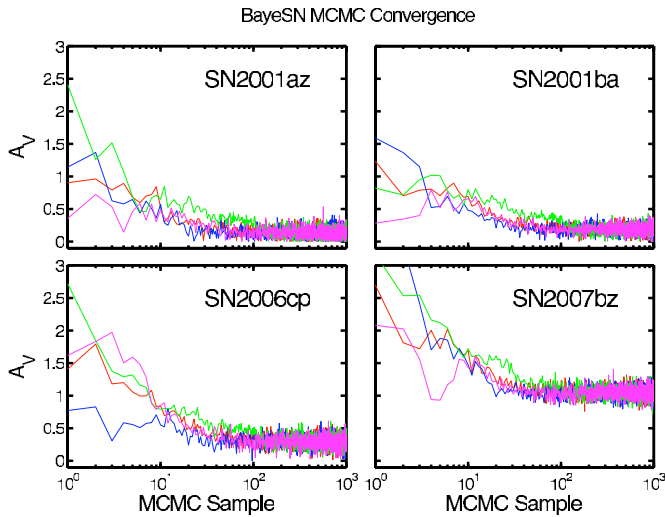
Figure 2. Example sample paths of Markov Chain Monte Carlo (MCMC) chains generated by the BAYESN MCMC sampling code. The full chain stochastically samples the parameter space of all individual SNe in the set, and the populations of SN Ia light curves and the dust. This plot focuses on the coordinates of the chain concerning the visual extinction A_V to particular SN. Each color represents an independent chain starting from a randomized initial guess. The chains explore the full parameter space and converge within a few hundred iterations upon the same global posterior distribution. The posterior uncertainty in the estimate is reflected in the distribution (variability) of the chain samples upon convergence. The plot depicts the simultaneous convergence of the chains, both for the estimate of a single SN and for estimates of the ensemble of SNe, ensuring the attainment of a consistent global solution for the SN population. Each color represents one of the four independent chains. For example, the blue line in each panel is a different coordinate (projection) of the same MCMC chain.

(A color version of this figure is available in the online journal.)

thinned out the chains by recording only every 40th value. This reduces the autocorrelation between successive recorded samples and saves memory. To assess convergence, we computed the Gelman–Rubin (G-R) statistic (Gelman & Rubin 1992) for each parameter in the chain to compare the coverages of the independent chains. We considered a maximum G-R ratio less than 1.10 to indicate convergence. We discarded the first 20% of each chain as burn-in, and the chains were concatenated for analysis.

5. RESULTS: POSTERIOR INFERENCES

In this section, we report the posterior inferences of light curves and the population when the training set consists of all the SNe and their redshifts $(\mathcal{D}, \mathcal{Z})$. We report the posterior inference obtained when adopting Case 5 ($m = 1$) for the (A_V, r_V) population model, which models linear trends between the dust slope r_V and the dust extinction A_V. Posterior inferences can be described in terms of light curve fits and dust estimates for individual SNe, intrinsic covariances in the population of SN light curves, and the population distribution and correlations of host galaxy dust properties.

5.1. Individual Supernovae

Optical and NIR light curve fits in the rest frame are shown for one SN, SN 2005eq, in Figure 3. The points are the measured magnitudes in the observer frame minus the estimated K-corrections and Milky Way extinction in each passband. The black curves represent the fitted apparent light curves in each rest-frame passband, with each light curve represented by the differential decline rates model (Appendix A). The peak
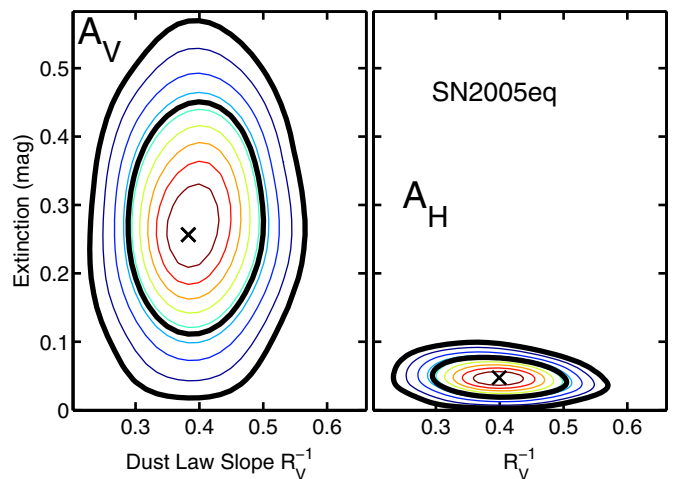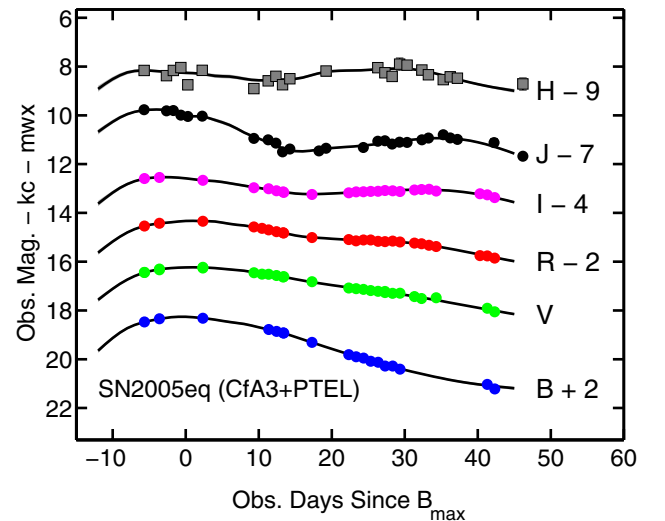

Figure 3. Top: optical (CfA3; Hicken et al. 2009a) and NIR (PAIRITEL; WV08) observations of nearby SN Ia 2005eq are fitted with a multi-band light curve model. The points are the observed magnitudes in each filter minus estimated K-corrections and Milky Way extinction. Bottom: optical and NIR light curves of SN 2005eq are used to infer the host galaxy dust extinction properties. The hierarchical model enables coherent inference of host galaxy dust properties (A_V, R_V) (assuming a CCM dust law), while marginalizing over the posterior uncertainties in the dust and SN light curve populations. The cross indicates the marginal bivariate mode, and the two black contours contain 68% and 95% of the posterior probability. The inferred NIR extinction A_H is much smaller than the optical extinction A_V and has much smaller uncertainty. This SN exhibits moderate extinction and reddening due to host galaxy dust.

(A color version of this figure is available in the online journal.)

apparent magnitudes for each SN and the decline rate $\Delta m_{15}(B)$ are listed in Table 4.

We also depict the posterior inferences of the dust properties: the visual extinction A_V, the NIR extinction, A_H, and the slope of the extinction law $r_V \equiv R_V^{-1}$. The bivariate marginal probability densities were estimated from the MCMC samples using kernel density estimation. The marginal distributions integrate over the posterior uncertainties in individual light curve fits and the population distribution. For SN 2005eq, we find a moderate amount of visual extinction, $A_V \sim 0.3$ mag. We can see from the side-by-side comparison that not only is the H-band extinction about five times smaller, but its uncertainty is also much smaller.

Since dust extinction is nonnegative, $A_V \geqslant 0$, the posterior probability densities of the dust parameters is highly non-Gaussian for SNe with low extinction. For example, from
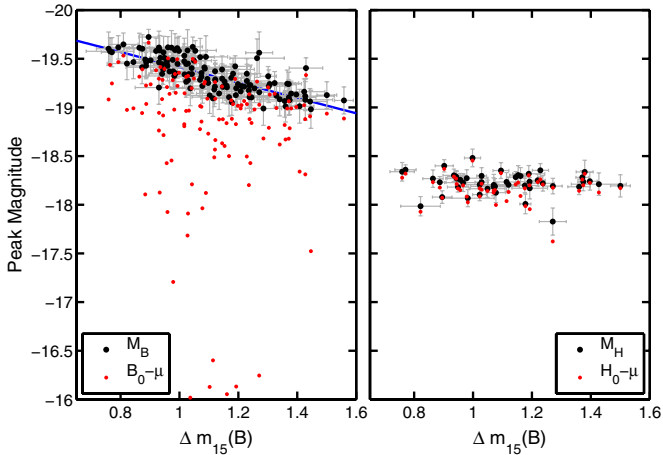
Figure 4. Left: post-maximum optical decline rate $\Delta m_{15}(B)$ vs. posterior estimates of the inferred optical absolute magnitudes M_B (black points) and the extinguished magnitudes $B_0 - \mu$ (red points). Each black point maps to a red point through optical dust extinction in the host galaxy. The intrinsic light-curve-width–luminosity Phillips relation is reflected in the trend of the black points, indicating that SNe brighter in B have slower decline rates. The blue line is the linear trend of Phillips et al. (1999). Right: inferred absolute magnitudes and extinguished magnitudes in the near-infrared H band. The extinction correction, depicted by the difference between red and black points, is much smaller in H than in B. The absolute magnitudes M_H have no correlation with $\Delta m_{15}(B)$. The standard deviation of peak absolute magnitudes is also much smaller for M_H compared to M_B.

(A color version of this figure is available in the online journal.)

Table 4, we infer that SN 2006ax has little host galaxy dust extinction with the most likely value being $A_V = 0.01$ mag. However, it is uncertain enough that $A_V = 0.12$ still lies within 68% highest posterior density (HPD) contour. By contrast, the A_H estimate is near zero, and the 68% contour lies within $A_H < 0.03$. Even SNe with low extinction benefit from observations in the H-band by reducing the uncertainty in the dust estimate. Table 4 lists summary statistics of the marginal posterior distribution of each host galaxy dust parameter for each SN, obtained from the MCMC samples.

5.2. Intrinsic Correlation Structure of SN Ia Light Curves in the Optical–NIR

We use the hierarchical model to infer the intrinsic correlation structure of the absolute SN Ia light curves. This correlation structure captures the statistical relationships between peak absolute magnitudes and decline rates of light curves in multiple filters at different wavelengths and phases. We summarize inferences about light curve shape and luminosity across the optical and near-infrared filters; a more detailed analysis of the intrinsic correlation structure of colors, luminosities, and light curve shapes will be presented elsewhere.

5.2.1. Intrinsic Scatter Plots

The hierarchical model fits the individual light curves with the differential decline rates model and infers the absolute magnitudes in multiple passbands, corrected for host galaxy dust extinction. For each individual SN light curve, we can use the inferred local decline rates $\boldsymbol{d}^F$ to compute the $\Delta m_{15}(F)$ of the light curve in each filter. In the left panel of Figure 4, we plot the posterior estimate of the peak absolute magnitude M_B versus its canonical $\Delta m_{15}(B)$ decline rate with black points. The error bars reflect measurement errors and the marginal uncertainties from the distance and inferred dust extinction.
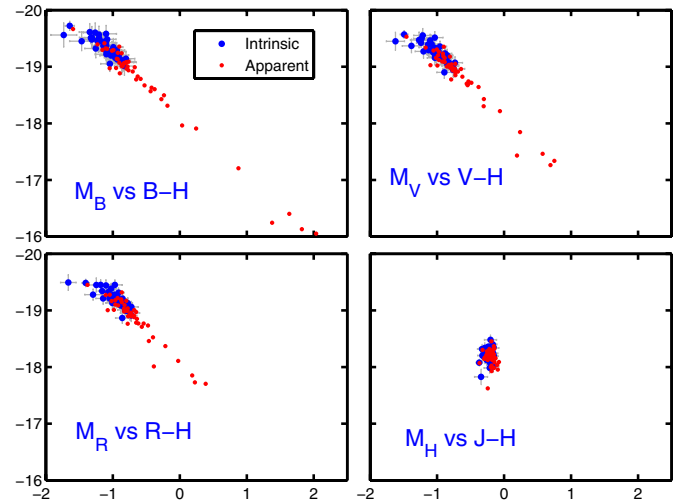

Figure 5. Inferred absolute magnitudes M_F (blue points) and the extinguished magnitudes $F_0 - \mu$ (red points) vs. colors relative to NIR H (intrinsic: blue points; apparent: red points). Only SNe with complete $BVRIJH$ data are plotted. The intrinsically optically bright SNe tend to be intrinsically bluer in the optical–NIR color. The H-band absolute magnitudes have no trend with intrinsic $J - H$ colors, which have a comparatively narrow distribution. Note that the magnitude and color axes have the same scale in each panel.

(A color version of this figure is available in the online journal.)

This set of points describes the well-known intrinsic light curve decline rate versus luminosity relationship (Phillips 1993). We also show the mean linear relation between M_B and $\Delta m_{15}(B)$ found by Phillips et al. (1999), who analyzed a smaller sample of SNe Ia. The statistical trend found by our model is consistent with that analysis. The red points are simply the peak apparent magnitudes minus the distance moduli, $B_0 - \mu$, which are the extinguished peak absolute magnitudes $M_B + A_B$. Whereas the range of extinguished magnitudes spans ~3 mag, the intrinsic absolute magnitudes lie along a narrow, roughly linear trend with $\Delta m_{15}(B)$.

In the right panel, we plot the intrinsic and extinguished absolute magnitudes of SNe Ia in the H band. In contrast to the left panel, the differences between the intrinsic absolute magnitudes and the extinguished magnitudes are nearly negligible. Notably, there is no correlation between the intrinsic M_H in the NIR and optical $\Delta m_{15}(B)$. This was noted previously by Krisciunas et al. (2004a) and WV08. The standard deviation of absolute magnitudes is much smaller in H than in B, demonstrating that the NIR SN Ia light curves are good standard candles (Krisciunas et al. 2004a, 2004c; WV08; Mandel et al. 2009). Theoretical models of Kasen (2006) indicate that NIR peak absolute magnitudes have relatively weak sensitivity to the input progenitor ^{56}Ni mass, with a dispersion of ~0.2 mag in J and K, and ~0.1 mag in H over models ranging from 0.4 to 0.9 solar masses of ^{56}Ni. The physical explanation may be traced to the ionization evolution of the iron group elements in the SN atmosphere.

These scatter plots convey some aspects of the population correlation structure of optical and near-infrared light curves that is captured by the hierarchical model. In the next section, we further discuss the multi-band luminosity and light curve shape correlation structure in terms of the estimated correlation matrices.

Figure 5 shows scatter plots of optical–near-infrared colors $(B - H, V - H, R - H, J - H)$ versus absolute magnitude (M_B, M_V, M_R, M_H) at peak. The blue points are the posterior estimates of the inferred peak intrinsic colors and absolute
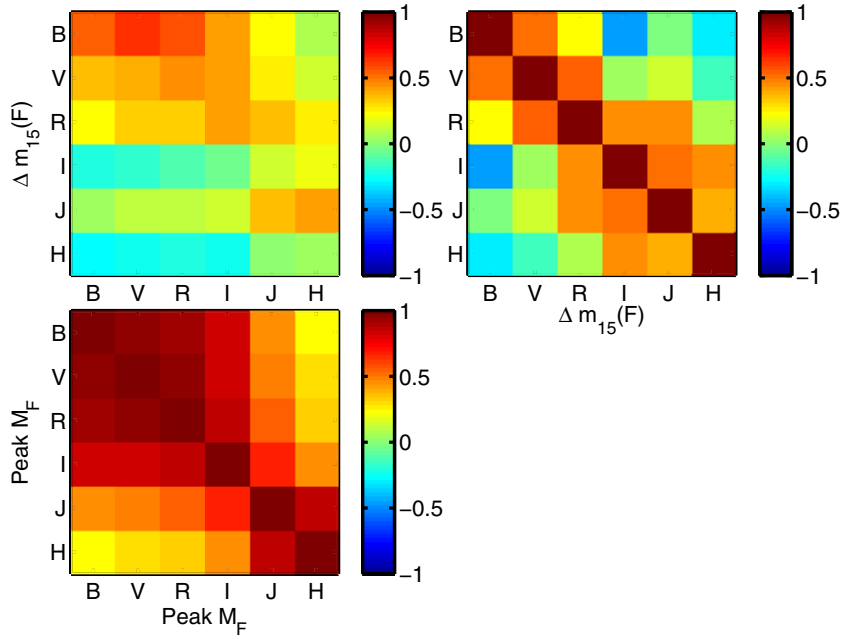
Figure 6. Intrinsic correlation structure of optical–near-infrared peak absolute magnitudes M_F and decline rates $\Delta m_{15}(F)$. The color scale encodes the strength of the correlation between any two intrinsic quantities. The marginal mode for each correlation coefficient is depicted, integrating over the uncertainties of individual SNe. The posterior uncertainties in the correlations are computed but not shown. Top left: correlations between the peak intrinsic absolute magnitude and decline rate. Top right: correlations between decline rates for each pair of passbands. Bottom: correlations between peak intrinsic absolute magnitudes for each pair of filters.

(A color version of this figure is available in the online journal.)

magnitudes of the SN, along with their marginal uncertainties. Red points are the peak apparent colors and extinguished absolute magnitudes, including host galaxy dust extinction and reddening. These plots show correlations between the peak optical–near-infrared colors and peak optical luminosity, in the direction of intrinsically brighter SNe having bluer peak colors. In contrast, the intrinsic $J - H$ colors have a relatively narrow distribution, and the near-infrared absolute magnitude M_H is uncorrelated with intrinsic $J - H$ color.

5.2.2. Intrinsic Correlation Matrices

Using the hierarchical model, we compute posterior inferences of the population correlations between the different components of the absolute light curves of SNe Ia. This includes population correlations between peak absolute magnitudes in different filters, $\rho(M_F, M_{F'})$, correlations between the peak absolute magnitudes and light curve shape parameters (differential decline rates) in different filters, $\rho(M_F, \boldsymbol{d}^{F'})$, and the correlations between light curve shape parameters in different filters, $\rho(\boldsymbol{d}^F, \boldsymbol{d}^{F'})$. They also imply correlations between these quantities and intrinsic colors. This information and its uncertainty are captured in the posterior inference of the population covariance matrix $\boldsymbol{\Sigma}_\psi$ of the absolute light curve parameters $\{\boldsymbol{\psi}_s\}$. The posterior estimate of the absolute light curve population integrates over the posterior uncertainties in the individual light curves and the host galaxy dust estimates.

In Figure 6, we have distilled some of the information in this intrinsic covariance matrix to show the inferred intrinsic correlations. For brevity, instead of depicting correlations with every differential decline rate d_t^F, we only show correlations with the canonical 15 day post-maximum decline rate in each filter, $\Delta m_{15}(F)$. The correlations range from -1 to 1 and are color coded according to strength. The joint uncertainties of the correlations are computed but not shown. The bottom matrix

shows the posterior inferences of the correlation matrix of peak absolute magnitudes. The optical luminosities and light curve shapes are strongly correlated with each other, but not with the NIR. The J and H luminosities are strongly correlated with each other, but not with the optical. Since the NIR luminosities have low intrinsic correlation with the optical luminosities, they provide independent information on the distance.

The top left matrix shows the posterior inferences of the correlation matrix of the Δm_{15} decline rates and the peak absolute magnitudes in each filter. The decline rates Δm_{15} in BVR exhibit strong correlations with peak optical absolute magnitudes, but they show low correlation with peak NIR absolute magnitudes in J and H. The decline rates Δm_{15} in IJH exhibit little correlation with luminosities in any of the optical or near-infrared filters.

The top right matrix shows the posterior estimates of the correlations between the Δm_{15} light curve decline rates in each filter. The correlation matrix exhibits a band structure, with the largest correlations neighboring the diagonal. The decline rate Δm_{15} in a particular filter is typically most strongly correlated with the Δm_{15} in filters at neighboring wavelengths. This can be seen by examining each row of the correlation matrix. The optical decline rates in B and V are strongly or moderately correlated with each other, but have low correlation with NIR decline rates. Similarly, the decline rates in IJH show strong or moderate correlation with each other, but have lower correlation with the decline rates in B, V. The lack of strong correlation across the whole matrix indicates that the light curve shapes across optical and NIR wavelengths are unlikely to be adequately modeled with one degree of freedom.

These matrix plots depict some of the salient population correlation information of the SN Ia absolute light curves captured by the hierarchical model. This inferred correlation structure is used by the model to estimate luminosities from the light curves and to make distance predictions.
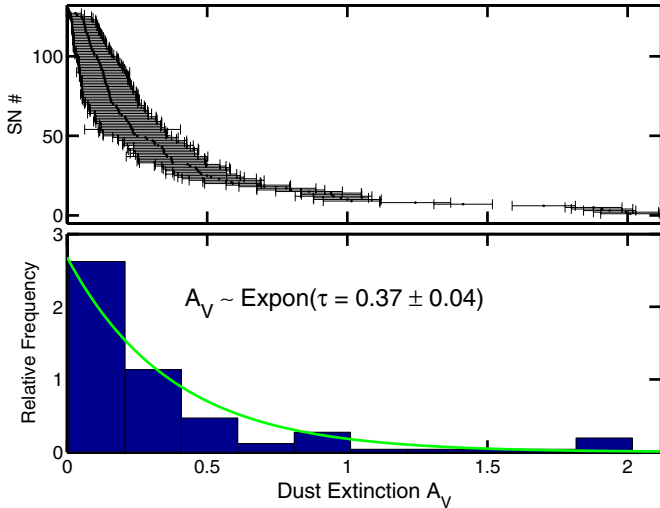
13

Figure 7. Distribution of inferred host galaxy dust extinction A_V. The hierarchical model estimates the extinction to each SN using the optical and near-infrared light curves, and models the dust population. Top: A_V estimates and uncertainties of each SN ranked from highest to lowest extinction. Bottom: histogram of the modal A_V estimates plotted against the fitted exponential distribution for the dust population.

(A color version of this figure is available in the online journal.)

5.3. Posterior Inference of the Host Galaxy Dust Population

5.3.1. Linear Correlation Dust Population Model

In this section, we describe posterior inferences for the host galaxy dust population. From the samples of the global posterior, Equation (19), we can estimate the host galaxy dust extinction, A_V, and the slope of the extinction law, $r_V = 1/R_V$, for each object and their uncertainties from their marginal distributions. We also estimate the characteristics of the dust population through the hyperparameters, τ_A, $\boldsymbol{\beta}$, and σ_r^2, while accounting for global uncertainties.

In Figure 7, we display the histogram of the A_V estimates for each SN in the sample, along with the individual marginal estimates and their uncertainties. This is compared against an exponential probability distribution with the marginal estimate of the extinction scale $\tau_A = 0.37 \pm 0.04$. There appears to be an overabundance of objects at $A_V > 1.5$ compared to the exponential distribution. We explore this further in Section 6.

In Figure 8, we display the estimates of (A_V, R_V) for the $m = 1$ population model, described in Section 2.6. This model assumes that the mean trend of r_V versus A_V is linear in A_V. Fitting the hierarchical model then entails computing posterior estimates of (A_V, r_V) for individual objects and the population trend, parameterized by $\boldsymbol{\beta}$, σ_r^2. For SNe at low A_V, the r_V parameter for each individual SN cannot be estimated precisely, since it only enters into the extinction model, Equation (9), and thus, the likelihood, multiplied by A_V. For these SNe, there is not enough information in individual light curves to distinguish between the individual r_V estimates, and so the model pools them toward the group mean or trend. At high A_V, the r_V parameter can be estimated more precisely for each SN, so they can be individually distinguished. In the top panel, we show the A_V, R_V values for each SN for three joint samples from the MCMC chain. A joint sample of $\{A_V^s, R_V^s\}$, $\boldsymbol{\beta}$, σ_r^2 represents a single probable realization of these parameters given the data, and is labeled by a single color. The R_V estimates at low A_V show considerable scatter between samples, reflecting the underlying uncertainty. At high A_V, there is less scatter between individual
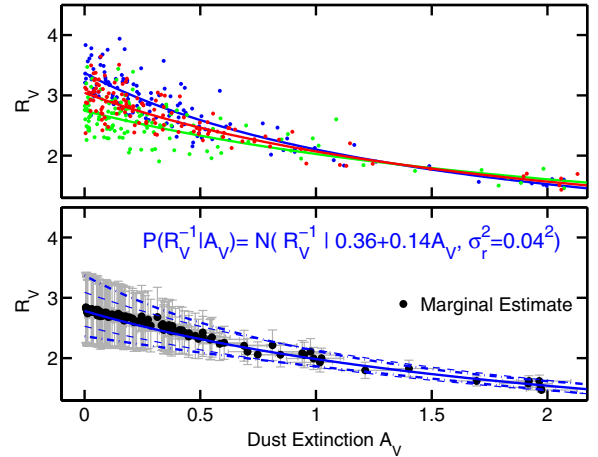

Figure 8. Apparent correlation between host galaxy dust visual extinction A_V and dust law slope R_V in the sample of SNe Ia. This model assumes a dust population where R_V^{-1} has a linear trend with A_V with some rms scatter σ_r. The linear regression coefficients and residual scatter $(\boldsymbol{\beta}, \sigma_r^2)$ are estimated from the marginal global posterior distribution. Top: the points of each color and the regression relation are different probable realizations of (A_V, R_V) for each SN and dust population hyperparameters, $\boldsymbol{\beta}$ and σ_r^2, obtained from snapshots of the MCMC. The R_V estimates at low extinctions have more uncertainty than those at high extinctions, as reflected by the scatter of points with different colors. Bottom: averaging over all probable realizations, we plot the inferred marginal posterior mode of (A_V, R_V) and their marginal uncertainties for each SN with the marginal estimates of the regression model. When the individual R_V estimates for single SN are very uncertain, they tend to be pulled toward the population mean value (for its extinction A_V) using partial pooling. The data favor an apparent non-zero correlation between A_V and the dust slope R_V^{-1}. SN Ia light curves with low to moderate extinction are consistent with the Milky Way average $R_V \sim 3.1$ for interstellar extinction, but for highly extinguished SNe, a low value of $R_V \lesssim 2$ is favored.

(A color version of this figure is available in the online journal.)

SNe and between samples, reflecting the increased precision for estimating R_V. In the bottom panel, each point and error bar represents the marginal estimate, averaging over all the MCMC samples, of (A_V, R_V) for each SN.

In Figure 9, we show the bivariate marginal probability density of the regression parameters $\boldsymbol{\beta} = (\beta_0, \beta_1)$, obtained from the MCMC samples. The joint mode and the 68% and 95% HPD contours are shown. The intercept β_0 represents the population mean value of r_V at vanishing $A_V \to 0$, and β_1 represents the population mean linear trend of r_V against A_V. Also shown is the value of r_V corresponding to the Milky Way interstellar average $R_V = 3.1$. The intercept β_0 at vanishing A_V is uncertain, but consistent with the Milky Way average within 1σ. The regression slope β_1 is positive with zero excluded from the 95% credible region. The marginal estimates of each of the regression parameters are listed in Table 1. The marginal estimate of $\beta_0 = 0.35 \pm 0.05$ can be compared against $r_V = 0.32$ for the Milky Way average. The characteristic value of R_V as $A_V \to 0$, β_0^{-1}, is uncertain because of the difficulty of determining R_V for low-extinction objects. The marginal posterior density of β_0^{-1} has a non-Gaussian profile: the mean is 2.9, the mode is 2.7, and the interval containing 68% of the highest probability density is [2.3, 3.3]. The marginal probability that $\beta_0^{-1} < 2$ is $p = 0.02$. The marginal estimate of the slope is $\beta_1 = 0.15 \pm 0.03$. This is a strong indication of a differential trend of r_V versus A_V in the host galaxy dust population of nearby SNe.

These results were consistent when changing the peculiar velocity dispersion σ_{pec} from 150 to 300 km s^{-1}. The posterior mean of β_0^{-1} was 2.8, the mode was 2.5, and the 68%
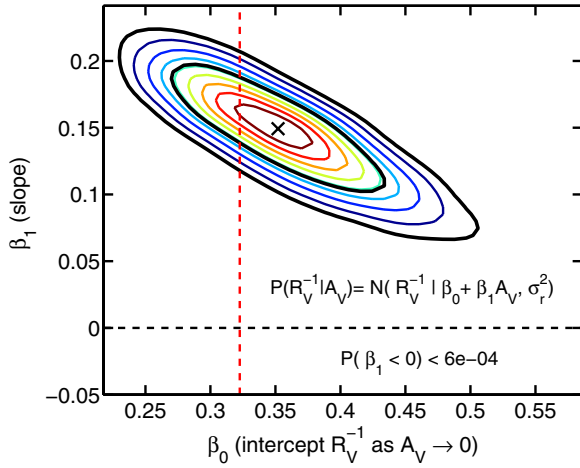
Figure 9. Marginal posterior distribution of the linear regression coefficients of the dust population model assuming a linear mean trend between R_V^{-1} and A_V. The parameter β_1 is the slope of R_V^{-1} against extinction A_V, and β_0 is the population mean value of R_V^{-1} in the low-extinction limit. The two-dimensional mode is marked, and the inner and outer solid black lines contain 68% and 95% of the marginal probability, respectively. The posterior estimate of β_0 is consistent with $R_V = 3.1$ (vertical red dashed line) and inconsistent with $R_V < 2$, $P(\beta_0 > 0.5) = 0.02$. The posterior estimate of the regression slope β_1 is extremely inconsistent with zero (horizontal dashed line).

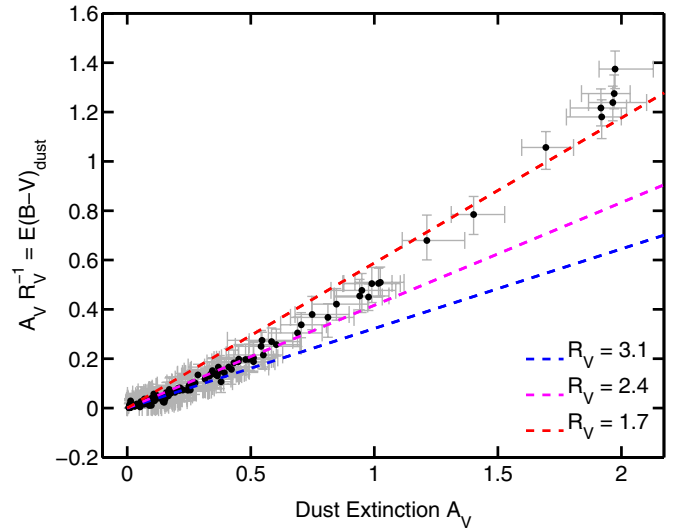(A color version of this figure is available in the online journal.)

Figure 10. Marginal posterior estimates of inferred color excess $E(B - V)$ due to host galaxy dust vs. inferred extinction A_V, assuming the linear correlation model. This model assumes a dust population where R_V^{-1} has a linear trend with A_V with some rms scatter σ_r. The SNe Ia at lower extinction ($A_V \lesssim 1$) have implied color excesses consistent with $R_V = 2$–3. The SNe Ia at higher extinction favor a dust law with $R_V < 2$.

(A color version of this figure is available in the online journal.)

Table 1
μ Prediction Errors for R_V Scenarios

Assumptions on R_V Population	Inferred Hyperparameters	Opt. (mag)	Opt+NIR (mag)
$R_V = 3.1$	$\cdots$	0.20	0.13
Complete pooling	$R_V = 1.6 \pm 0.1$	0.15	0.13
No pooling	$\cdots$	0.16	0.12
PP: $m = 0$	$\mu_r^{-1} = 1.7 \pm 0.1$, $\sigma_r = 0.04 \pm 0.02$	0.16	0.12
PP: $m = 1$	$\beta_0 = 0.35 \pm 0.05$ $\beta_1 = 0.15 \pm 0.03$ $\sigma_r = 0.04 \pm 0.02$	0.15	0.11
PP: 4-Steps	c.f. Table 3	0.15	0.11

Notes. Optical and optical+NIR rms prediction errors at $cz > 3000$ km s^{-1} for different dust population models. Estimates of hyperparameters are the marginal posterior means and standard deviations. The rms prediction errors are the 0.632 bootstrap cross-validation estimates. Sampling variance of prediction errors is typically ± 0.01 mag.

Table 2
Inference for the "2-Step" R_V Population Model

A_V Range	μ_r^{-1}	μ_r	σ_r	p_{tail}
[0, 0.8]	2.3 ± 0.3	0.45 ± 0.06	0.04 ± 0.02	0.008
>0.8	1.7 ± 0.1	0.59 ± 0.04	0.06 ± 0.03	$\cdots$

Notes. The hyperparameter μ_r is the population mean R_V^{-1} for SNe in each interval in A_V. The hyperparameter σ_r^2 is the population variance of R_V^{-1} in each interval. Estimates are marginal posterior means and standard deviations. The marginal posterior density of σ_r is highly non-Gaussian. The estimate of μ_r^{-1} is not precisely the inverse of the estimate of μ_r, because of uncertainty and the nonlinear transformation. The marginal probability that μ_r is larger than μ_r of the high extinction bin is p_{tail}.

interval containing highest probability density was [2.1, 3.3]. The marginal estimate of the slope is $\beta_1 = 0.14 \pm 0.04$.

Figure 10 plots the posterior estimate of the inferred optical reddening $E(B - V) \equiv A_B - A_V$ due to host galaxy dust versus the estimated dust extinction A_V, assuming the linear correlation model. The reddening estimates at $A_V > 1.5$ favor a $R_V = 1.7$ reddening law, whereas at lower extinction, $A_V < 1$, the reddening estimates are consistent with $2.4 \lesssim R_V \lesssim 3.1$.

5.3.2. Step Function Dust Population Model

The linear correlation model, $m = 1$, assumes that the mean trend of r_V with A_V is linear across the entire range of A_V. However, we do not know if this assumption is true. To test the sensitivity of the apparent differential trend in r_V versus A_V to the linear correlation assumption of the $m = 1$ model, we fit alternate models using the "Step" Case 6 of Section 2.6.

Instead of using all the SNe over the range of A_V to determine a linear correlation in the host galaxy dust population, this case groups together only SNe in the same interval of A_V to determine their group mean and variance of $r_V = R_V^{-1}$ in each bin. We have retrained the hierarchical model using the step function assumptions, first by dividing the range in A_V into "high" ($A_V > 0.8$) and "low" ($A_V < 0.8$) extinction bins. Second, we subdivided those bins and retrained the model using four bins in extinction, A_V.

Marginal posterior estimates for the "2-step" model are listed in Table 2. The low extinction bin, with $A_V < 0.8$, has a group R_V mean of 2.3 ± 0.3, while the high extinction bin, $A_V > 0.8$, has a group R_V mean of 1.7 ± 0.1. We examined the marginal posterior probability of the difference between the group means μ_r at low and high extinction. This calculation takes into account the posterior covariance between the estimates and marginalizes over uncertainties in the other parameters. The tail probability that μ_r of the low extinction bin is greater than μ_r of the high extinction bin is denoted p_{tail}, and is computed directly from the MCMC samples. We find less than 1% probability that the difference is positive, suggesting that the difference is significant.

Table 3
Inference for the "4-Step" R_V Population Model

A_V Range	μ_r^{-1}	μ_r	σ_r	p_{tail}
[0, 0.4]	2.9 ± 0.7	0.35 ± 0.08	0.03 ± 0.02	<0.001
[0.4, 0.8]	2.3 ± 0.3	0.45 ± 0.06	0.03 ± 0.03	0.005
[0.8, 1.25]	2.1 ± 0.2	0.48 ± 0.04	0.03 ± 0.03	0.004
>1.25	1.6 ± 0.1	0.63 ± 0.03	0.04 ± 0.03	⋯

Notes. The hyperparameter μ_r is the population mean R_V^{-1} for SNe in each interval in A_V. The hyperparameter σ_r^2 is the population variance of R_V^{-1} in each interval. Estimates are marginal posterior means and standard deviations. The marginal posterior density of σ_r is highly non-Gaussian. The estimate of μ_r^{-1} is not precisely the inverse of the estimate of μ_r, because of uncertainty and the nonlinear transformation. The marginal probability that the mean μ_r of each bin is larger than μ_r of the highest extinction bin is p_{tail}.

For the "4-step" model, the posterior inferences of the hyperparameters in each of the four intervals in A_V are listed in Table 3. The group means, μ_r, for each interval display the same trend of lower R_V for higher A_V. The bin with $0 \leqslant A_V < 0.4$ has a group mean consistent with $R_V \approx 3$, the interstellar average for the Milky Way. However, the group mean for the lowest extinction SN is uncertain due to the difficult of determining r_V at low A_V. The marginal posterior density of the characteristic R_V in the lowest extinction bin, μ_r^{-1}, is non-Gaussian: the mean is 2.9, the mode is 2.5 and the interval containing 68% of the highest probability density is [2.1, 3.3]. The marginal probability that μ_r^{-1} of the lowest extinction bin is less than 2 is $p = 0.04$. The highest extinction bin, $A_V > 1.25$, favors a group mean $R_V = 1.6 \pm 0.1$. We calculated p_{tail}, the marginal probability that the group mean μ_r of each bin is greater than μ_r of the highest extinction bin. Each of the lower extinction bins had significantly different group mean μ_r estimates than that of the highest extinction bin.

For an assumed peculiar velocity dispersion $\sigma_{pec} = 300$ km s^{-1}, the results were consistent. For the lowest extinction bin, the non-Gaussian marginal probability density of μ_r^{-1} had a peak at 2.5, with mean 2.7 and a 68% interval [2, 3]. The posterior mean and standard deviation of μ_r^{-1} in the highest extinction bin was 1.5 ± 0.1. The inferences for the alternate step function model confirm the trend seen in the $m = 1$ linear correlation model. We discuss the implications of these findings in Section 8.

5.3.3. Other Dust Population Models

Posterior inferences of the hyperparameters using the other models for the host galaxy dust population (Section 2.6) are listed in Table 1. In the case of CP, in which it is assumed that all SNe have the same value of R_V, the marginal estimate of that value is $R_V = 1.6 \pm 0.1$. In the population model $m = 0$, in which each r_V is drawn from a Gaussian with mean independent of A_V, we find a population mean with a similar value ($\mu_r^{-1} = 1.7 \pm 0.1$). These results indicate that the highly extinguished SNe dominate the estimate of the global constant or population mean in these cases, since their individual R_V estimates are the most precise. The CP and $m = 0$ are special cases of the $m = 1$ model. If the $m = 0$ model were favored then when fitting the $m = 1$ model we should have found $\beta_1 \approx 0$. If CP were favored then we would have also found that $\sigma_r^2 = 0$. We inferred none of those in the expanded $m = 1$ model; this illustrates the pitfalls of those simpler assumptions.
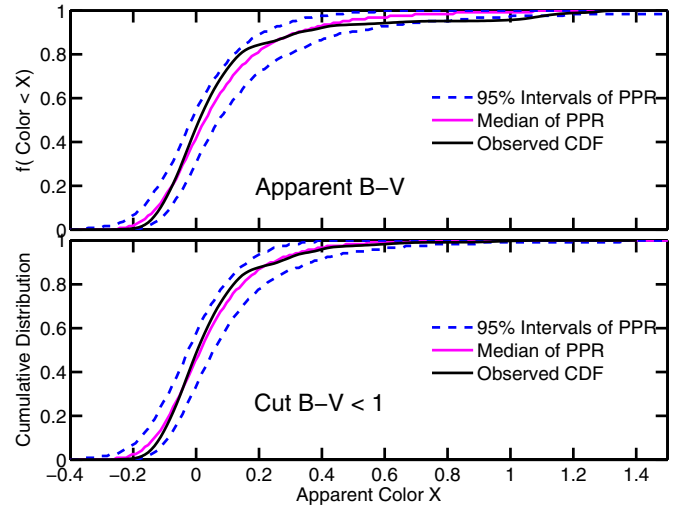

Figure 11. Comparison of the cumulative distribution of peak apparent $B - V$ colors to those of posterior predictive replication sets randomly generated from the trained model. Top: when the hierarchical model is trained on our full sample, the distribution of apparent colors of the observed sample has a thicker tail toward the extreme red ($B - V \gtrsim 0.8$–1 mag) than most of the replicated sets. Bottom: when the hierarchical model is retrained on a sample restricted to apparent $B - V < 1$, the apparent color distribution agrees well with replicated SN sets. This suggests that the exponential model population distribution for extinction inadequately accounts for the number of SNe observed at very high reddening.

(A color version of this figure is available in the online journal.)

6. MODEL CHECKS

After fitting the hierarchical model by computing the global posterior density, Equation (19), using our BAYESN code, we checked the model fit using several methods. We did this to ensure first that the MCMC code was fitting the assumed statistical model to the data set, and to diagnose technical or algorithmic errors. Second, we checked the fit of the hierarchical model to the observed sample to look for disagreements between the assumptions and the observed data. Third, we tested the robustness of the model to the training set and evaluated the distance prediction error by performing extensive CV (Section 7.2). For individual SNe, we inspected the fits of the light curve model to the photometric data (Section 5.1).

To check the fit of the model population distributions to the apparent distributions of the data set, we performed posterior predictive model checks (Rubin 1984; Gelman et al. 1996, 2003). From the trained hierarchical model, we generated a new random set of apparent light curves of the same size as the observed sample. The replicated set was generated by sampling forward through the DAG, Figure 1. The distribution of apparent properties of the replicated light curves was compared to those of the observed set. We illustrate such a comparison of peak apparent $B - V$ optical colors in Figure 11. We generated 1000 replications, each containing the same number of SNe as the observed sample. The apparent colors of the SNe within each replicated set have a cumulative distribution function. The distribution of apparent $B - V$ colors is the convolution of the intrinsic $B - V$ color distribution and the dust $E(B - V)$ color excess distribution implied by the extinction distribution. The set of replications form an ensemble of color distributions, reflecting random sampling variation and posterior uncertainty in the model. For each value of the $B - V$ color we show the median, 2.5% and 97.5% quantiles of the ensemble of cumulative distribution functions (CDFs) at that value. The

black curve is the CDF of the apparent colors of the observed data set. If the observed CDF lies outside the 95% range of the replications, then the observed distribution disagrees with the model's replications.

In the top panel of Figure 11, the observed distribution of peak apparent $B - V$ colors has a significantly thicker tail toward redder (positive) colors than the replicated distributions. This suggests that the number of very red SNe with $B - V > 1$ is large compared to what can be expected with the exponential model for the dust population. The abundance of very red SNe in the nearby sample might be a consequence of preferred selection of these events for follow-up observation.

To test whether the model distribution adequately describes the SNe with less reddening, we removed the 4% of SNe with apparent $B - V > 1$ from the training set, retrained the whole hierarchical model, and again generated posterior predictive replciations (bottom panel, Figure 11). There is good agreement between the apparent color distributions of the ensemble of replications and the observed data set. With the color cut, the estimated exponential scale of the extinction distribution decreased from $\tau_A = 0.37 \pm 0.04$ mag to $\tau_A = 0.28 \pm 0.04$, so that the model captures a dust extinction distribution with a thinner tail, which implies a narrower apparent color distribution. We found that β was consistent within the uncertainties with the values found by using the whole sample. This demonstrates that the trend is not determined just by the reddest outliers of the SN sample.

A key assumption of the model is that the two populations, the SN Ia light curves and the dust extinction, are statistically independent. This entails that an intrinsically faint or red SN has the same chance of encountering a particular level of host galaxy dust extinction as an intrinsically bright or blue SN. We expect that the amount of extinction to SNe should be uncorrelated with the intrinsic properties of the SN Ia light curves. A significantly non-zero relationship between the two might indicate a miscalibration of the model, possibly related to a confusion between intrinsic color variation and dust extinction. We tested this hypothesis, as shown in Figure 12, where we plot the fitted intrinsic $\Delta m_{15}(B)$ decline rates and the inferred intrinsic $B - I$ color of SN Ia light curves versus the inferred dust extinction. The plots show the expected lack of correlation between the parameters from the two populations, and is a consistency check on the model fit.

7. DISTANCE PREDICTION

7.1. Hubble Residuals under Resubstitution

After training the model on all the SNe in the sample $(\mathcal{D}, \mathcal{Z})$, the distance modulus for each SN can be estimated by resubstituting each light curve set into the model and computing the posterior predictive density $P(\tilde{\mu}_s | \tilde{\mathcal{D}}_s, \tilde{z}_s; \mathcal{D}, \mathcal{Z})$, which marginalizes over the uncertainty in the trained model. The expected value of this density is μ_{resub}^s. The Hubble residual is the difference between the resubstitution distance modulus and the distance modulus expected from the redshift and the Hubble law, $f(z_s) = \mathbb{E}(\mu_s | z_s)$. The estimates $\mathbb{E}(\mu_s | z_s)$, $\sigma_{\mu,s}$, and μ_{resub} are listed in Table 5. The uncertainty-weighted mean square resubstitution error, $\text{err}_{\text{resub}}^2$, is computed as a sum over all SNe, using Equation (31) of Mandel et al. (2009). For the $m = 1$ dust model, the error-weighted rms of the Hubble residuals at $cz > 3000$ km s^{-1} is 0.13 mag for the full sample. However, for the SNe with NIR data, the resubstitution error at
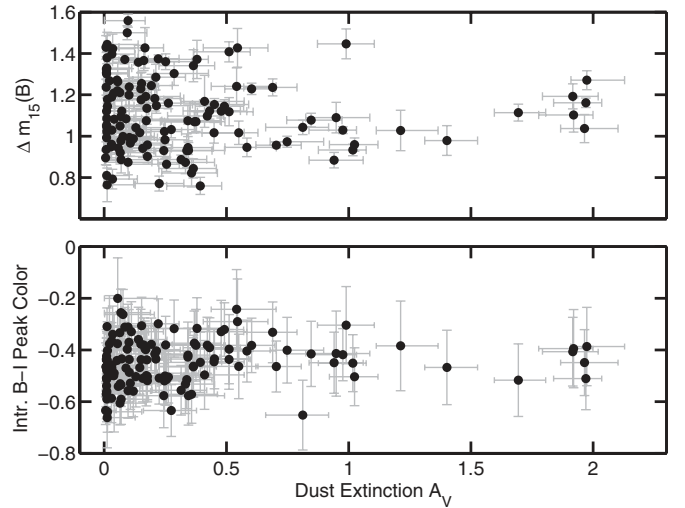

Figure 12. Top: fitted intrinsic $\Delta m_{15}(B)$ decline rate of SN Ia light curves vs. inferred dust extinction. There is no significant trend between $\Delta m_{15}(B)$ and A_V. Bottom: inferred intrinsic $B - I$ color at peak vs. the inferred dust extinction. There is no significant trend between peak intrinsic $B - I$ and A_V.

$cz > 3000$ km s^{-1} is 0.10 mag, and those with only optical data have a resubstitution error of 0.14 mag.

7.2. Cross-validation and Prediction Error

For finite samples, the rms Hubble diagram residual of the training set SN is an optimistic estimate of the ability of the statistical model to make accurate distance predictions given the SN observables. This is because it uses the SN data twice: first for estimating the model parameters (training) and second for evaluating the residual error. To evaluate predictive performance and guard against overfitting with a statistical model based on finite data, we should estimate the prediction error for SNe not included in the training set ("out-of-sample"). We use cross-validation to evaluate the utility of optical and NIR light curves for accurately predicting distances in the Hubble diagram, and to test the sensitivity of the model to the finite training set. The importance of cross-validating statistical models for predicting SN Ia distances has been discussed by Mandel et al. (2009) and Blondin et al. (2011).

To estimate the distance prediction error of our statistical model, we have performed bootstrap CV. This method was first used for assessing distance predictions of SN Ia light curve models by Mandel et al. (2009). From the full set of SNe, a new bootstrapped training set is created by sampling with replacement individual SNe up to the same size as the original set. The complement of this new training set forms a validation or prediction set. The training set light curves and redshifts are used to build the statistical model for SN Ia light curves, with the model hyperparameters estimated using hierarchical Bayesian inference and MCMC. The prediction set light curves are used to generate distance predictions for those SNe, which are then compared to the Hubble distances expected from their redshifts.

We randomly bootstrapped 30 training sets, so that on average each SN was held out for distance prediction 11 times. For each SN s, the expected value of the posterior predictive probability density, $\mu_{\text{pred},B}^s \equiv \mathbb{E}(\tilde{\mu}_s | \tilde{\mathcal{D}}_s, \tilde{z}_s; \mathcal{D}^B, \mathcal{Z}^B)$, is a point estimate of the distance modulus prediction under the training set data $\mathcal{D}^B$, $\mathcal{Z}^B$ for training set B. The .632 bootstrap estimate (Efron 1983; Efron & Tibshirani 1997) of rms prediction error

Table 4
Apparent Light Curve and Dust Estimates for Individual SNe Ia

SN	B_0[a]	$\Delta m_{15}(B)$	V_0[b]	R_0	I_0	J_0	H_0	$\hat{A}_V$[c]	68%(A_V)[d]	R_V[e]	Ref.[f]
SN1998bu	12.11 ± 0.01	1.03 ± 0.02	11.80 ± 0.01	11.65 ± 0.01	11.65 ± 0.01	11.74 ± 0.02	11.87 ± 0.03	0.97	[0.85, 1.07]	2.2 ± 0.3	J99,H00
SN1999cl	14.86 ± 0.03	1.17 ± 0.07	13.74 ± 0.02	13.27 ± 0.03	13.08 ± 0.03	12.96 ± 0.04	13.03 ± 0.04	1.95	[1.86, 2.10]	1.6 ± 0.1	K00
SN1999cp	13.93 ± 0.04	0.97 ± 0.07	13.98 ± 0.04	14.04 ± 0.04	14.32 ± 0.04	14.66 ± 0.05	14.90 ± 0.06	0.07	[0.00, 0.20]	2.7 ± 0.6	K00
SN1999ee	14.85 ± 0.01	0.96 ± 0.01	14.59 ± 0.01	14.51 ± 0.01	14.65 ± 0.02	14.95 ± 0.02	15.14 ± 0.03	0.77	[0.62, 0.81]	2.1 ± 0.2	K04b
SN1999ek	15.63 ± 0.02	1.14 ± 0.03	15.52 ± 0.02	15.41 ± 0.02	15.54 ± 0.02	15.82 ± 0.02	16.07 ± 0.02	0.46	[0.37, 0.58]	2.3 ± 0.3	K04c
SN1999gp	15.98 ± 0.03	0.86 ± 0.04	16.01 ± 0.02	16.02 ± 0.03	16.31 ± 0.03	16.83 ± 0.08	17.12 ± 0.09	0.23	[0.11, 0.33]	2.6 ± 0.5	K01
SN2000E	12.81 ± 0.04	1.08 ± 0.05	12.72 ± 0.03	12.66 ± 0.04	12.91 ± 0.05	13.41 ± 0.05	13.67 ± 0.06	0.43	[0.26, 0.53]	2.4 ± 0.4	V03
SN2000bh	15.89 ± 0.07	1.10 ± 0.09	15.90 ± 0.05	15.92 ± 0.05	16.22 ± 0.06	16.55 ± 0.06	16.79 ± 0.07	0.18	[0.08, 0.31]	2.7 ± 0.5	K04b
SN2000bk	16.92 ± 0.10	1.47 ± 0.10	16.80 ± 0.07	16.74 ± 0.07	16.78 ± 0.08	16.96 ± 0.10	17.17 ± 0.10	0.50	[0.36, 0.68]	2.4 ± 0.4	K01
SN2000ca	15.58 ± 0.03	0.95 ± 0.04	15.63 ± 0.03	15.69 ± 0.03	16.00 ± 0.03	16.52 ± 0.02	16.81 ± 0.03	0.01	[0.00, 0.10]	2.8 ± 0.6	K04b

Notes.
[a] Apparent magnitude at maximum light in the rest-frame B filter, corrected for Milky Way extinction and K-corrections. Estimates only listed if SN was observed in the filter.
[b] Apparent magnitude in rest-frame V at the time of maximum in B, corrected for Milky Way extinction and K-corrections.
[c] Marginal posterior mode of extinction A_V.
[d] Highest posterior density interval containing 68% of the marginal probability.
[e] Marginal posterior mean and standard deviation.
[f] Reference: CfA3: Hicken et al. 2009a; WV08: Wood-Vasey et al. PAIRITEL; 2008; WC3: WV08+CfA3; J99: Jha et al. 1999; H00: Hernandez et al. 2000; K00: Krisciunas et al. 2000; K01: Krisciunas et al. 2001; DP02: Di Paola et al. 2002; V03: Valentini et al. 2003; K03: Krisciunas et al. 2003; K04b: Krisciunas et al. 2004b; K04c: Krisciunas et al. 2004c; K07: Krisciunas et al. 2007; ER06: Elias-Rosa et al. 2006; ER07: Elias-Rosa et al. 2008; Pa07: Pastorello et al. 2007; St07: Stanishev et al. 2007; P08: Pignata et al. 2008.

(This table is available in its entirety in a machine-readable form in the online journal. A portion is shown here for guidance regarding its form and content.)

Table 5
Distance Modulus Predictions for SNe Ia

SN	cz (km s^{-1})	$\mu_{\mathrm{LCDM}}\|z$ (mag)	$\sigma_\mu\|z$ (mag)	μ_{resub} (mag)	$\bar{\mu}_{\mathrm{pred}}$ (mag)	s_{pred} (mag)	σ_{pred} (mag)	$\bar{A}_V^{\mathrm{pred}}$ (mag)
SN1998bu	708.90	29.97	0.46	30.00	29.95	0.02	0.10	0.97
SN1999cl	957.00	30.62	0.39	30.94	30.94	0.05	0.12	1.94
SN1999cp	2909.10	33.05	0.11	33.16	33.12	0.04	0.11	0.06
SN1999ee	3296.10	33.32	0.10	33.19	33.19	0.03	0.09	0.74
SN1999ek	5190.90	34.32	0.06	34.16	34.10	0.03	0.10	0.54
SN1999gp	8112.90	35.31	0.04	35.22	35.21	0.03	0.11	0.25
SN2000E	1803.00	32.00	0.18	31.59	31.61	0.04	0.12	0.46
SN2000bh	6765.00	34.90	0.05	34.98	34.94	0.03	0.12	0.14
SN2000bk	7976.10	35.27	0.04	35.34	35.34	0.06	0.17	0.50
SN2000ca	6989.10	34.98	0.05	34.88	34.89	0.03	0.10	0.01

Notes. $\mu_{\mathrm{LCDM}}\|z$ is the distance modulus expected from the redshift assuming $h = 0.72$, $\Omega_M = 0.27$, $\Omega_\Lambda = 0.73$, and $w = -1$. Its magnitude variance, assuming peculiar velocity dispersion $\sigma_{\mathrm{pec}} = 150$ km s^{-1}, is σ_μ^2. μ_{resub} is the distance modulus estimated under resubstitution. Under bootstrap cross-validation, the mean prediction over bootstraps is $\bar{\mu}_{\mathrm{pred}}$ and the standard deviation of predictions over boostraps is s_{pred}. Zero values of s_{pred} are less than 0.005 mag. The average standard deviation of uncertainty of a prediction is σ_{pred}. The marginal posterior mode of A_V under prediction, averaged over the prediction sets for each SN, is $\bar{A}_V^{\mathrm{pred}}$.

(This table is available in its entirety in a machine-readable form in the online journal. A portion is shown here for guidance regarding its form and content.)

is computed using the sum of uncertainty-weighted squared prediction errors over all bootstrapped sets, as described by Equations (32) and (33) of Mandel et al. (2009). For the $m = 1$ dust model, we list in Table 5 the predicted distance modulus for each SN, averaged over all the training sets B that do not include that SN, as $\bar{\mu}_{\mathrm{pred}}$. The standard deviation of predictive uncertainty, i.e., the square root of $\mathrm{Var}[\tilde{\mu}_s | \tilde{\mathcal{D}}_s, \tilde{z}_s; \mathcal{D}^B, \mathcal{Z}^B]$, averaged over the training sets B not containing that SN, is σ_{pred}. This measures the precision with which the trained model makes a distance prediction for a particular SN. The standard deviation

of the point estimates $\mu_{\mathrm{pred}, B}^s$ over all of those training sets, s_{pred}, is a measure of the sensitivity of the predicted distances to resampling the training set.

Figure 13 shows the predicted distances to the SN using bootstrap CV. For Hubble flow SN at $cz > 3000$ km s^{-1}, the cross-validated prediction error is 0.15 mag overall. For the SNe with optical and NIR data, the prediction error is estimated to be 0.11 mag, and for the SNe with optical light curves alone, the rms prediction error is 0.16 mag. The predicted distances to SNe with optical and NIR light curve measurements have a smaller scatter in the Hubble diagram than those with only optical data. These estimates of Hubble diagram scatter can be compared to the 0.18–0.22 mag rms found for the CfA3 sample using the MLCS2k2 and SALT2 methods (Hicken et al. 2009b).

The weighted rms prediction error measures the total Hubble diagram scatter, comprised of at least two components: a dispersion associated with unknown and random peculiar velocities with respect to the Hubble expansion, and an intrinsic variance that represents a floor to the precision of distance predictions. We compute this intrinsic component of the prediction error using the maximum likelihood estimator described in Appendix B of Blondin et al. (2011). Assuming a velocity dispersion $\sigma_{\mathrm{pec}} = 150$ km s^{-1}, the rms intrinsic prediction error was 0.15 ± 0.01 mag for SNe with optical data only, and 0.10 ± 0.02 for SNe with optical and near-infrared light curves. For $\sigma_{\mathrm{pec}} = 300$ km s^{-1}, the weighted rms prediction error remains the same, but the scatter attributed to intrinsic prediction error is 0.13 ± 0.02 mag for SNe with optical data only, or 0.08 ± 0.03 mag for SNe with optical and NIR light curve data. These estimates of intrinsic prediction error are smaller when a larger σ_{pec} is assumed because more of the Hubble diagram scatter is attributed to random galaxy motions.

The predictive variance σ_{pred}^2 measures the uncertainty with which the hierarchical model predicts the distance modulus of each individual SN, after marginalizing over the uncertainties in the training set, SN Ia, and dust populations. Figure 14 shows the distributions of predictive uncertainties (standard deviations) in
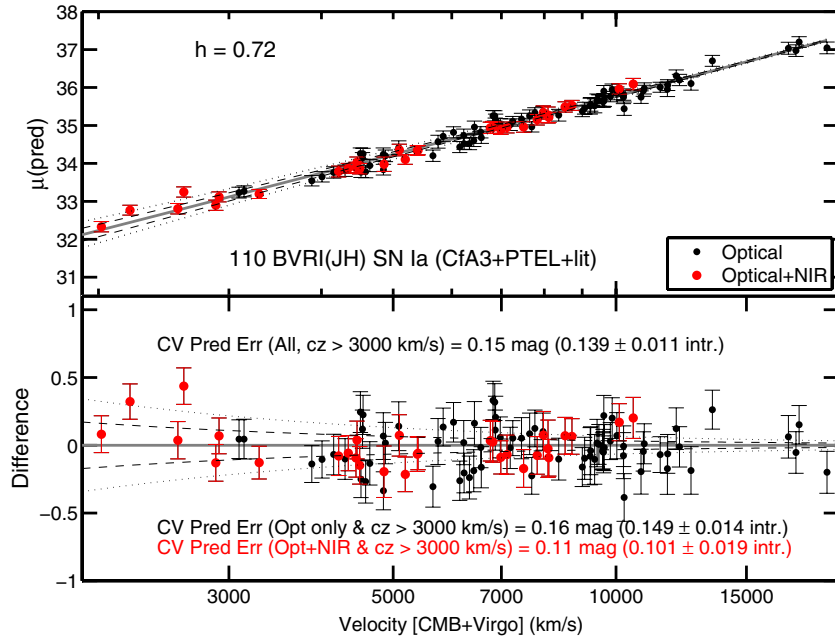
Figure 13. Cross-validated Hubble diagram computed with BAYESN for the low-z nearby set of CfA and literature SNe. Red points indicate the SN with joint optical *BVRI* and near-infrared *JH* data. Black points have only optical data. The dashed (dotted) line indicates the magnitude uncertainty in $\mu(z)$ for $\sigma_{pec} = 150\,(300)$ kms^{-1}. We perform CV with 30 bootstrapped training sets to estimate the out-of-sample prediction error and test the sensitivity of the model predictions to the finite sample. The rms prediction error in distance modulus for SNe with optical light curve data only at $cz > 3000$ km s^{-1} is 0.16 mag. The SNe with optical and near-infrared light curve data have an rms prediction error at $cz > 3000$ km s^{-1} of 0.11 mag. The maximum likelihood estimate of the rms intrinsic prediction error is shown in parentheses, assuming $\sigma_{pec} = 150$ km s^{-1}. For a velocity dispersion $\sigma_{pec} = 300$ km s^{-1}, the total rms prediction error remains the same, but the scatter attributed to the intrinsic prediction error is 0.129 ± 0.016 mag for SNe with optical data only, or 0.081 ± 0.026 mag for SNe with optical and NIR light curve data. These results indicate that one can make more accurate distance predictions with SNe Ia with combined optical and NIR data than with optical data alone.

(A color version of this figure is available in the online journal.)

the individual SN distance moduli predicted from this model. The cumulative distribution of the predictive posterior standard deviations for SNe with optical light curve data only is compared to that of SNe with optical and NIR data. For SNe Ia with joint optical and NIR data, the predictive uncertainties are typically between 0.10 and 0.12 mag, whereas for SNe Ia with optical data only, the uncertainties mostly lie between 0.12 and 0.16 mag. A simple Kolmogorov–Smirnov test verifies that these precision distributions are inconsistent. This demonstrates that the hierarchical model estimates the distances to SNe Ia with optical and NIR light curves with smaller uncertainty than those of SNe Ia with only optical data.

The sample variance of the distance predictions for a single SN over bootstrapped training sets, s^2_{pred} in Table 5, is always much smaller than the uncertainty variance σ^2_{pred} of a single prediction, and is smaller than the mean square error over the set of SNe in the Hubble diagram. The typical value of s_{pred} over the set of SNe is ~ 0.03 mag. This demonstrates that our model's distance predictions to individual SNe are fairly robust to perturbing the composition of the training set; the sensitivity of predictions to resampling is of order a few hundredths of a magnitude. With a larger set of optical and NIR light curves, this sensitivity could be reduced further.

We examined the CV prediction errors to check for systematic trends against observable or inferred quantities, as possible signs of model misfit. In Figure 15, we show a scatter plot of the prediction error for each SN versus an observable or inferred quantity. We find no significant trends of prediction error versus predicted dust extinction A_V, the apparent optical colors at peak (e.g., $B - V$), apparent optical–near-infrared colors at peak (e.g., $V - H$), or optical light curve shape, summarized by the
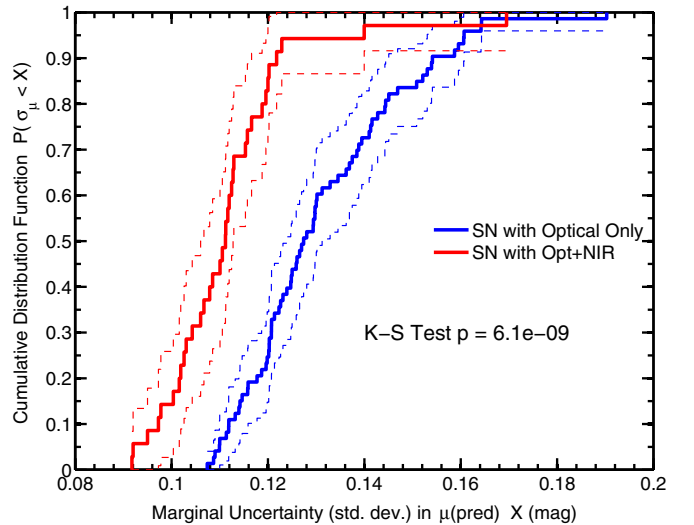

Figure 14. Cumulative distributions of the marginal predictive uncertainties of distance moduli of SNe Ia. The uncertainty in the predicted distance modulus is represented by a probability density after integrating over the other uncertainties in the dust estimates, light curve fits and K-corrections, and the population. We show the CDFs of the standard deviations of the predictive probability distributions of the individual SN distance moduli. The SNe with optical and NIR light curve measurements (red) typically have smaller distance uncertainties (higher precision) than those with only optical light curve data (blue). The dashed lines represent 95% confidence intervals of the respective CDFs. The two distributions are highly discrepant according to the Kolmogorov–Smirnov test. Using combined optical and NIR light curve data, the hierarchical model makes distance predictions with smaller estimated uncertainty and higher precision than it does with optical data alone.

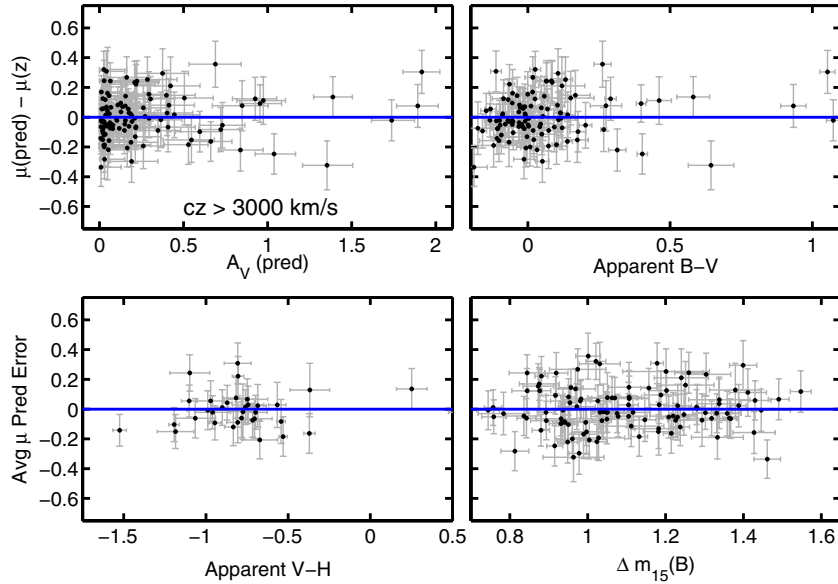(A color version of this figure is available in the online journal.)

Figure 15. Cross-validated distance prediction errors $\mu_{pred} - \mu(z)$ vs. observed and inferred SN quantities of interest. The distance modulus prediction errors of the model, averaged over 20 bootstrapped training sets, do not show statistically significant trends with respect to host galaxy dust extinction A_V, apparent optical color $B - V$, apparent optical–near-infrared color $V - H$, or optical decline rate $\Delta m_{15}(B)$. Fitted regressions have slopes consistent with zero (blue).

(A color version of this figure is available in the online journal.)
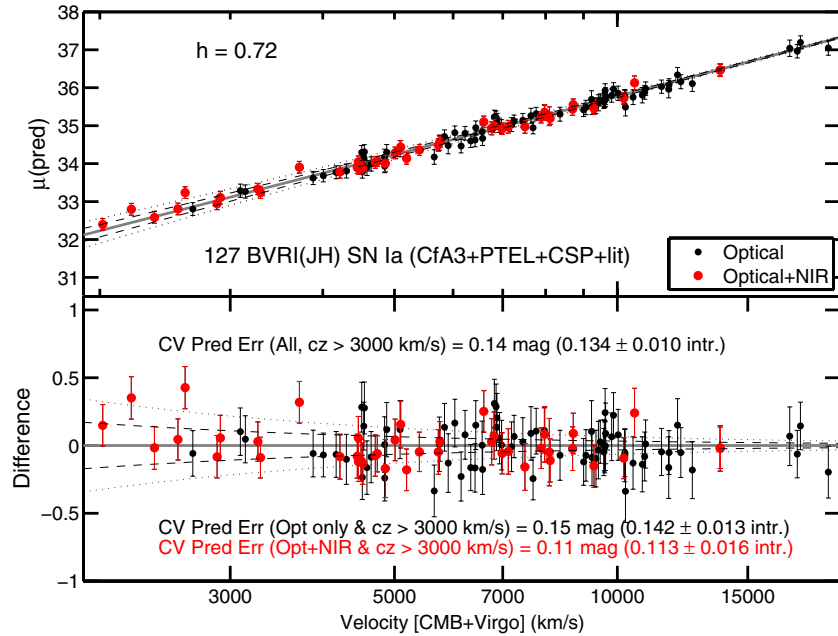
Figure 16. Cross-validated Hubble diagram computed with BAYESN for the low-z nearby training set using CfA, CSP, and literature SNe. Red points indicate the SNe with joint optical *BVRI* and near-infrared *JH* data. Black points are SNe with only optical data. The dashed (dotted) line indicates the magnitude uncertainty in $\mu(z)$ for $\sigma_{pec} = 150$ (300) km s^{-1}. We perform bootstrap CV to estimate the out-of-sample prediction error and test the sensitivity of the model predictions to the finite sample. The rms prediction errors are consistent with those of Figure 13.

(A color version of this figure is available in the online journal.)

canonical $\Delta m_{15}(B)$. Linear regressions fit to the prediction errors versus each quantity yield both slopes and intercepts that are statistically consistent with zero.

7.3. Distance Error Comparison with CSP Light Curves

We augmented our SN sample with 27 nearby SNe recently published by the CSP (Contreras et al. 2010), and again performed the CVs to produce predictions for each SN. There were 10 SNe that were contained both in the CSP sample and the CfA3+PAIRITEL sample. To avoid including duplicate light

curves for the same SN in the joint sample, we selected the CSP light curves in those cases, since this resulted in the retention of the most optical and NIR data. We recomputed training and prediction under the $m = 1$ dust population model. The Hubble diagram of these distance predictions is shown in Figure 16. The results are consistent with the previous Hubble diagram: the total rms dispersion at $cz > 3000$ km s^{-1} was 0.15 for SNe with optical data only and 0.11 for SNe with optical and near-infrared light curves. We compared the distribution of distance modulus errors $\bar{\mu}_{pred} - f(z)$ for Hubble flow SN with optical and NIR data
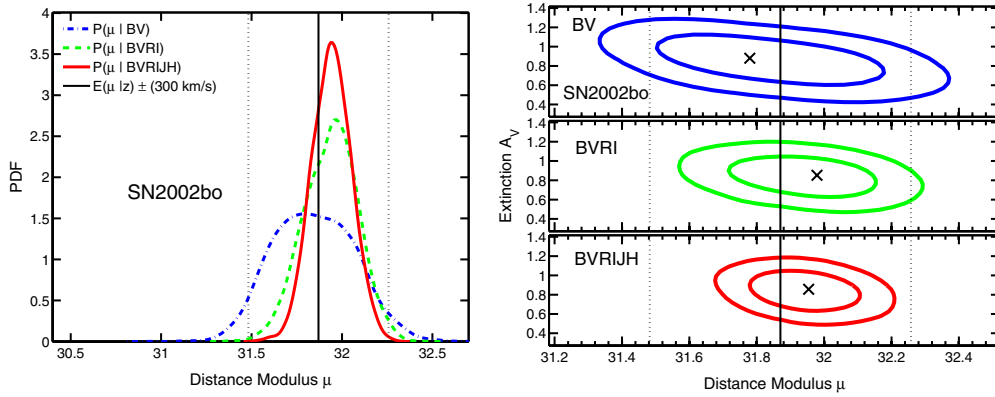
Figure 17. Effect of adding NIR light curve data on statistical uncertainties on distance modulus μ and extinction A_V for SN 2002bo. Left: predictive probability density of μ using BV light curve data only (blue), $BVRI$ data only (green), and $BVRIJH$ data (red), computed from the trained optical–near-infrared statistical model. The pdfs integrate to one, so that more precise predictions are taller, and less precise predictions are broader. The black vertical lines indicate the expected value of μ given the redshift and the associated error of ± 300 km s^{-1} peculiar velocity dispersion. Right: predictive joint probability density of (μ, A_V) using optical or optical and near-infrared data. The two-dimensional modes are marked, and the inner and outer contours contain 68% and 95% of the highest probability regions. Whereas with BV data only the distance is uncertain due to the uncertain extinction by host galaxy dust, with $BVRIJH$ data, the uncertainties in extinction, and thus in distance, are reduced significantly.

(A color version of this figure is available in the online journal.)

from the CSP sample with that of the "CfA+literature" sample. Using a two-sample Kolmogorov–Smirnov test we cannot rule out that they are from the same distribution ($p = 0.88$). The distance predictions of each set are statistically consistent, so it is reasonable to analyze the combined set.

7.4. Cross-validation with Different R_V Assumptions

In this section, we investigate the effect of different model assumptions about the dust population on distance predictions. For each case of Section 2.6, we computed cross-validated distance predictions for SNe in the Hubble flow at $cz > 3000$ km s^{-1} by generating 20 bootstrapped sets for training and predicting the distance moduli for the complementary validation set. For these computations, we used the "CfA+CSP+literature" sample of 127 SNe Ia. Table 1 displays the results of these calculations, including the marginal posterior estimates of the hyperparameters in each case, and the 0.632 estimate of total prediction error for the SNe with optical light curves only, and for the SNe with both optical and near-infrared data.

The case with fixed $R_V = 3.1$ (the Milky Way interstellar average) for all SNe leads to the worst distance predictions (0.20 mag for optical, 0.13 mag for optical and near-infrared). The cases of complete pooling (all SNe have R_V with the same value) or partial pooling with $m = 0$ (the R_V come from a population independent of the A_V value), or even no pooling (each R_V estimated independently for each SN), rms prediction errors are about 0.15 mag to 0.16 mag for optical light curves only, and 0.12–0.13 mag for optical plus NIR light curves.

If we model a potential population correlation between A_V and R_V^{-1}, using either linear or step function models, we find the smallest cross-validated distance prediction errors, both for the SNe with optical data only (0.15 mag) and for the SNe with optical and NIR light curves (0.11 mag). These are both significant improvements over the rms prediction errors under the assumption that $R_V = 3.1$ has the mean value for Milky Way interstellar dust, and are also marginally better than those of the other cases. However, the statistical sampling uncertainty is about ± 0.01 mag, so it is difficult to draw significant distinctions between the rms cross-validated prediction errors of the latter five cases. An analysis of a larger, future sample of optical and

NIR light curves will help to further discriminate between these competing cases.

It is notable that the change in the rms distance modulus prediction error for SNe with optical light curves alone is 0.05 mag between the worst case and best case dust population models, whereas this change for SNe with optical and NIR light curves is only 0.02 mag. This highlights the advantage of including the NIR data; since the H band provides a good standard candle, the model can rely mostly on the NIR light curves to provide distance estimates that are both less vulnerable to host galaxy dust and less sensitive to the assumptions about the dust.

7.5. Improving Constraints on Dust and Distance with Optical and NIR Data

In this section, we demonstrate the effect of using NIR light curve observations in conjunction with optical data for constraining extinction and for making more precise predictions. In Figure 17, we show the posterior predictive densities for the distance modulus and the joint probability densities for SN 2002bo, an event with high extinction. With the trained probability model, we computed the joint probability $P(\tilde{\mu}, \tilde{A}_V | \tilde{\mathcal{D}}_s, \tilde{z}_s; \mathcal{D}, \mathcal{Z})$ under prediction where the light curve data $\tilde{\mathcal{D}}_s$ alternately included the SN 2002bo observations in the BV, $BVRI$, or $BVRIJH$ filters. Recall that, under prediction, the tilded redshift $\tilde{z}_s$ is only used for K-corrections and Milky Way extinction, but not in the redshift–distance likelihood function. The data set used for training is denoted as $\mathcal{D}, \mathcal{Z}$. We also compute the marginal posterior predictive probability $P(\tilde{\mu} | \tilde{\mathcal{D}}_s, \tilde{z}_s; \mathcal{D}, \mathcal{Z})$ for each case. The probability density in $(\tilde{\mu}, \tilde{A}_V)$ integrates over the uncertainties in the SN light curve fit, and the dust and SN populations. The marginal density in $\tilde{\mu}$ additionally integrates over the uncertainty in $\tilde{A}_V$, as a "nuisance" parameter. These marginal probability densities were computed directly from the MCMC samples, obtained under prediction, using kernel density estimation. They are not Gaussian approximations of the posterior probability density.

For comparison we mark the expected distance modulus for the observed redshift and its expected magnitude uncertainty
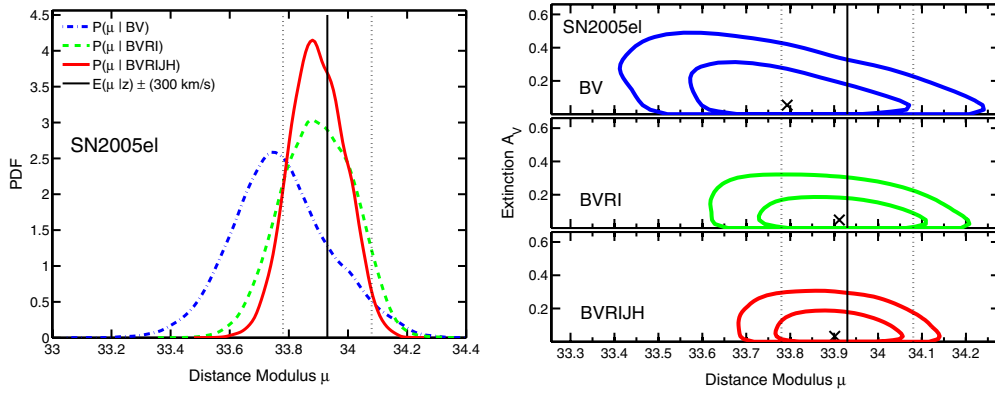
Figure 18. Effect of adding NIR light curve data on statistical uncertainties on distance modulus μ and extinction A_V for SN 2005el. Left: predictive probability density of μ using BV light curve data only (blue), $BVRI$ data only (green), and $BVRIJH$ data (red), computed from the trained optical–near-infrared statistical model. The pdfs integrate to one, so that more precise predictions are taller, and less precise predictions are broader. The black vertical lines indicate the expectation value of μ given the redshift and the associated error of ± 300 km s^{-1} peculiar velocity dispersion. Right: predictive joint probability density of (μ, A_V) using optical or optical and near-infrared data. The two-dimensional modes are marked, and the inner and outer contours contain 68% and 95% of the highest probability regions. Although this SN most likely has little extinction, the addition of the NIR data still helps to improve the constraints on A_V and the precision of distance predictions.

(A color version of this figure is available in the online journal.)

for a 300 km s^{-1} velocity dispersion. The marginal probability density in μ integrates to one, so that the taller pdfs make the most precise distance predictions, and the shorter pdfs make the most uncertain predictions. The figure demonstrates that adding observations in the redward filters greatly improves the precision of distance predictions, with the full optical and NIR data set yielding the greatest precision with this model. We also show the mode and 68% and 95% highest probability density regions of the joint probability $P(\tilde{\mu}, \tilde{A}_V | \tilde{\mathcal{D}}_s, \tilde{z}_s; \mathcal{D}, \mathcal{Z})$. When only the blue BV data are used, there is a strong degeneracy between the uncertainty in distance modulus and the uncertainty in host galaxy dust extinction. Both the uncertainty in the distance modulus and in dust extinction are reduced when we condition on the available NIR light curve observations. The predictive precision (the inverse variance) of the distance modulus of an individual SN is, on average, improved by a factor of 2.2 using $BVRI$ and by a factor of 3.6 using $BVRIJH$ data, compared to using BV light curves alone. The predictive precision is improved by 60%, on average, using optical and NIR $BVRIJH$ data versus optical $BVRI$ data alone, and can be improved by up to a factor of 2.6, based on the current sample.

In Figure 18, we illustrate these inferences for a different event, SN 2005el. This SN appears to have near zero host galaxy extinction. However, even with near zero dust extinction there is uncertainty in A_V due to the intrinsic variance of SN colors. This uncertainty is in the direction of positive extinction, and hence the joint distribution $P(\tilde{\mu}, \tilde{A}_V | \tilde{\mathcal{D}}_s, \tilde{z}_s; \mathcal{D}, \mathcal{Z})$ appears non-Gaussian. Although this event has close to zero extinction, there is still a strong degeneracy in the uncertainties between μ and A_V under prediction with the BV data alone. The combination of optical and near-infrared data constrains this joint uncertainty and yields improved precision of distance predictions even for low extinction events.

8. DISCUSSION AND CONCLUSION

We have constructed a comprehensive hierarchical model for SN Ia light curves in the optical and near-infrared ($BVRIJH$). We model the apparent light curves as the sum of random draws from an absolute light curve population distribution and from a host galaxy population distribution, plus the distance moduli. While fitting the individual SN Ia light curves, we also estimate

the characteristics of the two populations. These include the intrinsic correlation structure of the absolute light curves, and the joint distribution of extinction A_V and the slope of the dust law R_V in SN Ia host galaxies. The application of our new BAYESN MCMC algorithm enables coherent probabilistic inference of the unknown parameters and hyperparameters given the observed data. We also use it to generate distance predictions for SNe while marginalizing over the uncertainties in the population models and training set.

The inferred correlation matrices of the intrinsic light curve properties (Section 5.2.2) show that the peak optical absolute magnitudes ($BVRI$) are strongly correlated with each other, but have weaker correlation with the J and H near-infrared absolute magnitudes. Similarly, while the peak optical absolute magnitudes are correlated with optical decline rates (particularly $\Delta m_{15}(B)$), they have low correlation with the NIR decline rates. The near-infrared absolute magnitudes exhibit low correlation with the optical decline rates. This indicates that the NIR light curves provide independent information on the luminosities of SNe Ia, which can be leveraged to improve the precision of distance estimates.

We inferred the distribution of host galaxy extinction A_V, with an average value of $\tau_A = 0.37 \pm 0.04$ for our nearby sample. However, the exponential dust distribution does not adequately fit the fat tail of the peak apparent color distribution: excluding the 4% of SNe in the extreme red tail, the apparent colors and dust extinctions of the other 96% of SNe are well described by an exponential distribution in A_V with $\tau_A = 0.28 \pm 0.04$ and an intrinsic color distribution. Using both linear and step function models (Section 2.6), we modeled and inferred the joint distribution of A_V and the extinction law slope parameter, R_V, and found strong evidence for an apparent correlation. Under the assumption of a linear trend between R_V^{-1} and A_V, we found a positive slope. In the limit of low extinction, the marginal estimate of $R_V \approx 2.8 \pm 0.5$ is consistent with the Milky Way interstellar average, $R_V = 3.1$, and with independent measurements of dust in external galaxies (Finkelman et al. 2008, 2010, $R_V = 2.8$). For SNe with very high extinction, $A_V > 1$, values of $R_V \approx 1.7$ are favored. However, we do not know if the linear assumption is valid over the whole range of A_V, so we have explored alternative models for the potential differential behavior of R_V. Under the assumption

of a "step" function model that groups together SNe in four bins in A_V, the characteristic R_V in the lowest extinction bin ($A_V < 0.4$) had a modal value of 2.5 and mean values of 2.7–2.9 and for the highest extinction events ($A_V > 1.25$), we found $R_V = 1.6 \pm 0.1$. We find that these differences are statistically significant.

These results suggest that SNe at low extinction are seen through lines of sight with "normal" interstellar dust, but SNe at high extinctions are seen through dust with a steeper reddening law. This may indicate a circumstellar dust component dominating the absorption of light to high extinction events. Wang (2005) and Goobar (2008) suggested that scattering of SN light by circumstellar dust clouds could lead to low values of R_V. Goobar (2008) calculated that multiple scattering of SN light by dust in the locality of the SNe would attenuate short wavelength photons and steepen the extinction law to $R_V \sim 1.5$–2.5. Patat et al. (2007) reported the detection of spectroscopic signatures of circumstellar material around SN 2006X, a highly extinguished SN Ia. The effects of circumstellar dust might provide an explanation for the unusual colors of some high extinction events.

With consideration to the uncertainties of our inferences and model assumptions, a conservative conclusion is that most SNe in our sample are affected by host galaxy dust with R_V in the range of 2–3. These SNe are extinguished by $A_V \lesssim 1$ mag. At higher extinctions, $A_V \gtrsim 1$, the SNe are obscured by dust with R_V in the range of 1.5–2, although it is also possible that those SNe have different intrinsic colors than the general population. Notably, we do not find R_V values lower than 1.5. This is at odds with the R_V values between 1 and 2, many of which were below 1.5, fit for their whole sample by Folatelli et al. (2010) by minimizing the scatter in the Hubble diagram of CSP SNe. It is also at variance with similarly low R_V values found in the recent literature. However, those analyses assume that a single R_V value applies to all the SNe for a given fit, while we have allowed for distributions in R_V^{-1} that may be dependent on A_V. If we assume that every SN in our sample has exactly the same R_V (complete pooling), or that the distribution of R_V^{-1} has a single mean independent of A_V ($m = 0$), we also find $R_V = 1.6$–1.7. This suggests that these assumptions lead to estimates of R_V that are biased toward smaller values. This is not surprising, since R_V is best determined for SNe with high A_V, and as these SNe also have apparently low R_V, they dominate the estimate of a global constant or average.

Differences are also likely to arise from the treatment of the intrinsic covariance structure of the SN Ia light curves. In this paper, we have modeled the intrinsic covariances between the absolute light curves in optical and NIR wavelengths spanning -12 to 45 days in phase and estimated them by probabilistically deconvolving the apparent distributions using a hierarchical model. Posterior estimates of A_V, R_V for each SN and all other parameters were obtained via Equation (19) and by marginalizing over uncertainties. In particular, as estimates of R_V for individual SNe at low extinction are difficult to determine exactly, it is necessary to marginalize over that uncertainty when making inferences and predictions, as we do as part of the Bayesian inference. The inclusion of the NIR light curve data yields an added benefit. Since the peak H-band absolute magnitude is a good standard candle by itself, it is used to predict precise distances that are less susceptible to error from the dust estimate or the dust model. For example, for a single SN with average extinction, $A_V \approx 0.3$ mag, the change in A_H between $R_V = 1.7$ and $R_V = 3.1$ is about 0.02 mag.

The linear and step function models of the joint distribution of (A_V, r_V) both suggest that the average effective R_V at a given level of A_V decreases gradually with the increasing host galaxy extinction. We might speculate on the existence of two kinds of host galaxy dust with two different reddening laws over wavelength. One would correspond to "normal" interstellar dust as found in the Milky Way, $R_V \approx 3$, and the other would correspond to some kind of circumstellar dust with a reddening law with $R_V \approx 1.7$. If the dust affecting each SN is comprised of random amounts of these two types of dust, then the effective r_V would roughly be an extinction-weighted average of the characteristic r_V of their respective reddening profiles. If the circumstellar component was associated with highly dusty environments, then this mixture could generate an apparent trend of effective r_V against total extinction. This suggests an extension of our hierarchical model, which we will address in a future work.

Using bootstrap CV, we have randomized the optical and near-infrared training set to generate probabilistic estimates of the distance moduli to out-of-sample SN. Comparing these to the distances expected from the Hubble expansion, we found a total rms prediction error of 0.16 mag (at $cz > 3000$ km s^{-1}) for SNe with optical light curves (*BVRI*) only, but a total rms error of 0.11 mag for SNe with optical and near-infrared (*BVRIJH*) light curves. After accounting for the dispersion expected from random peculiar velocities with $\sigma_{pec} = (150, 300)$ km s^{-1} the rms intrinsic prediction errors for these subsets were $(0.15 \pm 0.01, 0.13 \pm 0.02)$ mag for optical and $(0.10 \pm 0.02, 0.08 \pm 0.03)$ mag for optical and NIR. This demonstrates that distances to SNe Ia observed in the optical and near-infrared can be estimated with about twice the accuracy ($\sim[0.15/0.10]^2$) of SNe Ia observed in the optical alone. By conditioning on light curve data subsets (*BV*, *BVRI*, *BVRIJH*) for individual SN, we show that including near-infrared light curve data tightens the constraints on host galaxy extinction and distance predictions (Section 7.5).

The number of published optical and near-infrared light curves of SNe Ia is still small compared to the sample of optically observed events. Future, larger samples of SNe Ia with accurate, joint optical and near-infrared photometry will help test and build the statistical strength of our conclusions on the utility of combining optical and NIR light curves for improving distance predictions and will help illuminate the nature of the dust in SN Ia host galaxies. In addition to estimating the intrinsic correlation structure of SN Ia light curves and the distribution of host galaxy dust, our hierarchical framework can be applied to distance prediction and analysis of a cosmological sample of SNe Ia. Cosmological samples of SNe Ia observed in the rest-frame NIR are possible. The improved precision and accuracy of the inferences about the history of cosmic expansion may justify the extra effort required to obtain these data now with the *Hubble Space Telescope*, soon with the *James Webb Space Telescope*, and eventually with the *WFIRST* mission.

APPENDIX A

DIFFERENTIAL DECLINE RATES LIGHT CURVE MODEL

The continuous normalized light curve is equivalent to the specification of the total decline rates of the apparent light curve. Let $D^F(t)$ be the total decline rate from phase zero to phase t: $D^F(t) \equiv \mathrm{LC}^F(t) - F_0 = l^F(t)$. For example, the well-known B-band decline rate from peak to 15 day past maximum is $\Delta m_{15}(B) = D^B(15)$ (Phillips 1993).

For practical purposes, it is necessary to parameterize the continuous model for the light curve with a discrete set of variables. Let $\boldsymbol{\tau}$ be a grid in rest-frame phase. The total decline rates to each discrete grid point, $\boldsymbol{D}^F = \{D_j^F \equiv D^F(\tau_j)\}$, define the normalized F-band light curve $l^F(t)$ at all times if we choose a suitable interpolation rule. If we choose a natural cubic spline we ensure continuity up to two derivatives, and the normalized light curve is then linear in the total decline rates $l^F(t) = \boldsymbol{s}(t, \boldsymbol{\tau}) \cdot \boldsymbol{D}^F$ and the linear smoother $\boldsymbol{s}(t, \boldsymbol{\tau})$ is specified. The differential decline rate is $d_j^F \equiv D_j^F - D_{j-1}^F$. The total decline rates at the knots of $\boldsymbol{\tau}$ are sums of the differential decline rates over the span in phase. There is a simple constant matrix $\boldsymbol{G}$ so that $\boldsymbol{D}^F = \boldsymbol{G}\boldsymbol{d}^F$, where $\boldsymbol{d}^F$ is the vector of differential decline rates d_j^F. The model for the normalized light curve in band F is linear in the differential decline rates: $l^F(t) = \boldsymbol{s}(t, \boldsymbol{\tau}) \cdot \boldsymbol{G}\boldsymbol{d}^F$.

This differential decline rates representation is a special case of Equation (1), with $l_0^F(t) \equiv 0$, $\boldsymbol{l}_1^F(t) \equiv \boldsymbol{s}(t, \boldsymbol{\tau}) \cdot \boldsymbol{G}$, $\boldsymbol{\theta}^F = \boldsymbol{\theta}_L^F \equiv \boldsymbol{d}^F$, and $\boldsymbol{\theta}_{NL}^F = \emptyset$. We use an irregular grid $\boldsymbol{\tau}$ in phase spanning -10 days to 45 days. The knots are placed more densely near phases where we expect the most observations (near phase zero) and where we expect more curvature of light curves in certain bands.

APPENDIX B

K-CORRECTIONS AND MILKY WAY EXTINCTION

The observed SED of an SN Ia changes relative to a fixed observer-frame passband due to the effect of cosmological redshift and extinction in the Milky Way varying with line of sight. To account for these differences in flux, we derive K-corrections and Galactic extinctions for SNe Ia in observer-frame optical and NIR filters. As a high-quality SED time sequence is seldom available for all SNe, particularly at higher redshifts, we use the average spectral template sequence from Hsiao et al. (2007).

We compute K-corrections for optical and NIR passbands following the method of Nugent et al. (2002). We model the $UBVRI$ filters with the "shifted" Bessell passbands (Kessler et al. 2009) and the JHK_s filters with the 2MASS passbands. We choose a standard color that includes each passband (U:$U - B$, B:$B - V$, V:$V - R$, R:$R - I$, I:$R - I$, J:$J - H$, H:$J - H$, K_s:$H - K_s$) and for each model filter, we warp the SED sequence using the $R_V = 3.1$ extinction law (Cardelli et al. 1989) to take on a wide range of the corresponding rest-frame color, measured with synthetic photometry. For a given observer-frame passband and redshift we determine the rest-frame passband with the nearest effective wavelength. The warped SED series is then redshifted and used to determine the K-correction as a function of rest-frame color. This procedure does not account for spectral features that vary with light curve shape and there is no

constraint on the SED blueward or redward of the bluest or reddest model filter.

To compute the Milky Way Galactic extinction, we follow a procedure, modified from that outlined in Jha et al. (2007). For a given model filter, we use the warped SED sequence constructed for the K-corrections and determine the unextinguished observer-frame magnitude with synthetic photometry as a function of phase and rest-frame color. The sequence is then reddened with an $R_V = 3.1$ law for a range of values of the Milky Way reddening $E_{\mathrm{MW}} \equiv E(B - V)$ and the extinguished observer-frame magnitude is computed. The difference between the two magnitudes is the Galactic extinction. We find that the Galactic extinction for a given phase t and passband X is well modeled by a quadratic in E_{MW}: $A_X(t) = [\alpha_X(t, c) + \beta_X(t, c)E_{\mathrm{MW}}]E_{\mathrm{MW}}$. We solve for the polynomial coefficients α_X and β_X for all the phases t and rest-frame apparent color c. Further, as the rest-frame color dependence was introduced by warping the same spectral sequence, we find the coefficients α_X and β_X to be smoothly varying functions of rest-frame apparent color c. We model the slope and intercept by polynomials of rest-frame color with degree 4 and 5, respectively. We can thereby reduce the Galactic extinction to a simple set of polynomial coefficients of color tabulated with phase. The value of E_{MW} for each SN is obtained from the Schlegel et al. (1998) maps.

APPENDIX C

SPECIFICATION OF THE HYPERPRIOR

There are two populations in this hierarchical model: the multi-band light curve distribution and the host galaxy dust extinction distribution for A_V and R_V. The hyperparameters of the SN Ia light curves are the population mean $\boldsymbol{\mu}_\psi$ and the covariances $\boldsymbol{\Sigma}_\psi$. The hyperparameters of the dust populations are τ_A, $\boldsymbol{\beta}$, and σ_r^2 (e.g., for Case 5 in Section 2.6). We must make explicit our priors on these hyperparameters, i.e., hyperpriors. At the highest level of the hierarchical model, we use diffuse, or "non-informative" prior distributions by default.

The dust population hyperprior is $P(\tau_A, \boldsymbol{\beta}, \sigma_r^2) = P(\tau_A)P(\boldsymbol{\beta}|\sigma_r^2)P(\sigma_r^2)$. We adopt uniform prior $P(\boldsymbol{\beta}|\sigma_r^2) \propto 1$. For τ_A and σ_r^2, we use the standard non-informative prior for positive scale parameters, $P(\log \tau_A) \propto 1$, $P(\log \sigma_r^2) \propto 1$.

The hyperprior on the absolute light curve distribution hyperparameters can be conditionally decomposed: $P(\boldsymbol{\mu}_\psi, \boldsymbol{\Sigma}_\psi) = P(\boldsymbol{\mu}_\psi|\boldsymbol{\Sigma}_\psi)P(\boldsymbol{\Sigma}_\psi)$. We assume a uniform $P(\boldsymbol{\mu}_\psi|\boldsymbol{\Sigma}_\psi) \propto 1$. For $P(\boldsymbol{\Sigma}_\psi)$, we require a diffuse density that has support only on the space of symmetric, positive definite, and invertible matrices. We employ the standard inverse Wishart distribution, which is conjugate to the normal covariance matrix: $P(\boldsymbol{\Sigma}_\psi) = \text{Inv-Wishart}_{\nu_0}(\boldsymbol{\Sigma}_\psi|\boldsymbol{\Lambda}_0) \times f(\boldsymbol{\sigma}_\psi)$. The inverse Wishart density is multiplied by a smooth density on the variances $f(\boldsymbol{\sigma}_\psi)$ described below. The degrees-of-freedom parameter is set to $\nu_0 = K + 1$ where $K \equiv \dim(\boldsymbol{\psi})$. This guarantees that the marginal prior density of any individual correlation $\rho(\psi_i, \psi_j) \equiv R_\psi^{ij}$ is uniform between -1 and 1 (Barnard et al. 2000). The scale matrix is set to $\boldsymbol{\Lambda}_0 = \epsilon_0 \boldsymbol{I}$, where $\boldsymbol{I}$ is a $K \times K$ identity matrix. The scale ϵ_0 has the effect of setting a floor on each σ_ψ^i so that it does not fall below the value of $\epsilon_0/\sqrt{N_{\mathrm{SN}}} \approx 0.02$. Since we do not realistically expect any standard deviation σ_ψ^i to be less than a few 0.01 mag, this limit is conservative and helps to prevent the MCMC chain from getting stuck in a region of parameter space with a near-zero variance, where the covariance matrix may be nearly singular.

We found it useful to stabilize the estimation of magnitude variances with a power-law density: $\log f(\sigma_\psi) = \log f[\sigma(M_B)] = pN_{SN} \log \sigma(M_B)$ with $p \approx 0.09$. We chose the smallest value of p for which the inferences were locally insensitive to its value. We have checked that the choice of p does not significantly impact the rms prediction error. We also used the scaled inverse Wishart distribution (Gelman & Hill 2006; O'Malley & Zaslavsky 2008) as an alternative hyperprior and obtained comparable results.

APPENDIX D

MATHEMATICAL DETAILS: BAYESN

In this section, we provide mathematical results for each step of the BAYESN algorithm.

1. The goal is to sample from $P(\boldsymbol{\mu}_\psi, \boldsymbol{\Sigma}_\psi | \cdot, \mathcal{D}, \mathcal{Z})$. This can be factored as $P(\boldsymbol{\mu}_\psi | \boldsymbol{\Sigma}_\psi, \{\boldsymbol{\psi}_s\}) P(\boldsymbol{\Sigma}_\psi | \{\boldsymbol{\psi}_s\})$. These densities only depend on $\{\boldsymbol{\psi}_s\}$ through the sufficient statistics: the sample mean $\bar{\boldsymbol{\psi}}$ and the matrix sum of squared deviations from the mean: $\boldsymbol{S}_\psi = \sum_{s=1}^{N_{SN}} (\boldsymbol{\psi}_s - \bar{\boldsymbol{\psi}}) (\boldsymbol{\psi}_s - \bar{\boldsymbol{\psi}})^T$. We generate a new $\boldsymbol{\Sigma}_\psi^*$ from the proposal density $q(\boldsymbol{\Sigma}_\psi^* | \{\boldsymbol{\psi}_s\}) = \text{Inv-Wishart}_{\nu_N}(\boldsymbol{\Sigma}_\psi^* | \boldsymbol{\Lambda}_N^{-1})$, where $\nu_N = \nu_0 + N_{SN}$ and $\boldsymbol{\Lambda}_N = \boldsymbol{S}_\psi + \boldsymbol{\Lambda_0}$. When the $(\epsilon_0/N_{SN})^2$ is negligible compared to the variances, the expectation of this distribution is just the standard maximum likelihood estimator of covariance, $\boldsymbol{S}_\psi/N_{SN}$. If $f(\sigma_\psi) \propto 1$ then the proposal is the same as $P(\boldsymbol{\Sigma}_\psi^* | \{\boldsymbol{\psi}_s\})$, and this is Gibbs sampling. If not, then the Metropolis–Hastings ratio r simplifies to $r = f(\sigma_\psi^*)/f(\sigma_\psi)$. The proposal is accepted ($\boldsymbol{\Sigma}_\psi \to \boldsymbol{\Sigma}_\psi^*$) with probability r. This method results in fast convergence since it allows for updating the entire covariance matrix at once. A new $\boldsymbol{\mu}_\psi$ is Gibbs sampled from $P(\boldsymbol{\mu}_\psi | \boldsymbol{\Sigma}_\psi, \{\boldsymbol{\psi}_s\}) = N(\boldsymbol{\mu}_\psi | \bar{\boldsymbol{\psi}}, \boldsymbol{\Sigma}_\psi/N_{SN})$.

2. The conditional density for τ_A is $P(\tau_A | \cdot, \mathcal{D}, \mathcal{Z}) = P(\tau_A | \{A_V^s\}) = \text{Inv-Gamma}(\tau_A | N_{SN}, \sum_{s=1}^{N_{SN}} A_V^s)$.

3. The conditional densities $P(\boldsymbol{\beta} | \sigma_r^2, \{r_V^s, A_V^s\})$ and $P(\sigma_r^2 | \{r_V^s, A_V^s\})$ are standard results of the Bayesian analysis of ordinary linear regression of r_V versus A_V (Gelman et al. 2003, chap. 14).

4. Since the subsequent steps concern only one SN at a time, we suppress the label s on individual SN parameters.

   (a) The conditional posterior density of the fit $(T_0, \boldsymbol{\phi})$ for a single SN is proportional to

   $$P(T_0, \boldsymbol{\phi} | \cdot, \mathcal{D}_s, z_s) \propto P(\boldsymbol{m} | T_0, \boldsymbol{\phi}, z_s) \times N(\boldsymbol{\phi} | \boldsymbol{\mu}_\phi, \boldsymbol{\Sigma}_\psi), \quad \text{(D1)}$$

   where $\boldsymbol{\mu}_\phi \equiv \boldsymbol{\mu}_\psi + \boldsymbol{A} + \boldsymbol{v}\mu$, and the first factor is Equation (5). We construct a proposal density for the new fit $(T_0^*, \boldsymbol{\phi}^*)$ given the current one: $q(T_0^*, \boldsymbol{\phi}^* | T_0, \boldsymbol{\phi}) = q(\boldsymbol{\phi}^* | T_0^*; T_0, \boldsymbol{\phi}) \times q(T_0^* | T_0; \boldsymbol{\phi})$. The proposal for the new T_0^* is $q(T_0^* | T_0; \boldsymbol{\phi}) = N(T_0^* | T_0, s_T^2)$. The proposal $q(\boldsymbol{\phi}^* | T_0^*; T_0, \boldsymbol{\phi})$ is an approximation to Equation (D1) with the K-correction and Milky Way extinction factors fixed at the current fit $(T_0, \boldsymbol{\phi})$. These depend on $\boldsymbol{\phi}$ only through the apparent colors. The proposal is

   $$q(\boldsymbol{\phi}^* | T_0^*; T_0, \boldsymbol{\phi}) \propto N[\boldsymbol{m} | \mathbf{KC}(T_0^*; z, \boldsymbol{\phi})$$
   $$+ \mathbf{GX}(T_0^*; z, \boldsymbol{\phi}, E_{MW}) + \boldsymbol{L}_2(T_0^*, z)\boldsymbol{\phi}^*, \boldsymbol{W}]$$
   $$\times N(\boldsymbol{\phi}^* | \boldsymbol{\mu}_\phi, \boldsymbol{\Sigma}_\psi) \quad \text{(D2)}$$

After algebraic simplifications, it can be shown that this is a Gaussian probability density on $\boldsymbol{\phi}^*$ and thus can be used to generate a random proposal. The joint proposal $(T_0^*, \boldsymbol{\phi}^*)$ is accepted with probability

$$r = \frac{P(T_0^*, \boldsymbol{\phi}^* | \cdot, \mathcal{D}_s, z_s)}{P(T_0, \boldsymbol{\phi} | \cdot, \mathcal{D}_s, z_s)} \times \frac{q(\boldsymbol{\phi} | T_0; T_0^*, \boldsymbol{\phi}^*)}{q(\boldsymbol{\phi}^* | T_0^*; T_0, \boldsymbol{\phi})}. \quad \text{(D3)}$$

The rejection step corrects the approximation of the conditional, Equation (D1), with the proposal Equation (D2). If the $\mathbf{KC}$ and $\mathbf{GX}$ factors are constant with respect to SN color and phase, then $r = 1$ and this is just Gibbs sampling. This scheme is efficient when $\mathbf{KC}$ and $\mathbf{GX}$ are slowly varying with phase and apparent color.

   (b) The conditional density for μ simplifies to $P(\mu | \cdot, \mathcal{D}_s, z_s) = N(\mu | \hat{\mu}, \hat{\sigma}_\mu^2)$, where $\tilde{\mu} = s_\mu^2 \boldsymbol{v}^T \boldsymbol{\Sigma}_\psi^{-1} (\boldsymbol{\phi} - \boldsymbol{A} - \boldsymbol{\mu}_\psi)$; $s_\mu^{-2} = \boldsymbol{v}^T \boldsymbol{\Sigma}_\psi^{-1} \boldsymbol{v}$; $\hat{\sigma}_\mu^{-2} = s_\mu^{-2} + \sigma_\mu^{-2}$; and $\hat{\mu} = \hat{\sigma}_\mu^2(\sigma_\mu^{-2} f(z) + s_\mu^{-2}\tilde{\mu})$. For prediction, we take $\sigma_\mu \to \infty$.

   (c) The conditional density for A_V is $P(A_V | \cdot, \mathcal{D}_s, z_s) = P(A_V | \boldsymbol{\phi}, \mu, r_V; \boldsymbol{\mu}_\psi, \boldsymbol{\Sigma}_\psi, \tau_A, \boldsymbol{\beta}, \sigma_r^2)$. This is a probability density on $A_V \geq 0$ proportional to $N(A_V | \hat{A}, s_A^2) \times N(r_V | \beta_0 + \beta_1 A_V, \sigma_r^2)$, where $\hat{A} = s_A^2 \boldsymbol{c}^T \boldsymbol{\Sigma}_\psi^{-1}[\boldsymbol{\phi} - \boldsymbol{v}\mu - \boldsymbol{\mu}_\psi] - s_A^2/\tau_A$; $\boldsymbol{c} = (\boldsymbol{\alpha} + \boldsymbol{\beta}r_V)$, and $s_A^{-2} = \boldsymbol{c}^T \boldsymbol{\Sigma}_\psi^{-1} \boldsymbol{c}$. This can be sampled using griddy Gibbs sampling.

   (d) The conditional posterior $P(r_V | \cdot, \mathcal{D}_s, z_s) = P(r_V | \boldsymbol{\phi}, \mu, A_V; \boldsymbol{\mu}_\psi, \boldsymbol{\Sigma}_\psi, \boldsymbol{\beta}, \sigma_r^2)$. Defining $\tilde{\sigma}_r^{-2} = A_V^2 \boldsymbol{\beta}^T \boldsymbol{\Sigma}_\psi \boldsymbol{\beta}$; $\tilde{\mu}_r = \tilde{\sigma}_r^2 \boldsymbol{\beta}^T A_V \boldsymbol{\Sigma}_\psi^{-1}[\boldsymbol{\phi} - \boldsymbol{v}\mu - A_V\boldsymbol{\alpha} - \boldsymbol{\mu}_\psi]$; $\hat{\sigma}_r^{-2} = \tilde{\sigma}_r^{-2} + \sigma_r^{-2}$; $\hat{r}_V = \hat{\sigma}_r^2[\tilde{\sigma}_r^{-2}\tilde{\mu}_r + \sigma_r^{-2}(\beta_0 + \beta_1 A_V)]$, this density is proportional to $N(r_V | \hat{r}_V, \hat{\sigma}_r^2)$ over the restricted range $0.18 < r_V < 0.7$. A new sample is generated by evaluating the pdf on a fine grid and using griddy Gibbs sampling.

   (e) (Optional) Generalized conditional sampling allows the MCMC to move along expected degeneracies between parameters in the posterior density that may be oblique with respect to the natural coordinate system defined by the chosen parameters (Liu & Sabatti 2000; Liu 2002). We expect there to be a tradeoff between dust extinction and distance to SNe, since both make SNe appear dimmer. Let $p(A_V, \mu) = P(A_V, \mu | \cdot, \mathcal{D}_s, z_s)$ be the conditional posterior of dust and distance. To perform the translation $(A_V, \mu) \to (A_V, \mu) + \gamma(1, -x)$, we first choose a scalar x which sets a direction in the (A_V, μ) plane to move along. To select an appropriate direction along the tradeoff between dust and distance, we find $\bar{x} = \min_x |(\boldsymbol{\alpha} + \boldsymbol{\beta}r_V) - x\boldsymbol{v}|^2$. For typical values of r_V, this was $\bar{x} \approx 0.7$. To select a translation vector near this direction, we sample $x \sim N(0.7, 0.05)$. Then we sample a random $\gamma \sim p(A_V + \gamma, \mu - x\gamma)$, where A_V and μ are the current values. The sample can be generated by evaluating the univariate density on a grid and using the inverse CDF method. Given γ the chain can be translated into the new position.

REFERENCES

Amanullah, R., et al. 2010, ApJ, 716, 712
Astier, P., et al. 2006, A&A, 447, 31

Bailey, S., et al. 2009, A&A, 500, L17
Barnard, J., McCulloch, R., & Meng, X.-L. 2000, Stat. Sin., 10, 1281
Blondin, S., Mandel, K. S., & Kirshner, R. P. 2011, A&A, 526, A81
Bloom, J. S., et al. 2006, in ASP Conf. Ser. 351, Astronomical Data Analysis Software and Systems XV, ed. C. Gabriel et al. (San Francisco, CA: ASP), 751
Branch, D., & Tammann, G. A. 1992, ARA&A, 30, 359
Cardelli, J. A., Clayton, G. C., & Mathis, J. S. 1989, ApJ, 345, 245
Conley, A., Carlberg, R. G., Guy, J., Howell, D. A., Jha, S., Riess, A. G., & Sullivan, M. 2007, ApJ, 664, L13
Conley, A., et al. 2008, ApJ, 681, 482
Contreras, C., et al. 2010, AJ, 139, 519
Di Paola, A., Larionov, V., Arkharov, A., Bernardi, F., Caratti o Garatti, A., Dolci, M., Di Carlo, E., & Valentini, G. 2002, A&A, 393, L21
Draine, B. T. 2003, ARA&A, 41, 241
Efron, B. 1983, J. Am. Stat. Assoc., 78, 316
Efron, B., & Tibshirani, R. 1997, J. Am. Stat. Assoc., 92, 548
Elias, J. H., Frogel, J. A., Hackwell, J. A., & Persson, S. E. 1981, ApJ, 251, L13
Elias, J. H., Matthews, K., Neugebauer, G., & Persson, S. E. 1985, ApJ, 296, 379
Elias-Rosa, N., et al. 2006, MNRAS, 369, 1880
Elias-Rosa, N., et al. 2008, MNRAS, 384, 107
Finkelman, I., et al. 2008, MNRAS, 390, 969
Finkelman, I., et al. 2010, MNRAS, 409, 727
Folatelli, G., et al. 2010, AJ, 139, 120
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
Freedman, W. L., et al. 2001, ApJ, 553, 47
Freedman, W. L., et al. 2009, ApJ, 704, 1036
Frogel, J. A., Gregory, B., Kawara, K., Laney, D., Phillips, M. M., Terndrup, D., Vrba, F., & Whitford, A. E. 1987, ApJ, 315, L129
Garnavich, P. M., et al. 1998, ApJ, 509, 74
Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. 2003, Bayesian Data Analysis (2nd ed.; Boca Raton, FL: CRC Press)
Gelman, A., & Hill, J. 2006, Data Analysis Using Regression and Multilevel/ Hierarchical Models (Cambridge: Cambridge Univ. Press)
Gelman, A., Meng, X.-L., & Stern, H. 1996, Stat. Sin., 6, 733
Gelman, A., & Rubin, D. B. 1992, Stat. Sci., 7, 457
Geman, S., & Geman, D. 1984, IEEE Trans. Pattern Anal. Mach. Intell., 6, 721
Goldhaber, G., et al. 2001, ApJ, 558, 359
Goobar, A. 2008, ApJ, 686, L103
Graham, J. R., Meikle, W. P. S., Longmore, A. J., & Williams, P. M. 1988, ApJ, 333, 743
Guy, J., Astier, P., Nobili, S., Regnault, N., & Pain, R. 2005, A&A, 443, 781
Guy, J., et al. 2007, A&A, 466, 11
Hamuy, M., et al. 1996a, AJ, 112, 2408
Hamuy, M., et al. 1996b, AJ, 112, 2438
Hastings, W. K. 1970, Biometrika, 57, 97
Hernandez, M., et al. 2000, MNRAS, 319, 223
Hicken, M., et al. 2009a, ApJ, 700, 331
Hicken, M., et al. 2009b, ApJ, 700, 1097
Höflich, P., et al. 2010, ApJ, 710, 444
Hogg, D. W., Myers, A. D., & Bovy, J. 2010, ApJ, 725, 2166
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
Jha, S., Riess, A. G., & Kirshner, R. P. 2007, ApJ, 659, 122
Jha, S., et al. 1999, ApJS, 125, 73
Jha, S., et al. 2006, AJ, 131, 527
Kasen, D. 2006, ApJ, 649, 939
Kelly, B. C. 2007, ApJ, 665, 1489

Kelly, B. C., & Bechtold, J. 2007, ApJS, 168, 1
Kelly, P. L., Hicken, M., Burke, D. L., Mandel, K. S., & Kirshner, R. P. 2010, ApJ, 715, 743
Kessler, R., et al. 2009, ApJS, 185, 32
Kirshner, R. P., Willner, S. P., Becklin, E. E., Neugebauer, G., & Oke, J. B. 1973, ApJ, 180, L97
Kowalski, M., et al. 2008, ApJ, 686, 749
Krisciunas, K., Hastings, N. C., Loomis, K., McMillan, R., Rest, A., Riess, A. G., & Stubbs, C. 2000, ApJ, 539, 658
Krisciunas, K., Phillips, M. M., & Suntzeff, N. B. 2004a, ApJ, 602, L81
Krisciunas, K., et al. 2001, AJ, 122, 1616
Krisciunas, K., et al. 2003, AJ, 125, 166
Krisciunas, K., et al. 2004b, AJ, 127, 1664
Krisciunas, K., et al. 2004c, AJ, 128, 3034
Krisciunas, K., et al. 2007, AJ, 133, 58
Liu, J. S. 2002, Monte Carlo Strategies in Scientific Computing (New York: Springer)
Liu, J. S., & Sabatti, C. 2000, Biometrika, 87, 353
Loredo, T. J., & Hendry, M. A. 2010, in Bayesian Methods in Cosmology, ed. M. Hobson et al. (Cambridge: Cambridge Univ. Press), 245
Mandel, K. S., Wood-Vasey, W. M., Friedman, A. S., & Kirshner, R. P. 2009, ApJ, 704, 629
Meikle, W. P. S. 2000, MNRAS, 314, 782
Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. 1953, J. Chem. Phys., 21, 1087
Mould, J. R., et al. 2000, ApJ, 529, 786
Nobili, S., & Goobar, A. 2008, A&A, 487, 19
Nugent, P., Kim, A., & Perlmutter, S. 2002, PASP, 114, 803
O'Malley, A. J., & Zaslavsky, A. M. 2008, J. Am. Stat. Assoc., 103, 1405
Pastorello, A., et al. 2007, MNRAS, 376, 1301
Patat, F., et al. 2007, Science, 317, 924
Perlmutter, S., et al. 1999, ApJ, 517, 565
Phillips, M. M. 1993, ApJ, 413, L105
Phillips, M. M., Lira, P., Suntzeff, N. B., Schommer, R. A., Hamuy, M., & Maza, J. 1999, AJ, 118, 1766
Pignata, G., et al. 2008, MNRAS, 388, 971
Prieto, J. L., Rest, A., & Suntzeff, N. B. 2006, ApJ, 647, 501
Radburn-Smith, D. J., Lucey, J. R., & Hudson, M. J. 2004, MNRAS, 355, 1378
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. 1999, AJ, 117, 707
Riess, A. G., et al. 2005, ApJ, 627, 579
Riess, A. G., et al. 2009a, ApJ, 699, 539
Riess, A. G., et al. 2009b, ApJS, 183, 109
Ritter, C., & Tanner, M. A. 1992, J. Am. Stat. Assoc., 87, 861
Robert, C. P., & Casella, G. 2005, Monte Carlo Statistical Methods (Secaucus, NJ: Springer)
Rubin, D. B. 1984, Ann. Stat., 12, 1151
Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, ApJ, 500, 525
Stanishev, V., et al. 2007, A&A, 469, 645
Sullivan, M., et al. 2010, MNRAS, 406, 782
Tripp, R. 1998, A&A, 331, 815
Tripp, R., & Branch, D. 1999, ApJ, 525, 209
Valentini, G., et al. 2003, ApJ, 595, 779
Wang, L. 2005, ApJ, 635, L33
Wang, L., Goldhaber, G., Aldering, G., & Perlmutter, S. 2003, ApJ, 590, 944
Wang, X., et al. 2008, ApJ, 675, 626
Wang, X., et al. 2009a, ApJ, 699, L139
Wang, X., et al. 2009b, ApJ, 697, 380
Wood-Vasey, W. M., et al. 2007, ApJ, 666, 694
Wood-Vasey, W. M., et al. 2008, ApJ, 689, 377 (WV08)