TYPE Ia SUPERNOVA LIGHT-CURVE INFERENCE: HIERARCHICAL BAYESIAN ANALYSIS IN THE NEAR-INFRARED

Kaisey S. Mandel, W. Michael Wood-Vasey[1], Andrew S. Friedman, and Robert P. Kirshner
Harvard-Smithsonian Center for Astrophysics, 60 Garden St., Cambridge, MA 02138, USA; kmandel@cfa.harvard.edu

ABSTRACT

We present a comprehensive statistical analysis of the properties of Type Ia supernova (SN Ia) light curves in the near-infrared using recent data from Peters Automated InfraRed Imaging TELescope and the literature. We construct a hierarchical Bayesian framework, incorporating several uncertainties including photometric error, peculiar velocities, dust extinction, and intrinsic variations, for principled and coherent statistical inference. SN Ia light-curve inferences are drawn from the global posterior probability of parameters describing both individual supernovae and the population conditioned on the entire SN Ia NIR data set. The logical structure of the hierarchical model is represented by a directed acyclic graph. Fully Bayesian analysis of the model and data is enabled by an efficient Markov Chain Monte Carlo algorithm exploiting the conditional probabilistic structure using Gibbs sampling. We apply this framework to the JHK_s SN Ia light-curve data. A new light-curve model captures the observed J-band light-curve shape variations. The marginal intrinsic variances in peak absolute magnitudes are $\sigma(M_J) = 0.17 \pm 0.03$, $\sigma(M_H) = 0.11 \pm 0.03$, and $\sigma(M_{Ks}) = 0.19 \pm 0.04$. We describe the first quantitative evidence for correlations between the NIR absolute magnitudes and J-band light-curve shapes, and demonstrate their utility for distance estimation. The average residual in the Hubble diagram for the training set SNe at $cz > 2000$ km s^{-1} is 0.10 mag. The new application of bootstrap cross-validation to SN Ia light-curve inference tests the sensitivity of the statistical model fit to the finite sample and estimates the prediction error at 0.15 mag. These results demonstrate that SN Ia NIR light curves are as effective as corrected optical light curves, and, because they are less vulnerable to dust absorption, they have great potential as precise and accurate cosmological distance indicators.

Key words: distance scale – supernovae: general

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). The cosmic acceleration may be caused by a dark energy component of the universe (Frieman et al. 2008 provide a recent review). 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). 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).

The effectiveness of SNe Ia as distance indicators has been improved greatly by the construction of empirical methods that exploit relationships between peak optical luminosities of SNe Ia and distance-independent measures such as light-curve shape or color that have been observed in the burgeoning sample of nearby low-z SNe Ia (Hamuy et al. 1996a; Riess et al. 1999; Jha et al. 2006; Hicken et al. 2009a). Methods have included $\Delta m_{15}(B)$ (Phillips 1993; Hamuy et al. 1996b; Phillips et al. 1999), MLCS (Riess et al. 1996a, 1998; Jha et al. 2007), "stretch" (Goldhaber et al. 2001), CMAGIC (Wang et al. 2003), and SALT (Guy et al. 2005, 2007). The largest systematic uncertainty that limits the precision of rest-frame optical light curves is dust extinction in the host galaxy and the entanglement of dust reddening with the intrinsic color variations of SNe (e.g., Conley 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). Studies of nearby SN Ia light curves in the NIR have found 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). Furthermore, the effect of dust extinction is significantly diminished at near-infrared wavelengths, relative to the optical. The combination of optical and near-infrared observations of SN Ia light curves could lead to even better SN Ia distances (Krisciunas et al. 2007).

Wood-Vasey et al. (2008; hereafter WV08) compiled the largest homogeneous sample of NIR SN Ia observations, taken with the Peters Automated InfraRed Imaging TELescope (PAIRITEL; Bloom et al. 2006). After combining these with NIR light-curve observations from the literature to yield a sample of 41 NIR SNe Ia, they constructed template light curves by interpolating and smoothing the data. They measured the scatter in the absolute magnitudes at the time of B maximum in each of the J, H, and K_s bands and found $\sigma(M_H) \approx 0.15$ mag, $\sigma(M_J) \approx 0.33$ mag, and $\sigma(M_{Ks}) \approx 0.26$ mag. This analysis did not take into account NIR light curve shape variations, but it was found, as in Krisciunas et al. (2004a), that the Hubble diagram residuals had no trend with the optical light-curve width.

The purpose of this paper is two-fold. First, we formulate the hierarchical Bayesian approach to probabilistic inference with SN Ia light curves in general. A proper Bayesian approach provides a principled, coherent framework for inference based on the joint probability density over all quantities of interest conditioned on the available data. It is natural to employ a simultaneous multi-level approach and derive joint probability densities over the parameters of individual supernova light curves, their

---

[1] Current address: Department of Physics & Astronomy, 100 Allen Hall, 3941 O'Hara St., University of Pittsburgh, Pittsburgh, PA 15260, USA.

distance moduli, and also the variables that describe the population, including those governing the joint probability distributions over multiband absolute magnitudes and light-curve shape parameters. This approach enables consistent statistical inference of all hierarchical parameters, by coherently incorporating several sources of uncertainty, including peculiar velocity uncertainties, photometric measurement errors, intrinsic randomness of light curves, and dust extinction into the global posterior probability density conditioned on the entire data set simultaneously. This framework leads to a natural and consistent method for probabilistic distance prediction with new SN Ia light-curve data. The logical structure of our hierarchical model for SN Ia light-curve inference is demonstrated by the equivalent directed acyclic graph (DAG), a graphical model that facilitates easy inspection of the probabilistic relationships.

Although the probabilities for fully Bayesian analysis of SN Ia light curves are simple to write down, the joint posterior distribution is generally non-Gaussian and difficult to evaluate. To enable probabilistic inference, we have developed a Markov Chain Monte Carlo (MCMC) algorithm, BAYESN, designed to exploit the conditional probabilistic structure using Gibbs sampling. We employ this code for both training the statistical model and using the model to predict distances. The use of advanced sampling methods facilitates the computation of marginal probabilities of parameters from the global joint posterior density.

In the second part of the paper, we apply this framework to the NIR SN Ia light-curve data from the compilation of WV08. We first construct model template light curves for the JHK_s bands. We compute fixed maximum likelihood template models between -10 and 20 days for the H and K_s using all available data. The J-band data are typically much less noisy than the H and K, so we construct an extensive J-band model between -10 and 60 days that accounts for light-curve variations, in particular the structure around the second maximum. Next, we apply the BAYESN method to simultaneously (1) fit the individual JHK_s light curves, (2) compute the population characteristics, especially the absolute magnitude variances and covariances with J-band light-curve shape, and (3) estimate the joint and marginal uncertainties over all hierarchical parameters. We construct a Hubble diagram for the training set SNe Ia and compute its residual errors.

The average Hubble diagram residual of the training set SNe is an optimistic assessment of the predictive ability of the statistical model for SN Ia light curves because it uses the SN data twice: first for estimating the model parameters (training), and second in evaluating the error of its "predictions." Hence, the residuals, or training errors, underestimate the expected prediction error. This effect is present for all models based on finite training data, and is particularly important for small sample sizes. We perform bootstrap cross-validation to realistically estimate the out-of-sample prediction error and to test the sensitivity to the finite NIR SN Ia sample. This technique ensures that the same SNe are not simultaneously used for training and prediction. It has not been used previously in SN Ia statistical modeling and inference.

This paper demonstrates hierarchical Bayesian modeling and distance estimation for SN Ia light curves in the NIR only. The application of these methods to combined optical and NIR light curves for the estimation of dust and distances will be described in a subsequent paper (K. S. Mandel et al. 2009, in preparation).

This paper is organized as follows: in Section 2, we describe the hierarchical Bayesian framework for SN Ia light-curve inference. The structure of the hierarchical model can be depicted formally as a directed acyclic graph presented in Section 2.3. In Section 2.4, we describe BAYESN, an MCMC algorithm designed for computing posterior inferences in the hierarchical framework. In Section 3, the construction of template light-curve models in JHK_s is described. In Section 4, we apply this approach to the NIR light-curve data, and summarize the posterior inferences for both individual SNe and the population. In Section 4.4, we construct Hubble diagrams by applying the statistical model and describe the application of bootstrap cross-validation to estimate the prediction error. In Section 4.5, we discuss the potential impact of dust in the NIR sample. We conclude in Section 5. In Appendix A, we briefly review the conditional independence properties of graphical models. Appendix B presents mathematical details of the BAYESN method, and Appendix C describes its use in practice.

2. HIERARCHICAL BAYESIAN FRAMEWORK FOR SN IA LIGHT-CURVE INFERENCE

Simple Bayesian analysis describes an inference problem in which a generative model $\mathcal{H}$ with a free parameter θ is assumed to underlie the observed data $\mathcal{D}$. The Bayesian paradigm is to derive inferences on θ from the posterior density of the parameter conditioned on the data: $P(\theta | \mathcal{D}, \mathcal{H}) \propto P(\mathcal{D} | \theta, \mathcal{H}) P(\theta | \mathcal{H})$, where the first factor is the likelihood function and the second factor is the prior on the model parameter.

Hierarchical, or multi-level, Bayesian analysis is a modern paradigm of statistical modeling, which enables the expression of rich probabilistic models with complex structure on multiple logical levels (Gelman et al. 2003). For example, if $\mathcal{D}_i$ represents the data on *individual i*, and θ_i is a model parameter describing i, the values of θ_i themselves may be drawn from a prior or *population* distribution $P(\theta_i | \alpha, \beta)$, which in turn depends on unknown variables that describe the *group* level probabilistic model. These unknown variables α, β are termed *hyperparameters* to distinguish them from the individual level parameters θ_i. The hierarchical Bayesian joint posterior distribution over all parameters $\{\theta_i\}$ and hyperparameters α, β conditioned on the data for many (N) individuals $\mathcal{D} = \{\mathcal{D}_i\}$ is then

$$P(\{\theta_i\}; \alpha, \beta | \mathcal{D}) \propto \left[\prod_{i=1}^{N} P(\mathcal{D}_i | \theta_i) P(\theta_i | \alpha, \beta) \right] P(\alpha, \beta), \quad (1)$$

where the last factor represents the *hyperprior* density on the hyperparameters. The fully Bayesian approach is to analyze the full joint posterior density of all parameters and hyperparameters simultaneously conditioned on the entire data set. This ensures the complete and consistent accounting of uncertainty over all the inferred parameters. We can build in complex probabilistic structure by layering single-level models, at the individual level and also at possibly multiple population levels, and expressing the conditional relationships that connect them. The hierarchical Bayesian paradigm is very well suited for combining information and uncertainties from many logical sources of randomness, interacting in non-trivial ways, in a principled, coherent and consistent statistical framework for studying structured data. This is the strategy we adopt in this paper.

In contrast to more classical methods of model fitting, the Bayesian approach is less concerned with the optimization problem, i.e., finding the "best" values or point estimates of

model parameters fitted to given data, and more concerned with the construction of the full joint probability model, consistent integration over uncertainties, and the simulation and sampling of the joint posterior density. For most non-trivial models, the integrations of interest are analytically intractable, so we employ modern computation techniques to enable probabilistic inference. We introduce an MCMC algorithm (BAYESN) that uses stochastic simulation to calculate posterior inferences from the hierarchical framework.

2.1. SN Ia Light-curve Models

A central task in the statistical analysis of SN Ia light-curve data is the fitting of empirical light-curve models to time-series photometric data in multiple passbands. A data set for an individual SN s, $\{\mathcal{D}^F\}$, consists of observations in n photometric filters, $F \in \{F^1, \ldots, F^n\}$ (e.g.,$\{B, V, J, H\}$). Let $\mathcal{D}^F = \{t_i, m_i^F, \sigma_{F,i}^2\}_{i=1}^{N_F}$ be the set of N_F observations in band F. We assume that these have already been corrected for time dilation (Blondin et al. 2008) and K-corrected to the SN rest frame, so that t_i is the rest-frame phase of the observation referenced to the time of maximum light in the B band (i.e., $t = 0$ corresponds to $B_{\max}$), m_i^F is the apparent magnitude in filter F at this phase, and $\sigma_{F,i}^2$ is the measurement variance.

We wish to fit to these data a light-curve model, $F_0 + l^F(t, \boldsymbol{\theta}^F)$, where F_0 is the F-band apparent magnitude of the model at a reference time $t = 0$, and $l^F(t, \boldsymbol{\theta}^F)$ is the *normalized* light-curve model such that $l^F(0, \boldsymbol{\theta}^F) = 0$. A vector of *light-curve shape parameters*, $\boldsymbol{\theta}^F$, governs the functional form of the light-curve models and may take on different values for each SN.

At the present time there is no known set of physical functions $l^F(t, \boldsymbol{\theta}^F)$ describing the temporal evolution of SNe light curves. Theoretical SN Ia models alone provide insufficient guidance, so these functions must be constructed from the light-curve data itself. Typical parameterizations are motivated by a combination of simplicity, intuition, mathematical convenience, and the empirical study of detailed light-curve data of a large sample of SNe. The ultimate utility of a functional form lies in its ability to fit the observed modes of variation in data and capture the observable information in the light curves. Examples of light-curve shape functions are the following.

1. "Stretch" template method (Perlmutter et al. 1999; Goldhaber et al. 2001): $l^F(t, \theta^F) = f(s^F t)$, where the shape parameter is $\theta^F = s^F$, the "stretch" factor in filter F, and $f(\cdot)$ is a fiducial light curve, e.g., the Leibundgut template (Leibundgut 1989), which is a function of the rest-frame SN phase with respect to t_0, the time of $B_{\max}$.

2. The $\Delta m_{15}(B)$ decline rate parameterization (Hamuy et al. 1996b; Phillips et al. 1999). Here the light-curve shape parameter is $\theta^B = \Delta m_{15}(B)$, the magnitude decline between the peak and 15 days after the peak in B-band. The light-curve function is defined at particular values of θ_i^B using *BVI* light-curve templates generated from observed SNe. Interpolation is used to fit light curves at intermediate values of $\Delta m_{15}(B)$. Prieto et al. (2006) presented an updated formulation.

3. MLCS (Riess et al. 1996a, 1998; Jha et al. 2007): the light-curve model in *UBRVI* is of the form: $l^F(t, \Delta) = F_0(t) + \Delta P_F(t) + \Delta^2 Q_F(t)$, where $F_0(t)$, $P_F(t)$, and $Q_F(t)$ are defined by templates. The light-curve shape parameter is $\theta^F = \Delta$.

In this section, we do not assume any particular form for the light-curve models $l^F(t; \boldsymbol{\theta}^F)$. The results will be applicable to a broad class of possible models. Without loss of generality, the light-curve model can depend on some parameters ($\boldsymbol{\theta}_L^F$) linearly and others ($\boldsymbol{\theta}_{NL}^F$) nonlinearly. Hence, a general form of a light-curve model in band F for the data is

$$m_i^F = F_0 + l_0^F\left(t_i; \boldsymbol{\theta}_{NL}^F\right) + \boldsymbol{l}_1^F\left(t_i; \boldsymbol{\theta}_{NL}^F\right) \cdot \boldsymbol{\theta}_L^F + \epsilon_i^F, \qquad (2)$$

where $\boldsymbol{l}_1^F(t; \boldsymbol{\theta}_{NL}^F)$ is a vector of coefficients to the linear parameters. It is convenient for computational purposes to separate the linear from nonlinear shape parameters. The parameter F_0 could be considered a linear parameter and included with $\boldsymbol{\theta}_L$. However, they are physically distinct quantities, as F_0 sets the apparent magnitude scale for the light curve whereas $\boldsymbol{\theta}_L^F$ generally models the shape of the light curve, so we keep them separate.

2.2. Constructing the Global Posterior Density

2.2.1. Light-curve Likelihood Function

Assuming Gaussian noise for ϵ_i^F, we can write the likelihood of the light-curve model parameters $\boldsymbol{\theta}^F = (\boldsymbol{\theta}_L^F, \boldsymbol{\theta}_{NL}^F)$ for a single SN data set $\mathcal{D}^F$ in one band F. Define the vectors and matrices:

$$\boldsymbol{m}^F = (m^F(t_1), \ldots, m^F(t_{N_F}))^T \qquad (3)$$

$$\boldsymbol{L}_0^F\left(\boldsymbol{\theta}_{NL}^F\right) = \left(l_0^F\left(t_1; \boldsymbol{\theta}_{NL}^F\right), \ldots, l_0^F\left(t_{N_F}; \boldsymbol{\theta}_{NL}^F\right)\right)^T \qquad (4)$$

$$\boldsymbol{L}_1^F\left(\boldsymbol{\theta}_{NL}^F\right) = \left(\boldsymbol{l}_1^F\left(t_1; \boldsymbol{\theta}_{NL}^F\right), \ldots, \boldsymbol{l}_1^F\left(t_{N_F}; \boldsymbol{\theta}_{NL}^F\right)\right)^T. \qquad (5)$$

Let us construct a vector of ones, $\boldsymbol{1}$, of the same length as the data $\boldsymbol{m}_s^F$, and a measurement error covariance matrix $\boldsymbol{W}^F$. Due to the practical difficulties of estimating the error covariances, the current standard assumption is that the error terms ϵ_i^F are independent, so that $W_{ii}^F = \sigma_{F,i}^2$ is diagonal. The likelihood function for the light curve can be compactly written as

$$P(\mathcal{D}^F | F_0, \boldsymbol{\theta}^F) = N\left(\boldsymbol{m}^F | \boldsymbol{1} F_0 + \boldsymbol{L}_0^F\left(\boldsymbol{\theta}_{NL}^F\right) + \boldsymbol{L}_1^F\left(\boldsymbol{\theta}_{NL}^F\right) \cdot \boldsymbol{\theta}_L^F, \boldsymbol{W}^F\right), \qquad (6)$$

where $N(\boldsymbol{x} | \boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x)$ denotes the multivariate normal density in the random vector $\boldsymbol{x}$ with mean $\boldsymbol{\mu}_x$ and covariance $\boldsymbol{\Sigma}_x$. Since the photometric observations in multiple filters are sampled with independent noise, the likelihood function of all light-curve parameters over all bands given the multiband data $\{\mathcal{D}^F\}$ is the simple product of n single-filter likelihoods:

$$P(\{\mathcal{D}^F\} | \boldsymbol{\phi}) = \prod_F P(\mathcal{D}^F | F_0, \boldsymbol{\theta}^F). \qquad (7)$$

We define the *observable* (or apparent) parameters vector $\boldsymbol{\phi} = (F_0^1, \ldots, F_0^n; \boldsymbol{\theta}^{F^1}, \ldots, \boldsymbol{\theta}^{F^n})$. This vector, with the light-curve model functions $l^F(t; \boldsymbol{\theta}^F)$, encodes all the information needed to reconstruct the *apparent* light curve of a single SN, i.e., the apparent magnitudes at the reference time and the light-curve shape parameters in each of n photometric bands F. Similarly, we define the *intrinsic* (or absolute) parameters vector $\boldsymbol{\psi} = (M_{F^1}, \ldots, M_{F^n}; \boldsymbol{\theta}^{F^1}, \ldots, \boldsymbol{\theta}^{F^n})$ encoding all the information describing the *absolute* light curves of the SN. The absolute magnitude at peak in filter F is $M_F = F_0 - \mu - A_F$, where μ is the distance modulus and A_F is the dust absorption in that filter, for a particular SN.

2.2.2. Redshift-distance Likelihood Function

Type Ia SNe can be used as distance indicators because we possess some knowledge of their *relative* distances in the low-redshift regime (where they are independent of the cosmological parameters Ω_M, Ω_Λ, and w) from the Hubble law and measured redshifts to the host galaxies of the SNe. However, inference of true luminosities and absolute distances requires external calibration (e.g., from Cepheids) or knowledge of the Hubble constant, H_o, which has not yet been independently measured to high precision. If we are concerned only with relative distance estimation, it is sufficient to fix the distance scale with an assumed $h = H_o/100 \, \text{km s}^{-1}$. The uncertainty in these local distances is then dominated by the peculiar velocity field, which we model probabilistically.

Let z_c be the cosmological redshift of an SN. The measured redshift is z, with measurement variance σ_z^2, corrected to the cosmic microwave background (CMB) and the local infall flow model of Mould et al. (2000). In a smooth cosmological model, the distance modulus is related to z_c: $\mu = f(z_c) = 25 + 5 \log_{10}[d_L(z_c) \, \text{Mpc}^{-1}]$, where $d_L(z_c)$ is the luminosity distance in Mpc. If we model the effect of random peculiar velocity as a Gaussian noise with variance σ_{pec}^2, then $z = z_c + N(0, \sigma_{\text{pec}}^2/c^2 + \sigma_z^2)$, and the likelihood function is $P(z|\mu) = N(z| f^{-1}(\mu), \sigma_{\text{pec}}^2/c^2 + \sigma_z^2)$. The posterior density of the distance modulus conditioning only on the redshift is $P(\mu|z) \propto P(z|\mu)$. We use a flat prior $P(\mu) \propto 1$, since we have no *a priori* knowledge about μ without the data. For recession velocities $cz \gg \sigma_{\text{pec}}$, $f(z_c)$ can be linearized about the fixed z to find $f^{-1}(\mu)$, so that, to a good approximation,

$$P(\mu \,|\, z) = N\left[\mu \,\middle|\, f(z), \sigma_\mu^2 = [f'(z)]^2 \left(\sigma_z^2 + \frac{\sigma_{\text{pec}}^2}{c^2}\right)\right]. \quad (8)$$

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

$$\sigma_\mu^2 = \left(\frac{5}{z \ln 10}\right)^2 \left[\sigma_z^2 + \frac{\sigma_{\text{pec}}^2}{c^2}\right]. \quad (9)$$

Under the assumption that peculiar velocity uncertainty amounts to Gaussian noise, at recession velocities $cz < 5\sigma_{\text{pec}}$, the approximation of $P(\mu|z)$ with a normal distribution breaks down, due to the nonlinearity of the logarithm. This effect is inconsequential for our analysis because even though the distribution becomes non-Gaussian, at such low recession velocities, its width in magnitudes is much larger than the dispersion in SN Ia absolute magnitudes. The redshift of a very low z SN Ia carries little information about the absolute magnitude, so that $P(\mu|z)$ is essentially flat over the width of the posterior density in μ conditioning on the light curves of the SN. Hence, the exact form of $P(\mu|z)$ is irrelevant in this regime. SN Ia light curves can be used to infer the distances of these near-field SNe and to measure the local velocity field (Riess et al. 1995; Haugbølle et al. 2007; Neill et al. 2007).

2.2.3. The SN Ia Population Distribution

The utility of SNe Ia for cosmological studies lies in the observed correlation of their peak luminosities with the shapes of their optical light curves. Although the peak optical luminosities of SNe Ia range over a factor of 3, using correlations with light-curve shape reduces the scatter about the Hubble line to less than ~ 0.20 mag. Physical modeling and simulation of SN Ia progenitors may provide useful explanations for the observed relationships between the observable properties of SNe Ia and their peak luminosities. Such work may also describe detailed probabilistic relationships between the two. For example, we can define the joint population distribution of absolute magnitudes (at the reference time) and the observable light-curve shapes in multiple passbands. In our notation this population distribution is $P(\psi | \text{Physical Parameters})$ where the "Physical Parameters" may include, for example, the mass of the progenitor, the chemical composition and distribution within the progenitor, and the details of the explosion mechanism. Hillebrandt & Niemeyer (2000) provide a review of progress in SN Ia explosion modeling.

In the absence of such detailed information, we learn the probabilistic relationships from the data. We describe the SN Ia population distribution as $P(\psi | \mu_\psi, \Sigma_\psi)$. Here, μ_ψ is a vector of hyperparameters that describe the average intrinsic characteristics, and Σ_ψ is a collection of hyperparameters describing the dispersion (variance) and correlations of the intrinsic characteristics of absolute light curves. Since we have no information on these hyperparameters *a priori*, we seek to estimate them (and their uncertainties) from the data.

We will include this distribution in the global posterior density in the mathematical form of a "prior" on the intrinsic parameters ψ of a single SN. However, it is better to think of this distribution as a "population" distribution from which the intrinsic parameters are randomly drawn. Its hyperparameters are unknown and must be estimated simultaneously from the data. It has a different interpretation in the context of the hierarchical model than the fixed prior of the simple Bayesian treatment (which has no hyperparameters to be estimated).

Since we have no *a priori* information on the functional form of $P(\cdot | \mu_\psi, \Sigma_\psi)$, we must make some assumptions. The simplest choice for a multivariate probability density that models correlations between parameters is the multivariate Gaussian. In the rest of this paper we will assume $P(\cdot | \mu_\psi, \Sigma_\psi) = N(\cdot | \mu_\psi, \Sigma_\psi)$ with an unknown mean vector $\mu_\psi = \mathbb{E}(\psi_s)$ and unknown covariance matrix $\Sigma_\psi = \text{Var}(\psi_s)$. The intrinsic parameters of individual SNe are independent, identically distributed random variables drawn from this probability density: $\psi \sim N(\mu_\psi, \Sigma_\psi)$. If the data indicate a different distribution from the one we have assumed, we can attempt another choice of the form of the intrinsic population distribution.

The population hyperparameters μ_ψ, Σ_ψ are the most important variables in this hierarchical framework. During the training process, they model the intrinsic statistical properties of the SN Ia light curves, including the average behavior, intrinsic variability, correlations between different modes of light-curve shape variation, correlations between absolute magnitudes in different filters, and cross-correlations between light-curve shape parameters and the absolute magnitudes. When the model is used to make predictions, they are crucial for using this information, and its uncertainty, to make distance estimates for new SN Ia light curves.

2.2.4. Incorporating Dust Information

Dust along the line of sight from the SN to the observer causes both extinction and reddening of the emitted light. These effects originate from Galactic dust, which has been measured and mapped (Schlegel et al. 1998), and dust in the SN's host galaxy, which is often more important, poorly understood and is currently the largest systematic uncertainty in cosmological

inference with SNe Ia (Conley et al. 2007). Previous efforts to estimate dust extinction from SN Ia color excesses include Riess et al. (1996b), Phillips et al. (1999), and Krisciunas et al. (2000).

We incorporate the effects of host galaxy dust on SN observations probabilistically within the full statistical model as follows. The extinction in a given passband is denoted A_F. Assuming a CCM reddening law (Cardelli et al. 1989), the extinction in a given band can be related to the visual extinction A_V by $A_F/A_V = a_F + b_F R_V^{-1}$, so that there are two free parameters: the magnitude of visual extinction A_V and the slope of the extinction law in the optical bands R_V. The fixed regression coefficients a_F and b_F are determined from the dust reddening analysis of SNe spectra. Jha et al. (2007) suggested using an exponential prior on the non-negative A_V extinction to a particular SN. Interpreted as a population distribution, this can be incorporated into our framework as

$$P(A_V, R_V|\tau_{A_V}, \boldsymbol{\alpha}_R) = \text{Expon}(A_V|\tau_{A_V})P(R_V|\boldsymbol{\alpha}_R), \quad (10)$$

where τ_{A_V} is a hyperparameter describing the exponential scale length or the average amount of visual extinction to the population of SNe Ia. The form of the population distribution of the R_V is unknown, but we suppose that it may be specified with hyperparameters $\boldsymbol{\alpha}_R$. For example, if R_V is fixed to a single known value, e.g., $\boldsymbol{\alpha}_R = 1.7$ or 3.1, we may set $P(R_V|\boldsymbol{\alpha}_R) = \delta(R_V - \boldsymbol{\alpha}_R)$. Or one may allow the R_V to vary within a population probability density, e.g., $R_V \sim N(\mu_R, \sigma_R^2)$, where the hyperparameters $\boldsymbol{\alpha}_R = (\mu_R, \sigma_R^2)$ may be fixed or we may attempt to learn them from the data, if the data are sufficiently informative. It is not known *a priori* whether A_V and R_V can be treated as independent random variables and if the population probability density of (A_V, R_V) is separable as indicated in Equation (10).

When modeling the near-infrared observations, it makes more sense to reference the extinction values to the H-band, rather than to A_V. For given A_H, R_V values, the extinction in any band, A_F, can be computed from the dust law. The ratio of near-infrared to visual extinction is roughly $A_H/A_V \sim 0.2$. Since the behavior of dust in the near-infrared is relatively insensitive to the slope of the reddening law in the optical bands, R_V, it makes sense to set it to a fixed representative global value $\boldsymbol{\alpha}_R$. With this choice, the extinction population distribution is

$$P(A_H, R_V|\tau_{A_H}, \boldsymbol{\alpha}_R) = \text{Expon}(A_H|\tau_{A_H})\delta(R_V - \boldsymbol{\alpha}_R), \quad (11)$$

where τ_{A_H} is the exponential scale length in magnitudes of the population distribution of H-band extinctions. To simplify the notation, we denote this hyperparameter as $\tau_A \equiv \tau_{A_H}$, which controls the dust scale for all filters through the reddening law.

2.2.5. Conditionally Conjugate Hyperpriors

To estimate the population hyperparameters $\boldsymbol{\mu}_\psi$ and $\boldsymbol{\Sigma}_\psi$, we make our priors on them explicit, called *hyperpriors*. If we lack any external motivating evidence, we should choose non-informative or diffuse hyperpriors. Additionally, it is convenient to choose the hyperprior from a parametric family that is conditionally conjugate to the parametric family of the population density. This means that the posterior density of the hyperparameters conditioned on the values of the other parameters and data is from the same parametric family of probability densities as the hyperprior. This property is advantageous because if one can sample directly (generate random numbers from) the

hyperprior density, then one can sample directly from the conditional posterior density of the hyperparameters. This is useful for constructing Markov chains for statistical computation of the posterior density using Gibbs sampling (Section 2.4).

The hyperparameters $\boldsymbol{\mu}_\psi$ and $\boldsymbol{\Sigma}_\psi$ describe a multivariate normal density on $\boldsymbol{\psi}_s$. The conjugate family to the multivariate normal with unknown mean and covariance matrix is the Normal-Inverse-Wishart. This hyperprior can be expressed as $P(\boldsymbol{\mu}_\psi, \boldsymbol{\Sigma}_\psi) = P(\boldsymbol{\mu}_\psi|\boldsymbol{\Sigma}_\psi)P(\boldsymbol{\Sigma}_\psi)$ such that

$$\boldsymbol{\Sigma}_\psi \sim \text{Inv-Wishart}_{\nu_0}(\boldsymbol{\Lambda}_0^{-1}) \quad (12)$$

$$\boldsymbol{\mu}_\psi|\boldsymbol{\Sigma}_\psi \sim N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_\psi/\kappa_0). \quad (13)$$

The non-informative or diffuse conditionally conjugate density is obtained in the limit as $\kappa_0 \to 0, \nu_0 \to -1, |\boldsymbol{\Lambda}_0| \to 0$ (Gelman et al. 2003; the conventions regarding the Wishart distributions differ: we choose the convention that if $W \sim \text{Inv-Wishart}_\nu(S^{-1})$ is a random matrix, then $\mathbb{E}(W) = S/(\nu - d - 1)$, where d is the dimension of the $d \times d$ covariance matrix). In this limit, the hyperprior of the population mean $\boldsymbol{\mu}_\psi$ becomes flat and the hyperprior of the covariance $\boldsymbol{\Sigma}_\psi$ is a diffuse distribution over the space of positive semi-definite matrices, so that the hyperprior does not favor any particular solution.

For the extinction exponential scale length $\tau_A > 0$, we choose a uniform positive hyperprior, expressing no prior preference for a particular value. If this is viewed as an Inv-Gamma$(-1, 0)$ density on τ_A, it is conditionally conjugate to the exponential distribution.

2.2.6. The Global Posterior Density

We now have the elements necessary to construct the full joint posterior density of the sample of SNe Ia. A single SN s with multiband light-curve data $\mathcal{D}_s = \{\mathcal{D}_s^F\}$, and redshift z_s, is described by intrinsic light-curve parameters $\boldsymbol{\psi}_s$, observable parameters $\boldsymbol{\phi}_s$, and distance modulus μ_s, with dust extinction modeled by A_H^s and R_V^s. The relations between these parameters can be encoded as follows. Let $\boldsymbol{v}$ be a constant indicator vector with the jth component v^j defined as

$$v^j \equiv \begin{cases} 1, & \text{if } \phi^j, \psi^j \text{ are magnitudes, e.g., } F_0 \text{ or } M_F \\ 0, & \text{if } \phi^j, \psi^j \text{ are shape parameters.} \end{cases} \quad (14)$$

Furthermore, define the vectors $\boldsymbol{A}_s$ with the jth component A_s^j: if ϕ_s^j and ψ_s^j are magnitudes in band F, then

$$A_s^j \equiv A_{F,s}^{\text{Gal}} + A_F(A_H^s, R_V^s), \quad (15)$$

otherwise, $A_s^j = 0$ if ϕ_s^j and ψ_s^j are shape parameters. The non-zero components depend on the host galaxy reddening law and H-band extinction, $A_{F,s}^{\text{Gal}}$ is the Galactic extinction, and $A_F(A_H^s, R_V^s)$ is the dust extinction in filter F as a function of A_H^s and R_V^s using the dust law. The relationship between the intrinsic and observable parameters of SN s can then be written compactly as

$$\boldsymbol{\phi}_s = \boldsymbol{\psi}_s + \boldsymbol{v}\mu_s + \boldsymbol{A}_s. \quad (16)$$

This equation encodes the relationship between apparent magnitudes, absolute magnitudes, extinction, and distance moduli. In this expression, neither dust nor distance modify the light-curve shape parameters, $\boldsymbol{\theta}_s^F$, common to both the observable $\boldsymbol{\phi}_s$ and intrinsic $\boldsymbol{\psi}_s$ vectors. The joint posterior probability density for the parameters of a single SN, conditioned on the values of the hyperparameters and the data, is proportional to the product of

1. the probability of observing the photometric data given the apparent light curve,
2. the probability of the distance modulus given the measured redshift,
3. the probability of an absolute light curve equal to the apparent light curve minus the distance modulus and extinction, and
4. the probability of the extinction value and the dust law,

conditioned on the population hyperparameters of the absolute light curves and dust properties:

$$
\begin{aligned}
P(\boldsymbol{\phi}_s, \mu_s, &A_H^s, R_V^s \,|\, \mathcal{D}_s, z_s; \boldsymbol{\mu}_\psi, \boldsymbol{\Sigma}_\psi, \tau_A, \alpha_R) \\
&\propto P(\mathcal{D}_s | \boldsymbol{\phi}_s) \times P(\mu_s | z_s) \\
&\times P(\boldsymbol{\psi}_s = \boldsymbol{\phi}_s - \mathbf{v}\mu_s - \mathbf{A}_s \,|\, \boldsymbol{\mu}_\psi, \boldsymbol{\Sigma}_\psi) \\
&\times P(A_H^s, R_V^s \,|\, \tau_A, \alpha_R).
\end{aligned}
\tag{17}
$$

Now consider the full database of SN Ia light curves $\mathcal{D} = \{\mathcal{D}_s\}$ with measured cosmological redshifts $\mathcal{Z} = \{z_s\}$. The global joint posterior density of all SN observables $\{\boldsymbol{\phi}_s\}$, distance moduli $\{\mu_s\}$, dust parameters $\{A_H^s, R_V^s\}$, and the population hyperparameters conditioned on the database $\mathcal{D}, \mathcal{Z}$ is proportional to the product of N_{SN} individual conditional posterior densities multiplied by the hyperpriors:

$$
\begin{aligned}
P(\{\boldsymbol{\phi}_s, &\mu_s, A_H^s, R_V^s\}; \boldsymbol{\mu}_\psi, \boldsymbol{\Sigma}_\psi, \tau_A, \alpha_R \,|\, \mathcal{D}, \mathcal{Z}) \\
&\propto \left[\prod_{s=1}^{N_{\mathrm{SN}}} P(\boldsymbol{\phi}_s, \mu_s, A_H^s, R_V^s \,|\, \mathcal{D}_s, z_s; \boldsymbol{\mu}_\psi, \boldsymbol{\Sigma}_\psi, \tau_A, \alpha_R) \right] \\
&\times P(\boldsymbol{\mu}_\psi, \boldsymbol{\Sigma}_\psi) \times P(\tau_A, \alpha_R).
\end{aligned}
\tag{18}
$$

For full generality we have derived the global joint posterior density in the case that we wish to estimate the probable values of the hyperparameters α_R from the data. If we fix the R_V^s to a fixed global value α_R, this is equivalent to evaluating the above joint density conditioned on $R_V^s = R_V = \alpha_R$. All fully Bayesian inferences on the remaining parameters and hyperparameters are based on mapping out this global joint posterior density. The marginal posterior density of the hyperparameters, $P(\boldsymbol{\mu}_\psi, \boldsymbol{\Sigma}_\psi, \tau_A | \mathcal{D}, \mathcal{Z}, R_V)$ or $P(\boldsymbol{\mu}_\psi, \boldsymbol{\Sigma}_\psi, \tau_A, \alpha_R | \mathcal{D}, \mathcal{Z})$, is obtained by integration over the individual SN parameters $\{\boldsymbol{\phi}_s, \mu_s, A_H^s, R_V^s\}$ or $\{\boldsymbol{\phi}_s, \mu_s, A_H^s\}$.

We refer to the set of SNe in $\mathcal{D}, \mathcal{Z}$ as a training set, and "training" means computing the marginal posterior density of the hyperparameters, conditioned on these data. In the Bayesian paradigm we are interested not just in point estimates of the hyperparameters, e.g., "best values" $\hat{\boldsymbol{\mu}}_\psi, \hat{\boldsymbol{\Sigma}}_\psi, \hat{\tau}_A$, but on their joint posterior probability density as a quantification of their uncertainties.

2.2.7. The Predictive Posterior Density

The ultimate purpose of SN Ia light-curve inference is to estimate luminosity distances to distant SN that are not included in the nearby, low-z training set. That is, given observations of a new SN's multiband light curve $\tilde{\mathcal{D}}_s$, we wish to fit the light-curve model and to predict the distance modulus. Prediction differs from training in the fact that we do not use any prior information on the distance (from, e.g., the redshift) in our probability calculus. The predictive posterior density for the

new SN $\tilde{s}$ (with parameters denoted by tilde) conditioned on the population hyperparameters and the new light-curve data $\tilde{\mathcal{D}}_s$ is

$$
\begin{aligned}
P(\tilde{\boldsymbol{\phi}}_s, &\tilde{\mu}_s, \tilde{A}_H^s, \tilde{R}_V^s \,|\, \tilde{\mathcal{D}}_s; \boldsymbol{\mu}_\psi, \boldsymbol{\Sigma}_\psi, \tau_A, \alpha_R) \\
&\propto P(\tilde{\mathcal{D}}_s | \tilde{\boldsymbol{\phi}}_s) \\
&\times P(\tilde{\boldsymbol{\psi}}_s = \tilde{\boldsymbol{\phi}}_s - \mathbf{v}\tilde{\mu}_s - \tilde{\mathbf{A}}_s \,|\, \boldsymbol{\mu}_\psi, \boldsymbol{\Sigma}_\psi) \\
&\times P(\tilde{A}_H^s, \tilde{R}_V^s \,|\, \tau_A, \alpha_R).
\end{aligned}
\tag{19}
$$

We must also incorporate our (joint) uncertainties of the hyperparameters. This is encapsulated in the marginal posterior density of the hyperparameters from the training set. The full predictive posterior probability density for the new SN $\tilde{s}$ is the previous expression multiplied by the training posterior density $P(\boldsymbol{\mu}_\psi, \boldsymbol{\Sigma}_\psi, \tau_A, \alpha_R | \mathcal{D}, \mathcal{Z})$ and integrated over the probability of the hyperparameters $\boldsymbol{\mu}_\psi, \boldsymbol{\Sigma}_\psi, \tau_A, \alpha_R$. The marginal predictive posterior density of the new SN's distance modulus $\tilde{\mu}_s$, $P(\tilde{\mu}_s | \tilde{\mathcal{D}}_s, \mathcal{D}, \mathcal{Z})$, is obtained by integrating this over the remaining parameters, $\tilde{\boldsymbol{\phi}}_s, \tilde{A}_H^s, \tilde{R}_V^s$.

2.3. Representation as a Directed Acyclic Graph

We have constructed the posterior density of all individual parameters and population-level hyperparameters conditioned on the observed data set of multiband SN Ia light curves. This was done by layering relationships of conditional probability. All hierarchical joint probability densities of data and parameters can be represented in terms of a probabilistic graphical model known as a directed acyclic graph. The graph consists of nodes representing parameters and data connected by arrows that represent probabilistic dependences. It obeys the restriction that there are no *directed cycles*, i.e., it is impossible to move from any node along the arrows and return to the same node. The acyclic requirement ensures that inference from the posterior density contains no loops of circular logic. It is useful to represent complex inference problems, involving many potential sources of randomness with an equivalent directed acyclic graphical model. Although all the information about the model and data is expressed by writing the joint probability density explicitly, probabilistic graphical models serve as a useful visual representation of the structure of the hierarchical model and their interface with data. Formal graphical models have not been used before in SN Ia inference, and they are not prevalent in astronomy, so we provide a basic introduction below. Further background and theory of graphical models can be found in Bishop ([2006]), Jensen ([2001]), and Pearl ([1988]).

The directed graph is constructed as follows. Each parameter or datum corresponds to a node (or vertex). Conditional relationships between nodes are encoded using directed links or arrows (edges). Hence the joint probability of two variables $P(x, y) = P(x)P(y|x)$ is represented by $x \rightarrow y$. For obvious reasons, the parameter x is termed the *parent* and y is termed the *child*. More generally, in a high-dimensional problem, if there exists a directed path of any length between node x and another node y, then y is a *descendant* of x. The joint probability distribution over N random variables θ_i represented by a directed graph can be written as the product of the conditional probabilities (the factorization):

$$
P(\{\theta_i\}) = \prod_{i=1}^{N} P(\theta_i | \{\text{Parents of } \theta_i\}).
\tag{20}
$$
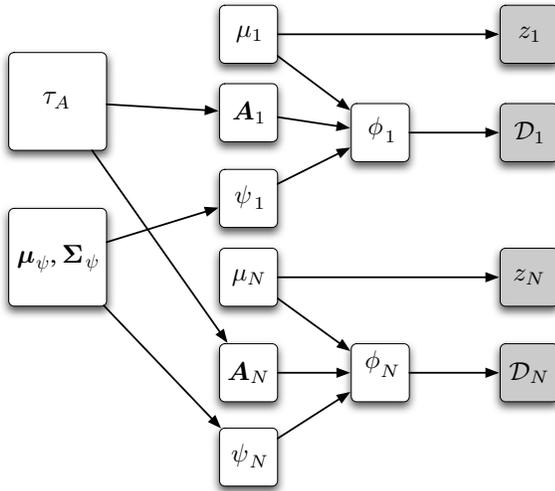
Figure 1. Directed acyclic graph for hierarchical Bayesian inference from a training set of SN Ia light curves. This is a graphical representation of the joint distribution of unknown parameters and observations for a training set of N SNe Ia. Each parameter is represented by a node, and the links between nodes indicate relationships of the conditional probability. The variables in the far left column are the hyperparameters, which describe the population probability distribution of SN characteristics, and the population distribution of extinction values. The variables in the middle left column describe the distances, extinctions, and absolute light curves of individual SNe. The variables in the middle right column are the observable parameters that describe the apparent light curves of individual SNe Ia. The final column contains the observations of the redshifts and multiband light curves of individual SN Ia. The open nodes describe unknown and hidden parameters, whereas the shaded nodes describe observed values that are conditioned upon in the posterior density.

If a parameter is observed (and thus conditioned on), its node is shaded. If the parameter is unknown and hidden, it is left open. The graph clearly distinguishes between the observed data and the hidden variables that are inferred. Graphical models are most useful in inference problems involving many potentially interacting components or sources of randomness. The complexity of the problem is reflected in the connectedness of the graph. The links between the nodes encode statements of conditional independence (Appendix A).

2.3.1. Directed Graph for Model Training

The directed graph corresponding to the global posterior density conditioned on the training set $\mathcal{D}$ of N SNe Ia is shown in Figure 1. The pathways from the roots to the data can be understood as a generative model for a data set of SN Ia light curves. At the highest hierarchical level (far left) the population distributions for SNe Ia and dust extinction are described by unknown hyperparameters $\boldsymbol{\mu}_\psi$, $\boldsymbol{\Sigma}_\psi$, and τ_A. At the next level, each SN s draws intrinsic light curves $\boldsymbol{\psi}_s$ and extinction values $\boldsymbol{A}_s$ from these populations as independently and identically distributed random samples. These hidden parameters combine with the SN distance modulus μ_s to produce the observable light curve $\boldsymbol{\phi}_s$, which is sampled with measurement noise to produce the photometric light-curve data $\mathcal{D}_s$. In addition, the (hidden) distance modulus is associated with a redshift z_s, observed with both measurement noise and peculiar velocity uncertainty. For a sample of SN Ia light curves, this process is replicated $N_{\rm SN}$ times. Our goal is to compute the joint posterior density of all the open nodes conditioned on the data in the shaded nodes.

The conditional independence properties of the graph and model imply that, although the individual parameters of one SN are *conditionally* independent from those of a different SN,

given the population hyperparameters, they are not *marginally* independent. The hidden hyperparameters are unknown *a priori*; they must be learned from the data jointly with the individual SN parameters. Thus, the full graph does not factor into independent $N_{\rm SN}$ subgraphs. We must condition the whole graph and the global joint density on a database of many SN Ia light curves simultaneously rather than on each SN individually.

Figure 1 shows that the SN Ia population hyperparameters $\boldsymbol{\mu}_\psi$, $\boldsymbol{\Sigma}_\psi$ are conditionally independent of every other parameters and the datum in the graph, given the intrinsic SN parameters $\{\boldsymbol{\psi}_s\}$: $P(\boldsymbol{\mu}_\psi, \boldsymbol{\Sigma}_\psi | \cdot, \mathcal{D}, \mathcal{Z}, \{\boldsymbol{\psi}_s\}) = P(\boldsymbol{\mu}_\psi, \boldsymbol{\Sigma}_\psi | \{\boldsymbol{\psi}_s\})$. Here we use $(\cdot)$ to indicate all the other parameters in the global joint density that have not been denoted explicitly. Similarly, given all the extinction values of the SNe, $\{\boldsymbol{A}_s\}$, the extinction population exponential scale is conditionally independent of all other parameters (and data), so that $P(\tau_A | \cdot, \mathcal{D}, \mathcal{Z}, \{\boldsymbol{A}_s\}) = P(\tau_A | \{\boldsymbol{A}_s\})$.

The graph also shows that $\boldsymbol{\psi}_s$, μ_s and $\boldsymbol{A}_s$ are conditionally dependent in the posterior distribution, because their descendant $\mathcal{D}_s$ is observed, even though they are *a priori* independent random variables. This dependence reflects the tradeoffs involved in explaining the observed light curves as a combination of random fluctuations due to dust, intrinsic randomness of the absolute light curves, and distance uncertainties attributed to peculiar velocities. The Bayesian approach is not to pick out just one possible combination of the separate factors, but to consider the probability distribution over the whole ensemble of hypotheses.

Another consequence of this conditional dependence is that there are unblocked paths between the SN Ia population hyperparameters, $\boldsymbol{\mu}_\psi$ and $\boldsymbol{\Sigma}_\psi$, and the dust extinction hyperparameter τ_A. These paths pass through the conditionally dependent parameters $\boldsymbol{A}_s$, $\boldsymbol{\psi}_s$, and $\boldsymbol{\phi}_s$ for each SN. Thus, the population hyperparameters are also conditionally dependent. This implies that posterior inferences of $\boldsymbol{\mu}_\psi$, $\boldsymbol{\Sigma}_\psi$ and those of τ_A cannot be separated. That is why we take the global approach, conditioning the global posterior density on the entire data set simultaneously and exploring the complete joint parameter space.

The conditional independence structure implied by the graph depends neither on the choices of distributions made in Section 2.2, nor on the particular functional light-curve model that is assumed. We depicted the directed graph for inference with fixed R_V. If we wish to learn about R_V, it would become a random variable with a population distribution. Hence the graph would include nodes for each R_V^s and a node for the hyperparameters α_R, with the appropriate links.

2.3.2. Directed Graph for Prediction

The directed graph for the prediction task using data from a new SN is presented in Figure 2. We depict the entire training set of SNe on a *plate* which is understood to represent $N_{\rm SN}$ different instances. The quantities relevant to the prediction SN are labeled with tildes. The essential difference between training and prediction is that in the training set we use distance information from the redshift, whereas in prediction we do not. The task of prediction is to infer the joint probability density of the hidden quantities $\tilde{\mu}$, $\tilde{A}$, and $\tilde{\boldsymbol{\psi}}$ by fitting the light-curve data $\tilde{\mathcal{D}}$ described by the observable parameters $\tilde{\boldsymbol{\phi}}$ plus measurement noise. The unblocked paths between the training set and the prediction set depict how information from the training set constrains the population hyperparameters (i.e., by informing the posterior density), which in turn pass that information (and its uncertainty) onto the prediction variables. The marginal
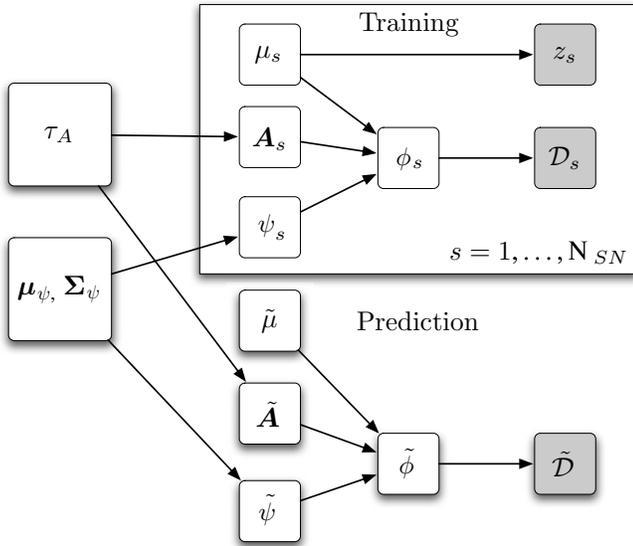
Figure 2. Directed acyclic graph for training and prediction with Type Ia SN light curves. The rectangle depicts a plate representing the N_{SN} SNe Ia in the training set. The tilde parameters describe a new SN for which we seek to predict the distance modulus. The open nodes describe unknown and hidden parameters, whereas the shaded nodes describe observed values that are conditioned upon in the predictive posterior density.

predictive posterior density for the new SN's distance modulus is obtained by integrating over the uncertainties in the population hyperparameters $\boldsymbol{\mu}_\psi$, $\boldsymbol{\Sigma}_\psi$, and τ_A, and over the extinction $\tilde{A}$, magnitudes, and the shape parameters, $\tilde{\boldsymbol{\phi}}$.

2.4. Statistical Computation of the Global Posterior Density

The global posterior probability density of all parameters and hyperparameters of the full model conditioned on the training set database of SN Ia observations, Equation (18), is a function of many variables. Consider a minimal model that does not account for dust extinction. We suppose that it has one shape parameter θ, and models light curves in three filters. There are four observable parameters, plus one for the distance modulus, for each SN. In addition, the hyperparameters $\boldsymbol{\mu}_\psi$ and $\boldsymbol{\Sigma}_\psi$ contain four plus ten variables (since the covariance matrix of the absolute magnitudes and light-curve shape parameters must be symmetric). Suppose a minimal training set contained observations of 40 SN Ia light curves in the three filters. The total number of variables, which is the dimensionality of the space over which the global posterior density is defined, is 214. Clearly, mapping the joint posterior on a rectangular multidimensional grid is intractable. Even a relatively crude grid, with only five points per dimension, would require more than 10^{149} evaluations of the posterior density.

To address the complexities of hierarchical inference with realistic data sets, it would be practically useful to construct a statistical inference approach without appealing to the asymptotic results from large-sample theory. The only way to account for all uncertainties in the model parameters consistently is to compute the full hierarchical joint density, Equation (18), and estimate all individual parameters and hyperparameters simultaneously, conditioned on the entire SN Ia database. Marginal estimates of parameters are obtained by integration over non-Gaussian uncertainties. However, the obstacles to this approach are two-fold: (1) we must compute the global posterior density in a parameter space with hundreds of dimensions, and (2) the marginal estimates of single parameters require integration of the posterior density over hundreds of other parameters.

We tackle both these problems by using stochastic simulation techniques to sample the full parameter space efficiently. In this section we describe the construction and operation of an MCMC algorithm that takes advantage of the conditional independence structure evident in the directed acyclic graph (Figure 1) of Section 2.3 to sample the global posterior density, Equation (18).

2.4.1. Metropolis Hastings and the Gibbs Sampler

MCMC is a general and well established technique for statistical analysis and is well suited for Bayesian computations. It is employed, for example, in CMB and joint cosmology analyses (Lewis & Bridle 2002; Tegmark et al. 2004), for fitting light-curve models (Mandel & Agol 2002) to planetary transit observations (Holman et al. 2006), and for radial velocity analysis of extrasolar planetary systems (Ford 2005). Since MCMC has not been used previously in SN Ia light-curve inference methods, we briefly review some basic elements of MCMC to establish terminology. More thorough treatments of MCMC methods and theory can be found elsewhere (Gilks et al. 1995; Liu 2002; Gelman et al. 2003).

The purpose of an MCMC algorithm is to generate a Markov chain stochastic process that is irreducible and ergodic, and converges in probability to a stationary distribution that is the same as the target distribution (the posterior density). Upon convergence, the probability that the chain is in a particular state is equal to the posterior density of the state, and the proportion of time the chain spends in a given region of parameter space is proportional to the posterior probability of that region. Hence, MCMC can be used to generate many samples from an arbitrary, complex probability distribution (which cannot be sampled from directly), and those samples can be used to represent the target distribution and compute characteristics of the distribution, such as means, modes, intervals, and integrals.

The cornerstone of many MCMC implementations is the Metropolis–Hastings algorithm. Suppose our target posterior probability density is $P(\boldsymbol{\theta}|\mathcal{D})$ for a vector of generic parameters $\boldsymbol{\theta}$, and can be computed up to a normalization constant. The MCMC algorithm generates a sequence of samples. Let $\boldsymbol{\theta}^t$ denote the tth sample. We start with some initial estimate $\boldsymbol{\theta}^{t=1}$, and generate subsequent values of the chain as follows. We select a proposal (or jumping) probability density $Q(\boldsymbol{\theta}^*|\boldsymbol{\theta})$, giving the probability of proposing $\boldsymbol{\theta}^*$ for the next value given that the current state is $\boldsymbol{\theta}$. This proposal density is chosen so that it can be directly sampled (e.g., a Gaussian). If the current state is $\boldsymbol{\theta}^t$, then we generate a proposal $\boldsymbol{\theta}^*$ from $Q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^t)$. We then compute the Metropolis–Hastings ratio:

$$r = \frac{P(\boldsymbol{\theta}^*|\mathcal{D})/Q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^t)}{P(\boldsymbol{\theta}|\mathcal{D})/Q(\boldsymbol{\theta}^t|\boldsymbol{\theta}^*)}. \tag{21}$$

The proposal $\boldsymbol{\theta}^*$ is accepted ($\boldsymbol{\theta}^{t+1} = \boldsymbol{\theta}^*$) with probability $\min(r, 1)$. If it is not accepted, the proposal is rejected and the next value of the chain is the same as the current value $\boldsymbol{\theta}^{t+1} = \boldsymbol{\theta}^t$. In the next iteration a new proposal is generated from $Q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{t+1})$ and the algorithm repeats.

A special case of Metropolis–Hastings is the random-walk Metropolis algorithm in which the proposal distribution is symmetric $Q(\boldsymbol{\theta}^*|\boldsymbol{\theta}) = Q(\boldsymbol{\theta}|\boldsymbol{\theta}^*)$ and the proposal is centered around the current position, e.g., $\boldsymbol{\theta}^* \sim N(\boldsymbol{\theta}, \sigma^2 \boldsymbol{I})$. The Gibbs sampling is another very useful case of the Metropolis–Hastings

rule and proceeds by simply drawing from the conditional probability of each block of parameters in turn, conditioning on the others as fixed, until all the parameters have been sampled. We can employ Gibbs sampling in our hierarchical SN Ia framework because the model is built up from conditional relations, and many of the conditional posterior distributions can be directly sampled. Our BAYESN algorithm uses a combination of these strategies to generate efficient MCMC chains.

2.4.2. The BAYESN Algorithm—Training

We describe the BAYESN MCMC algorithm in the context of computing the global posterior in Equation (18) for fixed R_V. Let $\mathcal{S} = (\{\boldsymbol{\phi}_s, \mu_s, A_H^s\}, \boldsymbol{\mu}_\psi, \boldsymbol{\Sigma}_\psi, \tau_A)$ be a vector containing all the current values of the parameters and hyperparameters in the model (the "state" or position). BAYESN utilizes a sequential Gibbs sampling structure that updates blocks of parameters in $\mathcal{S}$. After a full scan (after all parameters have been given a chance to update), the current state of $\mathcal{S}$ is recorded as an MCMC sample. To begin a chain we populate $\mathcal{S}$ with a set of initial positions. The BAYESN MCMC algorithm works in two stages: (a) sampling the population hyperparameters conditional on the individual parameters, and (b) sampling the individual SN parameters conditional on the population hyperparameters. An outline of the Gibbs scan follows; more details are presented in Appendices B and C.

1. Update the SN Ia population hyperparameters, $\boldsymbol{\mu}_\psi, \boldsymbol{\Sigma}_\psi$, conditional on the current values $\{\boldsymbol{\psi}_s\}$, obtained from the current parameters: $\boldsymbol{\psi}_s = \boldsymbol{\phi}_s - \boldsymbol{v}\mu_s - \boldsymbol{A}_s$. This is done by Gibbs sampling directly from the conditional posterior density $P(\boldsymbol{\mu}_\psi, \boldsymbol{\Sigma}_\psi | \cdot, \mathcal{D}, \mathcal{Z}) = P(\boldsymbol{\mu}_\psi, \boldsymbol{\Sigma}_\psi | \{\boldsymbol{\psi}_s\})$.

2. Gibbs sample the extinction population hyperparameter τ_A from the conditional density $P(\tau_A | \cdot, \mathcal{D}, \mathcal{Z}) = P(\tau_A | \{A_H^s\})$.

   Next we update the individual SN parameters, conditional on the population hyperparameters we have just sampled. The individual parameters of one SN are conditionally independent of those of another SN, given the hyperparameters. We cycle s through the list of SNe Ia, and for each SN s we repeat steps 3a to 3c to update observable parameters for each passband F. Let $\boldsymbol{\phi}_s^{-F_0}$, $\boldsymbol{\phi}_s^{-L,F}$, and $\boldsymbol{\phi}_s^{-NL,F}$ denote all the observable parameters in $\boldsymbol{\phi}_s$ other than the apparent magnitude, linear shape parameters, and nonlinear shape parameters in F, respectively.

3a. Gibbs sample the apparent magnitude $F_{0,s}$ by drawing directly from the conditional posterior $P(F_{0,s} | \boldsymbol{\phi}_s^{-F_0}, \mu_s, A_s; \boldsymbol{\mu}_\psi, \boldsymbol{\Sigma}_\psi, \tau_A, \mathcal{D}_s, z_s)$.

3b. Gibbs sample the linear shape parameters $\boldsymbol{\theta}_{L,s}^F$ in filter F by drawing directly from the conditional density, $P(\boldsymbol{\theta}_{L,s}^F | \boldsymbol{\phi}_s^{-L,F}, \mu_s, A_s; \boldsymbol{\mu}_\psi, \boldsymbol{\Sigma}_\psi, \tau_A, \mathcal{D}_s, z_s)$.

3c. Random-walk Metropolis update the nonlinear shape parameters in band F, $\boldsymbol{\theta}_{NL,s}^F$, using a jumping kernel $\boldsymbol{\Sigma}_{jump,s}^{NL,F}$ to move through the conditional density $P(\boldsymbol{\theta}_{NL,s}^F | \boldsymbol{\phi}_s^{-NL,F}, \mu_s, A_s; \boldsymbol{\mu}_\psi, \boldsymbol{\Sigma}_\psi, \mathcal{D}_s, z_s)$.

3d. Update the distance modulus μ_s using Metropolis–Hastings. We propose a new μ_s drawn from a Gaussian approximation to the conditional posterior density $P(\mu_s | \boldsymbol{\phi}_s, A_s; \boldsymbol{\mu}_\psi, \boldsymbol{\Sigma}_\psi, \tau_A, \mathcal{D}_s, \mathcal{Z}_s)$, and use Metropolis–Hastings rejection to correct for the approximation.

3e. Update the extinction A_H^s using a random-walk Metropolis step along the conditional density $P(A_H^s | \boldsymbol{\phi}_s, \mu_s; \boldsymbol{\mu}_\psi, \boldsymbol{\Sigma}_\psi, \tau_A)$, with a jumping scale $\sigma_{jump,s}^2$.

4. Steps 3a to 3e are repeated for all SNe Ia in the data set. After all parameters have been updated, we record the current state of $\mathcal{S}$ as an MCMC sample, and return to step 1. After we have iterated n times we finish with a Markov chain $\mathcal{S} = (\mathcal{S}_1, \ldots \mathcal{S}_t, \ldots \mathcal{S}_n)$.

2.4.3. BAYESN Algorithm—Prediction

The prediction mode of BAYESN follows essentially the same algorithm. We assume that the prediction set is sampled from the same population as the training set. This could be false, for example, if the SNe Ia in the prediction set had extremely different observed light curves. This would also be false if either observational selection effects or progenitor evolution caused a distant prediction set to sample a different portion of the SN Ia population, or a physically different population, that is not represented in the nearby training set. Training and prediction actually can be conducted simultaneously in a single run of the Gibbs sampler. The main distinction is we do not condition on the redshifts of the SNe in the prediction set, i.e., the factor $P(\mu_s | z_s)$ would be replaced by $P(\mu_s) \propto 1$ in step 3d above. With this change, the BAYESN algorithm will generate inferences on the graphical model in Figure 2, for both the training and prediction set SNe simultaneously.

In many cases, however, we may wish to train the model on the training set SNe once, and store the posterior inferences of the population hyperparameters. To make predictions for new SNe, we would recall this information and repeatedly apply it to the new data, without updating the training posterior inferences. We can do this by making two changes to the above algorithm. The goal is to generate a Markov chain $\mathcal{S}_P$ that samples the predictive posterior density. We assume that we have already done a training MCMC and have a chain $\mathcal{S}$ that samples the training posterior density conditioned on the training set $\mathcal{D}, \mathcal{Z}$. Steps 1, 2 and 3d change to:

1P & 2P. Draw the population hyperparameters $\boldsymbol{\mu}_\psi$, $\boldsymbol{\Sigma}_\psi$, and τ_A from the marginal posterior training density $P(\boldsymbol{\mu}_\psi, \boldsymbol{\Sigma}_\psi, \tau_A | \mathcal{D}, \mathcal{Z})$. This is easily done by picking a random sample from the training chain $\mathcal{S}$ and using the values of the hyperparameters in that sample.

3dP. Gibbs sample the predictive μ_s from $P(\mu_s | \boldsymbol{\phi}_s, A_s; \boldsymbol{\mu}_\psi, \boldsymbol{\Sigma}_\psi, \tau_A, \mathcal{D}, \mathcal{Z})$, omitting the factor $P(\mu_s | z_s)$ since we do not condition on the redshift for prediction SN s.

With these steps the algorithm is run to build up a Markov chain $\mathcal{S}_P$ of samples from the predictive posterior density.

3. CONSTRUCTING TEMPLATE MODELS FOR NEAR-INFRARED SN Ia LIGHT CURVES

To complete the statistical model we must specify functional models for the normalized light curves, $l^F(t; \boldsymbol{\theta}^F)$, such that $F_0 + l^F(t; \boldsymbol{\theta}^F)$ describes the apparent light curve. The function $l^F(t; \boldsymbol{\theta}^F)$ captures variations in the light-curve shapes in the observed filters. There is significant freedom in defining these functions and they will generally depend on the application. For the remainder of this paper, we apply the hierarchical model described above to the SN Ia data in the JHK_s NIR bands. In this section, we describe our methods for generating empirical template light-curve models in the JHK_s bands.

In the H and K_s bands, we model the light curves using maximum likelihood templates that assume the normalized light curves of all the SNe are identical, at least between -10 and 20 days from maximum light in B. In the J-band, where the PAIRITEL photometry is better, we construct a light-curve

model that captures the shape variations, specifically the timing and amplitudes of the initial decline, trough, second rise, second peak, and the subsequent decline. This Flexible Light-curve Infrared Template (FLIRT) model is constructed over the range -10 to 60 days past maximum light in B and depends on four light curve shape parameters to describe the features of the J-band light curve.

3.1. Maximum Likelihood Light-curve Templates

The simplest possible light-curve model assumes that the normalized light curves in filter F of all SNe are identical. The normalized light-curve data from all SNe are sampled with noise from an underlying light-curve function of the rest-frame phase $l^F(t)$. A strategy for describing this function is to construct a template that is defined by a set of knots $\{p^F, \tau\}$ and an interpolation method. The knots are defined such that $l^F(t = \tau_i) = p_i^F$, and $l^F(t) = s(t; \tau) \cdot p^F$, where the vector $s(t; \tau)$ is defined by an interpolation method that is linear in the knot values p^F. We choose a natural cubic spline rule, which ensures smoothness in the template up to two continuous derivatives (Press et al. 2007). It is convenient to choose the τ_i to lie on a regular grid with spacing $\Delta\tau$ such that one of the $\tau_j = 0$ coincides with the reference time (T_{Bmax}), forcing the corresponding $p_j^F = 0$, so that $l^F(t = \tau_j = 0) = p_j^F = 0$ for the normalized light curve.

The N_s^F photometric data points from SN s are sampled with Gaussian noise from the model template plus a constant apparent magnitude offset F_0^s. The joint posterior of the apparent magnitude offsets and the template for the data sets $\mathcal{D}_s^F$ consisting of measurements from N_{SN} SNe in one band F is proportional to the likelihood:

$$P(\{F_0^s\}, p^F \mid \{\mathcal{D}_s^F\}) \propto \prod_{s=1}^{N_{SN}} N(m_s^F \mid 1F_0^s + S p^F, W_s^F), \quad (22)$$

where S is a matrix with the ith row equal to $s(t_i^s; \tau)$. The joint maximum likelihood estimates (MLEs) of the apparent magnitudes $\{\hat{F}_0^s\}$ and the template $\hat{p}^F$ is obtained by maximizing the log joint likelihood function subject to the linear constraint $f(0) = p_j = 0$. This can be accomplished using the method of Lagrange multipliers. The quadratic optimization problem can be solved easily using non-iterative linear methods and yields a unique solution. The maximum likelihood template light-curve model produced this way is defined as $l^F(t) = s(t; \tau) \cdot \hat{p}^F$ for band F. The H and K_s band templates are depicted in Figure 3 and the values $\hat{p}^F$ are listed in Table 1.

3.2. Flexible Light-curve Infrared Template

3.2.1. Definition

Although the J-band light curves are very similar in shape near the peak, past 10 days after B_{max}, there are variations in the time and depth of the trough and the time and height of the second peak. This behavior may be explained by the changes in the ionization states of iron in the SN atmosphere (Kasen 2006). We describe an empirical model for J-band light curves that transform with simple scalings to mimic the individual decline timescales, trough depths, second rise timescales, and second-peak amplitudes of each SN light curve. These scalings are suggested by an inspection of the light curves.

We posit a J-band fiducial normalized light curve $f(\lambda)$ that can be transformed to accommodate the variations in
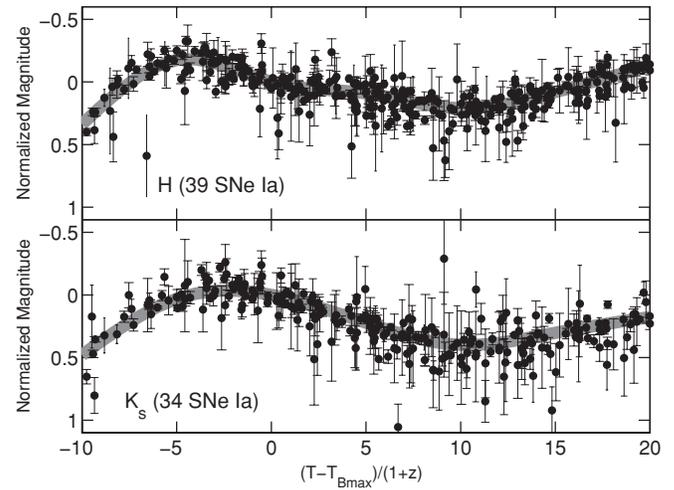
Figure 3. Maximum likelihood templates (gray curves) with the H and K_s band data. The H and K_s normalized light curves of different SNe are very similar between -10 and 20 days.

Table 1
JHK_s FLIRT and Max. Likelihood Templates

$T - T_0$	$J - J_0$	$H - H_0$	$K_s - K_{s0}$
-10	0.74	0.34	0.48
-5	-0.11	-0.18	0.03
0	0.00	0.00	0.00
5	0.44	0.09	0.22
10	1.36	0.21	0.40
15	1.63	0.08	0.30
20	1.44	-0.13	0.18
25	1.15	...	...
30	0.91	...	...
35	1.33	...	...
40	1.82	...	...
45	2.23	...	...
50	2.60	...	...
55	2.91	...	...
60	3.29	...	...

Notes.
All templates are interpolated using natural cubic splines.

characteristic timescales and amplitudes observed in individual J-band light curves. Here λ is a feature timescale describing the timings of features in the fiducial light curve. We can map this timescale to the chronological timescale t of a particular SN by introducing a time-warping function that allows the pre-trough phase to be scaled independently from the post-trough phase. The rate at which the feature time maps to the chronological time is

$$\frac{dt}{d\lambda} = \begin{cases} \alpha, & \text{if } \lambda \leqslant \lambda_t \\ \beta, & \text{if } \lambda > \lambda_t, \end{cases} \quad (23)$$

where the parameters α, β are positive constants and of order one and λ_t is the feature time of the trough in the fiducial light curve. The solution is $t(\lambda) = \alpha \min(\lambda, \lambda_t) + \beta(\lambda - \lambda_t)^+$ where $u^+ = \max(u, 0)$. This function can be inverted as

$$\lambda(t) = \begin{cases} \alpha^{-1}t, & \text{if } t \leqslant \alpha\lambda_t \\ \beta^{-1}t + \lambda_t(1 - \alpha/\beta), & \text{if } t > \alpha\lambda_t. \end{cases} \quad (24)$$

These equations represent simple transformations between the chronological and feature time axes, or the "horizontal" dimensions.

Even after adjusting for variations in the two timescales, there are still variations in the depth of the trough and the amplitude of the second peak. This suggests that an individual normalized light curve is related to the standard light curve $f(\lambda)$ as

$$
\begin{aligned}
& l^J(t; \alpha, \beta, d, r) \\
& = \begin{cases} d[f(\lambda(t)) - f_0], & \lambda(t) \leqslant \lambda_t \\ d[f(\lambda_t) - f_0] + r[f(\lambda(t)) - f(\lambda_t)], & \lambda(t) > \lambda_t, \end{cases}
\end{aligned}
\tag{25}
$$

where $f_0 = f(0) \equiv 0$, and the parameters d and r are positive constants of order 1. The decline parameter d controls the depth of the trough by scaling the decline from maximum light, "vertically" in the magnitude dimension. A larger d will produce a deeper trough and a faster decline rate (in magnitudes per day). At the trough the magnitude is $J(T_{tr}) = J_0 + d[f(\lambda_t) - f_0]$ and the rise parameter r controls the rise in flux toward the second maximum relative to the trough magnitude. A larger r will produce a higher second peak and a faster rise rate. This parameterization is constructed to preserve continuity in the light curve even as different phases are scaled in amplitude. This quantitative parameterization of two constants (α, β) to control timescales and two constants (d, r) to control amplitudes in two different regimes of the light curve is sufficient to describe the variation in J-band light curves. This is a simple transformation from the fiducial light curve to the realized light curves of individual SNe.

After the fiducial light curve $f(\lambda)$ is solved as a continuous function of feature time λ, one can easily measure any key features, such as the feature time of trough minimum (λ_t), the feature time of the second peak (λ_{p2}), and the normalized magnitudes at these points, $f(\lambda_t)$ and $f(\lambda_{p2})$. For any particular SN's J-band light curve, we can measure these features using the solved parameters. The chronological time of the trough is $T_{tr} = t(\lambda_t) = \alpha\lambda_t$, the trough-to-second-peak time is $T_2 - T_{tr} = \beta(\lambda_{p2} - \lambda_t)$, the depth of the trough is $J(T_{tr}) - J_0 = d[f(\lambda_t) - f_0]$ and the height of the second peak above the trough is $J(T_2) - J(T_{tr}) = r[f(\lambda_{p2}) - f(\lambda_t)]$. The parameters of the model can be directly related to the observable features of light curves.

3.2.2. Maximum Likelihood Construction of the FLIRT model

The FLIRT model described above is completely specified by the fiducial normalized light-curve function $f(\lambda)$. We represent this function in the same way as the Maximum Likelihood light-curve Template (Section 3.1). A set of knots (p, τ) is defined on a regular grid, such that $f(\lambda) = s(\lambda; \tau) \cdot p$, where the vector $s(\lambda; \tau)$ is fully specified by natural cubic spline interpolation. Once p is known, an individual SN light curve can be fitted with the FLIRT model by means of nonlinear maximization of the likelihood to get point estimates of the light-curve shape parameters $\theta_s^J = (d^s, r^s, \alpha^s, \beta^s)$ and apparent magnitude F_0^s. All that is now required is an estimate of the fiducial template p. The joint posterior over the SN parameters and the fiducial template is proportional to the likelihood function

$$
P(\{J_0^s, \theta_s^J\}, p \mid \{\mathcal{D}_s^J\}) \propto \prod_{s=1}^{N_{SN}} N(m_s^J \mid J_0^s + \tilde{S}(\theta_s^J) p, W_s^F),
\tag{26}
$$

where the matrix $\tilde{S}(\theta_s^J)$ is derived from the defining Equations (24) and (25) and the interpolation method $s(\lambda; \tau)$. If $\{J_0^s, \theta_s^J\}$ are estimated and fixed, then the conditional maximization of $\log P(p \mid \{J_0^s, \theta_s^J\}, \{\mathcal{D}_s^J\})$ with respect to the template p
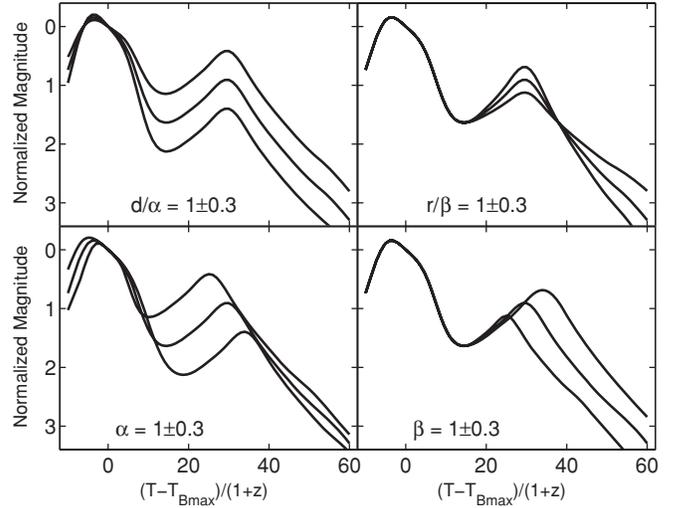
Figure 4. FLIRT model for J-band light-curve shape variations. In each panel, the fiducial FLIR template with parameters $(d/\alpha, r/\beta, \alpha, \beta) = (1, 1, 1, 1)$ is shown as the middle curve along with models with one parameter varied while keeping the others fixed. For example, the first panel depicts $(0.7, 1, 1, 1)$ and $(1.3, 1, 1, 1)$. The parameters correspond to the initial decline rate, the second rise rate, the time from the peak to the trough and the time from the trough to the second peak.

subject to the constraint that $f(\lambda = 0) = p_j = 0$ is a linear problem.

It is straightforward to solve for the FLIRT model template iteratively. We select a subset of SN light curves that are well sampled. First we pool the photometric data together and estimate the Maximum Likelihood light-curve Template $\hat{p}_0$ (Section 3.1) as a first approximation. Then we fit the FLIRT model using this template to each SN light curve by conditional maximization of $P(J_0^s, \theta_s^J, \mid \hat{p}_0, \{\mathcal{D}^J\})$ to get estimates $\hat{J}_0^s, \hat{\theta}_s^J$. Next we fix the SN parameters and update the fiducial template by constrained conditional maximization of $P(p \mid \{\hat{J}_0^s, \hat{\theta}_s^J\}, \{\mathcal{D}_s^J\})$ to get a new template $\hat{p}_1$. We iterate until the maximum likelihood template $\hat{p}$ converges. The template is rescaled so that the sample median values of the fitted SN shape parameters (d, r, α, β) are equal to 1 (so that the fiducial template model, which has shape parameters equal to 1, reflects a typical light curve).

In practice, it is convenient to measure the relative decline rates d/α and the relative rise rates r/β rather than the depths and heights directly. The light-curve shape parameters are then $\theta^J = (d/\alpha, r/\beta, \alpha, \beta)$. The maximum likelihood J-band FLIRT fiducial template $\hat{p}_J$ is listed in Table 1, and depicted in Figure 4, which also shows the effects of varying each of the parameters. The trough of the fiducial template is located at $(\lambda_t, f(\lambda_t)) = (14.43 \text{ days}, 1.64 \text{ mag})$ and the second peak is $(\lambda_{p2}, f(\lambda_{p2})) = (29.55 \text{ days}, 0.90 \text{ mag})$.

In Figure 5 we display the J-band FLIRT fiducial light curve, along with the J-band photometry for the 39 SNe listed in Table 2, shown in gray. We have also transformed each SN light-curve data set using the fitted light-curve parameters in Table 2 to the same scales as the fiducial light curve by inverting Equation (25). The dramatic reduction of dispersion from 5 to 60 days shows that the FLIRT model successfully captures the shape variations in the J-band SN Ia light curves. The double-peaked light-curve structure is also seen in the H, K, and I bands. In the future, it may be worth exploring FLIRT models in these bands.

Table 2
Posterior Summary of SN Ia JHK_s Light-Curve Parameters

SN	$T_{B\max}$[a] (MJD)	J_0 (mag)	H_0 (mag)	K_{s0} (mag)	d/α	r/β	α	β	References[b]
SN1998bu	50952.40	11.75 (0.01)	11.87 (0.01)	11.58 (0.01)	0.88 (0.03)	0.94 (0.08)	1.12 (0.02)	1.02 (0.13)	J99,H00
SN1999cl	51342.20	12.94 (0.01)	13.01 (0.01)	12.66 (0.01)	1.10 (0.03)	1.12 (0.08)	0.76 (0.02)	1.05 (0.13)	K00
SN1999cp	51363.20	14.66 (0.02)	14.93 (0.03)	14.64 (0.07)	0.85 (0.07)	0.86 (0.08)	1.15 (0.06)	1.06 (0.26)	K00
SN1999ee	51469.30	14.96 (0.01)	15.19 (0.01)	14.94 (0.13)	0.80 (0.01)	0.85 (0.02)	1.20 (0.01)	1.08 (0.01)	K04b
SN1999ek	51481.80	16.30 (0.01)	16.42 (0.01)	16.27 (0.09)	0.97 (0.02)	1.09 (0.07)	1.00 (0.02)	0.97 (0.26)	K04c
SN1999gp	51550.10	16.95 (0.17)	17.17 (0.12)	16.75 (0.12)	1.00 (0.18)	0.84 (0.11)	0.98 (0.17)	1.10 (0.27)	K01
SN2000E	51577.20	13.58 (0.01)	13.90 (0.02)	13.49 (0.03)	0.93 (0.02)	0.88 (0.04)	1.09 (0.01)	1.16 (0.03)	V03
SN2000bh	51636.00	16.53 (0.03)	16.88 (0.01)	16.69 (0.02)	1.03 (0.03)	0.97 (0.02)	1.06 (0.01)	1.09 (0.01)	K04b
SN2000bk	51647.00	17.15 (0.04)	17.43 (0.01)	17.52 (0.22)	0.84 (0.04)	1.23 (0.05)	1.04 (0.02)	0.58 (0.01)	K01
SN2000ca	51666.20	16.53 (0.01)	16.78 (0.02)	16.62 (0.20)	0.83 (0.02)	0.83 (0.09)	1.14 (0.03)	0.97 (0.25)	K04b
SN2000ce	51667.30	15.98 (0.12)	16.28 (0.02)	15.87 (0.03)	0.94 (0.11)	0.89 (0.07)	1.06 (0.04)	1.05 (0.25)	K01
SN2001ba	52034.20	17.19 (0.01)	17.51 (0.02)	17.27 (0.02)	1.03 (0.02)	1.00 (0.04)	1.06 (0.02)	1.10 (0.03)	K04b
SN2001bt	52062.90	15.55 (0.01)	15.82 (0.02)	15.51 (0.02)	1.03 (0.02)	0.96 (0.03)	0.99 (0.02)	0.96 (0.02)	K04c
SN2001cn	52071.00	15.64 (0.03)	15.91 (0.02)	15.61 (0.05)	0.96 (0.03)	1.01 (0.03)	1.09 (0.02)	0.98 (0.02)	K04c
SN2001cz	52103.40	15.53 (0.03)	15.91 (0.05)	15.63 (0.06)	0.96 (0.04)	0.89 (0.06)	1.16 (0.04)	1.08 (0.04)	K04c
SN2001el	52182.10	13.03 (0.01)	13.11 (0.01)	12.86 (0.02)	0.75 (0.02)	0.93 (0.01)	1.14 (0.01)	0.88 (0.01)	K03
SN2002bo	52356.00	13.78 (0.02)	14.08 (0.01)	13.99 (0.02)	0.90 (0.03)	0.92 (0.03)	1.02 (0.02)	0.97 (0.02)	K04c
SN2002dj	52450.60	14.68 (0.02)	14.91 (0.01)	14.64 (0.01)	0.86 (0.02)	0.88 (0.04)	1.13 (0.02)	0.92 (0.02)	P08
SN2003cg	52729.10	13.71 (0.04)	13.92 (0.01)	13.45 (0.01)	1.04 (0.04)	0.93 (0.04)	0.95 (0.04)	1.12 (0.03)	ER06
SN2003du	52765.90	14.29 (0.02)	14.66 (0.02)	14.35 (0.01)	0.95 (0.02)	0.96 (0.09)	1.20 (0.02)	1.04 (0.27)	St07
SN2004S	53038.70	14.82 (0.02)	15.00 (0.01)	14.71 (0.02)	0.70 (0.03)	0.85 (0.04)	1.32 (0.03)	0.83 (0.02)	K07
SN2004eo	53278.70	15.73 (0.04)	15.97 (0.04)	15.76 (0.08)	1.09 (0.09)	1.02 (0.04)	0.87 (0.05)	0.94 (0.05)	Pa07
SN2005ao	53442.00	17.98 (0.07)	18.13 (0.01)	18.34 (0.02)	0.73 (0.17)	1.06 (0.12)	1.07 (0.14)	0.88 (0.20)	WV08
SN2005cf	53533.60	13.93 (0.01)	14.08 (0.01)	13.99 (0.01)	0.81 (0.02)	0.91 (0.05)	1.13 (0.02)	1.12 (0.17)	WV08
SN2005ch	53536.00	17.03 (0.07)	17.28 (0.03)	17.07 (0.05)	1.10 (0.08)	1.05 (0.08)	1.00 (0.04)	0.99 (0.19)	WV08
SN2005el	53646.10	15.60 (0.01)	15.82 (0.01)	15.59 (0.01)	1.13 (0.01)	1.01 (0.02)	0.87 (0.01)	0.86 (0.01)	WV08
SN2005eq	53653.90	16.95 (0.01)	17.34 (0.02)	16.89 (0.03)	0.76 (0.02)	0.72 (0.03)	1.19 (0.02)	1.12 (0.02)	WV08
SN2005iq	53687.10	17.60 (0.05)	17.79 (0.14)	17.52 (0.20)	1.06 (0.11)	1.03 (0.10)	0.92 (0.12)	1.04 (0.26)	WV08
SN2005na	53740.50	16.66 (0.08)	17.12 (0.07)	16.85 (0.13)	0.99 (0.12)	0.89 (0.06)	0.89 (0.05)	1.21 (0.06)	WV08
SN2006D	53756.70	14.49 (0.01)	14.70 (0.01)	14.69 (0.01)	1.05 (0.02)	1.10 (0.04)	0.98 (0.01)	0.79 (0.02)	WV08
SN2006N	53760.60	15.69 (0.08)	15.97 (0.07)	15.81 (0.10)	1.11 (0.09)	1.10 (0.07)	0.87 (0.04)	0.93 (0.03)	WV08
SN2006X	53785.50	13.04 (0.02)	13.08 (0.01)	12.78 (0.01)	0.91 (0.02)	0.95 (0.01)	0.97 (0.01)	1.09 (0.01)	WV08
SN2006ac	53781.20	16.64 (0.07)	16.90 (0.09)	16.67 (0.09)	0.94 (0.12)	0.90 (0.08)	0.88 (0.09)	1.31 (0.19)	WV08
SN2006ax	53826.70	15.87 (0.01)	16.37 (0.03)	16.13 (0.03)	1.08 (0.02)	1.06 (0.06)	1.12 (0.02)	1.03 (0.02)	WV08
SN2006cp	53896.70	16.68 (0.04)	16.91 (0.05)	16.74 (0.08)	1.10 (0.06)	1.07 (0.07)	0.94 (0.08)	1.00 (0.26)	WV08
SN2006gr	54014.00	17.94 (0.10)	18.00 (0.14)	17.63 (0.18)	0.91 (0.09)	1.02 (0.10)	1.03 (0.12)	1.01 (0.27)	WV08
SN2006le	54048.00	16.35 (0.03)	16.64 (0.02)	16.22 (0.05)	0.90 (0.05)	0.96 (0.05)	1.16 (0.04)	1.14 (0.04)	WV08
SN2006lf	54044.80	15.78 (0.03)	15.84 (0.04)	15.56 (0.05)	1.12 (0.07)	1.12 (0.08)	0.85 (0.05)	0.79 (0.05)	WV08
SN2007cq	54280.00	16.50 (0.03)	17.01 (0.24)	16.80 (0.28)	0.96 (0.21)	0.89 (0.15)	0.97 (0.30)	1.10 (0.29)	WV08

Notes.

[a] Time of maximum light in B: Julian Date -2400000. J_0, H_0, K_{s0} are J, H, and K_s at $T_{B\max}$.

[b] Reference codes: WV08: Wood-Vasey et al. (2008, PAIRITEL photometry); J99: Jha et al. 1999; H00: Hernandez et al. 2000; K00: Krisciunas et al. 2000; K01: Krisciunas et al. 2001; 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; Pa07: Pastorello et al. 2007; St07: Stanishev et al. 2007; P08: Pignata et al. 2008.

4. APPLICATION AND RESULTS

4.1. Nearby SN Ia NIR Light Curves

A comprehensive data set of nearby SN Ia light curves in the near-infrared was compiled by WV08, including observations of 21 recent SNe with the PAIRITEL taken by the CfA Supernova Group and observations of 23 SNe from the literature (Jha et al. 1999; Hernandez et al. 2000; Krisciunas et al. 2000; Di Paola et al. 2002; Valentini et al. 2003; Krisciunas et al. 2001, 2003, 2004a, 2004b, 2007; Elias-Rosa et al. 2006, 2008; Pastorello et al. 2007; Stanishev et al. 2007; Pignata et al. 2008). Of these, three (SN 2005bl, SN 2005hk, and SN 2005ke) are omitted because they are fast-declining, peculiar SNe with "dromedary" H-band light curves that have only one peak, whereas most H-band light curves are "bactrian," having two peaks. We use

the remaining data set with two exceptions. The very late J-band secondary maximum of SN 2002cv and its extreme reddening and estimated optical extinction ($A_V > 8$) (Elias-Rosa et al. 2008) suggest that this light curve is unusual, so we have omitted it from the analysis. We have also omitted the PAIRITEL observations of SN 2005eu, because we judged the JHK_s image subtractions to be of poor quality. The final, edited set of observations covers 39 SNe Ia. We have only used photometric measurements with signal-to-noise ratio > 3. Extensive studies of two SNe in this set, SN 2005cf and SN 2006X, were presented by Wang et al. (2008, 2009).

To construct the H and K_s-band templates (Figure 3, Table 1) we used all the light-curve observations from the data set. For the J-band light curves we selected a subset of 16 well-sampled light curves to generate the fiducial FLIRT template (Figure 4, Table 1). This subset consisted of SN1998bu, SN 1999ee, SN
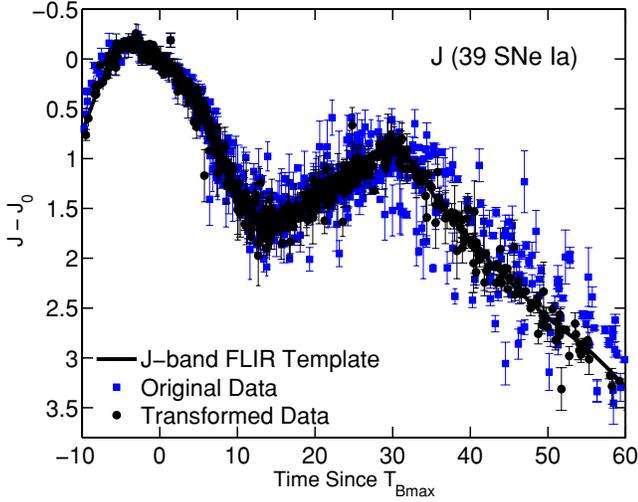
Figure 5. *J*-band light-curve data (blue squares) exhibit significant shape variation after the initial decline. After fitting for these variations using the FLIRT model, the data were transformed to the fiducial frame and overplotted (black dots) with the fiducial FLIR template light curve (black line). This demonstrates that the FLIRT model successfully captures these light-curve shape variations.

2001bt, SN 2001cn, SN 2001cz, SN 2002bo, SN 2005el, SN 2005eq, SN 2005na, SN 2006D, SN 2006N, SN 2006X, SN 2006ac, SN 2006ax, SN 2006le, and SN 2006lf.

All photometric data were *K*-corrected to the SN rest frame by linearly interpolating the tables of Krisciunas et al. (2004b), and registered to a common phase by subtracting from the Julian Day the time of *B*-band maximum, $T_{B\max}$, as determined by the MLCS2k2 fits to the optical light curves observed by the CfA Supernova Group (Hicken et al. 2009a). The phases were corrected for time dilation using the heliocentric redshifts. Recession velocities were corrected to the CMB+Virgo infall rest frame, as described in WV08. Furthermore, a peculiar velocity uncertainty $\sigma_{\mathrm{pec}} = 150$ km s^{-1} (Radburn-Smith et al. 2004) was assumed. Luminosity distances were computed from the redshifts assuming an LCDM model with $\Omega_M = 0.27$, $\Omega_\Lambda = 0.73$ and a Hubble constant scale of $h = 0.72$ (Freedman et al. 2001; Spergel et al. 2007). At the most distant end of the sample at $z \approx 0.04$, the relative difference between the luminosity distances in LCDM and in an Einstein-de Sitter universe is 2%.

4.2. JHK_s Light-curve Model Specification

The light-curve models we construct for the JHK_s, data set consist of maximum likelihood templates for H and K_s between -10 and 20 days (Section 3.1, Figure 3) and the *J*-band FLIRT model between -10 and 60 days (Section 3.2, Figure 4) with $\Delta\tau = 5$ days. The H and K_s models have no light-curve shape parameters, and the *J*-band FLIRT model has four: $\boldsymbol{\theta}^H = \emptyset$, $\boldsymbol{\theta}^{K_s} = \emptyset$, and $\boldsymbol{\theta}^J = (d/\alpha, r/\beta, \alpha, \beta)$. The multiband normalized light-curve models as defined in Equation (2) are then fully specified by

$$H(t) - H_0 = l_0^H(t) \equiv f_H(t) = \boldsymbol{S}(t; \tau) \cdot \hat{\boldsymbol{p}}^H, \qquad (27)$$

$$K_s(t) - K_{s0} = l_0^{K_s}(t) \equiv f_{K_s}(t) = \boldsymbol{S}(t; \tau) \cdot \hat{\boldsymbol{p}}^{K_s}, \qquad (28)$$

$$J(t) - J_0 = \boldsymbol{l}_1^J(t; \boldsymbol{\theta}_{\mathrm{NL}}^J) \cdot \boldsymbol{\theta}_{\mathrm{L}}^J, \qquad (29)$$

where the *J*-band linear parameters are $\boldsymbol{\theta}_{\mathrm{L}}^J = (d/\alpha, r/\beta)$, the nonlinear parameters are $\boldsymbol{\theta}_{\mathrm{NL}}^J = (\alpha, \beta)$, and the vector function $\boldsymbol{l}_1^J(t; \boldsymbol{\theta}_{\mathrm{NL}}^J)$ is determined by Equations (24) and (25).

In the notation of the hierarchical framework described in Section 2, the observable or apparent parameters are $\boldsymbol{\phi}_s = (J_0, H_0, K_{s0}, d/\alpha, r/\beta, \alpha, \beta)$ for each SN s, and the intrinsic or absolute parameters are $\boldsymbol{\psi}_s = (M_J, M_H, M_{K_s}, d/\alpha, r/\beta, \alpha, \beta)$ for each SN s. The population hyperparameters are $\boldsymbol{\mu}_\psi = \mathbb{E}[\boldsymbol{\psi}_s]$ and $\boldsymbol{\Sigma}_\psi = \mathrm{Cov}[\boldsymbol{\psi}_s, \boldsymbol{\psi}_s^T]$ with expectations with respect to the SN Ia NIR light-curve population randomness.

Since dust extinction and reddening have small effect on the NIR light curves in our sample, we omit the full modeling of the multiband extinctions $\boldsymbol{A}_s$ and dust population characteristic τ_A. The most optically reddened SNe in the sample (SN 1999cl, 2006X, and SN 2003cg) are also at low redshifts, where the adopted velocity model gives them little weight in the determinations of population means and covariances of the NIR absolute magnitudes. Hence we set all $\boldsymbol{A}_s$ to zero and use the one-population model for SN Ia NIR light-curve randomness only. In Section 4.5 we estimate the potential effect of dust on our posterior inferences. In the near future, we will use the full two-population model with NIR and optical data for a simultaneous hierarchical modeling of SN Ia light-curve shapes and dust extinction.

After plugging the specified JHK_s light-curve models and parameter dependence into the hierarchical framework of Section 2, we perform probabilistic inference using the BAYESN algorithm of Section 2.4 to compute the joint posterior density over all individual parameters for the 39 SNe and population hyperparameters. There is a total of 347 parameters and hyperparameters in the statistical model. Initial positions for the Markov chains were obtained by adding random noise to the MLEs of the SN parameters obtained in Section 3. It is not necessary to specify initial guesses for the hyperparameters. We set the scale of the inverse Wishart hyperprior, $\boldsymbol{\Lambda}_0 = \epsilon_0 \boldsymbol{I}$, by choosing a small value $\epsilon_0 = 10^{-4}$. We found that our inferences were insensitive to ten-fold changes in ϵ_0.

The BAYESN MCMC algorithm was run for five independent chains with 2×10^4 samples each. The Gelman–Rubin (G-R) statistic was computed for all parameters: the maximum value was 1.03 and 99% had values less than 1.02, with the mean and median values less than 1.005. Acceptable values of the G-R statistic are typically less than 1.10 (Gelman et al. 2003). The first 2000 samples of each of the chains were then discarded as burn-in and the chains were concatenated for posterior analysis. We found that our inferences were insensitive to the burn-in cutoff if it was greater than ∼ 1000 samples.

4.3. Posterior Inferences

The MCMC results produce samples from the global posterior density over all parameters and hyperparameters, Equation (18). We summarize the posterior density by examining marginal posterior densities over subsets of parameters. Inferences at the level of individual SNe can be summarized by the probability density $P(\boldsymbol{\phi}_s, \mu_s | \mathcal{D}, \mathcal{Z})$ for each SN s. Inferences at the SN Ia NIR population level are summarized by $P(\boldsymbol{\mu}_\psi, \boldsymbol{\Sigma}_\psi | \mathcal{D}, \mathcal{Z})$. This can be further broken down into the marginal densities over mean properties of absolute light curves $P(\boldsymbol{\mu}_\psi | \mathcal{D}, \mathcal{Z})$, the probability over covariances between multiband absolute magnitudes $P(\boldsymbol{\Sigma}[(M_J, M_H, M_{K_s}), (M_J, M_H, M_{K_s})] | \mathcal{D}, \mathcal{Z})$, marginal densities over the covariances in light-curve shape: $P(\boldsymbol{\Sigma}(\boldsymbol{\theta}, \boldsymbol{\theta}))$, and marginal posterior densities over covariances between
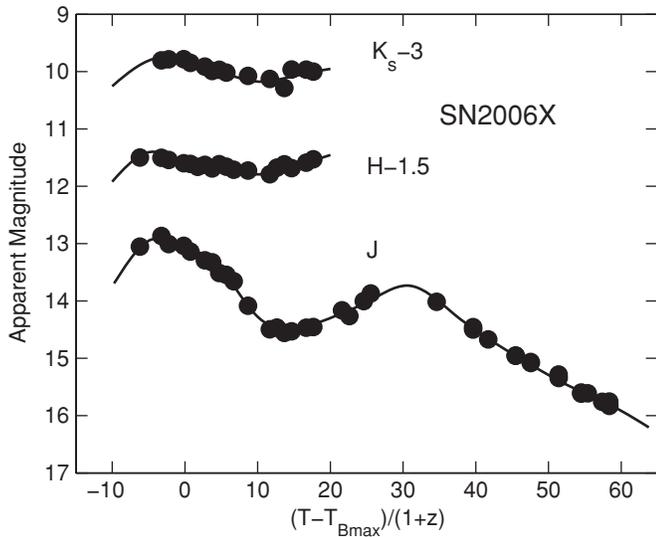
Figure 6. JHK_s light-curve data and model fits to SN 2006X. This is a very well sampled light curve, and the fit to the light-curve model (black curves) is excellent. The H and K_s bands are fit to the maximum likelihood templates, and the J- band is fit to the FLIRT model. The data of this SN tightly constrain the light-curve parameters.
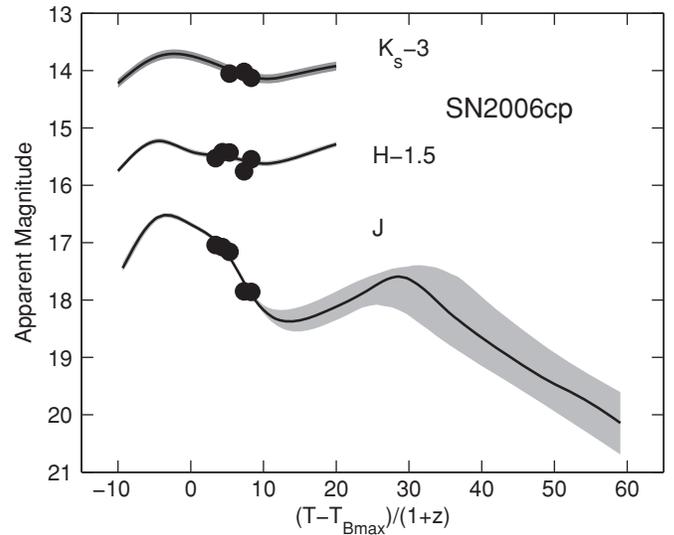

Figure 7. JHK_s light-curve data and model fits (black curves) to SN 2006cp. The light-curve data are sparse and incomplete. The BAYESN method estimates the J-band light curve where the data are missing using the information in the population distribution of the set of SNe and its uncertainty. For example, the correlation of the initial decline rate with the second rise rate provides some information. Since the population of J-band light curves exhibits significant late-time shape variations, the late-time model fit is very uncertain, as reflected by the gray error tube spanning the 16% and 84% quantiles of the posterior uncertainty in the light curve.

light-curve shape and absolute magnitudes: $P(\Sigma[(M_J, M_H, M_{K_s}), \boldsymbol{\theta}]|\mathcal{D}, \mathcal{Z})$. These posterior densities are integrals over the global posterior density and can all be computed easily and directly from the MCMC chain. We show example light-curve data and model fits in Figures 6–8.

4.3.1. SN Ia JHK_s Light Curves

The univariate marginal posterior median and standard deviations of the individual SN light-curve parameters $\boldsymbol{\phi}_s = (J_0, H_0, K_{s0}, d/\alpha, r/\beta, \alpha, \beta)$ for each of the 39 SNe are listed in Table 2. The light-curve fits are excellent, especially in the J- band, where the PAIRITEL photometry is the best. The MCMC chains quickly find the region of parameter space near the peak of the posterior probability distribution, especially if the data tightly constrain the light-curve fits.

The SN light-curve data in the training set are not homogeneously well sampled. Some SNe, e.g., SN 2006X, have extremely good sampling in JHK_s from before maximum to well past the secondary maximum. Such well-sampled, complete data sets constrain the observable light-curve parameters very well. Other SNe are sparsely sampled or have incomplete coverage over the range of the model, for example, SN 1999gp, SN 2007cq, SN 2005ao. Some SNe, for example, SN 1998bu, are well sampled in the early part of the J-band light curve but the measurements stop before the secondary maximum. Several SNe in our sample (SN 1999ee, SN 2000bk, SN 2000ca, SN 2005iq, and SN 2007cq) have no K_s-band data.

In these cases the advantages of the Bayesian approach are clear. Since we have defined a joint probability density over all data and parameters (both of which are considered random variables) in Equation (18), we have a probability distribution over the parameters that are not well constrained by the individual SN data because of missing observations. The Bayesian computation yields sensible numerical estimates of the poorly constrained parameters and their uncertainties using the joint probability over the observed parameters, the population distribution of individual SN parameters, and the uncertainty in the hyperparameters of the population distribution, all conditioned

on the actual observed data and its uncertainty. For the very well sampled, complete light curves, the posterior density over SN light-curve parameters will be dominated by the information from its own light-curve data. An example of this is SN 2006X, shown in Figure 6 along with the light-curve fits. For sparse light curves, for example, SN 2006cp (Figure 7), some of the parameters will be informed by the population distribution constrained by the whole training set of SNe. In an intermediate case (e.g., SN 2005cf, Figure 8, an incomplete light curve), a balance between the existing data, observed parameters, and the population distribution of the poorly constrained parameters is achieved, and an appropriate uncertainty is computed. These computations are already handled automatically by our sampling of the global posterior probability density.

4.3.2. NIR Absolute Magnitudes

A summary of posterior inferences of the SN Ia NIR light-curve population hyperparameters is presented in Table 3. The univariate expectation of the means $\boldsymbol{\mu}_\psi$ is shown along with the standard deviations of the univariate marginal densities. We list the modal values of the square root of the variances σ_ψ^2, which are the diagonal values of the covariance matrix $\boldsymbol{\Sigma}_\psi$, and the standard deviations of their univariate marginal posterior probability densities. We also list the modal values of the correlations $\rho(\cdot, \cdot)$ obtained from the off-diagonal terms of the covariance $\boldsymbol{\Sigma}_\psi$ after factoring out the variances. The marginal modes are estimated from the histogram of MCMC samples in each quantity. In addition, we list the tail probabilities of each correlation coefficient, defined as

$$p_{\text{tail}} = \begin{cases} P(\rho < 0), & \text{if mode}(\rho) > 0, \\ P(\rho > 0), & \text{if mode}(\rho) < 0. \end{cases} \quad (30)$$

The smaller the tail probability, the greater the evidence that the correlation is different from zero, either positively or negatively. The probability densities of correlation coefficients have support

Table 3
Summary of Posterior Inference: Population Hyperparameters

ψ^i	M_J	M_H	M_{Ks}	d/α	r/β	α	β
$\mu(\cdot)$	$-18.25\ (0.03)$	$-18.01\ (0.03)$	$-18.25\ (0.04)$	$0.95\ (0.03)$	$0.97\ (0.03)$	$1.04\ (0.03)$	$1.01\ (0.04)$
$\sigma(\cdot)$	$0.17\ (0.03)$	$0.11\ (0.03)$	$0.19\ (0.04)$	$0.15\ (0.03)$	$0.13\ (0.03)$	$0.15\ (0.02)$	$0.22\ (0.04)$
$\rho(M_J,\cdot)$	1.00	$0.73\ (0.03)$	$0.41\ (0.09)$	$-0.14\ (0.29)$	$0.52\ (0.03)$	$-0.07\ (0.46)$	$-0.28\ (0.18)$
$\rho(M_H,\cdot)$	$0.73\ (0.03)$	1.00	$0.53\ (0.04)$	$-0.03\ (0.47)$	$0.59\ (0.03)$	$0.24\ (0.20)$	$-0.21\ (0.35)$
$\rho(M_{Ks},\cdot)$	$0.41\ (0.09)$	$0.53\ (0.04)$	1.00	$-0.16\ (0.34)$	$0.76\ (0.01)$	$0.07\ (0.39)$	$-0.48\ (0.11)$
$\rho(d/\alpha,\cdot)$	$-0.14\ (0.29)$	$-0.03\ (0.47)$	$-0.16\ (0.34)$	1.00	$0.55\ (0.02)$	$-0.77\ (0.00)$	$0.07\ (0.41)$
$\rho(r/\beta,\cdot)$	$0.52\ (0.03)$	$0.59\ (0.03)$	$0.76\ (0.01)$	$0.55\ (0.02)$	1.00	$-0.50\ (0.02)$	$-0.43\ (0.07)$
$\rho(\alpha,\cdot)$	$-0.07\ (0.46)$	$0.24\ (0.20)$	$0.07\ (0.39)$	$-0.77\ (0.00)$	$-0.50\ (0.02)$	1.00	$-0.04\ (0.47)$
$\rho(\beta,\cdot)$	$-0.28\ (0.18)$	$-0.21\ (0.35)$	$-0.48\ (0.11)$	$0.07\ (0.41)$	$-0.43\ (0.07)$	$-0.04\ (0.47)$	1.00

Notes. Top: population means and variances of the absolute parameters. Values in the parentheses are the standard deviations of the marginal posterior density in each parameter. The estimates of the $\sigma(\cdot)$ are modal values. Bottom: population correlation matrix for the absolute parameters. Estimates of the correlations $\rho(\cdot,\cdot)$ are the modal values. The parentheses contain the tail probabilities as described in Equation (30).
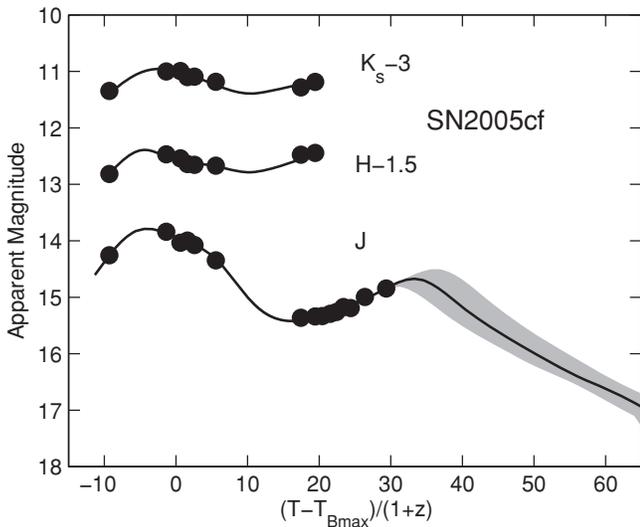
Figure 8. JHK_s light-curve data and model fits (black curves) to SN 2005cf. The light-curve data are adequately sampled in the early part of the light curve up to second rise, but ends before reaching the second peak and decline. The BayeSN method estimates the second maximum and late-time decline using a combination of the constraints imposed by the data and the population distribution of the training set. For example, the final data point provides a lower bound for the time of the second maximum. This makes the posterior distribution of the β parameter non-Gaussian.


Figure 9. Joint posterior probability densities in the population mean μ and population variance σ^2 of the peak absolute magnitudes in each NIR band. The crosses and the numbers in each panel denote the mode of the bivariate probability density. The contours contain the 68% and 95% highest posterior density regions. These estimates were obtained directly from the BayeSN MCMC chain of the trained statistical model. The univariate marginal estimates of the population variances are $\sigma(M_J) = 0.17 \pm 0.03$, $\sigma(M_H) = 0.11 \pm 0.03$, and $\sigma(M_{Ks}) = 0.19 \pm 0.04$.

between -1 and 1 and are typically asymmetric. The probability densities of variance parameters are also non-Gaussian, since they are forced to be positive, and have fat tails toward higher variance. This captures the intuition that for a finite sample with fixed scatter, it is more difficult to discount the hypothesis that it arose from a high-variance distribution rather than a low variance one.

The population mean absolute magnitudes are $\mu(M_J) = -18.25 \pm 0.03$, $\mu(M_H) = -18.01 \pm 0.03$, and $\mu(M_{Ks}) = -18.25 \pm 0.04$ mag (on the scale of $h = 0.72$), and the population standard deviations are $\sigma(M_J) = 0.17 \pm 0.03$, $\sigma(M_H) = 0.11 \pm 0.03$, and $\sigma(M_{Ks}) = 0.19 \pm 0.04$ mag. In Figure 9, we show the bivariate joint posterior density of the mean and variance for the absolute magnitude in each band, and the bivariate modal values. The skews in the posterior densities for the variances are visible. The absolute magnitude in the H-band clearly has much less intrinsic dispersion than in the J- and K-band and is the best constrained. We have used bivariate

kernel density estimation with the MCMC samples to compute the 68% and 95% highest posterior density contours and the mode, as shown in the figure.

Figure 10 shows the marginal posterior estimates of the individual SN H and J absolute magnitudes (obtained from, e.g., $P(J_0^s - \mu_s | \mathcal{D}, \mathcal{Z})$) plotted with contours representing the 68% and 95% probability contours of the bivariate population density $P(M_J, M_H | \boldsymbol{\mu}_\psi, \boldsymbol{\Sigma}_\psi)$ estimated using the modal values of the covariance matrix $\boldsymbol{\Sigma}_\psi$ (Table 3). We also show the marginal posterior density of the correlation coefficient for the pair of absolute magnitudes.

We see that the absolute magnitudes in J and H are highly correlated ($\rho \approx 0.73$) with strong evidence for positive correlation ($P(\rho > 0) > 0.97$). The data also suggest that intrinsically brighter SNe are typically bluer in the $J - H$ color. Interestingly, this parallels the "brighter-bluer" relation seen in optical light curves (e.g., Guy et al. 2005; Jha et al. 2006). There is also evi-
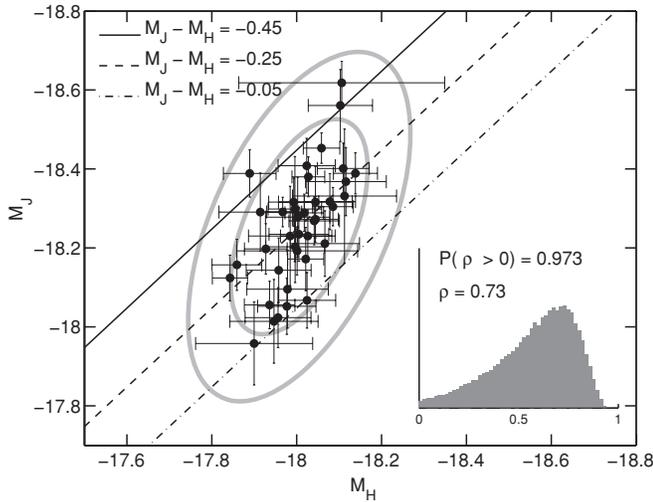
Figure 10. Strong correlation ρ between J- and H- band peak luminosities. The gray ellipses contain 95% and 68% of the bivariate population probability distribution using the modal values of the population covariance. The straight lines indicate sets of constant $J - H$ color. There appears to be a trend that bluer $J - H$ objects are also intrinsically brighter. The marginal posterior probability density of the correlation coefficient obtained via MCMC is shown in the inset. The mode is $\rho = 0.73$, and $P(\rho > 0) = 0.97$ is obtained by numerical integration of the marginal density.

dence for positive correlations between M_J, M_{Ks} and M_H, M_{Ks}, although the modal correlations are weaker ($\rho \approx 0.4 - 0.5$, Table 3).

4.3.3. Statistical Structure of J-band Light-curve Shapes

We examine the statistical relationships between the different features of the J-band light curve. We focus only on those correlations, which are significantly different from zero, as measured by the tail probabilities of the posterior distribution (Table 3).

The peak-to-trough initial decline rate, as measured by d/α, is moderately correlated ($\rho \approx 0.55$) with the trough-to-second peak rise rate, as measured by r/β. The posterior probability of a positive correlation is 98%. This trend indicates that faster pre-trough declines lead to faster post-trough rises. There is a strong correlation ($\rho \approx -0.77$) in the early light curve between the initial decline rate and the time to trough (α), demonstrating that slower declining light curves have later troughs. There is a moderate ($\rho \approx -0.43$) correlation between the post-trough rise rate and the trough-to-second peak time (β), suggesting that in the post-trough phase, slower risers have later secondary maxima. The posterior probability of a negative correlation is 93%. As shown in Table 3, there is no correlation between the early timescale (α) and the late, post-trough timescale (β).

4.3.4. Statistical Correlations between NIR Absolute Magnitudes and Light-curve Shape

Statistical correlations between peak SN Ia absolute magnitudes and light-curve shape are of paramount importance to cosmological studies, because they relate the intrinsic luminosity, a hidden, intrinsic parameter, to a distance-independent observable measure. Relations between optical light-curve shape and optical luminosity have been leveraged to improve the utility of SN Ia as standard candles and distance indicators. We present the first quantitative search and measurement of correlations between near-infrared absolute magnitudes and light-curve shape
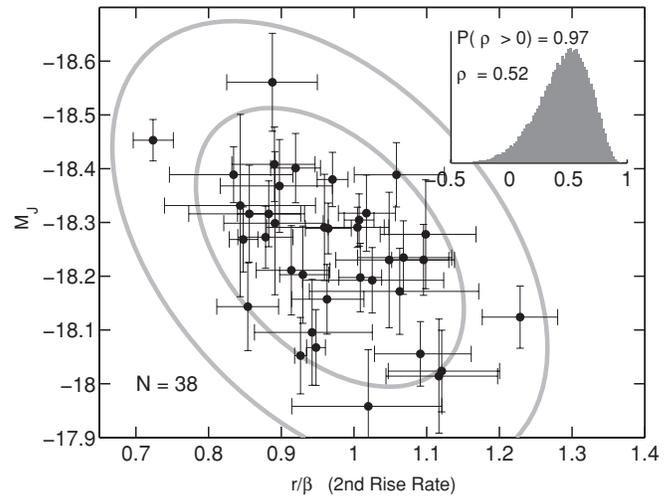
Figure 11. Evidence for moderate correlation between the J-band second rise rate and the J-band peak luminosity. The gray ellipses contain 95% and 68% of the bivariate population probability distribution using the modal values of the population covariance. The inset shows the marginal posterior probability density of the correlation coefficient obtained via MCMC, along with the mode and probability of positive correlation with the absolute magnitude.
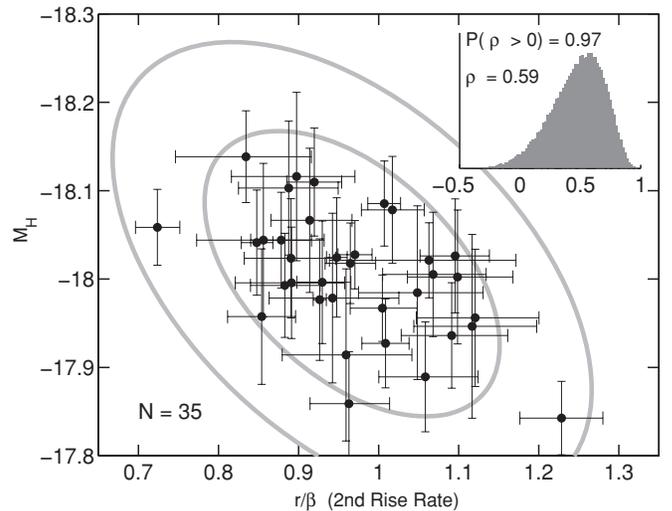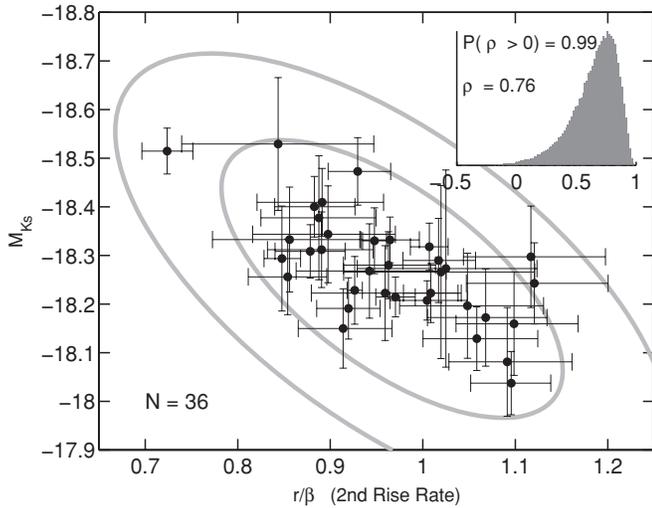

Figure 12. Moderate correlation between the J-band second rise rate and the H-band peak luminosity. The gray ellipses contain 95% and 68% of the bivariate population probability distribution using the modal values of the population covariance. The inset shows the marginal posterior probability density of the correlation coefficient obtained via MCMC, along with the mode and probability of positive correlation with the absolute magnitude.

as measured from the J-band light curves. Again, we only highlight correlations having the highest posterior probability of being non-zero as measured from the tail probability.

Figure 11 shows the M_J and r/β estimates for individual SNe together with 68% and 95% probability contours of the population density $P(M_J, r/\beta | \boldsymbol{\mu}_\psi, \boldsymbol{\Sigma}_\psi)$ using the expected posterior estimate of the population mean and the modal covariance matrix. The most likely correlation is moderate ($\rho \approx 0.52$). The evidence for a positive correlation is fairly strong: $P(\rho > 0) = 97\%$. This demonstrates that the brighter SN Ia J-band light curves are likely to rise more slowly to the second maximum.

Figure 12 shows a moderate correlation ($\rho \approx 0.59$) of the J-band rise rate with the H-band luminosity. There is good evidence for a positive correlation ($P(\rho > 0) = 0.97$).

Figure 13. Evidence for correlation between the *J*-band second rise rate and the K_s-band peak luminosity. The gray ellipses contain 95% and 68% of the bivariate population probability distribution using the modal values of the population covariance. The inset shows the marginal posterior probability density of the correlation coefficient obtained via MCMC, along with the mode and probability of positive correlation with the absolute magnitude.

Figure 13 shows that the K_s-band luminosity has a fairly strong correlation with the *J*-band rise rate ($\rho \approx 0.76$) with strong evidence ($P(\rho > 0) = 99\%$) for a positive relation. In these bivariate plots, we have only shown individual SNe with posterior uncertainty smaller than the population width in each parameter. The fully Bayesian calculation properly accounts for the uncertainties in the parameters of individual SNe when determining the posterior density of the population correlation.

Taken together these correlations suggest that SN Ia light curves brighter in the NIR at peak have slower rates of evolution at later times, as measured from the *J*-band light curve. A larger sample of SN Ia light curves in the NIR is needed to confirm these correlations. The best measured light curves tend to be at lower redshifts, where peculiar velocity uncertainties make the absolute magnitudes highly uncertain. SNe farther out in the Hubble flow have better determined absolute magnitudes but are likely to have poorer quality measurements of the whole light-curve evolution. Continued monitoring of local SNe in the NIR over a wide range of distances and redshifts will help to solidify our estimates of these correlations (by narrowing their posterior probability densities and providing better estimates of the modal correlation coefficients).

Using the trained statistical model, we can estimate the expected precision of distance prediction for different types of light-curve observations represented by subsets of the observable light-curve parameters in $\boldsymbol{\phi}_s$. Suppose the observable vector $\tilde{\boldsymbol{\phi}}_s$ of a new hypothetical SN can be partitioned into the observed parameters and the unobserved parameters $\tilde{\boldsymbol{\phi}}_s = (\tilde{\boldsymbol{\phi}}_s^o, \tilde{\boldsymbol{\phi}}_s^u)$. We have computed the variance of the predictive $\tilde{\mu}_s$ conditional on $\tilde{\boldsymbol{\phi}}_s^o$, and marginalizing over $\tilde{\boldsymbol{\phi}}_s^u$ and the posterior uncertainty of the hyperparameters $\boldsymbol{\mu}_\psi$, $\boldsymbol{\Sigma}_\psi$, for various partitions of $\tilde{\boldsymbol{\phi}}_s$. We find that the statistical model implies the following properties: (1) if one only observes the light curve around $T_{B\max}$, then the single most valuable measurement is the *H*-band apparent magnitude, providing a distance modulus precision of ~ 0.14 mag; the *J*- and K_s-bands do not add much more statistical power. (2) If one monitors the *J*-band light curve at late times to measure the second rise to the secondary maximum, the moderate corre-

lation of the rise rate with absolute magnitudes can be used to decrease the uncertainty toward ~ 0.1 mag. In the next section, we will test the predictive performance of the model using the NIR SN Ia sample.

4.4. The Hubble Diagram of JHK_s SN Ia Light Curves

4.4.1. Hubble Residuals and Training Error

In Table 4, we list the redshifts and several estimates of the distance moduli to the SNe Ia in our training set. We list the redshift-based LCDM Hubble flow distance and its uncertainty, described by Equation (8), on the $H_o = 72$ km s^{-1} Mpc^{-1} scale. This describes the factor $P(\mu_s|z_s)$, conditioning on the redshift only, and incorporating the assumed $\sigma_{\rm pec} = 150$ km s^{-1} peculiar velocity uncertainty and the redshift measurement error. As a product of the Bayesian treatment, we obtain the posterior estimate of the distance modulus, combining the redshift information with the statistical light-curve model and conditioning on the entire data set to generate an "information update." This is expressed as the marginal posterior probability $P(\mu_s|\mathcal{D}, \mathcal{Z})$. The mean and standard deviation of this probability density are listed as $\mu_{\rm post}$ and $\sigma_{\rm post}$ for each SN.

The typical measure of the quality of a model for SN Ia standard candles is the average residual in the Hubble diagram. First, the redshift-based distance moduli and photometric light curves are used to "train" a statistical model, by determining the mean absolute magnitude and variance, and perhaps relationships between absolute magnitude and light-curve shape. Once these parameters of the statistical model are determined from the training set data, the photometric light curves are fed into the model *without* the redshift-based distances to "predict" standard candle distances using the model. These new distances are compared to the Hubble flow distances to calculate the average residual error.

This measure of "training error" may be called the *resubstitution error* because it involves using the redshifts and light-curve data of the training set to train the model parameters, and then resubstituting the light-curve data back into the model to produce distance "predictions" *as if* the light-curve data were new.

In our fully Bayesian formulation, the process of "training" corresponds to computing the (non-Gaussian) posterior density over the hyperparameters $P(\boldsymbol{\mu}_\psi, \boldsymbol{\Sigma}_\psi|\mathcal{D}, \mathcal{Z})$ obtained by integrating over Equation (18). This is to be contrasted with simpler approaches that merely find point estimates of the model parameters. The process of prediction uses this posterior probability together with new light curve data $\tilde{\mathcal{D}}_s$ to compute the predictive posterior $P(\tilde{\mu}_s|\tilde{\mathcal{D}}_s, \mathcal{D}, \mathcal{Z})$. The marginalization over the hyperparameters correctly incorporates the uncertainties in the means, variances, and correlations of the absolute magnitudes and light-curve shape parameters. Recall that the training set is $\mathcal{D} = \{\mathcal{D}_s\}$ and $\mathcal{Z} = \{\mathcal{Z}_s\}$. The resubstitution distance estimate is obtained by setting $\tilde{\mathcal{D}}_s = \mathcal{D}_s$ and computing the predictive probability $P(\mu_s|\mathcal{D}_s, \mathcal{D}, \mathcal{Z})$ for each SN s in the training set.

The uncertainty-weighted mean resubstitution error is then a sum over all resubstituted predicted distances for each SN:

$$\text{err}_{\rm resub}^2 = \sum_{s=1}^{N_{\rm SN}} w_s \times \left[\mu_{\rm resub}^s - \mathbb{E}(\mu_s|z_s)\right]^2 \bigg/ \sum_{s=1}^{N_{\rm SN}} w_s, \quad (31)$$

where the expected predictive distance is $\mu_{\rm resub}^s$, the variance is $\sigma_{\rm resub,s}^2$, and the weights are $w_s^{-1} = \sigma_{\mu,s}^2 + \sigma_{\rm resub,s}^2$. The

Table 4
SN Ia NIR Distance Modulus Estimates

SN	cz^a (km s^{-1})	cz_{err} (km s^{-1})	$\mu_{LCDM}|z$ (mag)	$\sigma_\mu|z$ (mag)	μ_{post}^b (mag)	σ_{post} (mag)	μ_{resub} (mag)	σ_{resub} (mag)	$\bar{\mu}_{pred}$ (mag)	s_{pred} (mag)	σ_{pred} (mag)
SN1998bu	709	20	30.00	0.46	29.84	0.10	29.85	0.09	29.84	0.09	0.12
SN1999cl	957	86	30.62	0.39	30.95	0.10	30.99	0.08	31.00	0.12	0.14
SN1999cp	2909	14	33.05	0.11	32.98	0.08	32.94	0.14	32.94	0.04	0.15
SN1999ee	3296	15	33.32	0.10	33.23	0.06	33.20	0.05	33.16	0.07	0.09
SN1999ek	5191	10	34.32	0.06	34.36	0.06	34.45	0.14	34.47	0.06	0.15
SN1999gp	8113	18	35.30	0.04	35.28	0.04	35.10	0.20	35.05	0.06	0.20
SN2000E	1803	19	32.00	0.18	31.89	0.06	31.89	0.07	31.86	0.09	0.11
SN2000bh	6765	21	34.90	0.05	34.91	0.04	34.90	0.04	34.88	0.04	0.06
SN2000bk	7976	20	35.27	0.04	35.27	0.04	35.34	0.08	35.57	0.13	0.15
SN2000ca	6989	62	34.97	0.05	34.92	0.05	34.78	0.10	34.75	0.05	0.11
SN2000ce	5097	15	34.28	0.06	34.28	0.06	34.28	0.12	34.24	0.09	0.14
SN2001ba	8718	22	35.46	0.04	35.48	0.03	35.53	0.08	35.51	0.06	0.10
SN2001bt	4220	13	33.87	0.08	33.84	0.05	33.83	0.05	33.82	0.06	0.08
SN2001cn	4454	250	33.97	0.14	33.84	0.05	33.83	0.06	33.86	0.05	0.08
SN2001cz	4506	20	34.00	0.07	33.94	0.06	33.88	0.12	33.85	0.04	0.12
SN2001el	978	10	30.70	0.33	31.08	0.07	31.10	0.03	31.17	0.08	0.08
SN2002bo	1696	20	31.88	0.19	32.19	0.06	32.19	0.05	32.21	0.11	0.12
SN2002dj	2880	22	33.03	0.11	32.95	0.05	32.94	0.05	32.92	0.07	0.09
SN2003cg	1340	24	31.37	0.25	31.92	0.07	31.97	0.06	31.97	0.18	0.19
SN2003du	2206	14	32.44	0.15	32.58	0.10	32.61	0.17	32.60	0.08	0.19
SN2004S	2607	16	32.81	0.13	32.96	0.08	33.01	0.06	33.07	0.12	0.14
SN2004eo	4859	17	34.17	0.07	34.05	0.05	33.93	0.10	33.92	0.05	0.11
SN2005ao	11828	126	36.14	0.04	36.15	0.04	36.20	0.10	36.40	0.20	0.23
SN2005cf	2018	11	32.24	0.16	32.14	0.08	32.13	0.06	32.10	0.09	0.11
SN2005ch	8094	1499	35.30	0.40	35.26	0.10	35.27	0.11	35.25	0.06	0.12
SN2005el	4349	8	33.93	0.08	33.91	0.05	33.90	0.03	33.85	0.08	0.08
SN2005eq	8535	25	35.41	0.04	35.40	0.04	35.39	0.05	35.26	0.15	0.16
SN2005iq	10102	40	35.79	0.03	35.79	0.03	35.86	0.17	35.81	0.06	0.18
SN2005na	7826	26	35.23	0.04	35.22	0.04	35.21	0.14	35.16	0.10	0.17
SN2006D	2560	18	32.76	0.13	32.73	0.06	32.72	0.06	32.76	0.12	0.14
SN2006N	4468	27	33.99	0.07	33.97	0.06	33.98	0.15	33.95	0.06	0.16
SN2006X	1091	20	30.88	0.30	31.10	0.07	31.12	0.03	31.12	0.09	0.10
SN2006ac	7123	17	35.01	0.05	35.01	0.05	34.98	0.14	34.99	0.09	0.17
SN2006ax	4955	20	34.21	0.07	34.26	0.06	34.31	0.09	34.38	0.06	0.11
SN2006cp	6816	14	34.92	0.05	34.92	0.05	34.92	0.14	34.90	0.04	0.14
SN2006gr	10547	22	35.89	0.03	35.90	0.03	36.00	0.19	36.02	0.09	0.21
SN2006le	5403	12	34.40	0.06	34.50	0.06	34.63	0.08	34.69	0.04	0.09
SN2006lf	4048	10	33.77	0.08	33.80	0.07	33.83	0.11	33.85	0.17	0.20
SN2007cq	7501	50	35.13	0.05	35.11	0.05	35.01	0.19	34.91	0.03	0.19

Notes.
[a] Corrected to the CMB+Virgo frame.
[b] μ_{post} and σ_{post} are the mean and standard deviation of the trained posterior density in the distance modulus. μ_{resub} and σ_{resub} are the mean and standard deviation of the resubstituted predictive posterior of the distance modulus. $\bar{\mu}_{pred}$ and s_{pred} are the average and scatter over multiple bootstrapped training sets of the expected predictive distance moduli. σ_{pred}^2 is the quadrature sum of predictive uncertainty and the scatter over bootstrap predictions.

resubstitution predictive distance moduli and uncertainties are listed in Table 4. Figure 14 shows the Hubble diagram constructed from the resubstitution distance moduli from our JHK_s statistical model. We compute the training resubstitution error over the training set SNe with recession velocities $cz > 2000$ km s^{-1}, and find err$_{resub} = 0.10$ mag.

4.4.2. Cross-validation and Prediction Error

The resubstitution error is an optimistic estimate of the predictive, or generalization, error arising from predicting the distances of new SN Ia light curves that were not in the training set ("out of sample"). The resubstitution prediction $P(\mu_s|\mathcal{D}_s, \mathcal{D}, \mathcal{Z})$ conditions on the data $\mathcal{D}_s$ twice: once

during training when it is included in the training set $\mathcal{D}$, $\mathcal{Z}$ and once in resubstitution to assess the training error. Double use of the data for both training and evaluation is likely to lead to optimistic measures of predictive performance, underestimating the true predictive error. It is always possible to reduce the residuals of a fit to a finite, noisy sample by introducing arbitrarily more complex relations, but arbitrarily complex models will typically not generalize well to new data. We should compute the generalization error for out-of-sample cases to assess predictive performance of the statistical model. The distinction between the training error, or Hubble residual, and the expected prediction error has not been fully addressed in the literature on SN Ia light-curve inference methods. In this section, we describe the novel application of a cross-validation procedure to assess
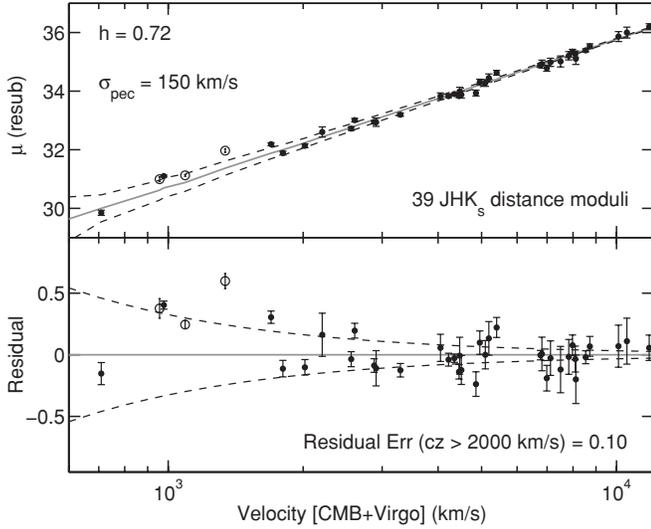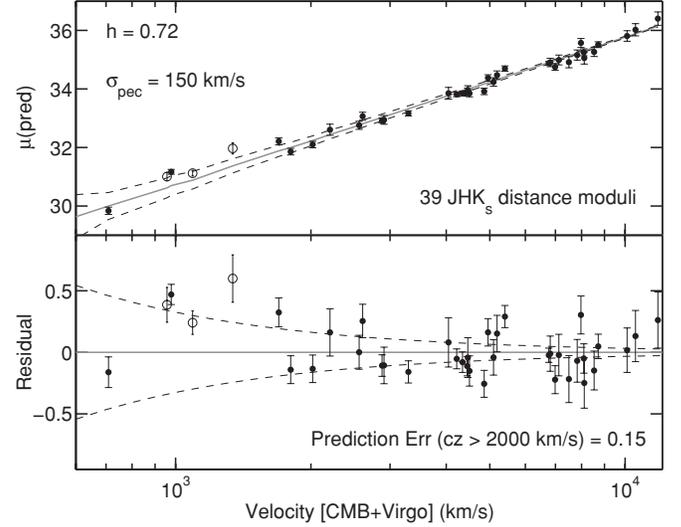
Figure 14. Hubble diagram constructed by resubstitution of training set NIR SN Ia light curves into the trained statistical model. The dotted lines indicate the uncertainty in distance modulus due to peculiar velocities. The average residual at $cz > 2000$ km s^{-1} is an excellent 0.10 mag. The three open circles are the SNe with $A_V > 2$ as measured from the optical light curves with MLCS2k2.

Figure 15. Hubble diagram constructed using predicted distances of NIR SN Ia light curves obtained by inferring the statistical model from 50 boot-strapped training sets. The error bars include both the predictive posterior uncertainty and the scatter over multiple bootstrapped predictions. The estimate of the prediction error for $cz > 2000$ km s^{-1} is an excellent 0.15 mag. The three open circles are the SNe with $A_V > 2$ as measured from the optical light curves with MLCS2k2.

the prediction error and to test the sensitivity of the statistical model to the training set SNe.

To estimate the out-of-sample prediction error and to avoid using the light-curve data twice for training and evaluation, we performed bootstrap cross-validation (Efron 1983; Efron & Tibshirani 1997). We sample SNe with replacement from the original training set to simulate the generation of alternative training sets of the same size. Because of the random resampling, each bootstrapped training set will typically contain duplicate SN data sets and will be missing the others. Each bootstrapped training set of size n SNe will be missing approximately $(1 - 1/n)^n \approx 37\%$ of the SNe in the original training set. The SNe missing from the bootstrap training set form a prediction set, on which we assess the predictive error of a model trained on the complementary training set. Let $\mathcal{D}^B$, $\mathcal{Z}^B$ be a training set bootstrapped from the original $\mathcal{D}$, $\mathcal{Z}$. Then the prediction set is $\mathcal{D}\backslash\mathcal{D}^B$. To train the statistical model we compute $P(\boldsymbol{\mu}_\psi, \boldsymbol{\Sigma}_\psi | \mathcal{D}^B, \mathcal{Z}^B)$ as in Equation (18). For each SN light curve $\mathcal{D}_s \in \{\mathcal{D}\backslash\mathcal{D}^B\}$, we compute the predictive density $P(\mu_s | \mathcal{D}_s, \mathcal{D}^B, \mathcal{Z}^B)$. This random process is repeated so that each SN distance is predicted several times from different bootstrapped training sets. This process avoids using each SN light curve simultaneously for both prediction and training.

We repeated this process 50 times for the original training set in Table 2. On average, each SN is held out of the training set and its distance modulus μ_s is predicted about 18 times. For each SN, the average over all predictions $\bar{\mu}_{\text{pred}}$ and the standard deviation s_{pred} over all predictions are listed in Table 4. We also list the sum of the variance over predictions s_{pred}^2 and the average uncertainty of a prediction (the variance of $P(\mu_s | \mathcal{D}_s, \mathcal{D}^B, \mathcal{Z}^B)$) as σ_{pred}^2. Often the uncertainty of a single prediction is larger than the scatter of the predictions from different bootstrapped training sets, although this is not always true. In Figure 15 we show the Hubble diagram of mean predicted distance moduli $\bar{\mu}_{\text{pred}}$ and their total scatter σ_{pred}. Because we do not use the data twice for training and prediction, the scatter about the Hubble line is less tight than in the resubstitution Hubble diagram, Figure 14.

The "leave-one-out" bootstrap error is computed as an uncertainty-weighted average of squared prediction errors:

$$\text{Err}_{(1)}^2 = \frac{\sum_{B=1}^{50} \sum_{s \in \{\mathcal{D}\backslash\mathcal{D}^B\}} w_s^B \times \left[\mu_{\text{pred}, B}^s - \mathbb{E}(\mu_s | z_s)\right]^2}{\sum_{B=1}^{50} \sum_{s \in \{\mathcal{D}\backslash\mathcal{D}^B\}} w_s^B}, \quad (32)$$

where $\mu_{\text{pred}, B}^s \equiv \mathbb{E}(\mu_s | \mathcal{D}_s, \mathcal{D}^B, \mathcal{Z}^B)$ and the weights are $(w_s^B)^{-1} = \sigma_{\mu, s}^2 + \text{Var}[\mu_s | \mathcal{D}_s, \mathcal{D}^B, \mathcal{Z}^B]$. This bootstrap error estimate is known to be upwardly biased. Efron (1983) and Efron & Tibshirani (1997) have shown that a better estimate of prediction error is obtained by averaging the bootstrap error with the resubstitution error, using the "0.632 bootstrap estimator":

$$\text{Err}_{.632}^2 = 0.632 \times \text{Err}_{(1)}^2 + 0.368 \times \text{err}_{\text{resub}}^2. \quad (33)$$

For the Hubble flow SNe ($cz > 2000$ km s^{-1}) in our sample, we compute this estimate of prediction error: $\text{Err}_{.632} = 0.15$ mag. This is a larger error than the resubstitution error computed above, as expected. However, it is a more realistic estimate of predictive performance of distance estimation with our JHK_s light-curve model and the current SN sample. This result confirms that NIR SNe Ia are excellent standard candles.

This process of resampling of alternative training set tests how sensitive the predictions are to the composition of the finite training set. If the statistical model is reasonable, and we had an infinite training set, we would expect the original training set to be representative of the population of NIR SNe Ia and we would expect the resampled training sets (and also the complementary held out sets) to look like the original, and also be representative of the population. We would expect that the prediction error and the resubstitution error to be almost the same. Our actual training set is finite, so the resampled sets will not look exactly like the original set. This procedure tests the sensitivity to the finite sample in addition to making predictions without double use of the light-curve data.

The gap between the estimated prediction error (0.15 mag) and the resubstitution error (0.10 mag) tells us that the trained

statistical model is sensitive to the finite sampling of the training set. A larger training set would be more robust to resampling and we expect that the future predictive uncertainty will be in between the current resubstitution error and the estimated predictive error. A naïve argument would suggest that if this gap of 0.05 mag between the prediction error and the training error decreases with the square root of the number of SNe Ia in the sample, then a set of ~ 200 SNe Ia would reduce it to about 0.02 mag, and a few hundred would be needed to reduce it to ~ 0.01 mag. To build up statistical strength and further solidify our knowledge of the properties of SNe Ia in the NIR, we continue our campaign to observe SNe Ia in the near-infrared with PAIRITEL.

4.5. The Effect of Dust on the SN Ia NIR Sample

The results presented thus far have ignored the effects of dust extinction in the NIR sample. We can examine the possibility of extracting information about the dust distribution from the NIR by looking at the colors $J_0 - H_0$, $J_0 - K_{s0}$, and $H_0 - K_{s0}$ at the peak. From the population hyperparameters we can compute the mean and standard deviation of these colors. The mean colors are -0.25, 0.0, and 0.25, and their population dispersions are 0.14, 0.20, and 0.17, respectively. This means that, for optical extinctions that are less than $A_V \sim 2 - 3$, dust extinction in the near infrared cannot be clearly distinguished from intrinsic color variations, because of the diminished effects of dust absorption in the NIR. Only SNe with much greater NIR reddening carry information on the dust distribution from their NIR data alone. Estimates of the optical A_V extinctions of the SNe in our sample from MLCS analysis of optical light curves (Hicken et al. 2009a) were reported in WV08. There are only three highly reddened SNe with $A_V > 2$: SN 1999cl, 2006X, and SN 2003cg ($A_V = 3.49$, 3.83, and 4.20, respectively) in the sample of 39 SNe, and they are depicted in Figures 14 and 15 with open circles. Although they are redder in their NIR colors than the population mean, their colors are only about $\sim 1 - 2\sigma$ redder, so they are barely distinguishable from the intrinsic color variations. This conclusion does not change if we calculate the mean and standard deviation of the colors by including or excluding the highly optically reddened SNe.

If we take the A_V estimates from the optical data at face value, we can estimate the likely effect of dust extinction on our posterior estimates of absolute magnitudes. The relative weight of a particular SN in posterior inferences about absolute magnitude-related quantities (means, variances and correlations) is inversely proportional to its magnitude uncertainty due to peculiar velocities: $w_M^s = c/\sigma_{\mu,s}^2$ and $c^{-1} = \sum_s \sigma_{\mu,s}^{-2}$. Comparing these weights to the A_V for each SN, we find that 87% of the magnitude weight lies with SNe with $A_V < 0.5$, 97% of the weight is in $A_V < 1$, and 99.7% of the weight is in $A_V < 2$. The three SNe with $A_V > 2$ have a total weight of 0.32% in magnitude calculations. Although the highly reddened SNe have large Hubble residuals in Figures 14 and 15, since they are at low redshifts where the contribution of peculiar velocity uncertainties to their distance uncertainty is large, they have little influence on the posterior inferences about the NIR absolute magnitudes. Furthermore, they have no effect on the estimates of the training error or prediction error, because only the Hubble flow SNe at $cz > 2000$ km s^{-1} are useful for validation of the statistical model.

The weighted mean A_V value of the sample is $\sum_s w_M^s A_V^s = 0.23$ mag. Assuming a Cardelli et al. (1989) law with $R_V = 2$, this means that the estimated NIR absolute magnitudes would be

impacted by mean extinctions of about $A_J = 0.05$, $A_H = 0.03$, and $A_K = 0.02$. The weighted scatters in the NIR extinctions implied by the A_V values are about $\sigma(A_J) = 0.08$, $\sigma(A_H) = 0.05$, $\sigma(A_K) = 0.03$. If these are the dust contributions to the measured dispersions $\sigma(M_X)$, then subtracting them in quadrature yields intrinsic dispersions of $\sigma(M_J) = 0.15$, $\sigma(M_H) = 0.10$, and $\sigma(M_{Ks}) = 0.19$. These rough estimates do not substantially change our results.

We conclude that the NIR sample alone contains little if any information about the dust distribution and hence it is not worthwhile to use the full model described in Section 2 at this time to infer the dust properties. Additionally, if we extrapolate the A_V estimates obtained from the optical data to the near-infrared, the estimated effect of dust extinction is fairly small. These rough estimates do not take into account the non-Gaussianity of the dust distribution and a full Bayesian analysis of the directed graph in Figure 1, conditioned on both the NIR and optical data simultaneously will be required to obtain informative inferences about the dust properties (K. S. Mandel et al. 2009, in preparation).

5. CONCLUSION

We have constructed the hierarchical Bayesian formulation of statistical inference with SN Ia light curves, and represented the probabilistic structure using formal graphical models. Furthermore, we have presented an MCMC algorithm that uses the conditional independence structure of the equivalent directed acyclic graph to efficiently sample the global posterior probability distribution over individual light-curve parameters and population hyperparameters for training the statistical model on the low-z data set, and for prediction on future SN Ia data. We have applied this approach and computational method to the JHK_s light-curve data set compiled by WV08, including a recent homogeneous set of light curves from PAIRITEL, and computed the joint posterior probabilities over all individual light-curve parameters (Table 2) and the statistical characteristics of the population, including the covariance of absolute magnitudes and J-band light-curve shape parameters (Table 3).

We summarize the assumptions of our statistical model. First, we have assumed that the normalized H- and K_s- band light curves of different SNe are identical between -10 and 20 days around maximum. Furthermore, we have posited a parametric light-curve model for the J-band between -10 and 60 days that captures the variations in the double-peaked structure. A quick look at the data and the template models we have constructed in Figures 3 and 5 reveals that these are reasonable models for the JHK_s data. The major assumption in the application of our hierarchical model is that the parameters governing the multiband absolute light curves are drawn from a jointly multivariate Gaussian population distribution. This is the simplest multivariate distribution that models correlations, and its use is reasonable in the absence of other guiding information. Our results (Figures 9–13) reveal no obvious deviations from this assumption, but this is certainly not a proof, especially with a small sample. This assumption must be constantly re-evaluated in applications of the hierarchical framework to larger or different data sets or with other light-curve models. In this paper, we have not estimated the dust-related aspects of Figure 1, because the effects of dust are small for our NIR sample. However, in future studies in conjunction with optical data, the full graph can be computed using BAYESN to perform probabilistic inference of the SN Ia population and dust distribution.

The marginal intrinsic scatter in peak absolute magnitudes were found to be $\sigma(M_J) = 0.17 \pm 0.03$, $\sigma(M_H) = 0.11 \pm 0.03$, and $\sigma(M_{Ks}) = 0.19 \pm 0.04$. We have presented the first quantitative measurements of the correlations of NIR absolute magnitudes with J-band light-curve shape. We showed that with greater than 95% probability there are positive correlations between peak JHK_s absolute magnitudes and the J-band post-trough rise rate. Intrinsically dimmer SN Ia light curves tend to rise to the second J-band maximum faster. Since in our J-band model, the post-second-peak decline rate is linked to the rise rate, this also suggests that the late-time slopes of J-band light curves are steeper for dimmer SNe. We have also quantitatively measured correlations of the rise rate with other aspects of the light-curve shape (Table 3), which show that faster decline rates go with faster rise rates, shorter times to trough and shorter times to the second maximum. These results suggest that NIR SNe Ia are excellent standard candles at the peak, and they can be improved by using the information in the late-time light curve.

These relations may be useful for better understanding of SN Ia progenitor explosions in conjunction with physical modeling. The theoretical models of Kasen (2006) suggest that the structure of the secondary maximum in the NIR is related to the ionization evolution of the iron group elements in the SN atmosphere. They also 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 optical and bolometric peak magnitudes have much larger variations over the same range of mass. Further observational studies of SNe Ia in the NIR may place valuable constraints on theoretical explosion models.

We constructed a Hubble diagram with the training set SNe Ia, and found an average residual of 0.10 mag for $cz > 2000$ km s^{-1}. We have also performed bootstrap cross-validation to estimate the out-of-sample prediction error, which was found to be an excellent 0.15 mag. The gap between these estimates suggests that a larger sample is needed to solidify our inferences about the population of near-infrared SN Ia light curves. Our group continues to collect an extensive set of nearby NIR SN Ia light curves using PAIRITEL. With an ever-growing data set, in the near future, we will be able to expand the model considered here to include more extensive light-curve models in H and K_s, and to combine the NIR and optical data to gain a better understanding of SN colors and dust extinction (K. S. Mandel et al. 2009, in preparation).

It is worth considering whether the propitious properties of SNe Ia in the NIR can be leveraged by future space missions for SN Ia cosmology. The measurement of dark energy properties by the NASA/DOE Joint Dark Energy Mission will be limited by systematic effects, in particular dust extinction. The diminished absorption by dust and the narrow dispersion of peak luminosities in the NIR, particularly in the H-band, may be crucial to the precise measurement of dark energy, if observations of high-z SNe can be conducted in the rest-frame NIR.

APPENDIX A

CONDITIONAL INDEPENDENCE AND d-SEPARATION

Let $P(\{\theta_i\})$ be a joint distribution of random variables represented by a directed acyclic graph. Consider three disjoint subsets of the random variables $\{\theta_i\}$: $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$. Two sets are *marginally independent* if $P(\mathcal{A}, \mathcal{B}) = P(\mathcal{A})P(\mathcal{B})$. Two sets, $\mathcal{A}$ and $\mathcal{B}$, are *conditionally independent* given a third set $\mathcal{C}$ if $P(\mathcal{A}, \mathcal{B}|\mathcal{C}) = P(\mathcal{A}|\mathcal{C})P(\mathcal{B}|\mathcal{C})$. Marginal independence between $\mathcal{A}$ and $\mathcal{B}$ can be seen in a directed graph because there will be no links between the nodes in set $\mathcal{A}$ and the nodes in set $\mathcal{B}$. Conditional independence indicates that if the values of the nodes in $\mathcal{C}$ are known, then the variables in $\mathcal{A}$ and those in $\mathcal{B}$ are statistically independent. Graphically, this means that all directed or undirected paths (ignoring the arrows) from one set to the other are "blocked" by nodes in $\mathcal{C}$.

Conditional independence between two sets of nodes given a third set can be ascertained from a directed graph using the d-separation property (Pearl 1988): a path between a node in $\mathcal{A}$ and a node in $\mathcal{B}$ is blocked at node θ_i if (1) the intermediate node θ_i is in set $\mathcal{C}$ and the arrows meet at θ_i in a tail-to-tail or head-to-tail fashion (not convergent), or (2) the arrows meet head-to-head (convergent) and the intermediate node θ_i is not in $\mathcal{C}$, and neither are any of its descendants. The nodes $\mathcal{A}$ are d-separated from the nodes $\mathcal{B}$ given set $\mathcal{C}$ if all paths between elements in $\mathcal{A}$ and $\mathcal{B}$ are blocked. If the nodes $\mathcal{A}$ are d-separated from the nodes $\mathcal{B}$ by $\mathcal{C}$, then $\mathcal{A}$ is conditionally independent of $\mathcal{B}$ given $\mathcal{C}$.

APPENDIX B

THE BAYESN ALGORITHM—MATHEMATICAL APPENDIX

In this appendix, we present mathematical details of the BAYESN algorithm. Let $\boldsymbol{\psi}_s^{-M_0^F}$, $\boldsymbol{\psi}_s^{-L, F}$, and $\boldsymbol{\psi}_s^{-NL, F}$ indicate all the intrinsic parameters in $\boldsymbol{\psi}_s$ other than the peak absolute magnitude, the linear shape parameters, and the nonlinear shape parameters in filter F, respectively.

1. We have used the conjugate hyperprior density $P(\boldsymbol{\mu}_\psi, \boldsymbol{\Sigma}_\psi)$ defined in Equations (12) and (13) and choose the noninformative limit by setting $\kappa_0 = 0$, $\nu_0 = -1$, and $\boldsymbol{\Lambda}_0 = \epsilon_0 \boldsymbol{I}$ for small ϵ_0. Let $\bar{\boldsymbol{\psi}}$ be the sample mean of the $\{\boldsymbol{\psi}_s\}$, and let $\boldsymbol{S}_\psi = \sum_{s=1}^{N_{SN}}(\boldsymbol{\psi}_s - \bar{\boldsymbol{\psi}})(\boldsymbol{\psi}_s - \bar{\boldsymbol{\psi}})^T$ be the matrix sum of squared deviations from the mean. The conditional posterior density $P(\boldsymbol{\mu}_\psi, \boldsymbol{\Sigma}_\psi | \{\boldsymbol{\psi}_s\})$ can be decomposed as $\boldsymbol{\Sigma}_\psi | \{\boldsymbol{\psi}_s\} \sim \text{Inv-Wishart}_{N-1}\left([\boldsymbol{\Lambda}_0 + \boldsymbol{S}_\psi]^{-1}\right)$ and

$\boldsymbol{\mu}_\psi | \boldsymbol{\Sigma}_\psi, \{\boldsymbol{\psi}_s\} \sim N(\bar{\boldsymbol{\psi}}, \boldsymbol{\Sigma}_\psi / N_{\text{SN}})$ (Gelman et al. 2003). We first directly sample a new covariance matrix $\boldsymbol{\Sigma}_\psi$ from the inverse Wishart distribution. The matrix drawn in this way is guaranteed to be a proper covariance matrix (i.e., positive semi-definite). Conditional on that matrix we directly sample a new population mean $\boldsymbol{\mu}_\psi$ from the multivariate normal distribution.

2. Let $\bar{A}$ be the sample mean of the $\{A_H^s\}$. The conditional posterior density is $P(\tau_A | \{A_H^s\}) = \text{Inv- Gamma}(N_{\text{SN}} - 1, N_{\text{SN}}\bar{A})$.

3a. Let $N^{-1} = \mathbf{1}^T W^{-1} \mathbf{1}$ and $\bar{F}_0 = N\mathbf{1}^T(W_s^F)^{-1}[\boldsymbol{m}_s^F - \boldsymbol{L}_0^F(\boldsymbol{\theta}_{\text{NL},s}^F) - \boldsymbol{L}_1^F(\boldsymbol{\theta}_{\text{NL},s}^F)\boldsymbol{\theta}_{\text{L},s}^F]$. We can compute the population conditional expectation: $\tilde{F}_0 = \mu_s + A_s^F + \mathbb{E}[M_{0,s}^F | \boldsymbol{\psi}_s^{-M_0^F}, \boldsymbol{\mu}_\psi, \boldsymbol{\Sigma}_\psi]$ and the population conditional variance $C = \text{Var}[M_{0,s}^F | \boldsymbol{\psi}_s^{-M_0^F}, \boldsymbol{\mu}_\psi, \boldsymbol{\Sigma}_\psi]$, using the conditioning property of the multivariate Gaussian distribution. Then the conditional density of $F_{0,s}$ is normal $N(F_{0,s} | \hat{F}_0, \Lambda)$ with variance $\Lambda = (N^{-1}+C^{-1})^{-1}$ and mean $\hat{F}_0 = \Lambda(N^{-1}\bar{F}_0 + C^{-1}\tilde{F}_0)$.

3b. Compute $N^{-1} = \boldsymbol{L}_1^{F,T}(\boldsymbol{\theta}_{\text{NL},s}^F)(W_s^F)^{-1}\boldsymbol{L}_1^F(\boldsymbol{\theta}_{\text{NL},s}^F)$ and $\bar{\boldsymbol{\theta}}_{\text{L}}^F = N\boldsymbol{L}_1^{F,T}(\boldsymbol{\theta}_{\text{NL},s}^F)(W_s^F)^{-1}[\boldsymbol{m}_s^F - \mathbf{1}F_{0,s} - \boldsymbol{L}_0^F(\boldsymbol{\theta}_{\text{NL},s}^F)]$. The conditional population expectation and variance are $\tilde{\boldsymbol{\theta}}_{\text{L}}^F = \mathbb{E}[\boldsymbol{\theta}_{\text{L},s}^F | \boldsymbol{\psi}_s^{-L,F}, \boldsymbol{\mu}_\psi, \boldsymbol{\Sigma}_\psi]$ and $C = \text{Var}[\boldsymbol{\theta}_{\text{L},s}^F | \boldsymbol{\psi}_s^{-L,F}, \boldsymbol{\mu}_\psi, \boldsymbol{\Sigma}_\psi]$. The conditional posterior density of $\boldsymbol{\theta}_{\text{L},s}^F$ is normal $N(\boldsymbol{\theta}_{\text{L},s}^F | \hat{\boldsymbol{\theta}}_{\text{L}}^F, \Lambda)$ with covariance matrix $\Lambda = (N^{-1}+C^{-1})^{-1}$ and mean $\hat{\boldsymbol{\theta}}_{\text{L}}^F = \Lambda(N^{-1}\bar{\boldsymbol{\theta}}_{\text{L}}^F + C^{-1}\tilde{\boldsymbol{\theta}}_{\text{L}}^F)$. Note that steps 3a and 3b could be combined by Gibbs sampling from $P(F_{0,s}, \boldsymbol{\theta}_{\text{L},s}^F | \boldsymbol{\phi}_s^{-L,F}, \mu_s, A_s; \boldsymbol{\mu}_\psi, \boldsymbol{\Sigma}_\psi, \tau_A, \mathcal{D}_s, z_s)$.

3c. Compute the expectation and covariance of the conditional population density: $\tilde{\boldsymbol{\theta}}_{\text{NL}}^F = \mathbb{E}[\boldsymbol{\theta}_{\text{NL},s}^F | \boldsymbol{\psi}_s^{-\text{NL},F}, \boldsymbol{\mu}_\psi, \boldsymbol{\Sigma}_\psi]$ and $C = \text{Var}[\boldsymbol{\theta}_{\text{NL}}^F | \boldsymbol{\psi}_s^{-\text{NL},F}, \boldsymbol{\mu}_\psi, \boldsymbol{\Sigma}_\psi]$. The conditional posterior density of the nonlinear parameters in band F, $\boldsymbol{\theta}_{\text{NL}}^F$ is proportional to $N(\boldsymbol{m}_s^F | \mathbf{1}F_{0,s} + \boldsymbol{L}_0^F(\boldsymbol{\theta}_{\text{NL},s}^F) + \boldsymbol{L}_1^F(\boldsymbol{\theta}_{\text{NL},s}^F)\boldsymbol{\theta}_{\text{L}}^F, W_s^F) \times N(\boldsymbol{\theta}_{\text{NL},s}^F | \tilde{\boldsymbol{\theta}}_{\text{NL}}^F, C)$. We obtain a proposal $\boldsymbol{\theta}_{\text{NL},s}^{F,*} \sim N(\boldsymbol{\theta}_{\text{NL},s}^F, \boldsymbol{\Sigma}_{\text{jump},s}^{\text{NL},F})$, and apply the Metropolis rejection rule.

3d. We allow for the possibility that the probability density of the distance modulus conditioned on the redshift only, $P(\mu_s | z_s)$, may be mildly non-Gaussian. The conditional posterior density is $P(\mu_s | \boldsymbol{\phi}_s, A_s; \boldsymbol{\mu}_\psi, \boldsymbol{\Sigma}_\psi, \mathcal{D}_s, z_s) \propto N(\boldsymbol{\phi}_s - A_s - \boldsymbol{v}\mu_s | \boldsymbol{\mu}_\psi, \boldsymbol{\Sigma}_\psi) \times P(\mu_s | z_s)$, and generally cannot be sampled directly. However, we can approximate $P(\mu_s | z_s) \approx N(\mu_s | \mu_g \equiv f(z_s), \sigma_\mu^2)$ with a Gaussian using Equation (8). We choose the Metropolis–Hasting proposal distribution to be $Q(\mu_s^* | \boldsymbol{\phi}^s, A_s, \boldsymbol{\mu}_\psi, \boldsymbol{\Sigma}_\psi) \propto N(\boldsymbol{\phi}_s - A_s - \boldsymbol{v}\mu_s^* | \boldsymbol{\mu}_\psi, \boldsymbol{\Sigma}_\psi) \times N(\mu_s^* | \mu_g, \sigma_\mu^2) = N(\mu_s^* | \hat{\mu}, \hat{\sigma}_\mu^2)$, where $\hat{\sigma}_\mu^2 = (\sigma_\mu^{-2} + s^{-2})^{-1}$, and $\hat{\mu} = \hat{\sigma}_\mu^{-2}(\sigma_\mu^{-2}\mu_g + s^{-2}\tilde{\mu})$ is a weighted average of the distance information from the redshift and the light curves. The mean $\tilde{\mu} = s^2 \boldsymbol{v}^T \boldsymbol{\Sigma}_\psi^{-1}(\boldsymbol{\phi}_s - \boldsymbol{\mu}_\psi - A_s)$ and variance $s^2 = (\boldsymbol{v}^T \boldsymbol{\Sigma}_\psi^{-1}\boldsymbol{v})^{-1}$ describe the distance information from the individual SN light curves only. We draw a proposed μ_s^* from Q. The Metropolis–Hastings ratio is computed from Equation (21) using the above conditional posterior density and the proposal density. After cancellation of terms, this simplifies to $r = P(\mu_s^* | z_s)N(\mu_s | \mu_g, \sigma_\mu^2)/P(\mu_s | z_s)N(\mu_s^* | \mu_g, \sigma_\mu^2)$. If

$P(\mu_s | z_s)$ is actually close to the Gaussian, Equation (8), then the M-H ratio is identically one, and the proposal μ_s^* is always accepted, as this is the same as the Gibbs sampling. If $P(\mu_s | z_s)$ is mildly non-Gaussian, then r and the acceptance rate will be slightly less than 1.

3dP. For distance prediction, the distance modulus μ_s is Gibbs sampled from the conditional posterior density $N(\mu_s | \tilde{\mu}, s^2)$.

3e. Sample a proposed extinction $A_H^{s,*} \sim N(A_H^s, \sigma_{\text{jump},A,s}^2)$. The conditional posterior density of the extinction $A_H^{s,*}$ is proportional to $N(\boldsymbol{\phi}_s - \boldsymbol{v}\mu_s - A_s(A_H^{s,*}) | \boldsymbol{\mu}_\psi, \boldsymbol{\Sigma}_\psi) \times \text{Expon}(A_H^{s,*} | \tau_A)$. Apply Metropolis rejection.

APPENDIX C

BAYESN—PRACTICAL CONSIDERATIONS

The chain is seeded with initial starting estimates for all the parameters. It is useful before running the MCMC to obtain rough point estimates of the light-curve parameters using, e.g., the MLE. Point estimates of the distance moduli can be obtained from $\mathbb{E}(\mu_s | z_s)$. The extinction values A_H^s can be chosen to be small random numbers. Random noise is added to these point estimates to generate different starting positions of each chain, to ensure that each chain begins in a different region.

The Metropolis steps within the Gibbs scan use jumping kernels that must be tuned to generate efficient MCMC chains. The scalar kernels $\sigma_{\text{jump},A,s}$ are tuned to generate $\sim 40\%$ acceptance rates for their respective Metropolis steps. This is easily done by running a few preliminary short chains to compute the average acceptance rates and adjusting the jumping sizes accordingly. The nonlinear jumping kernel, $\boldsymbol{\Sigma}_{\text{jump},s}^{\text{NL},F}$ is a matrix if there are more than one nonlinear parameters in the light-curve model for band F. This can be estimated from the inverse of the Fisher information matrix at the MLE, or from the sample covariance of the $\boldsymbol{\theta}_{\text{NL},s}^F$ chain values, to reflect the shape of the underlying density. The overall size of the matrix is then scaled to produce acceptance rates of $\sim 40\%$ in preliminary short runs, or $\sim 23\%$ if the dimensionality of $\boldsymbol{\theta}_{\text{NL},s}^F$ is high (Gelman et al. 2003). Once the jumping kernels have been set to appropriate values, long chains are run.

To assess the convergence of the MCMC, a few independent long chains with different initial positions are run. The BAYESN computation is easily be parallelized as each independent MCMC chain can be run on a separate processor. The G-R statistic (Gelman & Rubin 1992) compares the between-chain variances with the within-chain variances in each parameter to compare the coverages of the chains. If the chains have converged, the G-R ratio should be close to 1. The sample paths of representative parameters are inspected visually to ascertain that the chains are well mixed. Upon convergence, the initial portions of each chain are discarded as "burn-in," and the chains are concatenated for final inferences.

REFERENCES

Astier, P., et al. 2006, A&A, 447, 31
Bishop, C. M. 2006, Pattern Recognition and Machine Learning (New York: Springer)
Blondin, S., et al. 2008, ApJ, 682, 724
Bloom, J. S., et al. 2006, in ASP Conf. Ser. 351, Astronomical Data Analysis Software and Systems XV, ed. C. Gabriel, C. Arviset, D. Ponz, & S. Enrique (San Francisco, CA: ASP), 751
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

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
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
Ford, E. B. 2005, AJ, 129, 1706
Freedman, W. L., et al. 2001, ApJ, 553, 47
Frieman, J. A., Turner, M. S., & Huterer, D. 2008, ARA&A, 46, 385
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: Chapman & Hall)
Gelman, A., & Rubin, D. B. 1992, Stat. Sci., 7, 457
Gilks, W. R., Richardson, S., & Spiegelhalter, D. 1995, Markov Chain Monte Carlo in Practice (Boca Raton, FL: Chapman & Hall)
Goldhaber, G., et al. 2001, ApJ, 558, 359
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
Haugbølle, T., Hannestad, S., Thomsen, B., Fynbo, J., Sollerman, J., & Jha, S. 2007, ApJ, 661, 650
Hernandez, M., et al. 2000, MNRAS, 319, 223
Hicken, M., et al. 2009a, ApJ, 700, 331
Hicken, M., et al. 2009b, ApJ, 700, 1097
Hillebrandt, W., & Niemeyer, J. C. 2000, ARA&A, 38, 191
Holman, M. J., et al. 2006, ApJ, 652, 1715
Jensen, F. V. 2001, Bayesian Networks and Decision Graphs (New York: Springer)
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
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
Leibundgut, B. 1989, PhD thesis, Univ. of Basel
Lewis, A., & Bridle, S. 2002, Phys. Rev. D, 66, 103511
Liu, J. S. 2002, Monte Carlo Strategies in Scientific Computing (New York: Springer)
Mandel, K., & Agol, E. 2002, ApJ, 580, L171
Meikle, W. P. S. 2000, MNRAS, 314, 782
Mould, J. R., et al. 2000, ApJ, 529, 786
Neill, J. D., Hudson, M. J., & Conley, A. 2007, ApJ, 661, L123
Pastorello, A., et al. 2007, MNRAS, 376, 1301
Pearl, J. 1988, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference (San Mateo, CA: Morgan Kaufmann)
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
Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 2007, Numerical Recipes: The Art of Scientific Computing (New York: Cambridge Univ. Press)
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. 1995, ApJ, 445, L91
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
Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, ApJ, 500, 525
Spergel, D. N., et al. 2007, ApJS, 170, 377
Stanishev, V., et al. 2007, A&A, 469, 645
Tegmark, M., et al. 2004, Phys. Rev. D, 69, 103501
Valentini, G., et al. 2003, ApJ, 595, 779
Wang, L., Goldhaber, G., Aldering, G., & Perlmutter, S. 2003, ApJ, 590, 944
Wang, X., et al. 2008, ApJ, 675, 626
Wang, X., et al. 2009, ApJ, 697, 380
Wood-Vasey, W. M., et al. 2007, ApJ, 666, 694
Wood-Vasey, W. M., et al. 2008, ApJ, 689, 377