CrossMark

Steve: A Hierarchical Bayesian Model for Supernova Cosmology

S. R. Hinton[1], T. M. Davis[1], A. G. Kim[2], D. Brout[3], C. B. D'Andrea[3], R. Kessler[4,5], J. Lasker[4,5], C. Lidman[6], E. Macaulay[7], A. Möller[6], M. Sako[3], D. Scolnic[5], M. Smith[8], R. C. Wolf[9], M. Childress[8], E. Morganson[10], S. Allam[11], J. Annis[11], S. Avila[7], E. Bertin[12,13], D. Brooks[14], D. L. Burke[15,16], A. Carnero Rosell[17,18], M. Carrasco Kind[10,19], J. Carretero[20], C. E. Cunha[15], L. N. da Costa[18,21], C. Davis[15], J. De Vicente[17], D. L. DePoy[22], P. Doel[14], T. F. Eifler[23,24], B. Flaugher[11], P. Fosalba[25,26], J. Frieman[5,11], J. García-Bellido[27], E. Gaztanaga[25,26], D. W. Gerdes[28,29], R. A. Gruendl[10,19], J. Gschwend[18,21], G. Gutierrez[11], W. G. Hartley[14,30], D. L. Hollowood[31], K. Honscheid[32,33], E. Krause[23], K. Kuehn[34], N. Kuropatkin[11], O. Lahav[14], M. Lima[18,35], M. A. G. Maia[18,21], M. March[3], J. L. Marshall[22], F. Menanteau[10,19], R. Miquel[20,36], R. L. C. Ogando[18,21], A. A. Plazas[24], E. Sanchez[17], V. Scarpine[11], R. Schindler[16], M. Schubnell[29], S. Serrano[25,26], I. Sevilla-Noarbe[17], M. Soares-Santos[37], F. Sobreira[18,38], E. Suchyta[39], G. Tarle[29], D. Thomas[7], V. Vikram[40], and Y. Zhang[11]

[1] School of Mathematics and Physics, University of Queensland, Brisbane, QLD 4072, Australia
[2] Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA
[3] Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, PA 19104, USA
[4] Department of Astronomy and Astrophysics, University of Chicago, Chicago, IL 60637, USA
[5] Kavli Institute for Cosmological Physics, University of Chicago, Chicago, IL 60637, USA
[6] The Research School of Astronomy and Astrophysics, Australian National University, ACT 2601, Australia
[7] Institute of Cosmology and Gravitation, University of Portsmouth, Portsmouth, PO1 3FX, UK
[8] School of Physics and Astronomy, University of Southampton, Southampton, SO17 1BJ, UK
[9] Graduate School of Education, Stanford University, 160, 450 Serra Mall, Stanford, CA 94305, USA
[10] National Center for Supercomputing Applications, 1205 West Clark Street, Urbana, IL 61801, USA
[11] Fermi National Accelerator Laboratory, P.O. Box 500, Batavia, IL 60510, USA
[12] CNRS, UMR 7095, Institut d'Astrophysique de Paris, F-75014, Paris, France
[13] Sorbonne Universités, UPMC Univ Paris 06, UMR 7095, Institut d'Astrophysique de Paris, F-75014, Paris, France
[14] Department of Physics & Astronomy, University College London, Gower Street, London, WC1E 6BT, UK
[15] Kavli Institute for Particle Astrophysics & Cosmology, P.O. Box 2450, Stanford University, Stanford, CA 94305, USA
[16] SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA
[17] Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), Madrid, Spain
[18] Laboratório Interinstitucional de e-Astronomia - LIneA, Rua Gal. José Cristino 77, Rio de Janeiro, RJ - 20921-400, Brazil
[19] Department of Astronomy, University of Illinois at Urbana-Champaign, 1002 West Green Street, Urbana, IL 61801, USA
[20] Institut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology, Campus UAB, E-08193 Bellaterra (Barcelona), Spain
[21] Observatório Nacional, Rua Gal. José Cristino 77, Rio de Janeiro, RJ - 20921-400, Brazil
[22] George P. and Cynthia Woods Mitchell Institute for Fundamental Physics and Astronomy, and Department of Physics and Astronomy, Texas A&M University, College Station, TX 77843, USA
[23] Department of Astronomy/Steward Observatory, 933 North Cherry Avenue, Tucson, AZ 85721-0065, USA
[24] Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91109, USA
[25] Institut d'Estudis Espacials de Catalunya (IEEC), E-08034 Barcelona, Spain
[26] Institute of Space Sciences (ICE, CSIC), Campus UAB, Carrer de Can Magrans, s/n, E-08193 Barcelona, Spain
[27] Instituto de Fisica Teorica UAM/CSIC, Universidad Autonoma de Madrid, E-28049 Madrid, Spain
[28] Department of Astronomy, University of Michigan, Ann Arbor, MI 48109, USA
[29] Department of Physics, University of Michigan, Ann Arbor, MI 48109, USA
[30] Department of Physics, ETH Zurich, Wolfgang-Pauli-Strasse 16, CH-8093 Zurich, Switzerland
[31] Santa Cruz Institute for Particle Physics, Santa Cruz, CA 95064, USA
[32] Center for Cosmology and Astro-Particle Physics, The Ohio State University, Columbus, OH 43210, USA
[33] Department of Physics, The Ohio State University, Columbus, OH 43210, USA
[34] Australian Astronomical Optics, Macquarie University, North Ryde, NSW 2113, Australia
[35] Departamento de Física Matemática, Instituto de Física, Universidade de São Paulo, CP 66318, São Paulo, SP, 05314-970, Brazil
[36] Institució Catalana de Recerca i Estudis Avançats, E-08010 Barcelona, Spain
[37] Brandeis University, Physics Department, 415 South Street, Waltham, MA 02453, USA
[38] Instituto de Física Gleb Wataghin, Universidade Estadual de Campinas, 13083-859, Campinas, SP, Brazil
[39] Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA
[40] Argonne National Laboratory, 9700 South Cass Avenue, Lemont, IL 60439, USA

Abstract

We present a new Bayesian hierarchical model (BHM) named Steve for performing Type Ia supernova (SN Ia) cosmology fits. This advances previous works by including an improved treatment of Malmquist bias, accounting for additional sources of systematic uncertainty, and increasing numerical efficiency. Given light-curve fit parameters, redshifts, and host-galaxy masses, we fit Steve simultaneously for parameters describing cosmology, SN Ia populations, and systematic uncertainties. Selection effects are characterized using Monte Carlo simulations. We demonstrate its implementation by fitting realizations of SN Ia data sets where the SN Ia model closely follows that used in Steve. Next, we validate on more realistic SNANA simulations of SN Ia samples from the Dark Energy Survey and low-redshift surveys (DES Collaboration et al. 2018). These simulated data sets contain more than 60,000 SNe Ia, which we use to evaluate biases in the recovery of cosmological parameters, specifically the equation of state of dark energy, w. This is the most rigorous test of a BHM method applied to SN Ia cosmology fitting and reveals small w biases that depend on the simulated SN Ia properties, in particular the intrinsic SN Ia scatter model. This w bias is less than 0.03 on average, less than half the statistical uncertainty on w. These

simulation test results are a concern for BHM cosmology fitting applications on large upcoming surveys; therefore, future development will focus on minimizing the sensitivity of Steve to the SN Ia intrinsic scatter model.

Key words: dark energy – methods: data analysis

1. Introduction

Two decades have passed since the discovery of the accelerating universe (Riess et al. 1998; Perlmutter et al. 1999). Since that time, the number of observed Type Ia supernovae (SNe Ia) has increased by more than an order of magnitude, with contributions from modern surveys at both low redshift (Bailey et al. 2008; Freedman et al. 2009; Hicken et al. 2009b; Contreras et al. 2010; Conley et al. 2011) and higher redshift (Astier et al. 2006; Wood-Vasey et al. 2007; Frieman et al. 2008; Balland et al. 2009; Amanullah et al. 2010; Chambers et al. 2016; Sako et al. 2018). Cosmological analyses of these supernova samples (Kowalski et al. 2008; Kessler et al. 2009b; Conley et al. 2011; Suzuki et al. 2012; Betoule et al. 2014; Rest et al. 2014; Scolnic et al. 2018) have been combined with complementary probes of large-scale structure and the cosmic microwave background. For a recent review, see Huterer & Shafer (2018). While these efforts have reduced the uncertainty on the equation of state of dark energy (w) by more than a factor of 2, it is still consistent with a cosmological constant, and the nature of dark energy remains an unsolved mystery.

In attempts to tease out the nature of dark energy, active and planned surveys are continually growing in size and scale. The Dark Energy Survey (DES; Bernstein et al. 2012; Abbott et al. 2016) has discovered thousands of SNe Ia, attaining both spectroscopically and photometrically identified samples. The Large Synoptic Survey Telescope (LSST; Ivezic et al. 2008; LSST Science Collaboration et al. 2009) will discover tens of thousands of photometrically classified supernovae. Such increased statistical power demands greater fidelity and flexibility in modeling supernovae for cosmological purposes, as we will require reduced systematic uncertainties to fully utilize these increased statistics (Betoule et al. 2014; Scolnic et al. 2018).

As such, considerable resources are aimed at developing more sophisticated supernova cosmology analyses. The role of simulations mimicking survey observations has become increasingly important in determining biases in cosmological constraints and validating specific supernova models. First used in SNLS (Astier et al. 2006) and ESSENCE analyses (Wood-Vasey et al. 2007) and then refined and improved for the Sloan Digital Sky Survey (SDSS; Kessler et al. 2009b), simulations are a fundamental component of modern supernova cosmology. Betoule et al. (2014) quantized and corrected observational bias using simulations, and, more recently, Scolnic & Kessler (2016) and Kessler & Scolnic (2017) explored simulations to quantify observational bias in SN Ia distances as a function of multiple factors to improve bias correction. Approximate Bayesian computation (ABC) methods also make use of simulations, trading traditional likelihoods and analytic approximations for more robust models with the cost of increased computational time (Weyant et al. 2013; Jennings et al. 2016). Bayesian hierarchical models (BHMs) abound (Mandel et al. 2009; March et al. 2011, 2014; Rubin et al. 2015; Shariff et al. 2016; Roberts et al. 2017) and either use simulation-determined distance corrections to correct for biases or attempt to find analytic approximations for effects such as Malmquist bias to model the biases inside the BHM itself.

In this paper, we lay out a new hierarchical model that builds off the past work of Rubin et al. (2015). We include additional sources of systematic uncertainty, including an analytic formulation of selection efficiency that incorporates parametric uncertainty. We also implement a different model of intrinsic dispersion to both incorporate redshift-dependent scatter and increase numerical efficiency, allowing our model to perform rapid fits to supernova data sets.

Section 2 is dedicated to a quick review of supernova cosmology analysis methods, and Section 3 outlines some of the common challenges faced by analysis methods. In Section 4, we outline our methodology. Model verification on simulated data sets is given in Section 5, along with details on potential areas of improvement. We summarize our methodology in Section 6.

2. Review

While supernova observations take the form of photometric time-series brightness measurements in many bands and a redshift measurement of the supernova (or its assumed host), most analyses do not work from these measurements directly. Instead, most techniques fit an observed redshift and these photometric observations to a supernova model, with the most widely used being that of the empirical SALT2 model (Guy et al. 2007, 2010). This model is trained separately before fitting the supernova light curves for cosmology (Guy et al. 2010; Betoule et al. 2014). The resulting output from the model is, for each supernova, an amplitude x_0 (which can be converted into apparent magnitude, $m_B = -2.5\log(x_0)$), a stretch term x_1, and a color term c, along with a covariance matrix describing the uncertainty on these summary statistics ($\mathcal{C}_\eta$). As all supernovae are not identical, an ensemble of supernovae form a redshift-dependent, observed population of $\hat{m}_B$, $\hat{x}_1$, and $\hat{c}$, where the hat denotes an observed variable.

This ensemble of $\hat{m}_B$, $\hat{x}_1$, and $\hat{c}$ represents an observed population, which—due to the presence of various selection effects—may not represent the true, underlying supernova population. Accurately characterizing this underlying population, its evolution over redshift, and effects from the host-galaxy environment is one of the challenges of supernova cosmology. Given some modeled underlying supernova population that lives in the redshift-dependent space M_B (absolute magnitude of the supernova, traditionally in the Bessell B band), x_1, and c, the introduction of cosmology into the model is simple: it translates the underlying population from absolute magnitude space into the observed population in apparent magnitude space. Specifically, for any given supernova, our map between absolute and apparent magnitude is given by the distance modulus

$$\mu_{\rm obs} = m_B + \alpha x_1 - \beta c - M_B + \Delta M \cdot m \\ + \text{other corrections,} \qquad (1)$$

where M_B is the mean absolute magnitude for all SNe Ia given $x_1 = c = 0$, α is the stretch correction (Phillips 1993;

Phillips et al. 1999), and β is the color correction (Tripp 1998) that respectively encapsulate the empirical relation that broader (longer-lasting) and bluer supernovae are brighter. Here $\Delta M \cdot m$ refers to the host-galaxy mass correlation discussed in Section 4.4.3. The distance modulus $\mu_{\rm obs}$ is a product of our observations; however, a distance modulus μ_C can be precisely calculated given cosmological parameters and a redshift. The "other corrections" term often includes bias corrections for traditional χ^2 analyses. Bias corrections can take multiple forms, such as a redshift-dependent function (Betoule et al. 2014) or a 5D function of c, x_1, α, β, and z (Kessler & Scolnic 2017; Scolnic et al. 2018).

2.1. Traditional Cosmology Analyses

Traditional χ^2 analyses, such as those found in Riess et al. (1998), Perlmutter et al. (1999), Wood-Vasey et al. (2007), Kowalski et al. (2008), Kessler et al. (2009b), Conley et al. (2011), and Betoule et al. (2014), minimize the difference in distance modulus between the observed distance modulus attained from Equation (1) and the cosmologically predicted distance modulus. The χ^2 function minimized is

$$\chi^2 = (\mu_{\rm obs} - \mu_C)^\dagger C_\mu^{-1} (\mu_{\rm obs} - \mu_C), \qquad (2)$$

where C_μ^{-1} is an uncertainty matrix that combines statistical and systematic uncertainties (see Brout et al. 2019 for a review of these uncertainties for the DES supernova analysis). The predicted μ_C is defined as

$$\mu_C = 5 \log \left[\frac{d_L}{10 \; {\rm pc}} \right], \qquad (3)$$

$$d_L = (1 + z) \frac{c}{H_0} \int_0^z \frac{dz'}{E(z')}, \qquad (4)$$

$$E(z) = \sqrt{\Omega_m (1 + z')^3 + \Omega_k (1 + z')^2 + \Omega_\Lambda (1 + z')^{3(1+w)}}, \qquad (5)$$

where d_L is the luminosity distance for redshift z given a specific cosmology; H_0 is the current value of Hubble's constant in km s^{-1} Mpc^{-1}; and Ω_m, Ω_k, and Ω_Λ represent the energy density terms for matter, curvature, and dark energy, respectively.

The benefit this analysis methodology provides is speed; for samples of hundreds of supernovae or less, efficient matrix inversion algorithms allow the likelihood to be evaluated quickly. The speed comes with several limitations. First, formulating a χ^2 likelihood results in a loss of model flexibility by assuming Gaussian uncertainty. Second, the method of creating a covariance matrix relies on computing partial derivatives; thus, any uncertainty estimated from this method loses information about correlation between sources of uncertainty. For example, the underlying supernova color population's mean and skewness are highly correlated; however, this correlation is ignored when determining population uncertainty using numerical derivatives of population permutations. While correlations can be incorporated into a covariance matrix, it requires human awareness of the correlations, and thus methods that can automatically capture correlated uncertainties are preferable. Third, the computational efficiency is dependent on both creating the global covariance matrix and inverting a covariance matrix with dimensionality
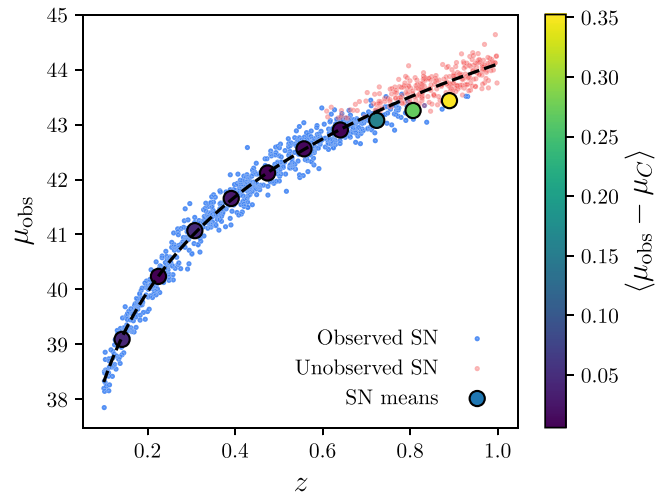

Figure 1. Example of the effects of Malmquist bias. Here we show 1000 simulated supernova redshifts and distance modulus given fiducial cosmology. The simulated survey is magnitude-limited; all supernovae brighter than magnitude 24 are successfully observed (blue dots), and all supernovae dimmer than 24th mag are not successfully observed (red dots). By binning the supernovae along the redshift and taking the mean distance modulus of the supernovae in each bin, we can see that at higher redshift, where Malmquist bias kicks in, the population mean drops and becomes biased. This source of bias must be corrected either by adjusting the data (such as subtracting the found bias) or by incorporating Malmquist bias explicitly in the cosmological model.

linearly proportional to the number of supernovae. As this number increases, the cost of inversion rises quickly and is not viable for samples with thousands of supernovae. A recent solution to this computational cost problem is to bin the supernovae in redshift bins, which produces a matrix of manageable size and can allow fitting without matrix inversion at every step. While binning data results in some loss of information, recent works tested against simulations show that this loss does not result in significant cosmological biases (Scolnic & Kessler 2016; Kessler & Scolnic 2017).

Selection efficiency, such as the well-known Malmquist bias (Malmquist 1922), is accounted for by correcting the determined $\mu_{\rm obs}$ from the data or, equivalently, adding a distance bias to the μ_C prediction. Specifically, Malmquist bias is the result of losing the fainter tail of the supernova population at high redshift. An example of Malmquist bias is illustrated in Figure 1, which simulates supernovae according to Equation (1). Simulations following survey observational strategies and geometry are used to calculate the expected bias in distance modulus, which is then subtracted from the observational data. When using traditional fitting methods such as that found in Betoule et al. (2014), these effects are not built into the likelihood and instead are formed by correcting data. This means that the bias uncertainty is not captured fully in the χ^2 distribution, and subtle correlations between cosmological or population parameters and the bias correction is lost. Recent developments such as the BBC method (Kessler & Scolnic 2017) incorporate corrections dependent on α and β, improving their capture of uncertainty on bias corrections in the χ^2 likelihood.

2.2. ABC

To avoid the limitations of the traditional approaches, several recent methods have adopted ABC, where supernova samples are forward-modeled in parameter space and compared to

observed distributions. Weyant et al. (2013) provided an introduction to ABC methods for supernova cosmology in the context of the SDSS-II results (Sako et al. 2018) and flat ΛCDM cosmology, while Jennings et al. (2016) demonstrated their superABC method on simulated first-season DES samples. In both examples, the supernova simulation package SNANA (Kessler et al. 2009a) is used to forward-model the data at each point in parameter space.

Simulations provide great flexibility and freedom in how to treat the systematic uncertainties and selection effects associated with supernova surveys. By using forward modeling directly from these simulations, the data do not need to be corrected, analytic approximations do not need to be applied, and we are free to incorporate algorithms that simply cannot be expressed in a tractable likelihood, such as those found in traditional analyses from Section 2.1. This freedom comes with a cost: computation. The classical χ^2 method's most computationally expensive step in a fit is matrix inversion. For ABC methods, we must instead simulate an entire supernova population, drawing from underlying supernova populations, modeling light curves, applying selection effects, fitting light curves, and applying data cuts. This is an intensive process.

One final benefit of ABC methods is that they can move past the traditional treatment of supernovae with summary statistics (m_B, x_1, and c). Jennings et al. (2016) presented two metrics, which are used to measure the distance between the forward-modeled and observed populations and are minimized in fitting. The first metric compares forward-modeled summary statistic populations (denoted the "Tripp" metric), and the second utilizes the observed supernova light curves themselves, moving past summary statistics. However, we note that evaluation of systematic uncertainty was only performed using the Tripp metric.

2.3. BHMs

Sitting between the traditional model's simplicity and the complexity of forward modeling lies BHM. Hierarchical models utilize multiple layers of connected parameters, with the layers linked via well-defined and physically motivated conditional probabilities. For example, an observation of a parameter from a population will be conditioned on the true value of the parameter, which itself will be conditioned on the population distribution of that parameter. We can thus incorporate different population distributions and parameter interdependence that cannot be found in traditional analyses, where uncertainty must be encapsulated in a covariance matrix. Unlike ABC methods, which can model arbitrary probability distributions, BHM methods are generally constrained to representing probabilities using analytic forms.

With the introduction of multiple layers in our model, we can add more flexibility than a traditional analysis while still maintaining most of the computational benefits that come from having a tractable likelihood. Mandel et al. (2009, 2011, 2017) constructed a hierarchical model that they applied to supernova light-curve fitting. March et al. (2011) derived a hierarchical model and simplified it by analytically marginalizing over nuisance parameters to provide increased flexibility with reduced uncertainty over the traditional method, but they did not incorporate bias corrections. March et al. (2014) and Karpenka (2015) improved upon this by incorporating redshift-dependent magnitude corrections from Perrett et al. (2010) to remove bias and validate on 100 realizations of SNLS-like

simulations. The recent BAHAMAS model (Shariff et al. 2016) builds on this and reanalyzes the JLA data set (using redshift-dependent bias corrections from Betoule et al. 2014) while including extra freedom in the correction factors α and β, finding evidence for redshift dependence on β. Ma et al. (2016) performed a reanalysis of the JLA data set within a Bayesian formulation, finding significant differences in α and β values from the original analysis from Betoule et al. (2014). Notably, these methods rely on data that is bias-corrected, or the methods ignore biases; however, the UNITY framework given by Rubin et al. (2015) incorporates selection efficiency analytically in the model and is applied to the Union 2.1 data set (Suzuki et al. 2012). The assumption made by the UNITY analysis is that the bias in real data is perfectly described by an analytic function. They validated their model to be free of significant biases using fits to 30 realizations of supernova data sets that are constructed to mimic the UNITY framework. The well-known Bayesian estimation applied to multiple species (BEAMS) methodology from Kunz et al. (2007) has been extended and applied in several works (Hlozek et al. 2012), most recently to include redshift uncertainty for photometric redshift application as zBEAMS (Roberts et al. 2017) and simulated bias corrections in Kessler & Scolnic (2017). For the latter case, by inferring biases using Bayesian models, sophisticated corrections can be calculated and then applied to more traditional χ^2 models.

While there are a large number of hierarchical models available, none of them have undergone comprehensive tests using realistic simulations to verify each models' respective bias. Additionally, testing has generally been performed on supernova simulations with either ΛCDM or flat ΛCDM cosmology. However, quantifying the biases on wCDM cosmology simulations with realistic simulations is becoming critically important as precision supernova cosmology comes into its own, and the focus shifts from determination of Ω_m to both Ω_m and w.

The flexibility afforded by hierarchical models allows for investigations into different treatments of underlying supernova magnitude, color and stretch populations, host-galaxy corrections, and redshift evolution, each of which will be discussed further in the outline of our model below. Our model is designed to increase the numerical efficiency of prior works while incorporating the flexibility of hierarchical models. We reduce our dependence on an assumed scatter model in simulations by not utilizing bias corrections in an effort to provide a valuable cross-check on analysis methodologies that utilize scatter model–dependent bias corrections.

3. Challenges in Supernova Cosmology

The diverse approaches and implementations applied to supernova cosmology are a response to the significant challenges and complications faced by analyses. In this section, we outline several of the most prominent challenges.

Foremost among these challenges is our ignorance of the underlying Type Ia intrinsic dispersion. Ideally, analysis of the underlying dispersion would make use of an ensemble of time-series spectroscopy to characterize the diversity of SNe Ia. However, these data are difficult to obtain, and recent efforts to quantify the dispersion draw inference from photometric measurements. The underlying dispersion model is not a solved problem, and we therefore test against two dispersion models in this work. The first is based on the Guy et al. (2010,

hereafter G10) scatter model, and the second is sourced from Chotard et al. (2011, hereafter C11). As the SALT2 model does not include full treatment of intrinsic dispersion, each scatter model results in different biases in m_B, x_1, and c when fitting the SALT2 model to light-curve observations and results in increased uncertainty on the summary statistics that is not encapsulated in the reported covariance $\mathcal{C}_\eta$. These two scatter models are currently assumed to span the possible range of scatter in the underlying supernova population. We have insufficient information to prefer one model over the other; thus, we have to account for both possible scatter models.

The underlying supernova population is further complicated by the presence of outliers. Non-SNe Ia often trigger transient follow-up in surveys and can easily be mistaken for SNe Ia and represent outliers from the standardized SN Ia sample. This contamination is not just a result of non-SNe Ia being observed but can also arise from host-galaxy misidentification causing incorrect redshifts to be assigned to supernovae. Different optimizations to the host-galaxy algorithm can result in misidentification of the host at the 3%–9% level (Gupta et al. 2016), resulting in a broad population of outliers. For spectroscopic surveys, where both supernova type and redshift can be confirmed through the supernova spectra, this outlier population is negligible. However, for photometric surveys, which do not have the spectroscopic confirmation, it is one of the largest challenges: how to model, fit, and correct for contaminants.

Finally, one of the other persistent challenges facing supernova cosmology analyses are the high number of systematics. Because of the rarity of SN Ia events, data sets are commonly formed from the SN Ia discoveries of multiple surveys in order to increase the number of supernovae in a data set. However, each survey introduces additional sources of systematic error, from sources within each survey, such as band calibration, to systematics introduced by calibration across surveys. Peculiar velocities, different host environments, and dust extinction represent additional sources of systematic uncertainty that must all be modeled and accounted for.

4. Our Method

We construct our hierarchical Bayesian model Steve with several goals in mind: creation of a redshift-dependent underlying supernova population, treatment of an increased number of systematics, and analytic correction of selection effects, including systematic uncertainty on those corrections. We also desire Steve to be more computationally efficient than prior works, such that cosmological results from thousands of supernovae are obtainable on the order of hours, rather than days. As this is closest to the UNITY method from Rubin et al. (2015, hereafter R15), we follow a similar model scaffold and construct the model in the programming language Stan (Carpenter et al. 2017; Stan Development Team 2017). The primary challenge of fitting hierarchical models is their large number of fit parameters, and Stan, which uses automatic differentiation and the no-U-turn Sampler (NUTS, a variant of Hamiltonian Monte Carlo), allows us to efficiently sample high-dimensional parameter space.

At the most fundamental level, a supernova cosmology analysis is a mapping from an underlying population onto an observed population, where cosmological parameters are encoded directly in the mapping function. The difficulty arises both in adequately describing the biases in the mapping
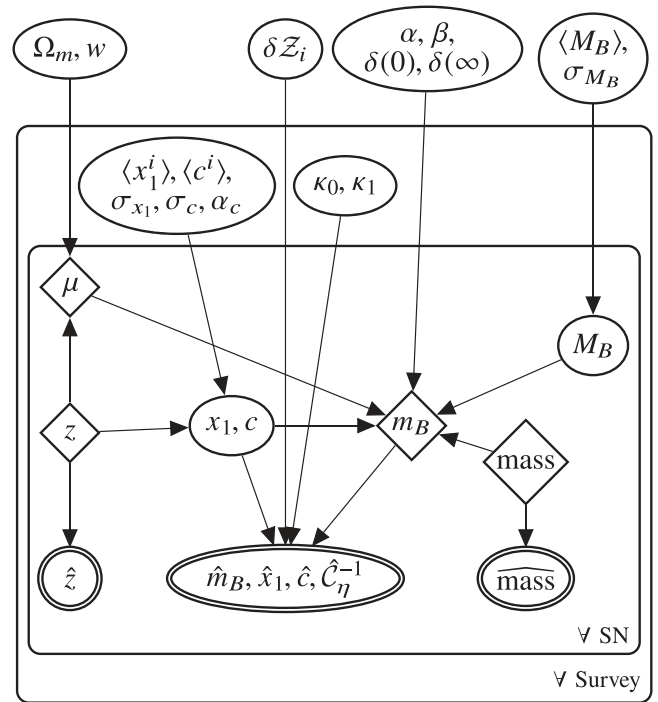

Figure 2. Probabilistic graphical model for our likelihood without selection effects. Double-lined nodes represent observed variables, diamond nodes represent deterministic variables, and single-lined ellipse nodes represent fit variables. The SN box represents observed and latent variables for each individual supernova, while the survey box represents survey-specific variables, which in general describe the supernova population for the survey and the systematics associated with it. Variables that appear outside both boxes represent top-level model parameters. We note that we have shown the model to have latent variables $\{M_B, x_1\, c\}$, which uniquely determines m_B, given μ and other parameters. Thus, the two nodes M_B and m_B make up a single layer in our hierarchy, not two layers. In the code implementation, m_B is more efficiently parameterized than M_B; however, the mathematics remains constant regardless of whether we parameterize M_B or m_B, as one can determine the other. We write out mass instead of m to reduce possible confusion with magnitude in the diagram. Finally, as we take the redshift measurement $\hat{z}$ and mass probability mâss as exact, they are not conditioned on underlying distributions and are top-level parameters.

function and in adding sufficient, physically motivated flexibility in both the observed and underlying populations while not adding too much flexibility, such that model fitting becomes pathological due to increasing parameter degeneracies within the model. In the analysis of this article, the underlying model universe maps to the observed universe as sketched in the BHM of Figure 2. The dependencies between the model and observations can be tracked following the arrows of the BHM, and a summary of all of the conditional probabilities can be found in Section 4.5.

In the following sections, we will describe the model parameters, the mapping functions that connect them to the data, the effect of sample selection (in Equation (12)), and the pathologies that can occur when evaluating the model. Summaries of observables and model parameters are shown in Table 1 for easy reference.

4.1. Observed Populations

4.1.1. Observables

Like most of the BHM methods introduced previously, we work from the summary statistics, where each observed supernova has a brightness measurement $\hat{m}_B$ (which is

Table 1
Model Parameters

Parameter	Description
	Global Parameters
Ω_m	Matter density
w	Dark energy equation of state
α	Stretch standardization
β	Color standardization
$\delta(0)$	Scale of the mass-magnitude correction
$\delta(\infty)/\delta(0)$	Redshift dependence of the mass-magnitude correction
$\delta\mathcal{Z}_i$	Systematics scale
$\langle M_B \rangle$	Mean absolute magnitude
	Survey Parameters
δS	Selection effect deviation
$\langle x_1^i \rangle$	Mean stretch nodes
$\langle c^i \rangle$	Mean color nodes
α_c	Skewness of color population
σ_{M_B}	Population magnitude scatter
σ_{x_1}	Population stretch scatter
σ_c	Population color scatter
κ_0	Extra color dispersion
κ_1	Redshift dependence of extra color dispersion
	Supernova Parameters
m_B	True flux
x_1	True stretch
c	True color
z	True redshift
M_B	Derived absolute magnitude
μ	Derived distance modulus
	Input Data[a]
$\hat{m}_B$	Measured flux
$\hat{x}_1$	Measured stretch
$\hat{c}$	Measured color
C	Covariance on flux, stretch, and color
$\hat{z}$	Observed redshift
$\hat{m}$	Observed mass probability

Note.
[a] Not model parameters but shown for completeness.

analogous to apparent magnitude), stretch $\hat{x}_1$, and color $\hat{c}$, with uncertainty on those values encoded in the covariance matrix C_η. Additionally, each supernova has an observed redshift $\hat{z}$ and a host-galaxy stellar mass associated with it, $\hat{m}$, where the mass measurement is converted into a probability of being above $10^{10}\ M_\odot$. Our set of observables input into Steve is therefore given as $\{\hat{m}_B, \hat{x}_1, \hat{c}, \hat{z}, \hat{m}, C_\eta\}$, as shown in the probabilistic graphical model in Figure 2.

As we are focused on the spectroscopically confirmed supernovae for this iteration of the method, we assume that the observed redshift $\hat{z}$ is the true redshift z such that $P(\hat{z}|z) = \delta(\hat{z} - z)$. Potential sources of redshift error (such as peculiar velocities) are taken into account, not via uncertainty on redshift (which is technically challenging to implement, as varying redshifts introduce computational complexity in computing the distance modulus integral by reducing the amount of precomputation that can be utilized) but instead via uncertainty on distance modulus. Similarly, we take the mass probability estimate $\hat{m}$ as correct and do not model a latent variable to represent uncertainty on the probability estimate. One of the strengths of Steve (and the R15 analysis) is that for future data sets, where supernovae have been classified photometrically and we expect some misclassification and

misidentification of the host galaxies, these misclassifications can naturally be modeled and taken into account by introducing additional populations that supernovae have a nonzero probability of belonging to.

4.1.2. Latent Variables for Observables

The first layer of hierarchy is the set of true (latent) parameters that describe each supernova. In contrast to the observed parameters, the latent parameters are denoted without a hat. For example, c is the true color of the supernova, while $\hat{c}$ is the observed color, which, as it has a measurement error, is different from c.

For the moment, let us consider a single supernova and its classic summary statistics, m_B, x_1, and c. For convenience, let us define $\eta \equiv \{m_B, x_1, c\}$. A full treatment of the summary statistics would involve determining $p(\hat{y}|\eta)$, where $\hat{y}$ represents the observed light-curve fluxes and uncertainties. However, this is computationally prohibitive, as it would require incorporating SALT2 light-curve fitting inside our model fitting. Due to this computational expense, we rely on initially fitting the light-curve observations to produce a best-fit $\hat{\eta}$ along with a 3×3 covariance matrix C_η describing the uncertainty on $\hat{\eta}$. Using this simplification, our latent variables are given by

$$p(\hat{\eta}|\eta) \sim \mathcal{N}(\hat{\eta}|\eta, C_\eta). \tag{6}$$

As discussed in Section 3, the SALT2 model does not include full treatment of intrinsic dispersion; thus, this approximation does not encapsulate the full uncertainty introduced from this dispersion.

4.2. Underlying Population

4.2.1. Type Ia Population

Unlike many previous formalisms that utilize M_B as a singular number and model magnitude scatter on the apparent magnitude m_B, we incorporate this scatter into the underlying rest-frame population by having a population in absolute magnitude space. This is mathematically equivalent; however, it allows us to model the underlying population and intrinsic scatter distinctly. To denote this difference, we refer to the mean of our absolute magnitude population with $\langle M_B \rangle$.

In addition to absolute magnitude, the underlying supernova population is also characterized by distributions in color and stretch. We follow the prior work of R15 and model the color population as an independent redshift-dependent skew-normal distribution for each survey. For the stretch population, we adopt a redshift-dependent normal distribution, and magnitude dispersion is modeled as a normal distribution. We also tested a skew-normal approach for these parameters, reverting to the normal distributions, as they are computationally easier to evaluate; we found no reduction in cosmological bias with the skew-normal distributions for stretch and magnitude. Following R15, we allow the mean color and stretch to vary over redshift, anchoring four equally spaced redshift nodes spanning the redshift range of each survey, linearly interpolating between the nodes to determine the mean stretch and color for a given redshift. These nodes are represented as $\langle x_1^i \rangle$ and $\langle c^i \rangle$. Both the color and stretch means are modeled with normal priors. Initial versions of our model adopted a fully covariant multivariate skew normal (with skewness set to zero only for

the magnitude component) to capture correlations between m_B and c; however, pathological fitting complications required us to simplify our treatment. We parameterize the color skewness α_c by sampling $\delta_c = \alpha_c / \sqrt{1 + \alpha_c^2}$, which itself is given a uniform prior $\mathcal{U}(0, 0.98)$ that allows α_c to span positive values in the realm of physical plausibility, as determined from constraints in Scolnic & Kessler (2016). We sample δ_c in log space for efficiency in sampling close to zero. The width of the population, represented by the vector $\{\sigma_{M_B}, \sigma_{x_1}, \sigma_c\}$, is subject to Cauchy priors with mean zero and width 1, following recommendations from the Stan user guide.

The only constant between survey populations is the absolute magnitude $\langle M_B \rangle$. We model the color skewness and redshift-dependent means and width of the color and skew distributions individually for each survey. The probability for a supernova to have the true values M_B, x_1, and c given the underlying population is thus given as

$$P(M_B, x_1, c, z|\theta) = \mathcal{N}(M_B|\langle M_B \rangle, \sigma_{M_B}) \, \mathcal{N}(x_1|\langle x_1(z) \rangle, \sigma_{x_1})$$
$$\times \, \mathcal{N}^{\text{skew}}(c|\langle c(z) \rangle, \sigma_c, \alpha_c), \qquad (7)$$

where $\theta = \{\langle M_B \rangle, \langle x_1(z) \rangle, \langle c(z) \rangle, \sigma_{m_B}, \sigma_{x_1}, \sigma_c, \alpha_c\}$ for legibility.

4.2.2. Outlier Populations

For the spectroscopic paper, we do not consider outlier populations; however, we ensure that our model has flexibility for such populations for future use with photometrically classified surveys. We thus include a simplistic outlier population model. We follow R15 (and therefore Kunz et al. 2007) by implementing a Gaussian mixture model, where an additional observable of the SN Ia probability would be needed in order to inform the weights of the mixture model. For surveys with low rates of contamination, it is not possible to fit a contamination population, and the mean of the outlier population has been fixed to the SN Ia population in prior works. However, with the increased number of contaminants expected in the DES photometric sample, we seek a more physically motivated outlier population. We find that an acceptable parameterization is to model the outlier population with a mean absolute magnitude of $\langle M_B^{\text{outl}} \rangle = \langle M_B \rangle + \delta_{M_B}^{\text{outl}}$, where $\delta_{M_B}^{\text{outl}}$ is constrained to be positive or even greater than a small positive number to reduce degeneracy between the two populations. We note that this represents the mean brightness of outliers, so outliers could be both brighter and dimmer than the mean SN Ia absolute magnitude. We set the population width to $\sigma^{\text{outl}} = 1$ in M_B, x_1, and c in our tests. The probability of each supernova falling into either population is determined by the observed Type Ia probability $\hat{p}$. For the spectroscopic survey, we set this to unity; thus, it is not included in Figure 2 or Table 1. For the photometric proof of concept, we provide an accurate probability estimate. Further investigation of the effect of inaccurate estimates will be left for future improvements during the analysis of the DES photometric sample.

4.3. Correcting Biased Summary Statistics

With the fitted summary statistics $\hat{\eta}$ being biased and their uncertainty underreported, we face a significant challenge utilizing these statistics naively in supernova cosmology. We must either correct the observables to remove the biases introduced by the intrinsic dispersion of the underlying population or incorporate this dispersion into our model. We should also avoid assuming a specific dispersion model, either the G10 or C11 model, or utilize the results of computing the bias from both models.

We model the extra dispersion only in color, and we do so by adding independent uncertainty on the color observation $\hat{c}$. We note that extra dispersion in magnitude $\hat{m}_B$ (from coherent scatter) is absorbed completely by the width of the underlying magnitude population (discussed in Section 4.2.1) without introducing cosmological bias, which is not true of the color term, hence the requirement for modeling additional color dispersion. Tests of incorporating extra dispersion on stretch also show that stretch is less biased than color and causes negligible bias in cosmology.

As shown in Kessler et al. (2013), the extra color dispersion shows heavy redshift dependence, increasing with redshift. This is an artifact of different filters; however, as we may be subject to similar effects in our observational data, we decide to incorporate redshift dependence in our extra uncertainty. We thus add $\kappa_0 + \kappa_1 z$ to our observed color uncertainty (in quadrature). The κ parameters are highly degenerate with the width of the intrinsic color population σ_c. We subject them to Cauchy priors centered on zero and with width 0.05, where κ is bounded between 0 and 0.05. We pick this maximum value to allow extra dispersion without completely subsuming the intrinsic population widths due to the severe degeneracy, where this maximum value easily encapsulates the determined dispersion according to the results of Kessler et al. (2013). As such, our combined covariance on the observation $\hat{\eta}$ is given by $\mathcal{C}_{\text{tot}} = \mathcal{C}_\eta + \text{DiagMatrix}[0, 0, (\kappa_0 + \kappa_1 z)^2]$.

Fully covariant extra dispersion on $\{m_B, x_1, c\}$ (rather than just dispersion on c) was also tested by modeling the dispersion as a multivariate Gaussian, but it showed negligible improvement in recovering unbiased cosmology over just color dispersion and was far more computationally inefficient. We note here that we model dispersion in magnitude, but this is done at the level of underlying populations, not observed populations. This magnitude dispersion is modeled with redshift independence.

4.4. Mapping Function

4.4.1. Cosmology

We formulate our model with three different cosmological parameterizations: flat ΛCDM, flat wCDM, and standard ΛCDM (without a flatness prior). Here Ω_m is given the prior $\mathcal{U}(0.05, 0.99)$, Ω_Λ was treated with $\mathcal{U}(0, 1.5)$, and the equation of state w was similarly set to a flat prior $\mathcal{U}(-0.4, -2.0)$. For calculating the distance modulus, we fix $H_0 = 70\,\text{km s}^{-1}\,\text{Mpc}^{-1}$. If the Hubble constant has a different value, the absolute magnitude is $M_B + 5\log(H_0/70\,\text{km s}^{-1}\,\text{Mpc}^{-1})$, with the other cosmological parameters unaffected.

4.4.2. Supernova Standardization Parameters

With increasingly large data sets and more nuanced analyses, the choice of how to handle α and β becomes an important consideration when constructing a model. A broken linear relationship is employed by R15 for both color and stretch, where different values of α and β are adopted depending on whether x_1 and c are positive or negative (although the cut could be placed at a location other than zero). Shariff et al. (2016) instead modeled β as redshift-dependent, testing two

phenomenological models: $\beta(z) = \beta_0 + \beta_1 z$ and a second model that effects a rapid but smooth change in β at a turnover redshift z_t.

We tested two models with varying β against simulated supernova sets: $\beta(c) = \beta_0 + \beta_1 c$ and $\beta(z) = \beta_0 + \beta_1 z$. See Section 5.2 for details on simulation generation. We found for both models that nonzero values for β_1 are preferred even with constant β used in simulation, due to severe degeneracy with selection effects. This degeneracy resulted in a significant bias in recovered cosmology. Due to the recovery of nonzero β_1, we continue to adopt the constant α and β found in traditional analyses. As such, our calculation of distance modulus μ mirrors that found in Equation (3).

4.4.3. Host-galaxy Environment

There are numerous results showing statistically significant correlations between host-galaxy environment and supernova properties (Kelly et al. 2010; Lampeitl et al. 2010; Sullivan et al. 2010; D'Andrea et al. 2011; Gupta et al. 2011; Johansson et al. 2013; Rigault et al. 2013). The latest sample of over 1300 spectroscopically confirmed SNe Ia shows $>5\sigma$ evidence for correlation between host mass and luminosity (Uddin et al. 2017). The traditional correction, as employed in analyses such as Suzuki et al. (2012) and Betoule et al. (2014), invokes a step function such that $\Delta M = \gamma \mathcal{H}(\log(M) - 10))$, where $\mathcal{H}$ is the Heaviside step function, M is the galaxy mass in solar masses, and γ represents the size of the magnitude step. The scale of this step function varies from analysis to analysis and is treated as a fit parameter. In this work, we adopt the model used in R15, which follows the work from Rigault et al. (2013), such that we introduce two parameters to incorporate a redshift-dependent host-galaxy mass correction,

$$\Delta M = \delta(0)\left[\frac{1.9\left(1 - \frac{\delta(0)}{\delta(\infty)}\right)}{0.9 + 10^{0.95z}} + \frac{\delta(0)}{\delta(\infty)}\right], \qquad (8)$$

where $\delta(0)$ represents the correction at redshift zero, and $\delta(\infty)$ is a parameter allowing the behavior to change with increasing redshift. We take flat priors on $\delta(0)$ and $\delta(0)/\delta(\infty)$. Finally, we assume that the observed mass probability $\hat{m}$ supplied to the model is perfectly determined and thus set $P(\hat{m}|m) = \delta(\hat{m} - m)$.

4.4.4. Uncertainty Propagation

The chief difficulty with including systematic uncertainties in supernova analyses is that they have difficult-to-model effects on the output observations. As such, the traditional treatment for systematics is to compute their effect on the supernova summary statistics—computing the numerical derivatives $\frac{d\hat{m}_B}{d\mathcal{Z}_i}$, $\frac{d\hat{x}_1}{d\mathcal{Z}_i}$, and $\frac{d\hat{c}}{d\mathcal{Z}_i}$ for each supernova light-curve fit, where $\mathcal{Z}_i$ represents the ith systematic.

Assuming that the gradients can be linearly extrapolated—which is a reasonable approximation for modern surveys with high-quality control of systematics—we can incorporate into our model a deviation from the observed values by constructing a $(3 \times N_{\text{sys}})$ matrix containing the numerical derivatives for the N_{sys} systematics and multiplying it with the row vector containing the offset for each systematic. By scaling the gradient matrix to represent the shift over 1σ of systematic uncertainty, we can simply enforce a unit normal prior on the systematic row vector to increase computational efficiency.

This method of creating a secondary covariance matrix using partial derivatives is used throughout the traditional and BHM analyses. For each survey and band, we have two systematics: the calibration uncertainty and the filter wavelength uncertainty. We include these in our approach, in addition to including the *Hubble Space Telescope* (*HST*) Calspec calibration uncertainty, 10 SALT2 model systematic uncertainties, a dust systematic, a global redshift bias systematic, and the systematic peculiar velocity uncertainty. A comprehensive explanation of all systematics is given in Brout et al. (2019); see Table 4 for details. This gives 13 global systematics shared by all surveys that apply globally to all supernova summary statistics plus two systematics per band in each survey. Systematics with known correlations are shifted together to produce covariant deviations, and we thus assume that the numerical derivatives input into our model represent independent systematics. Full details can be found in Brout et al. (2019). With $\eta \equiv \{m_B, x_1, c\}$, our initial conditional likelihood for our observed summary statistics shown in Equation (6) becomes

$$P\left(\hat{\eta}, \frac{\partial\hat{\eta}}{\partial\mathcal{Z}_i}|\eta, \delta\mathcal{Z}_i, C_\eta\right) = \mathcal{N}\left(\hat{\eta} + \delta\mathcal{Z}_i\frac{\partial\hat{\eta}}{\partial\mathcal{Z}_i}|\eta, C_\eta\right). \qquad (9)$$

4.4.5. Selection Effects

One large difference between traditional and BHM methods is that we treat selection effects by incorporating selection efficiency into our model, rather than relying on simulation-driven data corrections. We describe the probability that the possible events we observe are drawn from the distribution predicted by the underlying theoretical model and that those events, given that they happened, are observed and pass cuts. To make this extra conditional explicit, we can write the likelihood of the data given an underlying model, θ, and that the data are included in our sample, denoted by S, as

$$\mathcal{L}(\theta; \text{data}) = P(\text{data}|\theta, S). \qquad (10)$$

As our model so far describes components of a basic likelihood $P(\text{data}|\theta)$, and we wish to formulate a function $P(S|\text{data}, \theta)$ that describes the chance of an event being successfully observed, we rearrange the likelihood in terms of those functions and find

$$\mathcal{L}(\theta; \text{data}) = \frac{P(S|\text{data}, \theta)P(\text{data}|\theta)}{\int P(S|D, \theta)P(D|\theta)dD}, \qquad (11)$$

where the denominator represents an integral over all potential data D, and θ represents top-level parameters. In the case that our selection effects are best characterized by latent variables instead of data, we can add them to our formulation, and our likelihood becomes

$$\mathcal{L}(\theta; \text{data}) = \frac{\int P(S|L, \theta)P(\text{data}|L)P(L|\theta)dL}{\iint P(S|L, \theta)P(L|\theta)dL}, \qquad (12)$$

where L represents our latent parameters. This is derived in Appendix A.1. To evaluate the effect of our selection effects, we need to evaluate both the selection effect terms in the numerator and the integral in the denominator. The numerator represents the probability that we caught the supernova and it was selected into the cosmology sample. The integral

represents our global selection efficiency at a location in parameter space, rather than the probability of our data being selected into our sample. As θ represents the vector of all top-level model parameters, and L represents a vector of all latent parameters, this is not a trivial integral. Techniques to approximate this integral, such as Monte Carlo integration or high-dimensional Gaussian processes, failed to give tractable posterior surfaces that could be sampled efficiently by a Hamiltonian Monte Carlo, and post-fitting importance sampling failed due to high dimensionality (a brief dismissal of many months of struggle). We therefore simplify the integral and approximate the selection effects from their full expression in all of θ-space to apparent magnitude and redshift space independently (not dependent on x_1 or c), such that the denominator of Equation (12), now denoted d for simplicity, is given as

$$d = \int \left[\int P(S|m_B)P(m_B|z, \theta)dm_B \right] P(S|z)P(z|\theta)dz, \quad (13)$$

where $P(m_B|z, \theta)$ can be expressed by translating the underlying M_B, x_1, and c population to m_B given the cosmological parameters. A full derivation of this can be found in Appendix A.2.

We now apply two further approximations similar to those made in R15: that the redshift distribution of the observed supernovae samples the $P(S|z)P(z|\theta)$ distribution reasonably well, and that the survey color and stretch populations can be treated as Gaussian for the purposes of evaluating $P(m_B|z, \theta)$. We found that discarding the color population skewness entirely resulted in highly biased population recovery (see Figure 12 to see the populations), so we instead characterize the skew-normal color distribution with a Gaussian that follows the mean and variance of a skew normal, with a mean given by $\langle c(z)\rangle + \sqrt{\frac{2}{\pi}}\sigma_c\delta_c$ and variance $\sigma_c^2(1 - 2\delta_c^2/\pi)$. This shifted Gaussian approximation for color completely removes the unintended bias when simply discarding skewness. This shift was not required for the stretch population, so it was left out of the stretch population for numerical reasons. The impact of this approximation on the calculated efficiency is shown in Figure 3, and more detail on this shift and the resulting population recovery can be found in Appendix A.3.

The population $P(m_B|z, \theta)$ becomes $\mathcal{N}(m_B|m_B^*(z), \sigma_{m_B}^*)$, where

$$m_B^*(z) = \langle M_B\rangle + \mu(z) - \alpha\langle x_1(z)\rangle + \beta\langle c(z)\rangle, \quad (14)$$

$$\sigma_{m_B}^{*2} = \sigma_{M_B}^2 + (\alpha\sigma_{x_1})^2 + (\beta\sigma_c)^2. \quad (15)$$

What then remains is determining the functional form of $P(S|m_B)$. For the treatment of most surveys, we find that the error function, which smoothly transitions from some constant efficiency down to zero, is sufficient. Formally, this gives

$$P(S|m_B) = \Phi^c(m_B|\mu_{CDF}, \sigma_{CDF}), \quad (16)$$

where Φ^c is the complementary cumulative distribution function and μ_{CDF} and σ_{CDF} specify the selection function. The appropriateness of an error function has been found by many past surveys (Dilday et al. 2008; Barbary et al. 2010; Perrett et al. 2012; Graur et al. 2013; Rodney et al. 2014). However, for surveys that suffer from saturation and thus rejection of low-redshift supernovae, or for groups of surveys
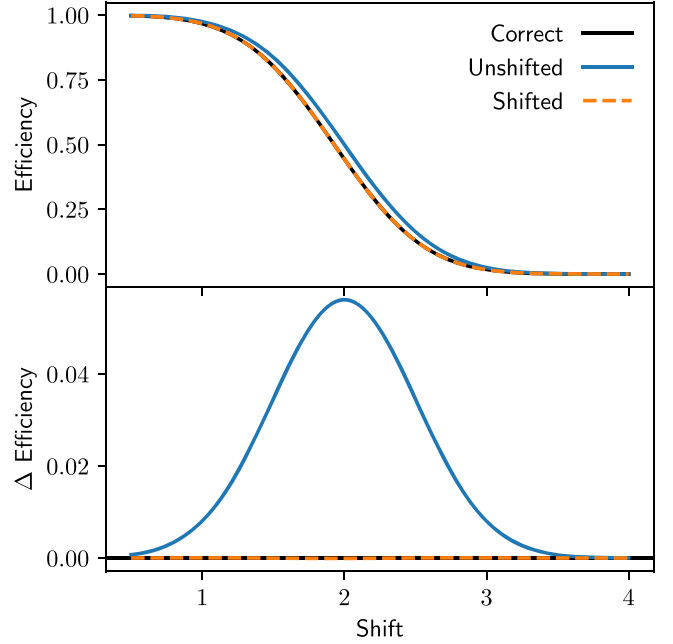

Figure 3. Testing the correctness of our normal approximation to the skewed color distribution. The "correct" line (shown in black) represents the exact integral $w = \int P(S|x)P(x)dx$, where $P(S|x)$ is an error function (following our high-redshift surveys) and $P(x) = \mathcal{N}^{Skew}(x, 0.1, 2)$, calculated numerically. The x-axis is analogous to m_B in cosmological context. As expected, all efficiencies drop toward zero as the shift increases (as objects get fainter). The unshifted normal approximation shows a significant discrepancy in the calculated efficiency as it transitions from 1 to zero, while the shifted normal approximation shows a negligible error to the correct solution. From these plots, further refinement of the normal approximation (such as including kurtosis or higher powers) is unnecessary.

treated together (as is common with low-redshift surveys), we find that a skew normal is a better analytic form, taking the form

$$P(S|m_B) = \mathcal{N}^{Skew}(m_B|\mu_{Skew}, \sigma_{Skew}, \alpha_{Skew}). \quad (17)$$

The selection functions are fit to apparent magnitude efficiency ratios calculated from SNANA simulations by taking the ratio of supernovae that are observed and passed cuts over the total number of supernovae generated in that apparent magnitude bin. That is, we calculate the probability that we would include a particular supernova in our sample, divided by the number of such supernovae in our simulated fields. To take into account the uncertainty introduced by the imperfection of our analytic fit to the efficiency ratio, uncertainty was uniformly added in quadrature to the efficiency ratio data from our simulations until the reduced χ^2 of the analytic fit reached 1, allowing us to extract an uncertainty covariance matrix for our analytic fits to either the error function or the skew normal. This is mathematically identical to fitting the efficiency ratio with a second "intrinsic dispersion" parameter, which adds uncertainty to the efficiency ratio data points.

We thus include parameterized selection effects in our model by including the covariance matrix of selection effect uncertainty. Formally, we include deviations from the determined mean selection function parameters with parameter vector δS and apply a normal prior on this parameter as per the determined uncertainty covariance matrix. While this uncertainty encapsulates the potential error from the simulations not matching the analytic approximations, it does not cover
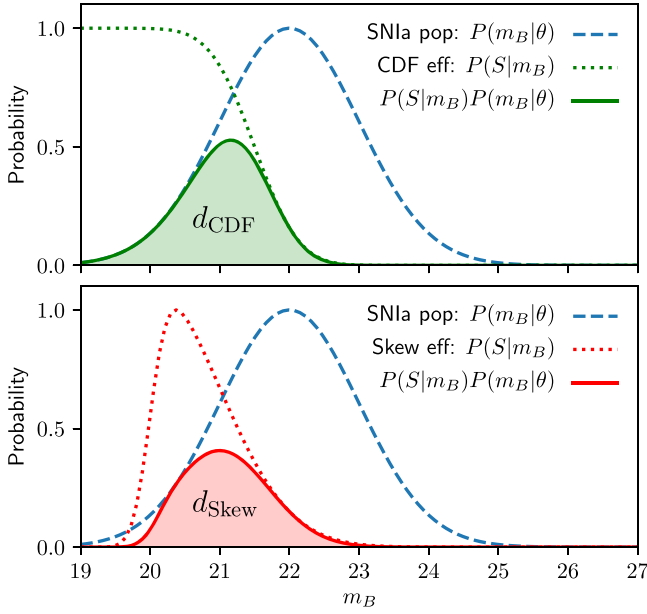
Figure 4. Efficiency of supernova discovery at an arbitrary redshift. Shown in both panels with a blue dashed line is the SN Ia population distribution, which takes the form of a normal distribution. The top panel shows a CDF-based survey efficiency (green dotted line), while the bottom panel shows a skew normal–based survey efficiency (red dotted line), as functions of apparent magnitude. The survey efficiency, given the SN Ia population, is shown as a solid line in both panels, and the probability of observing a SN Ia is found by integrating over the population detection efficiency as described in Equation (13) and is shown by shading the area integrated. This area is what is analytically given by Equations (18) and (19).

potential variations of the selection function at the top level: varying cosmology or spectroscopic efficiency. Tests with changing the intrinsic scatter model used in the selection efficiency simulations show that the uncertainty introduced is negligible.

With the well-sampled redshift approximation, we can remove the redshift integral in Equation (13) and replace it with a correction for each observed supernova. For the error function (denoted with the subscript "CDF") and skew-normal selection functions (denoted with a subscript "Skew"), respectively, the correction per SN Ia becomes

$$d_{\mathrm{CDF}} = \Phi^c\left(\frac{m_B^* - \mu_{\mathrm{CDF}}}{\sqrt{\sigma_{m_B}^{*2} + \sigma_{\mathrm{CDF}}^2}}\right), \tag{18}$$

$$d_{\mathrm{Skew}} = 2\mathcal{N}\left(\frac{m_B^* - \mu_{\mathrm{Skew}}}{\sqrt{\sigma_{m_B}^{*2} + \sigma_{\mathrm{Skew}}^2}}\right)$$

$$\times \Phi\left(\frac{\mathrm{sign}(\alpha_{\mathrm{Skew}})(m_B^* - \mu_{\mathrm{Skew}})}{\frac{\sigma_{m_B}^{*2} + \sigma_{\mathrm{Skew}}^2}{\sigma_{\mathrm{Skew}}^2}\sqrt{\frac{\sigma_{\mathrm{Skew}}^2}{\alpha_{\mathrm{Skew}}^2} + \frac{\sigma_{m_B}^{*2}\sigma_{\mathrm{Skew}}^2}{\sigma_{m_B}^{*2} + \sigma_{\mathrm{Skew}}^2}}}\right), \tag{19}$$

and is incorporated into our likelihood. Note that the above efficiencies utilize the common form of the normal distribution rather than the conditional probability notation found previously in this work. This is illustrated in Figure 4. Our corrections for the DES spectroscopic data utilize the CDF functional form, with the combined low-redshift surveys being

modeled with the skew-normal efficiency. Further details on this choice are given in Section 5.2.

4.5. Model Summary

Having laid out each individual aspect of the model, the relationships between variables, and our treatment of uncertainty, here we summarize the relationships in our model mathematically. In this summary of $P(\mathrm{data}|\theta)$, we leave out sample selection for simplicity. Referring to Equation (12), the relationships with $P(\mathrm{data}|\theta)$ are as follows:

$$P(\mathrm{data}|\theta)$$
$$\propto \int d\mu \ dm_B \ dx_1 \ dc \ dz, \ dm \ P(\hat{z}|z)P(\hat{m}|m)$$
$$\times P(\hat{m}_B, \hat{x}_1, \hat{c}, \hat{\mathcal{C}}_\eta^{-1}|m_B, x_1, c, \delta\mathcal{Z}_i, \kappa_0, \kappa_1)$$
$$\times P(x_1, c|\langle x_1^i\rangle, \langle c^i\rangle, \sigma_{x_1}, \sigma_c, z)$$
$$\times P(m_B|\langle M_B\rangle, \mu, \delta(0), \delta(\infty), \sigma_{M_B}, \sigma_{x_1}, \sigma_c, \alpha, \beta, z, m)$$
$$\times P(\mu|z, \Omega_m, w). \tag{20}$$

The denominator of Equation (12) is then given by either Equation (18) or (19), depending on the survey; similarly, $P(S|\mathrm{data}, \theta)$ is given by Equation (17) or (16), respectively. Combining all of these gives us our full model likelihood with selection effects accounted for.

5. Model Verification

In order to verify our model, we run it through stringent tests. First, we validate on toy models, verifying that we recover accurate cosmology when generating toy supernova data constructed to satisfy the assumptions of the BHM construction. We then validate our model on SNANA simulations based on a collection of low-redshift surveys and the DES 3 yr spectroscopic sample, termed the DES-SN3YR sample.

5.1. Applied to Toy Spectroscopic Data

We generate simple toy data to validate the basic premise of the model. The data generation algorithm is described below.

1. Draw a redshift from a power-law distribution. For the low-redshift survey, this is $\mathcal{U}(0.0004, 0.01)^{0.5}$, and for the DES-like survey, this is $\mathcal{U}(0.008, 1.0)^{0.33}$. For the low-redshift survey, this is equivalent to sampling $y = z$ from 0.02 to $\sqrt{0.1}$, and for the high-redshift survey, this is equivalent to sampling $y = z^2$ from 0.2 to 1.0. These distributions are arbitrary, and this test has been performed with various flat and power-law distributions.
2. Draw a random mass probability from $\mathcal{U}(0, 1)$ and calculate the mass-brightness correction using $\delta(0) = 0.08$, $\delta(0)/\delta(\infty) = 0.5$, and Equation (8).
3. Draw an absolute magnitude, stretch, and color from the respective distributions $\mathcal{N}(-19.3, 0.1)$, $\mathcal{N}(0, 1)$, and $\mathcal{N}(0, 0.1)$.
4. Calculate $\mu(z)$ given the drawn redshift and cosmological parameters $\Omega_m = 0.3$, $w = -1$ under flat ΛCDM cosmology. Use this to determine the true apparent magnitude of the object m_B using $m_B = \mu + M_B - \alpha x_1 + \beta c$.
5. Determine if the SN Ia is detected using the detection probability $P(S|m_B) = \mathcal{N}^{\mathrm{skew}}(13.72, 1.35, 5.87)$ for the low-redshift survey (numeric values obtained by fitting to existing low-redshift data). For the DES-like survey,
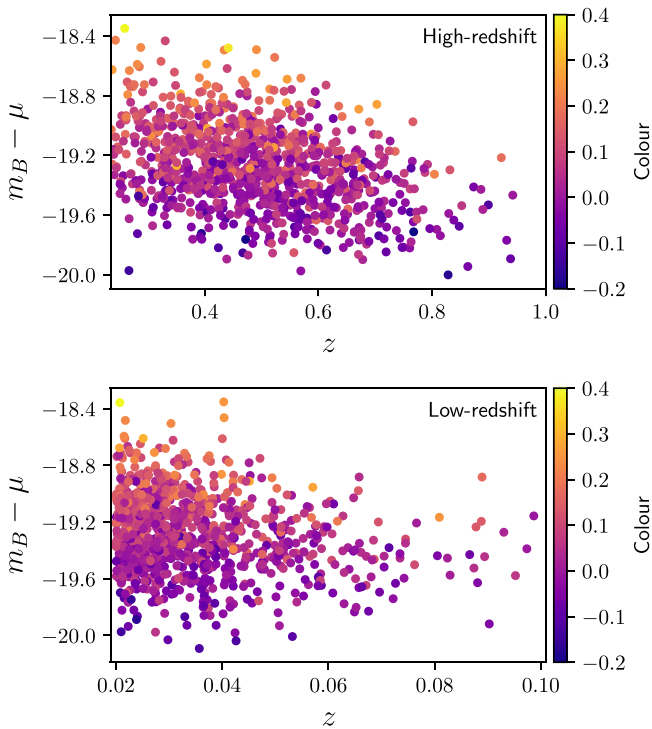
Figure 5. Population distributions shown in redshift and uncorrected absolute magnitude $m_B - \mu$ for 1000 supernovae in both high- and low-redshift surveys. Selection effects are visible in both samples, where red supernovae are often cut as redshift increases, creating a skewed color population. The color of the data points is representative of the supernova color itself, with a negative color value showing bluer supernovae and positive color values representing redder supernovae.

accept with probability $P(S|m_B) = \Phi^C(23.14, 0.5)$. Repeat from step 1 until we have a supernova that passes. We use realistic values for the selection probability to ensure that our model is numerically stable with highly skewed selection functions.

6. Add independent, Gaussian observational error onto the true m_B, x_1, and c using Gaussian widths of 0.04, 0.2, and 0.03, respectively (following the mean uncertainty for DES-like SNANA simulations). Add extra color uncertainty in quadrature of $\kappa_0 + \kappa_1 z$, where $\kappa_0 = \kappa_1 = 0.03$.

The selection function parameters (a skew normal for low-redshift and a complementary error function for high-redshift) are all given an independent uncertainty of 0.01 (mean and width for the CDF selection function and mean, width, and skewness for the skew-normal selection function). Draw from each survey simulation until we have 1000 low-z and 1000 DES-like supernovae, representing a statistical sample of greater power than the estimated 350 supernovae for the DES-SN3YR sample. Sample data for 1000 high- and low-redshift supernovae are shown in Figure 5, confirming the presence of strong selection effects in both toy surveys, as designed.

We test four models: flat ΛCDM, flat wCDM, ΛCDM, and flat wCDM with a prior $\Omega_m \sim \mathcal{N}(0.3, 0.01)$, with the latter included to allow sensitive tests on bias for w. To achieve statistical precision, we fit 100 realizations of supernova data sets. Cosmological parameters are recovered without significant bias. Combined posterior surfaces of all 100 realization fits for ΛCDM are shown in Figure 6, and fits for flat wCDM are shown in Figure 7. By utilizing the Stan framework and several
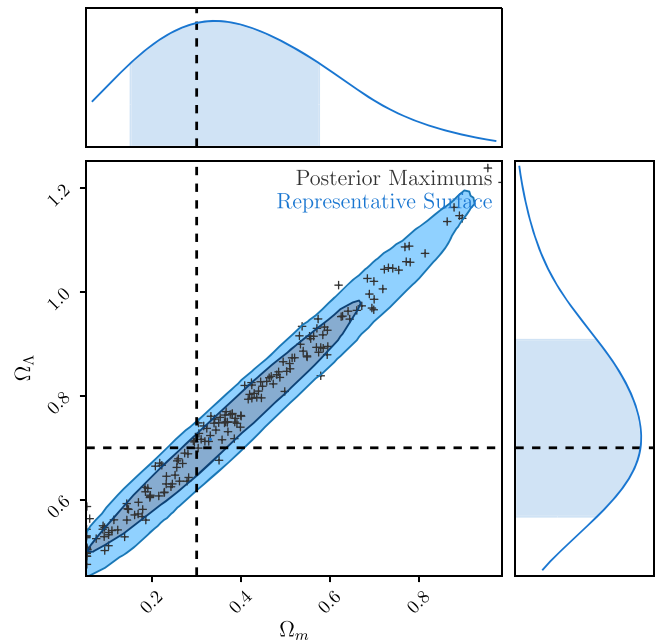
Figure 6. Maximal posterior points for 100 realizations of supernova data with the flat ΛCDM model, with a representative contour from a single data realization shown for context. Even a large supernova sample, when treated robustly, is insufficient to provide tight constraints on either Ω_m or Ω_Λ separately due to the severe degeneracy between the parameters.
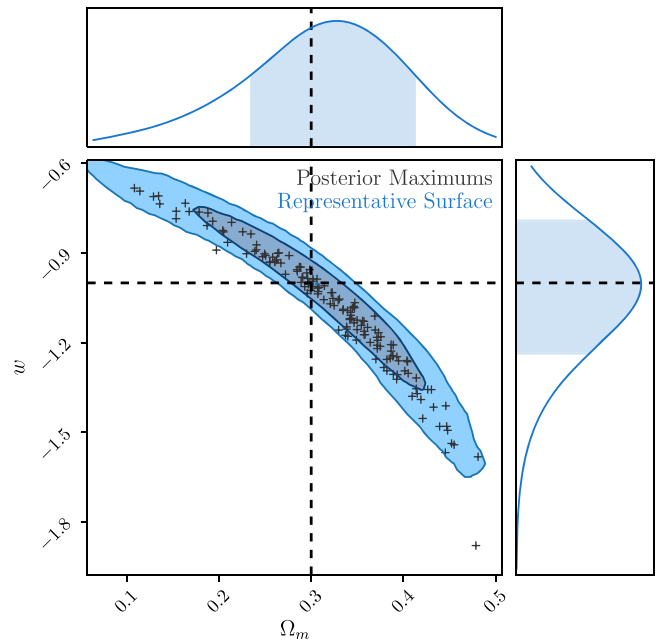
Figure 7. Maximal posterior points for 100 realizations of supernova data with the flat wCDM model, with a representative contour from a single data realization shown for context. The well-known banana-shaped contour is recovered, with the marginalized distributions in Ω_m and w shown to recover input cosmology. For contours that are non-Gaussian due to the curved degeneracy between Ω_m and w, the marginalized distributions can provide misleading statistics where maximum marginalized distribution can disagree with the maximum likelihood in multiple dimensions. For our contours, the non-Gaussianity is small, and the marginalized distributions still provide a valuable metric. The recovered posterior maxima show the same degeneracy direction as the representative surface and scatter around the truth values input into the simulation, which are shown with dashed lines.

Table 2
Cosmological Parameters for Toy Supernova Data

Model	$\Omega_m \langle\mu\rangle, \langle\sigma\rangle$ (scatter)	$w \langle\mu\rangle, \langle\sigma\rangle$ (scatter)
Flat ΛCDM	0.301, 0.015(0.012)	…
Flat wCDM	…	-1.00, 0.042(0.030)

Note. Cosmological parameters determined from the surfaces of 100 fits to independent realizations of toy supernova data. As described in the main text, each data set comprised 1000 low-redshift and 1000 high-redshift supernovae. For each chain, we record the mean and standard deviation and then show the average mean and average standard deviation in the table. The scatter introduced by simulation variance (the standard deviation of the 100 mean parameter values) is shown in parentheses. Model bias would appear as shifts away from the simulation values of $\Omega_m = 0.3$, $w = -1$.

Table 3
Supernova Population Distributions

Model	$\langle x_1 \rangle$	σ_{x_1}	$\langle c \rangle$	σ_c
SK16 low-z	0.55 and -1.5	$^{+0.45}_{-1.0}$ and $^{+0.5}_{-0.5}$	-0.055	$^{+0.15}_{-0.023}$
SK16 DES	0.973	$^{+0.222}_{1.472}$	-0.054	$^{+0.101}_{0.043}$

Note. The SK16 low-z stretch distribution is formed as the sum of two bifurcated Gaussians, with the mean and spread of each component given, respectively.

efficient parameterizations (discussed further in Appendix B), fits to these simulations of 2000 supernovae take only on order of a single CPU hour to run.

To investigate biases in the model in fine detail, we look for systematic bias in Ω_m in the flat ΛCDM cosmology test and bias in w for the flat wCDM test with strong prior $\Omega_m \sim \mathcal{N}(0.3, 0.01)$. This allows us to investigate biases without the investigative hindrances of non-Gaussian or truncated posterior surfaces. The strong prior on Ω_m cuts a slice through the traditional "banana" posterior surface in the w–Ω_m plane of Figure 7. Without making such a slice, the variation in w is larger due to a shift along the degeneracy direction of the "banana." By focusing the slice at an almost fixed Ω_m, we can see the variation in the mean value of w approximately perpendicular to the lines of degeneracy, instead of along them. The results of the analysis are detailed in Table 2 and demonstrate the performance of our model in recovering the true cosmological parameters. As we are using 100 independent realizations, the precision of our determination of the mean simulation result is approximately a tenth of the quoted scatter (as a degree of non-Gaussianity of our fits will make this relationship inexact). The deviation from truth values is below this threshold, and no significant bias is detected in either the flat ΛCDM or flat wCDM models. With these simple data, we also correctly recover underlying supernova populations, which can be seen in Figure 12.

5.2. DES Supernova Data Validation

Many BHM methods have previously been validated on data constructed explicitly to validate the assumptions of the model. This is a useful consistency check that the model implementation is correct, efficient, and free of obvious pathologies. However, the real test of a model is its application to realistic data sets that mimic expected observational data in as many ways as possible. To this end, we test using simulations (using the SNANA package) that follow the observational schedule and observing conditions for the DES and low-z surveys, where the low-z sample is based on observations from CfA3 (Hicken et al. 2009a, 2009b), CfA4 (Hicken et al. 2012), and CSP (Contreras et al. 2010; Folatelli et al. 2010; Stritzinger et al. 2011). Simulation specifics can be found in Kessler et al. (2019). The primary differences from the toy data of the previous section are the different underlying color and stretch, inclusion of spectroscopic data and light-curve cuts, and inclusion of intrinsic dispersion models.

Prior analyses often treated intrinsic dispersion simply as scatter in the underlying absolute magnitude of the underlying population (Conley et al. 2011; Betoule et al. 2014), but recent analyses require a more sophisticated approach. In our development of this model and tests of intrinsic dispersion, we analyze the effects of two different scatter models via simulations: the G10 and C11 models described in Section 3. The G10 model dispersion has 70% contribution from coherent variation and 30% from chromatic variation, while the C11 model has 25% coherent scatter and 75% from chromatic variation. These two broadband scatter models are converted to spectral energy distribution models for use in simulations in Kessler et al. (2013).

In addition to the improvements in testing multiple scatter models, we also include peculiar velocities for the low-z sample and our full treatment of systematics as detailed in Brout et al. (2019). Our simulated populations are sourced from Scolnic & Kessler (2016, hereafter SK16) and shown in Table 3. Initial tests were also done with a second, Gaussian population with color and stretch populations centered on zero and respective widths of 0.1 and 1; however, the cosmological parameters were not impacted by the choice of the underlying population, and we continue using only the SK16 population for computational efficiency. The selection effects were quantified by comparing all of the generated supernovae to those that pass our cuts, as shown in Figure 8. It is from this simulation that our analytic determination of the selection functions for the low-z and DES survey are based. We run two simulations to determine the efficiency using the G10 and C11 scatter models and find no difference in the functional form of the Malmquist bias between the two models. Uncertainty on the analytic selection function is incorporated into our fits, mitigating the imperfection of our analytic form by allowing it to vary in our fits.

Each realization of simulated SN Ia light curves contains the SALT2 light-curve fits and redshifts to 128 low-z and 204 DES-like supernovae, such that the uncertainties found when combining chains is representative of the uncertainty in the DES-SN3YR sample. As our primary focus is dark energy, we now focus specifically on the flat wCDM model with matter prior.

Points of maximum posterior for 100 data realizations are shown in Figure 9. The parameter bounds and biases for w are listed in Table 5, and secondary parameters are shown in Table 4.

Table 5 shows that the G10 model is consistent with $w = -1$, while the C11 model shows evidence of bias on w, scattering high. However, their deviation from the truth value represents a shift of approximately 0.5σ when taking into account the uncertainty on fits to w. The bias is subdominant to both the size of the uncertainty for each fit and the scatter
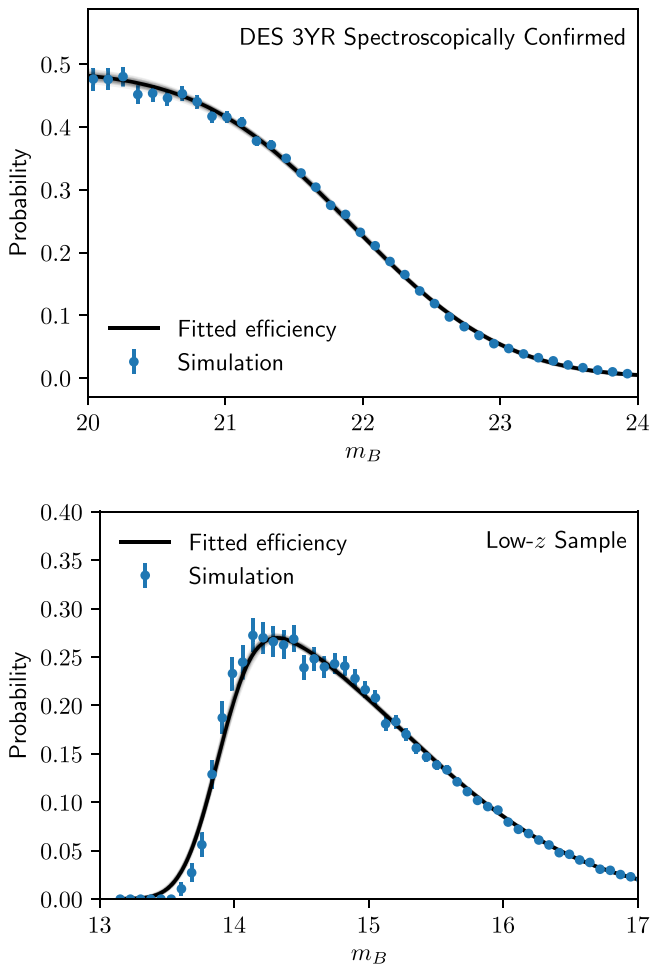
Figure 8. Fitting the selection function for both the DES-SN3YR spectroscopically confirmed supernova sample and the low-z sample. Blue error bars represent the efficiency calculated by determining the ratio of discovered to generated supernovae in apparent magnitude bins for SNANA simulations. The black line represents the best-fit analytic function for each sample, and the light gray lines surrounding the best-fit value represent random realizations of analytic function taking into account uncertainty on the best-fit value.

Figure 9. Maximum posterior points for 100 realizations of supernova data for two intrinsic dispersion models: the G10 model for the top panel and the C11 model for the bottom panel. Points are shown for parameters Ω_m, w, α, and β, with the other fit parameters being marginalized over. As we are unable to fully correct observed summary statistics, a step required by the lack of intrinsic scatter in the SALT2 model, we expect to see an offset in α and β. This, in turn, affects cosmology, resulting in small biases in w.

induced by statistical variance in the simulations. We also note that the simulations do not vary cosmological parameters or population. As our model does include uncertainty on those values, the simulation scatter is expected to be less than the model uncertainty and represents a minimum bound on permissible uncertainty values.

Table 4 shows a clear difference in both β and σ_{m_B} across the G10 and C11 simulations. As expected, the C11 simulations recover a far smaller intrinsic magnitude scatter, giving a result of approximately 0.025 when compared to the result of 0.070 for the G10 simulations. The extra smearing of the C11 model does not result in a significantly biased β value compared to the average uncertainty on β, with recovery of $\beta \approx 3.76$ close to the input truth value of 3.8; however, the β recovery for the G10 simulations is biased high, finding $\beta \approx 3.44$ with an input of 3.1. Interestingly, w bias is only found for the C11 simulations. A measure of the significance of the parameter bias can be calculated by comparing the bias to a tenth of the scatter (as our Monte Carlo estimate uncertainty is $\sqrt{100}$ of the scatter). From this, we can see that most biases are detected with high statistical significance due to the large number of simulations tested against.

We investigate the cosmological bias and find its source to be a bias in the observed summary statistics (i.e., the $\hat{m}_B$, $\hat{x}_1$, and $\hat{c}$ output from SALT2 light-curve fitting), in addition to an incorrect reported uncertainty on the summary statistics. To confirm this, we run two tests. In the first, we replace the SALT2-fitted $\hat{m}_B$, $\hat{x}_1$, and $\hat{c}$ with random numbers drawn from a Gaussian centered on the true SALT2 m_B, x_1, and c values with a covariance as reported by initial light-curve fits. With this test, both the G10 and C11 fits recover $w = -1.00$ exactly. Our second test aims to test our model while allowing biases in the summary statistics not caused by intrinsic scatter through.

Table 4
Realistic Simulation Standardization Parameters

Model	$\alpha - \alpha_{\text{True}}$	$\beta - \beta_{\text{True}}$	$\langle M_B \rangle - \langle M_B \rangle_{\text{True}}$	$\sigma_{m_B}^{\text{DES}}$	$\sigma_{m_B}^{\text{low}-z}$
G10 Stat + Syst	0.022 [0.009 (0.008)]	0.34 [0.19 (0.18)]	−0.002 [0.028 (0.015)]	0.070 [0.022 (0.018)]	0.073 [0.025 (0.022)]
G10 Stat	0.000 [0.008 (0.008)]	0.33 [0.20 (0.17)]	0.001 [0.016 (0.013)]	0.069 [0.023 (0.019)]	0.072 [0.026 (0.023)]
C11 Stat + Syst	0.002 [0.009 (0.007)]	−0.04 [0.15 (0.13)]	0.014 [0.030 (0.018)]	0.024 [0.016 (0.011)]	0.029 [0.020 (0.014)]
C11 Stat	0.000 [0.008 (0.007)]	−0.05 [0.16 (0.13)]	0.006 [0.016 (0.015)]	0.025 [0.016 (0.012)]	0.027 [0.020 (0.015)]

Note. Standardization parameters and base intrinsic scatter parameter results for the 100 fits to G10 and C11 simulations. We show the average parameter mean and average standard deviation, respectively, with the simulation scatter shown in brackets, such that each cell shows $\langle \mu \rangle [\langle \sigma \rangle (\text{scatter})]$. The width of the intrinsic scatter (σ_{m_B}) does not have an input truth value, as it is determined from the scatter model.

Table 5
Realistic Simulation Determination of w

Model	$w \ \langle \mu \rangle [\langle \sigma_w \rangle (\text{scatter})]$	w Bias
G10 Stat + Syst	−0.998 [0.097 (0.073)]	$(0.02 \pm 0.07)\sigma$
G10 Stat	−1.008 [0.080 (0.068)]	$(−0.10 \pm 0.08)\sigma$
C11 Stat + Syst	−0.945 [0.098 (0.077)]	$(0.55 \pm 0.08)\sigma$
C11 Stat	−0.948 [0.079 (0.066)]	$(0.65 \pm 0.08)\sigma$

Note. Investigating the combined 100 fits to the G10 and C11 simulations, fitting with both statistics only and when including systematics. The quoted value for w represents the average mean of the fits, with the average uncertainty being shown in brackets and the simulation scatter (the standard deviation of the mean of 100 fits) shown in parentheses. The bias significance represents our confidence that the deviation in the mean w away from −1 is not due to statistical fluctuation.

That is, the first test ascertained that biases in the summary statistics are the cause of cosmological bias. It is thus important to determine the source of those biases, whether they are from the intrinsic scatter model or another aspect of the simulation. To this end, we test a set of 100 simulations generated using an intrinsic dispersion model of only coherent magnitude scatter. We find $w = -1.00$, showing that the source of the biases in the summary statistics is the underlying intrinsic scatter model. From this, the main challenge of improving our methodology is to handle the fact that the observational uncertainty reported from fitting the SALT2 model to light curves is incorrect, non-Gaussian, and biased. Our current model and techniques can quantify the effect of different scatter models on biasing the observed summary statistics, but being unable to constrain the "correct" (simulated) scatter model in our model fit means we cannot fully correct for the bias introduced by an unknown scatter model.

Unfortunately, adding extra fit parameters to allow for shifting observables washes out our ability to constrain cosmology, and applying a specific bias correction requires running a fiducial simulation (assuming cosmology, population, and scatter model), which presents difficulties when trying to account for correlations with population and scatter model. This is compounded by the fact that bias corrections do not, in general, improve fits (increase the log posterior) and so are difficult to fit parametrically. Works such as Kessler & Scolnic (2017) show that bias corrections can be applied to supernova data sets that can robustly handle multiple intrinsic scatter models, and future work will center on uniting these methodologies, incorporating better bias corrections that separate intrinsic scatter bias and non-Gaussian summary

statistic bias from Malmquist bias without having to pre-compute standardization parameters and populations.

Difficulty in providing an adequate parameterization for realistic intrinsic dispersion and the simplification of Malmquist bias to only apparent magnitude also lead to biases in the population parameters. To determine if the underlying population mischaracterization was a cause of cosmological bias, we ran fits wherein the underlying population was fixed to the known distributions used for the simulation. These fits did not change the bias in the C11 simulation. We conclude that the biased population recovery is not the cause of cosmological bias. As the population parameters recovered using the simplistic toy supernova data in the previous section do not exhibit significant bias, future work will focus on intrinsic dispersion and Malmquist bias rather than alternate parameterizations of the underlying supernova population.

Table 6 lists the fit correlations between our model fit parameters (excluding the low-z band systematics and Malmquist bias uncertainty parameters that had negligible correlation), showing (in order) cosmological parameters, standardization parameters, population width and skewness parameters, intrinsic dispersion parameters, mass-step parameters, population mean parameters, SALT2 model systematics, dust systematics, global *HST* calibration systematics, peculiar velocity systematics, global redshift systematics, and DES band magnitude and wavelength systematics. Figure 10 shows the full correlations between all nonsystematic model parameters. Other interesting correlations are shown and discussed in Figure 10. The band systematics for DES filters *g*, *r*, and *i* also show a significant correlation with *w*, highlighting the importance of minimizing instrumental uncertainty.

For the sample size of the DES + low-z supernova samples (332 supernovae), the bias from intrinsic scatter models is subdominant to the statistical uncertainty, as shown in Figure 9. For our full systematics model, the bias represents a deviation between 0σ and 0.5σ, depending on scatter model, and given that they remain subdominant, we will leave more complicated treatment of them for future work.

5.3. Uncertainty Analysis

With the increased flexibility of BHMs over traditional models, we expect to find an increased uncertainty on parameter inference. This increased uncertainty is one of the strengths of hierarchical models, as it represents a more thorough accounting of model uncertainty. To characterize the influence of the extra degrees of freedom in our model, we analyze the uncertainty on w averaged across 10 nominal simulations of the DES-SN3YR sample with various model

Table 6
Reduced Parameter Correlations with w

Parameter	G10 Stat+Syst	C11 Stat+Syst
Ω_m	−0.19	−0.21
α	−0.17	−0.20
β	−0.29	−0.23
$\langle M_B \rangle$	0.68	0.66
$\delta(0)$	0.00	0.00
$\delta(\infty)/\delta(0)$	0.00	0.00
$\sigma_{\mathrm{m}B}^0$	0.04	0.07
$\sigma_{\mathrm{m}B}^1$	0.23	0.18
$\sigma_{x_1}^0$	0.04	0.03
$\sigma_{x_1}^1$	0.05	0.01
σ_c^0	0.01	0.11
σ_c^1	0.08	0.04
α_c^0	−0.04	0.04
α_c^1	0.03	0.01
κ_{c0}^0	−0.10	−0.05
κ_{c0}^1	−0.20	−0.17
κ_{c1}^0	−0.05	−0.01
κ_{c1}^1	−0.01	0.01
$\langle x_1^0 \rangle$	−0.01	−0.05
$\langle x_1^1 \rangle$	−0.02	0.02
$\langle x_1^2 \rangle$	−0.04	−0.04
$\langle x_1^3 \rangle$	−0.03	−0.06
$\langle x_1^4 \rangle$	−0.06	−0.06
$\langle x_1^5 \rangle$	0.04	0.02
$\langle x_1^6 \rangle$	0.04	0.04
$\langle x_1^7 \rangle$	0.08	0.03
$\langle c^0 \rangle$	−0.05	−0.12
$\langle c^1 \rangle$	0.11	0.03
$\langle c^2 \rangle$	0.11	0.06
$\langle c^3 \rangle$	0.14	0.04
$\langle c^4 \rangle$	−0.11	−0.11
$\langle c^5 \rangle$	−0.15	−0.08
$\langle c^6 \rangle$	−0.12	−0.13
$\langle c^7 \rangle$	−0.12	−0.06
$\delta[\mathrm{SALT}_0]$	0.05	0.05
$\delta[\mathrm{SALT}_1]$	−0.01	0.02
$\delta[\mathrm{SALT}_2]$	−0.10	−0.09
$\delta[\mathrm{SALT}_3]$	−0.03	−0.03
$\delta[\mathrm{SALT}_4]$	0.08	0.09
$\delta[\mathrm{SALT}_5]$	0.01	0.02
$\delta[\mathrm{SALT}_6]$	0.05	0.07
$\delta[\mathrm{SALT}_7]$	−0.11	−0.10
$\delta[\mathrm{SALT}_8]$	0.01	0.02
$\delta[\mathrm{SALT}_9]$	0.02	0.02
$\delta[\mathrm{MWE}_{B-V}]$	0.03	0.02
$\delta[HST\ \mathrm{Calib}]$	−0.07	−0.07
$\delta[v_{\mathrm{pec}}]$	0.00	−0.01
$\delta[\delta z]$	0.01	0.00
$\delta[\Delta g]$	0.05	0.11
$\delta[\Delta r]$	0.16	0.10
$\delta[\Delta i]$	−0.16	−0.18
$\delta[\Delta z]$	−0.26	−0.26
$\delta[\Delta \lambda_g]$	0.16	0.20
$\delta[\Delta \lambda_r]$	0.05	0.06
$\delta[\Delta \lambda_i]$	0.00	−0.01
$\delta[\Delta \lambda_z]$	0.09	0.07

Note. Correlations determined from the combined 100 simulation fits. Correlations for the low-z band systematics and the latent parameters representing selection function uncertainty are not shown but have negligible correlation. Zero superscripts indicate the DES, and a superscript 1 represents the low-z survey.

parameters allowed to either vary or stay locked to a fixed value. By taking the difference in uncertainty in quadrature, we can infer the relative contribution for each model feature to the uncertainty error budget.

The error budget detailed in Table 7 shows that our uncertainty is dominated by statistical error, as the total statistical uncertainty on w is ±0.08. With the low number of supernovae in the DES-SN3YR sample, this is expected. We note that the label "Systematics" in Table 7 represents all numerically computed systematics (as discussed in Section 4.4.4) and systematic uncertainty on the selection function.

5.4. Methodology Comparison

We compare the results of our model against those of the BBC+CosmoMC method (Kessler & Scolnic 2017). This method has been used in prior analyses, such as the Pantheon sample analysis of Scolnic et al. (2018), and is being used in the primary analysis of the DES-SN3YR sample (DES Collaboration et al. 2018). The BBC method is a two-part process: BBC computes bias corrections for observables, and then the corrected distances are fit using CosmoMC (Lewis & Bridle 2002). For shorthand, we refer to this combined process as the BBC method hereafter in this paper, as we are concerned with the results of cosmological parameter inference. As a leading supernova cosmology method, it provides a good consistency check as to the current levels of accuracy in recovering cosmological parameters.

To this end, we take the results of the BBC method that were also run on the same set of 200 validation simulations and compare the recovered w values to those of our method. The results are detailed in Table 8, and a scatter plot of the simulation results is presented in Figure 11.

As shown in Brout et al. (2019), the BBC method recovers cosmological parameters without bias so long as the intrinsic scatter model is known. As we do not know the correct intrinsic scatter model, the BBC method averages the results when using bias corrections from G10 and C11. As such, we expect the BBC method to have a w bias in one direction for G10 simulations and the other direction for C11 simulations. These results are consistent with those displayed in Table 8. Both the BBC method and Steve are sensitive to the intrinsic scatter model, finding differences of ∼0.066 and 0.053, respectively, in w when varying the scatter model. The BBC method finds w biased low for G10 and high for C11 (by about ±0.03), so taking the average result only results in a small bias of −0.01 in w. Our method shows a small improvement in the insensitivity to the intrinsic scatter model (having a decrease in difference in w between the G10 and C11 models), finding no bias for G10 but a w biased high for C11. This decrease in error is not statistically significant, as we have statistical uncertainty of ∼0.01 for 100 simulation realizations. The average bias over the two scatter models is +0.028, representing a larger bias than the BBC method.

When comparing both the G10 and C11 sets of simulations independently, our model differs from BBC in its average prediction of w by +0.044 and +0.033, respectively. For the G10 model, this difference is a result of bias in the BBC results; however, for the C11 simulations, this is a result of both bias from BBC and a larger bias from our method. These results also allow us to state the expected values for w when run on the DES-SN3YR sample. When using Planck priors, our uncertainty on w is reduced compared to using our simulation Gaussian prior on Ω_m, shrinking the average w difference from 0.06 to 0.04. After factoring this into our uncertainty, we expect our BHM method to, on average, recover $w_{\mathrm{BHM}} = w_{\mathrm{BBC}} + 0.04 \pm 0.04$.
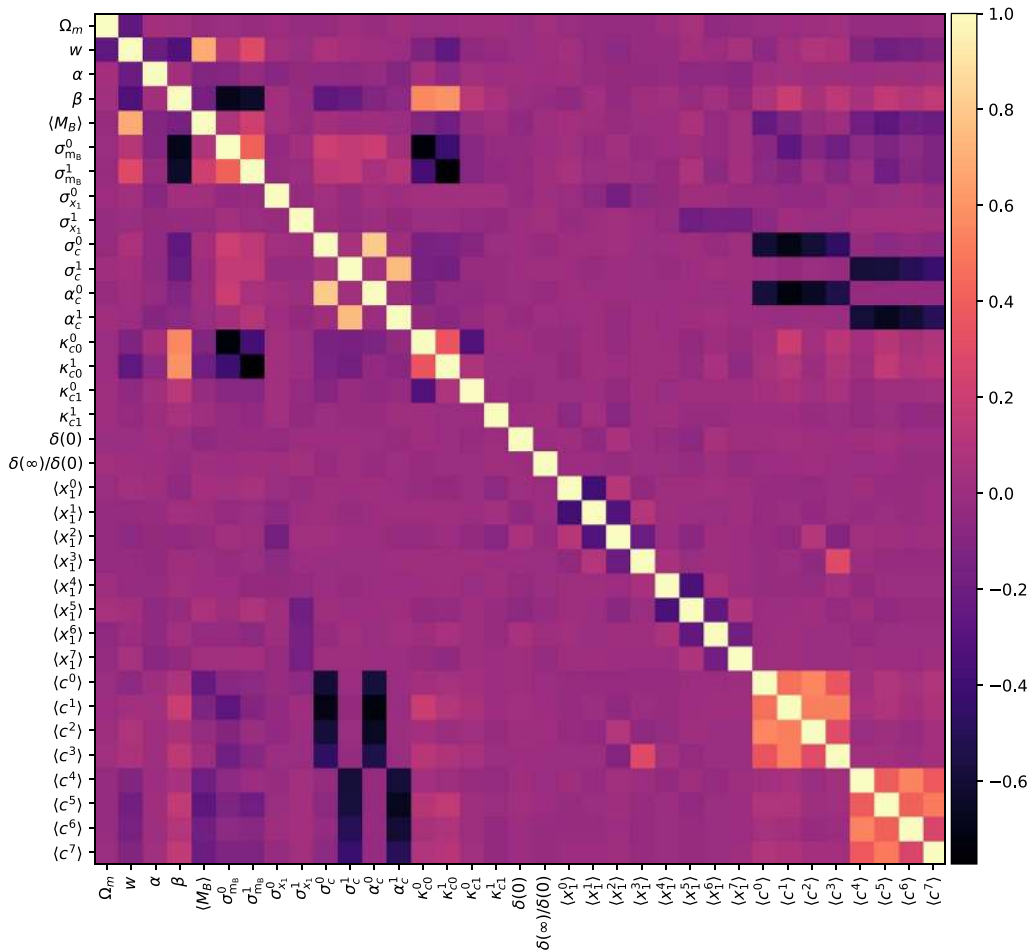
Figure 10. Parameter correlations for the combined fits to the 100 G10 scatter model simulations. We see that the primary correlations with w enter through α, β, and $\langle M_B \rangle$, as shown in Table 6. While $\langle M_B \rangle$ is generally thought to be a nuisance parameter, we find a cosmological correlation. We note that, by fixing H_0 in our distance modulus calculation, $\langle M_B \rangle$ absorbs any cosmological uncertainty on this term. Additionally, $\langle M_B \rangle$ also affects the selection efficiency, which was computed from simulations with a fixed M_B value, introducing a second plausible source of correlation. Also visible in this figure are several other interesting relationships. Here β is strongly anticorrelated with intrinsic dispersion σ_{m_B} for both surveys (DES-like and low-z), with σ_{m_B} showing strong anticorrelation with κ_c^0. This relationship is indeed expected; as κ_c^0 grows larger (more unexplained dispersion on the color observation), the width of the supernova population in apparent magnitude space increases. As the fit prefers it to conform to the observed width of the distribution, the extra width in color causes the inherent magnitude smearing amount to decrease. And, with extra freedom on the observed color from κ_c^0, β shifts in response. The other striking feature in the plot is the strong correlation blocks in the bottom right and the anticorrelation stripes on the edges. These too are expected, for they show the relationship between the color distribution's mean value, width, and skewness. As skewness or population width increases, the effective mean of the population shifts (see Appendix A.3 for details), creating an anticorrelation between skewness and the (Gaussian) mean color population. Strong anticorrelation between κ_{c0}^0 and κ_{c0}^1 with σ_{m_B} reveals the strong population degeneracy, and—for the C11 simulation results—a constrained positive value shows that a finite nonzero extra color dispersion is indeed preferred by our model.

Having established that our method exhibits similar shifts in the recovery of w compared to BBC, future work will focus on improving the parameterization of the intrinsic scatter model into our framework, with the goal of minimizing the effect of the underlying scatter model on the recovery of cosmology.

6. Conclusions

In this paper, we have outlined the creation of a hierarchical Bayesian model for supernova cosmology. The model takes into account selection effects and their uncertainty, fits underlying populations and standardization parameters, incorporates unexplained dispersion from intrinsic scatter color smearing, and incorporates uncertainty from peculiar velocities, survey calibration, *HST* calibration, dust, a potential global redshift offset, and SALT2 model uncertainty. Furthermore, our uncertainties in standardization, population, mass step, and more, being explicitly parameterized in our model, are captured with covariance intact, an improvement on many previous

methods. The model has been optimized to allow for hundreds of supernovae to be modeled fully with latent parameters. It runs in under 1 hr of CPU time and scales linearly with the number of supernovae, as opposed to the polynomial complexity of matrix inversion of other methods.

The importance of validating models using high-precision statistics gained by performing fits to hundreds of data realizations cannot be overstated; however, this validation is lacking in many earlier BHM models for supernova cosmology. We have validated this model against many realizations of simplistic simulations with well-known and well-defined statistics and found no cosmological bias. When validating using SNANA simulations, we find evidence of cosmological bias that is traced back to light-curve fits reporting biased observables and incorrect covariance. Allowing fully parameterized corrections on observed supernova summary statistics introduces too many degrees of freedom and is found to make cosmology fits too weak. Allowing simulation-based

Table 7
w Error Budget

Feature	Parameters	σ_w	Cumulative
Cosmology	Ω_m, w	0.051	0.051
Standardization	α, β, $\langle M_B \rangle$,	0.046	0.068
	$\delta(0)$, $\delta(\infty)/\delta(0)$		
Intrinsic scatter	κ_0, κ_1	0.020	0.071
Redshift-independent populations	σ_{M_B}, σ_c, σ_{x_1}, α_c	0.022	0.074
Redshift-dependent populations	$\langle c_i \rangle$, $\langle x_{1,i} \rangle$	0.030	0.080
Systematics	$\delta \mathcal{Z}_i$, δS	0.054	0.096

Note. Error budget determined from analyzing uncertainty on simulation data while progressively enabling model features. We start from the top of the table, only varying cosmological parameters Ω_m and w, and then progressively unlock parameters and let them fit as we progress down the table. The cumulative uncertainty shows the total uncertainty on w on the fit for all, where the σ_w term is derived by taking the quadrature difference in cumulative uncertainty as we progress.

Table 8
w Bias Comparison

Model	G10	C11	(G10 + C11)
Steve $\langle w \rangle$	-0.998 ± 0.007	-0.945 ± 0.007	-0.972 ± 0.006
BBC $\langle w \rangle$	-1.044 ± 0.006	-0.978 ± 0.007	-1.010 ± 0.005[a]
$\Delta \langle w \rangle$	$+0.044 \pm 0.006$	$+0.033 \pm 0.006$	$+0.038 \pm 0.004$
$\Delta \sigma_w$	0.057 ± 0.004	0.062 ± 0.004	0.060 ± 0.003

Note. We characterize the bias on w using the 100 simulations for the G10 scatter model and 100 simulations for the C11 scatter model. We also show the results when combining the G10 and C11 models into a combined set of 200 simulations. The mean w value for our method and BBC are presented, along with the mean when averaging the difference between our method and BBC for each individual simulation. Averages are computed giving each simulation sample the same weight. In the model, Δ represents Steve – BBC. The final row shows the scatter between Steve and BBC for the different simulations.
[a] This value is computed with each simulation having the same weight. It disagrees with the value provided in Brout et al. (2019, Table 10, row 3), which uses inverse variance–weighted averages. We do not utilize this weight because the variance is correlated with the value of w due to the Ω_m prior applied in the fitting process. We note that if inverse variance weighting is applied to both data sets, they both shift by $\Delta w \approx 0.005$, and thus the predicted difference between the BBC method and Steve remains the same.

corrections to vary in strength is found to give minor reductions in w bias; however, the uncertainty on the intrinsic scatter model itself limits the efficacy of the bias corrections. For the data size represented in the DES 3 yr spectroscopic survey, the determined biases should be subdominant to other sources of uncertainty; however, this cannot be expected for future analyses with larger data sets. Stricter bias corrections calculated from simulations are required to reduce bias. Ideally, this would include further work on the calculation of the intrinsic dispersion of the SN Ia population such that we can better characterize this bias.

With our model being validated against hundreds of simulation realizations representing a combined data set of more than 60,000 simulated supernovae, we have been able to accurately determine the biases in our model and trace their origin. With the current biases being subdominant to the total uncertainty, we now prepare to analyze the DES 3 yr data set.
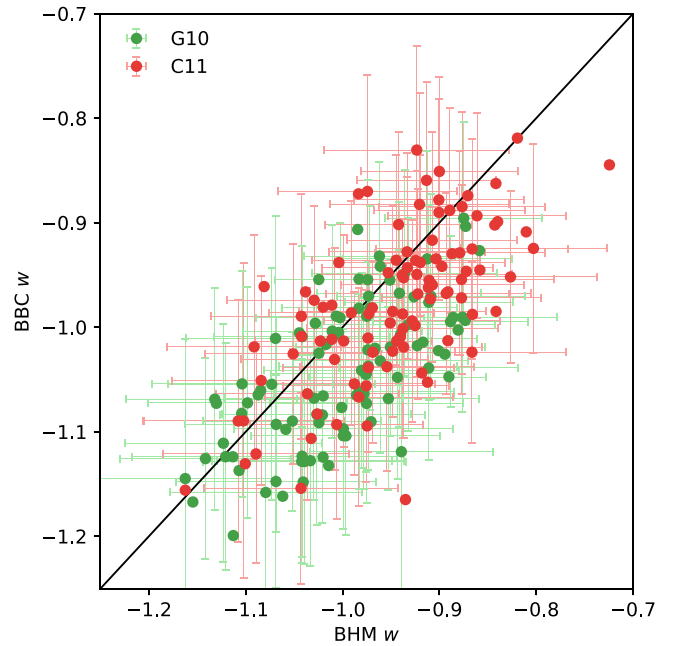
Figure 11. Recovered w for the 200 validation simulations with full treatment of statistical and systematic errors. Uncertainty on the recovered w value is shown for every second data point for visual clarity.

Plots of posterior surfaces and parameter summaries were created with `ChainConsumer` (Hinton 2016).

Appendix A
Selection Effect Derivation

A.1. General Selection Effects

When formulating and fitting a model using a constraining data set, we wish to resolve the posterior surface defined by

$$P(\theta|\text{data}) \propto P(\text{data}|\theta)P(\theta), \qquad (21)$$

which gives the probability of the model parameter values (θ) given the data. Prior knowledge of the allowed values of the model parameters is encapsulated in the prior probability $P(\theta)$. Of primary interest to us is the likelihood of observing the data given our parameterized model, $\mathcal{L} \equiv P(\text{data}|\theta)$. When dealing with experiments that have imperfect selection efficiency, our likelihood must take that efficiency into account. We need to describe the probability that the events we observe are both drawn from the distribution predicted by the underlying theoretical model and that those events, given that they happened, are subsequently successfully observed. To make this extra conditional explicit, we write the likelihood of the data given an underlying model, θ, and that the data are included in our sample, denoted by S, as

$$\mathcal{L}(\theta; \text{data}) = P(\text{data}|\theta, S). \qquad (22)$$

A variety of selection criteria are possible, and in our method, we use our data in combination with the proposed model to determine the probability of particular selection criteria. That is, we characterize a function $P(S|\text{data}, \theta)$, which colloquially can be stated as the probability of a potential observation passing selection cuts, given our observations and the underlying model. We can introduce this expression in a few lines due to symmetry of joint probabilities and utilizing that $P(x, y, z) = P(x|y, z)P(y, z) = P(y|x, z)P(x, z)$:

$$P(\text{data}|S, \theta)P(S, \theta) = P(S|\text{data}, \theta)P(\text{data}, \theta) \qquad (23)$$

$$P(\text{data}|S, \theta) = \frac{P(S|\text{data}, \theta)P(\text{data}, \theta)}{P(S, \theta)} \qquad (24)$$

$$= \frac{P(S|\text{data}, \theta)P(\text{data}|\theta)P(\theta)}{P(S|\theta)P(\theta)} \qquad (25)$$

$$= \frac{P(S|\text{data}, \theta)P(\text{data}|\theta)}{P(S|\theta)}, \qquad (26)$$

which is equal to the likelihood $\mathcal{L}$. At this point, our derivation depends on whether our selection effects are best modeled as a function of observables or latent parameters. In most cases, selection effects will be best modeled directly as a function of observables, and we would introduce an integral over all possible events D, so we can evaluate $P(S|\theta)$,

$$\mathcal{L}(\theta; \text{data}) = \frac{P(S|\text{data}, \theta)P(\text{data}|\theta)}{\int P(S, D|\theta)dD}, \qquad (27)$$

$$\mathcal{L}(\theta; \text{data}) = \frac{P(S|\text{data}, \theta)P(\text{data}|\theta)}{\int P(S|D, \theta)P(D|\theta)dD}. \qquad (28)$$

The second option, wherein selection effects can be more computationally efficiently modeled with the inclusion of latent parameters (for example, in the case where we do not have direct access to the observables upon which our data selection is determined), we can introduce our latent parameters in addition to an integral over all possible data,

$$\mathcal{L}(\theta; \text{data}) = \frac{\int P(S|\text{data}, L, \theta)P(\text{data}|L)P(L|\theta)dL}{\iint P(S|D, L, \theta)P(D|L, \theta)P(L|\theta)dD\,dL}, \qquad (29)$$

where L represents our latent parameters that model the true underlying values of our observables, such that our data are conditioned directly on them. In this formulation, the selection effects can depend on both data and latent variables.

A.2. Supernova Selection Effects

To turn the generalized equations from the previous sections into selection effects relevant for this model, we need to highlight that θ represents only our top-level parameters (Ω_m, w, α, β, etc.), and that our parameterization of the true underlying values (for example, m_B) takes the form of latent parameters. We thus continue from Equation (29) and write out the denominator d:

$$d = \iint P(S|D, L, \theta)P(D|L, \theta)P(L|\theta)dD\,dL. \qquad (30)$$

In our formulation, we assume that our selection effects can be sufficiently encapsulated by latent parameters. That is, we simplify $P(S|D, L, \theta) \rightarrow P(S|L)$. This allows us to isolate our integral $\int P(D|L, \theta)$ in the above equation, integrate it to unity,

and come to

$$d = \int P(S|L)P(L|\theta)dL. \quad (31)$$

Here we can now move from the generic label L to something specific to our model. Specifically, we assume that our selection effects can be quantified using the true apparent magnitude m_B and redshift z, so $L \rightarrow \{m_B, z\}$, or, formally, $P(S|D, L, \theta) = P(S|z)P(S|m_B)$. We do not write out other latent parameters found in the model, as they do not impact selection effects—we would simply find them integrated to unity. Note that this assumption does not capture biases engendered by Poisson noise fluctuations. We advocate that future analyses with higher statistical precision use a precisely determined $P(S|D)$. Writing out the latent variables, our denominator becomes

$$d = \int P(S|z)P(S|m_B)P(z, m_B|\theta) \, dz \, dm_B. \quad (32)$$

We can express $P(z, m_B|\theta)$ as $P(m_B|z, \theta)P(z|\theta)$, where the first term requires us to calculate the magnitude distribution of our underlying population at a given redshift, and the second term is dependent on survey geometry and supernova rates. We can thus state

$$d = \int \left[\int P(S|m_B)P(m_B|z, \theta)dm_B \right] P(S|z)P(z|\theta)dz. \quad (33)$$

By assuming that the distribution $P(S|z)P(z|\theta)$ is well sampled by the observed supernova redshifts, we can approximate the integral over redshift by evaluating

$$\int P(S|m_B)P(m_B|z, \theta)dm_B \quad (34)$$

for each supernova in the data set, i.e., Monte Carlo integration with assumed perfect importance sampling.

As stated in Section 4.4.5, the underlying population in apparent magnitude, when we discard skewness, can be represented as $\mathcal{N}(m_B|m_B^*(z), \sigma_{m_B}^*)$, where

$$m_B^*(z) = \langle M_B \rangle + \mu(z) - \alpha\langle x_1(z)\rangle + \beta\left(\langle c(z)\rangle + \sqrt{\frac{2}{\pi}}\sigma_c\delta_c\right), \quad (35)$$

$$\sigma_{m_B}^* = \sigma_{M_B}^2 + (\alpha\sigma_{x_1})^2 + \left(\beta\sigma_c\sqrt{1 - \frac{2\delta_c^2}{\pi}}\right)^2. \quad (36)$$

Then, modeling $P(S|m_B)$ as either a normal or a skew normal, we can analytically perform the integral in Equation (34) and reach Equations (18) and (19).

A.3. Approximate Selection Effects

In this section, we investigate the effect of approximating the skew normal underlying color distribution as a normal. Specifically, Equations (35) and (36) make the assumption that, for our color distribution, $\mathcal{N}^{\text{Skew}}(\mu, \sigma, \alpha)$ is well approximated by $\mathcal{N}(\mu, \sigma)$. We sought to improve on this approximation by adjusting the mean and standard deviation of the approximated normal to more accurately describe the actual mean and standard deviation of a skew normal. With

$\delta \equiv \alpha/\sqrt{1 + \alpha^2}$, the correct mean and standard deviation are

$$\mu_1 = \mu_0 + \sqrt{\frac{2}{\pi}}\delta\sigma_0, \quad (37)$$

$$\sigma_1 = \sigma_0\sqrt{1 - \frac{2\delta^2}{\pi}}, \quad (38)$$

where we highlight that μ here represents the mean of the distribution, not distance modulus. We can then test the approximation $\mathcal{N}^{\text{Skew}}(\mu_0, \sigma_0, \alpha) \rightarrow \mathcal{N}(\mu_1, \sigma_1)$. Unfortunately, this shift to the mean and standard deviation of the normal approximation, where we treat m_B, x_1, and c as a multivariate skew normal, did not produce stable posterior surfaces. Due to this, we treat the underlying m_B, x_1, and c populations as independent.

We tested a fixed σ_c in the shift correction such that $\mu_1 = \mu_0 + \sqrt{2/\pi}\,\delta k$, where we set $k = 0.1$ to mirror the width of the input simulation population. This resulted in stable posterior surfaces; however, this introduced recovery bias in several population parameters, so we do not fix σ_c. Comparing whether we shift our normal in the approximation or simply discard skewness, Figure 3 shows that the calculated efficiency is significantly discrepant to the actual efficiency if the normal approximation is not shifted. The biases when using shifted or unshifted normal approximations when we fit our model on Gaussian and skewed underlying populations are shown in Figure 12, and only the shifted normal approximation correctly recovers the underlying population parameters.

Appendix B
Numerical Optimizations

Not many fitting methodologies and algorithms can handle the thousands of fit parameters our model requires. By using Stan, we are able to take advantage of automatic differentiation and the NUTS sampler, which is a class of Hamiltonian Monte Carlo samplers. Even with these advantages, early implementations of our model still had excessive fit times, with our desired sub-hour running time being far exceeded.

The simplest and most commonly found optimization we employed was to precompute as much as possible. This is in a bid to reduce the complexity of the mathematical graph our model is translated into by Stan to simplify the computation of surface derivatives. For example, when computing the distance modulus, redshift is encountered to various powers. Instead of computing those powers in Stan, we simply pass in several arrays of redshift values already raised to the correct power. Small changes like this, however, only give small improvements.

The primary numerical improvement we made on existing frameworks was to remove costly probability evaluations of multivariate normals. To increase efficiency, the optimum way to sample a multivariate normal is to parameterize it such that instead of sampling $\mathcal{N}(x|\mu, \Sigma)$, we sample $\mathcal{N}(\delta|0, 1)$, where $x = \mu + L\delta$ and L is the Cholesky decomposition of Σ. In this way, we can efficiently sample the unit normal probability distribution instead of sampling a multivariate normal probability distribution. Switching to this parameterization resulted in a computational increase of an order of magnitude, taking fits for a sample of approximately 500 supernovae from roughly 4 hr down to 30 minutes.
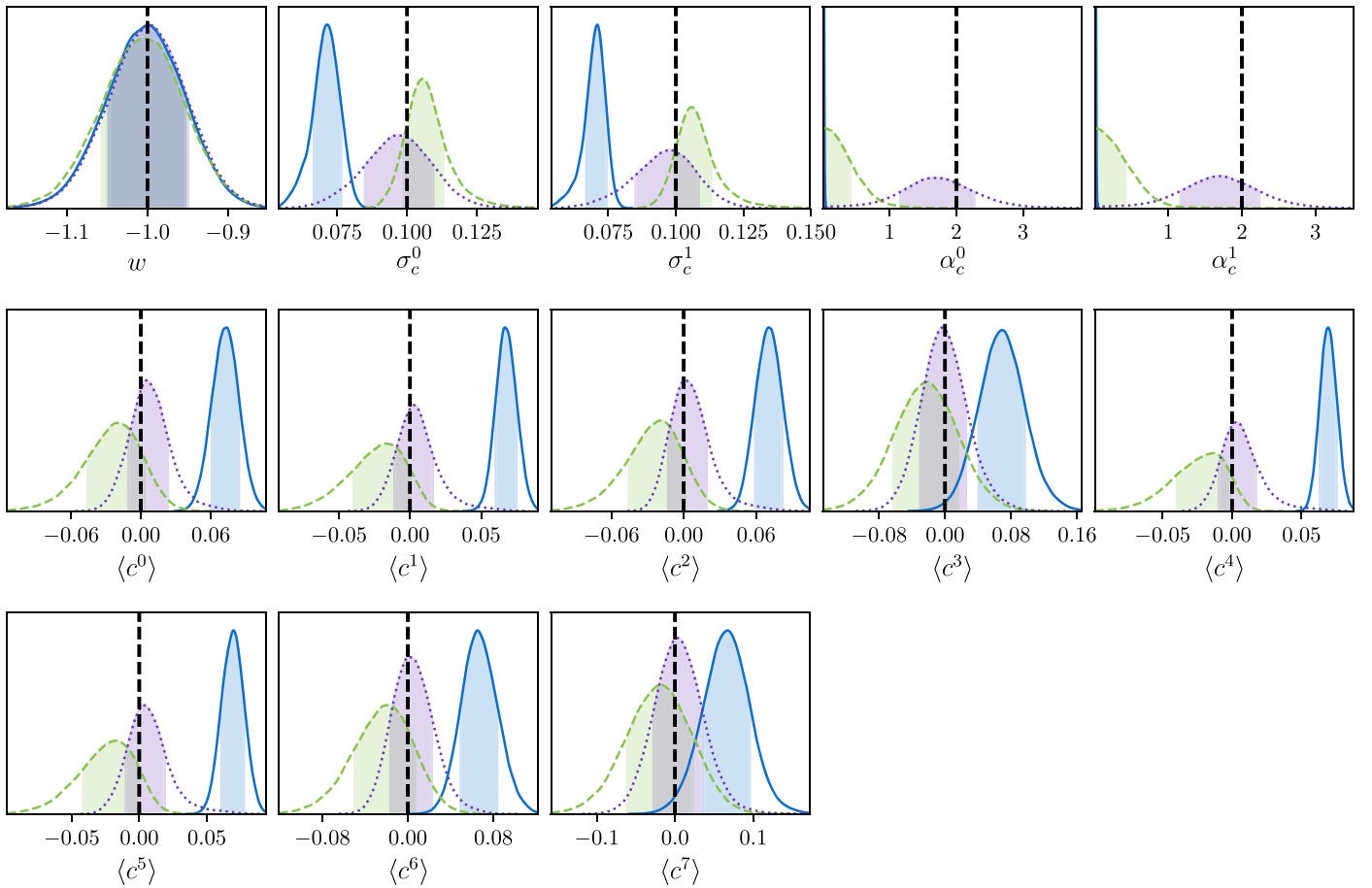
Figure 12. Marginalized probability distributions for 100 realizations of cosmology, fit to a flat wCDM with prior $\Omega_m \sim \mathcal{N}(0.3, 0.01)$, each containing 1000 simulated high-z and 1000 simulated low-z supernovae. The dashed green distribution represents a fit to an underlying Gaussian color population with the unshifted model. The blue solid surface represents fits to a skewed color population with the unshifted model, and the purple dotted surface represents a fit to a skewed color population with the shifted model. The superscripts 0 and 1 denote the two different surveys (high- and low-z, respectively); similarly, the first four $\langle c^i \rangle$ parameters represent the four redshift nodes in the high-z survey, and the last four represent the nodes for the low-z survey. We can see that the shifted model is far better able to recover skewed input populations than the unshifted, performing better in terms of recovering skewness α_c, mean color $\langle c \rangle$, and width of the color distribution σ_c. The unshifted model recovers the correct color mean and width if you approximate a skew normal as a normal, $\Delta\mu = \sqrt{2/\pi}\,\sigma_c \delta_c \approx 0.071$, which is approximately the deviation found in the fits to the color population mean. Importantly, the unshifted model, when run on skewed data (solid blue), shows extreme bias in α_c, where it fits strongly around zero regardless, showing it to be a poor approximation. Based on these results and the good performance in correctly recovering underlying populations of the shifted normal approximation, we adopt the shifted normal approximation in our model.

This parameterization does come with one significant downside: inflexibility. For each step the algorithm takes, we do not recompute the Cholesky decomposition of the covariance of the summary statistics—that happens once at the beginning of the model setup. If we had kept the full covariance matrix parameterization, we could modify the matrix easily; for example, when incorporating intrinsic dispersion, we could simply add on a secondary matrix to create an updated covariance. However, as the Cholesky decomposition of a sum of matrices is not equal to the sum of the Cholesky decomposition of each individual matrix, we would need to recompute the decomposition for each step, which discards most of the computational benefit just gained.

Considering a 3×3 matrix with Cholesky decomposition

$$L = \begin{pmatrix} a & 0 & 0 \\ b & c & 0 \\ d & e & f \end{pmatrix}, \tag{39}$$

the original covariance matrix Σ is given by

$$\Sigma = \begin{pmatrix} a^2 & ab & ad \\ ab & b^2 + c^2 & bd + ce \\ ad & bd + ce & d^2 + e^2 + f^2 \end{pmatrix}. \tag{40}$$

Now the primary source of extra uncertainty in the intrinsic dispersion models comes from chromatic smearing, which primarily influences the recovered color parameter, which is placed as the last element in the observables vector $\{m_B, x_1, c\}$. We can now see that it is possible to add extra uncertainty to the color observation on the diagonal without having to recompute the Cholesky decomposition—notice that f is unique in that it is the only element of L that appears in only one position in the covariance matrix. To take our covariance and add on the diagonal uncertainty for

color an extra σ_e term, we get

$$C = \begin{pmatrix} \sigma^2_{m_B} & \rho_{0,1}\sigma_{m_B}\sigma_{x_1} & \rho_{0,2}\sigma_{m_B}\sigma_c \\ \rho_{0,1}\sigma_{m_B}\sigma_{x_1} & \sigma^2_{x_1} & \rho_{1,2}\sigma_{x_1}\sigma_c \\ \rho_{0,2}\sigma_{m_B}\sigma_c & \rho_{1,2}\sigma_{x_1}\sigma_c & \sigma^2_c + \sigma^2_e \end{pmatrix}. \tag{41}$$

The Cholesky decomposition of this, in terms of the original Cholesky decomposition, is

$$L = \begin{pmatrix} a & 0 & 0 \\ b & c & 0 \\ d & e & f+g \end{pmatrix}, \tag{42}$$

where $g = \sqrt{f^2 + \sigma^2_e} - f$. This allows an easy update to the Cholesky decomposition to add extra uncertainty to the independent color uncertainty. For both the G10 and C11 models, we ran fits without the Cholesky parameterization to allow for extra correlated dispersion (instead of just dispersion on c), but we found no decrease in bias or improved fit statistics, allowing us to use the more efficient Cholesky parameterization.

ORCID iDs

S. R. Hinton ⓘ https://orcid.org/0000-0003-2071-9349
D. Brout ⓘ https://orcid.org/0000-0001-5201-8374
C. B. D'Andrea ⓘ https://orcid.org/0000-0002-8198-0332
R. Kessler ⓘ https://orcid.org/0000-0003-3221-0419
C. Lidman ⓘ https://orcid.org/0000-0003-1731-0497
E. Morganson ⓘ https://orcid.org/0000-0001-7180-109X
J. Annis ⓘ https://orcid.org/0000-0002-0609-3987
D. W. Gerdes ⓘ https://orcid.org/0000-0001-6942-2736
R. A. Gruendl ⓘ https://orcid.org/0000-0002-4588-6517
D. L. Hollowood ⓘ https://orcid.org/0000-0002-9369-4157
R. Miquel ⓘ https://orcid.org/0000-0002-6610-4836
M. Soares-Santos ⓘ https://orcid.org/0000-0001-6082-8529
G. Tarle ⓘ https://orcid.org/0000-0003-1704-0781

References

Abbott, T. M. C., Allam, S., & Andersen, P. 2018, ApJL, 872, 30
Amanullah, R., Lidman, C., Rubin, D., et al. 2010, ApJ, 716, 712
Astier, P., Guy, J., Regnault, N., et al. 2006, A&A, 447, 31
Bailey, S., Aldering, G., Antilogus, P., et al. 2008, arXiv:0810.3499
Balland, C., Baumont, S., Basa, S., et al. 2009, A&A, 507, 85
Barbary, K., Aldering, G., Amanullah, R., et al. 2010, ApJ, 745, 27
Bernstein, J. P., Kessler, R., Kuhlmann, S., et al. 2012, ApJ, 753, 152
Betoule, M., Kessler, R., Guy, J., et al. 2014, A&A, 568, 32
Brout, D., Scolnic, D., Kessler, R., et al. 2019, ApJ, 874, 150
Carpenter, B., Gelman, A., Hoffman, M. D., et al. 2017, J. Statistical Software, 76, 1
Chambers, K. C., Magnier, E. A., Metcalfe, N., et al. 2016, arXiv:1612.05560
Chotard, N., Gangler, E., Aldering, G., et al. 2011, A&A, 529, 6
Conley, A., Guy, J., Sullivan, M., et al. 2011, ApJS, 192, 1
Contreras, C., Hamuy, M., Phillips, M. M., et al. 2010, AJ, 139, 519
D'Andrea, C. B., Gupta, R. R., Sako, M., et al. 2011, ApJ, 743, 172
Dark Energy Survey Collaboration, Abbott, T., Abdalla, F. B., et al. 2016, MNRAS, 460, 1270
Dilday, B., Kessler, R., Frieman, J. A., et al. 2008, ApJ, 682, 262
Folatelli, G., Phillips, M. M., Burns, C. R., et al. 2010, AJ, 139, 120
Freedman, W. L., Burns, C. R., Phillips, M. M., et al. 2009, ApJ, 704, 1036
Frieman, J. A., Bassett, B., Becker, A., et al. 2008, AJ, 135, 338
Graur, O., Rodney, S. A., Maoz, D., et al. 2013, ApJ, 783, 28
Gupta, R. R., D'Andrea, C. B., Sako, M., et al. 2011, ApJ, 740, 92
Gupta, R. R., Kuhlmann, S., Kovacs, E., et al. 2016, AJ, 152, 154
Guy, J., Astier, P., Baumont, S., et al. 2007, A&A, 466, 11
Guy, J., Sullivan, M., Conley, A., et al. 2010, A&A, 523, 34
Hicken, M., Challis, P., Jha, S., et al. 2009b, ApJ, 700, 331
Hicken, M., Challis, P., Kirshner, R. P., et al. 2012, ApJS, 200, 12
Hicken, M., Wood-Vasey, W. M., Blondin, S., et al. 2009a, ApJ, 700, 1097
Hinton, S. 2016, JOSS, 1, 45
Hlozek, R., Kunz, M., Bassett, B., et al. 2012, ApJ, 752, 79
Huterer, D., & Shafer, D. L. 2018, RPPh, 81, 016901
Ivezic, Z., Kahn, S. M., Tyson, J. A., et al. 2008, ApJ, 873, 111
Jennings, E., Wolf, R., & Sako, M. 2016, 1, arXiv:1611.03087
Johansson, J., Thomas, D., Pforr, J., et al. 2013, MNRAS, 435, 1680
Karpenka, N. V. 2015, arXiv:1503.03844
Kelly, P. L., Hicken, M., Burke, D. L., Mandel, K. S., & Kirshner, R. P. 2010, ApJ, 715, 743
Kessler, R., Becker, A. C., Cinabro, D., et al. 2009b, ApJS, 185, 32
Kessler, R., Bernstein, J. P., Cinabro, D., et al. 2009a, PASP, 121, 1028
Kessler, R., Brout, D., D'Andrea, C. B., et al. 2019, MNRAS, 485, 1171
Kessler, R., Guy, J., Marriner, J., et al. 2013, ApJ, 764, 48
Kessler, R., & Scolnic, D. 2017, ApJ, 836, 56
Kowalski, M., Rubin, D., Aldering, G., et al. 2008, ApJ, 686, 749
Kunz, M., Bassett, B., & Hlozek, R. 2007, PhRvD, 75, 103508
Lampeitl, H., Smith, M., Nichol, R. C., et al. 2010, ApJ, 722, 566
Lewis, A., & Bridle, S. 2002, PhRvD, 66, 103511
LSST Science Collaboration et al. 2009, arXiv:0912.0201
Ma, C., Corasaniti, P.-S., & Bassett, B. A. 2016, MNRAS, 463, 1651
Malmquist, K. G. 1922, Lund Medd. Ser. I, 100, 1
Mandel, K. S., Narayan, G., & Kirshner, R. P. 2011, ApJ, 731, 120
Mandel, K. S., Scolnic, D., Shariff, H., Foley, R. J., & Kirshner, R. P. 2017, ApJ, 842, 26
Mandel, K. S., Wood-Vasey, W. M., Friedman, A. S., & Kirshner, R. P. 2009, ApJ, 704, 629
March, M. C., Karpenka, N. V., Feroz, F., & Hobson, M. P. 2014, MNRAS, 437, 3298
March, M. C., Trotta, R., Berkes, P., Starkman, G. D., & Vaudrevange, P. M. 2011, MNRAS, 418, 2308
Perlmutter, S., Aldering, G., Goldhaber, G., et al. 1999, ApJ, 517, 565
Perrett, K., Balam, D., Sullivan, M., et al. 2010, AJ, 140, 518
Perrett, K., Sullivan, M., Conley, A., et al. 2012, AJ, 144, 59
Phillips, M. M. 1993, ApJL, 413, L105
Phillips, M. M., Lira, P., Suntzeff, N. B., et al. 1999, AJ, 118, 1766
Rest, A., Scolnic, D., Foley, R. J., et al. 2014, ApJ, 795, 44
Riess, A. G., Filippenko, A. V., Challis, P., et al. 1998, AJ, 116, 1009
Rigault, M., Copin, Y., Aldering, G., et al. 2013, A&A, 560, A66
Roberts, E., Lochner, M., Fonseca, J., et al. 2017, JCAP, 10, 36
Rodney, S. A., Riess, A. G., Strolger, L. G., et al. 2014, AJ, 148, 13
Rubin, D., Aldering, G., Barbary, K., et al. 2015, ApJ, 813, 137
Sako, M., Bassett, B., Becker, A. C., et al. 2018, PASP, 130, 064002
Scolnic, D., & Kessler, R. 2016, ApJL, 822, L35
Scolnic, D. M., Jones, D. O., Rest, A., et al. 2018, ApJ, 859, 101
Shariff, H., Jiao, X., Trotta, R., & van Dyk, D. A. 2016, ApJ, 827, 1
Stan Development Team 2017, PyStan: the interface to Stan, http://mc-stan.org/
Stritzinger, M. D., Phillios, M. M., Boldt, L. N., et al. 2011, AJ, 142, 156
Sullivan, M., Conley, A., Howell, D. A., et al. 2010, MNRAS, 406, 782
Suzuki, N., Rubin, D., Lidman, C., et al. 2012, ApJ, 746, 85
Tripp, R. 1998, A&A, 331, 815
Uddin, S. A., Mould, J., Lidman, C., Ruhlmann-Kleider, V., & Zhang, B. R. 2017, ApJ, 848, 56
Weyant, A., Schafer, C., & Wood-Vasey, W. M. 2013, ApJ, 764, 116
Wood-Vasey, W. M., Miknaitis, G., Stubbs, C. W., et al. 2007, ApJ, 666, 694