SN-BHM: A hierarchical Bayesian model for Supernova Cosmology

Samuel R. Hinton,[1,2]★ Alex G. Kim,[3] Tamara M. Davis[1,2]
[1]*School of Mathematics and Physics, The University of Queensland, Brisbane, QLD 4072, Australia*
[2]*ARC Centre of Excellence for All-sky Astrophysics (CAASTRO)*
[3]*Physics Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA*

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT
Abstract Abstract

Key words: keyword1 – keyword2 – keyword3

1 INTRODUCTION

Almost two decades have passed since the discovery of the accelerating universe (Riess et al. 1998; Perlmutter et al. 1999). Since that time, of the number of observed Type Ia supernovae (SN Ia) have increased by more than an order of magnitude thanks to modern surveys at both low redshift (Bailey et al. 2008; Freedman et al. 2009; Hicken et al. 2009; Contreras et al. 2010; Conley et al. 2011), and higher redshift (Astier et al. 2006; Wood-Vasey et al. 2007; Balland et al. 2009; Amanullah et al. 2010). Cosmological analysis of these supernova samples (Kowalski et al. 2008; Conley et al. 2011; Suzuki et al. 2012; Betoule et al. 2014; Rest et al. 2014) have been combined with complimentary probes of large scale structure (Alam et al. 2017) and the CMB (Hinshaw et al. 2013; Planck Collaboration et al. 2013), and yet, despite these prodigious efforts, the nature of dark energy remains an unsolved mystery.

In attempts to tease out the nature of dark energy, currently running and planned surveys are once again ramping up their statistical power. The Dark Energy Survey (DES, Bernstein et al. 2012; Abbott et al. 2016) will be observing thousands of Type Ia supernova, attaining both spectroscopic and photometric confirmation. The Large Synoptic Survey Telescope (LSST, Ivezic et al. 2008; LSST Science Collaboration et al. 2009) will produce scores of thousands of photometrically classified supernovae. Such increased statistical power demands a similarly increased fidelity and flexibility in modelling the supernovae for cosmological purposes, as systematic uncertainty will prove to be the limiting factor in our analyses.

As such, staggering effort is being put into developing more sophisticated supernovae analyses. Scolnic & Kessler (2016) and Kessler & Scolnic (2017) explore sophisticated simulation corrections to traditional analyses. Approximate Bayesian computation methods also make use of simulations, trading traditional likelihoods and analytic approximations for more robust models with only the cost of increased computational time (Weyant et al. 2013; Jennings et al. 2016). Hierarchical Bayesian Models abound (Mandel et al. 2009; March et al. 2011, 2014; Rubin et al. 2015; Shariff et al. 2016; Roberts et al. 2017), however often face difficulties finding sufficient analytic approximations for complicated effects such as Malmquist bias.

In this paper, we lay out a new hierarchical model that advances the past work of Rubin et al. (2015) to include more robust treatment of systematics and selection effects whilst improving computational performance. Section 2 is dedicated to a quick review of the supernovae cosmology. In Section 3 we outline our methodology. Model verification on simulated datasets is given in Section 4. Forecasts for the impending DES three year spectroscopic supernova survey are contained in Section **??**. Section **??** investigates the effect of various systematics on our model, and Section **??** provides details on potential areas of improvement and unsuccessful methodologies.

2 REVIEW

Whilst supernova observations take the form of time-series photometric measurements of brightness in many photomet-

★ E-mail: samuelreay@gmail.com

ric bands, most analyses do not work from these measurements of apparent magnitude and colour. Instead, most techniques fit these observations of magnitude (along with redshift) 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 supernovae light curves for the cosmology selected supernova sample (Guy et al. 2010; Mosher et al. 2014). The resulting output from the model is, for each supernova, a characterised amplitude x_0 (which can be converted into apparent magnitude $m_B = -2.5\log(x_0)$), a stretch term x_1 and colour term c, along with a covariance matrix describing the uncertainty on these summary statistics. As such, the product at the end is a (redshift-dependent) population of m_B, x_1 and c.

The underlying actual supernova population is not as clear cut, and indeed accurately characterising this population, its evolution over redshift and effects from environment is one of the challenges of supernova cosmology. However, given some modelled underlying population that lives in the redshift-dependent space M_B, x_1 and c, the introduction of cosmology into the model is simple – it is encoded in the functional map between those two populations, from apparent magnitude space to absolute magnitude. Specifically, for any given supernova our functional map may take the traditional form:

$$M_B = m_B + \alpha x_1 - \beta c - \mu(z) + \text{corrections}, \tag{1}$$

where α is the stretch correction (Phillips 1993), and β is the colour correction (Tripp 1998) that respectively encapsulate the empirical relation that broader and bluer supernovae are brighter. The corrections term at the end often includes corrections for host galaxy environment, as this has statistically significant effects on 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; Uddin et al. 2017). The cosmological term, $\mu(z)$ represents the distance modulus, and is precisely known given cosmological parameters and an input redshift.

2.1 Traditional Analyses

Traditional χ^2 analyses such as that found in Kowalski et al. (2008); Conley et al. (2011); Betoule et al. (2014), minimise the difference in distance modulus between the cosmologically predicted values μ_C and the observed distance modulus $\mu_{\rm obs}$, shown respectively below:

$$\mu_C = 5\log\left[\frac{(1+z)r}{10}\right] \tag{2}$$

$$r = \frac{c}{H_0}\int_0^z \frac{dz'}{\sqrt{\Omega_m(1+z')^3 + \Omega_k(1+z')^2 + \Omega_\Lambda(1+z')^{3(1+w)}}} \tag{3}$$

$$\mu_{\rm obs} = m_B + \alpha x_1 - \beta c - M_B \tag{4}$$

The minimising function is then given as

$$\chi^2 = (\mu_{\rm obs} - \mu_C)^\dagger C^{-1} (\mu_{\rm obs} - \mu_C) \tag{5}$$

where C^{-1} is an uncertainty matrix that combines the uncertainty from the SALT2 fits, intrinsic dispersion, calibration, dust, peculiar velocity and many other factors (see Betoule et al. (2014) for a review). The benefit this analysis methodology provides is speed - for samples of hundreds of supernovae, efficient matrix inversion algorithms allow the likelihood to be evaluated quickly. The speed comes with two costs. Firstly, formulating a χ^2 likelihood requires a loss of model flexibility by building into the model assumptions of uncertainty Gaussianity. Secondly, the computational efficiency is dependent on inverting a covariance matrix with dimensionality 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.

Selection efficiency, such as the well known Malmquist bias (Malmquist K. G. 1922) is accounted for by correcting data. Simulations following survey observational strategies and geometry are used to calculate the expected bias in distance modulus, which is then added onto the observational data. As these effects are not built into the likelihood, their influence on the error budget is not captured fully in the χ^2 distribution, and any subtle correlations between cosmological or population parameters and the bias is lost.

2.2 Approximate Bayesian Computation

To try and escape the limitations of the traditional analysis methodology, several recent methods have adopted Approximate Bayesian Computation, where supernova samples are forward modelled in parameter space and compared to observed distributions. Weyant et al. (2013) provides an introduction into ABC methods for supernova cosmology in the context of the SDSS-II results (Sako et al. 2014) and Flat ΛCDM cosmology, whilst Jennings et al. (2016) demonstrates their *superABC* method on simulated first season Dark Energy Survey samples, described in Kessler et al. (2015). In both examples, the supernova simulation package SNANA (Kessler et al. 2009) is used to forward model the data at each point in parameter space.

By building the systematic uncertainties and selection effects into the simulation package, there is vastly more freedom in how to treat and model those effects. Data does not need to be corrected, analytic approximations do not need to be applied, we are free to incorporate algorithms that simply cannot be expressed in a tractable likelihood. 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 in its entirety – drawing from underlying supernova populations, modelling light curves, applying selection effects, fitting light curves and applying data cuts. This is an intensive process. Luckily, efficient sampling algorithms that have walkers that rely on the Markov properties of groups instead of individual walkers, such as Ensemble sampling (Foreman-Mackey et al. 2013) allow easy parallelisation of parameter fits and can help ease the computational burden.

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) presents both a metric used to compare forward modelled summary statistic populations (denoted the 'Tripp' metric) and a metric directly applicable to the observed supernova light curves themselves, however evaluation of systematic uncertainty was only performed using the Tripp metric.

2.3 Hierarchical Bayesian Models

Sitting comfortably between the traditional models simplicity and the complexity of forward modelling lies Hierarchical Bayesian Models. With the introduction of multiple layers in our model, we can add far more flexibility than a traditional analysis whilst still maintaining most of the computational benefits that come from having a tractable likelihood. Mandel et al. (2009) and Mandel et al. (2011) construct a hierarchical model that they apply to the light curve fitting for supernova. March et al. (2011, 2014); Karpenka (2015) derive a hierarchical model and simplify it by analytically marginalising over nuisance parameters to provide a model that offers increased flexibility with reduced uncertainty over the traditional method. The recent BAHAMAS model (Shariff et al. 2016) builds off this and reanalyses the JLA dataset, whilst including extra freedom in the correction factors α and β, finding evidence for redshift dependence on β. Ma et al. (2016) also performed a reanalysis of the JLA dataset, finding significant differences in α and β values from the original as well. Notably, these methods rely on data that is bias corrected, 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 dataset (Suzuki et al. 2012). The well known BEAMS (Bayesian estimation applied to multiple species) methodology from Kunz et al. (2007) has been extended and applied in several works (Hlozek et al. 2012), mostly lately to include redshift uncertainty for photometric redshift application as zBEAMS (Roberts et al. 2017) and to include simulated bias corrections in Kessler & Scolnic (2017). However, whilst there are a large amount of hierarchical models available, none of them have undergone high-statistics simulation verification to quantify each models respective bias, which is becoming critically important as precision supernovae cosmology comes into its own.

The flexibility afforded by a hierarchical model allows for investigations into different treatments of underlying supernova magnitude, colour and stretch populations, supernovae type rates and redshift distributions, host-galaxy corrections and redshift evolution, each of which will be discussed further in the outline of our model below.

3 OUR METHOD

We construct our Bayesian Hierarchical Model with several goals in mind: creation of a redshift-dependent correlated underlying supernova population, increased treatment of systematics, and analytic correction of selection effects. As this is closest to the UNITY method from Rubin et al. (2015, hereafter denoted R15), we follow a similar model scaffold, and construct the model in Stan (Carpenter et al. 2017; Stan Development Team 2017) which uses automatic differentiation and the no-U-turn Sampler (NUTS), which is a variant of Hamiltonian Monte Carlo, to efficiently sample high dimensional parameter space.

At the most fundamental level, a supernova analysis is simply a mapping from an underlying population onto an observed population, where cosmology is encoded directly in the mapping function. The difficulty arises both in adequately describing the biases in the mapping function, and in adding sufficient, physically motivated flexibility in both of these populations whilst not adding *too* much flexibility, such that model fitting becomes pathological due to increasing parameter degeneracies within the model.

3.1 Observables Nature Provides: Populations

Like most of the BHM methods introduced previously, we work from the summary statistics as well, where each observed supernova has an apparent magnitude $\hat{m_B}$, stretch $\hat{x_1}$ and colour $\hat{c}$, with uncertainty $\mathcal{C}$ on those values. Additionally, each supernova has an observed redshift $\hat{z}$ and a host galaxy mass associated with it, $\hat{m}$, where the mass measurement takes the form of a probability of being above 10^10 solar masses. Given machine learning classifiers to type supernovae, we will also have a probability of being a Type Ia, $\hat{p}$. Our set of observables is therefore given as $\{\hat{m_B}, \hat{x_1}, \hat{c}, \hat{z}, \hat{p}, \hat{m}, \mathcal{C}\}$, as shown in the probabilistic graphical model (PGM) in Figure 1.

The first layer of the hierarchy represents the latent population of m_B, x_1 and c. This can be thought of as the 'true' values for the supernova, and with correct summary statistics and uncertainty would be normally distributed around the observed values such that

$$P(\hat{m_B}, \hat{x_1}, \hat{c} | m_B, x_1, c, \mathcal{C}) = \mathcal{N}(\{\hat{m_B}, \hat{x_1}, \hat{c}\} | \{m_B, x_1, c\}, \mathcal{C}). \quad (6)$$

However, the presence of chromatic smearing in the intrinsic supernova population (Guy et al. 2010; Chotard et al. 2011) means the summary statistics report from light curve fitting are not Gaussian. We model the extra dispersion included in the observations from colour smearing by adding extra independent uncertainty on the colour observation. As shown in (Kessler et al. 2013), the extra dispersion shows heavy redshift dependence. As this may represent underlying physics more than the effects of Malmquist bias, we decide to incorporate redshift dependence in our extra uncertainty. We thus add $\kappa_0 + \kappa_1 z$ to our observed colour uncertainty (in quadrature), where κ_0 and κ_1 are subject to strong Cauchy priors centered on zero and with width 0.002, to enforce minimum possible extra uncertainty in our model. Each survey has its own set of κ parameters. Furthermore, we use fiducial simulations (discussed further in Section 3.3.5) to characterise the bias in the mean colour measured as a function of redshift, and subtract this bias from our observed data. We allow the amount of bias to be parametrised to account for different intrinsic scatter models, where parameter s_c controlled the fraction of bias subtraction - 0 for no bias, up to 1 for the maximum bias found in simulations (which corresponded to the intrinsic scatter model from Chotard et al. (2011)). Unlike κ_0 and κ_1, which are redefined for each survey, s_c represents a universal shift in supernovae biases. We place a uniform prior on s_c.

As we are focused on the spectroscopically confirmed supernovae for this iteration of the method, we assume the observed redshift $\hat{z}$ is the true redshift z such that $P(\hat{z}|z) = \delta(\hat{z} - z)$. Similarly, we take the mass probability estimate $\hat{m}$ as correct, and do not model a latent variable.

3.2 Underlying Population

The underlying supernova population is often treated with two components - a population distribution in colour and

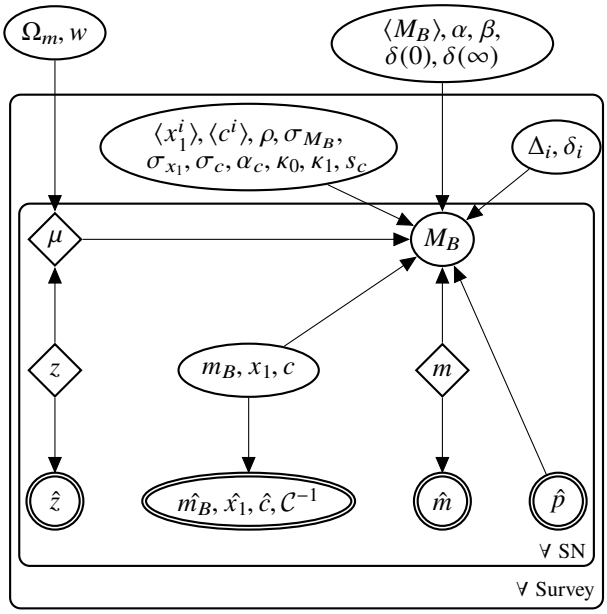

Figure 1. Probabilistic graphical model for our likelihood. Double-lined nodes represent observed variables, diamond nodes represent deterministic variables, and standard nodes represent fit variables.

stretch, and intrinsic dispersion. Analytic models often treat the colour and stretch populations with a skew normal and normal, respectively, and have them as independent. Intrinsic dispersion is treated in a variety of manners in other models, from representing it simply as (Gaussian) scatter on the absolute magnitude of the supernova population, to correlated multivariate normal scatter on the combined magnitude, stretch and colour distribution. Additionally, redshift drift of populations' mean colour and stretch will introduce a cosmological bias in our fits unless the population possesses similar ability to change as a function of redshift.

We follow prior work and model the colour population as an independent redshift-dependent skew normal for each survey. For the stretch population, we adopt a redshift-dependent normal, and magnitude dispersion is modelled as a normal. Following R15 we allow the mean colour and stretch to vary over redshift, anchoring four equally spaced redshift nodes spanning the redshift range of each survey, linearly interpolating between the nodes. These nodes are represented as $\langle x_1^i\rangle$ and $\langle c^i\rangle$. Both the colour and stretch means are modelled 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), however pathological fitting complications required us to simplify our treatment. We parameterise the standard skew normal skewness α_c by sampling $\delta_c = \alpha_c/\sqrt{1+\alpha_c^2}$ which itself is given a uniform prior $\mathcal{U}(0, 0.98)$. The width of the population, represented by the vector $\{\sigma_{M_B}, \sigma_{x_1}, \sigma_c\}$ is subject to Cauchy priors, however are sampled in log space for efficiency in sampling close to zero.

As such, the only constant between survey populations is the absolute magnitude M_B, with the skewness, redshift-dependent means and width fit individually for each survey. The probability for a supernova to have true values M_B, x_1, c given the underlying population is thus given as

$$P(M_B, x_1, c, z|\langle M_B\rangle, \langle x_1(z)\rangle, \langle c(z)\rangle, \rho, \sigma_{m_B}, \sigma_{x_1}, \sigma_c, \alpha_c) = \mathcal{N}(M_B|\langle M_B\rangle, \sigma_{M_B})\mathcal{N}(x_1|\langle x_1(z)\rangle\sigma_{x_1})\mathcal{N}^{\text{skew}}(c|\langle c(z)\rangle, \sigma_c, \alpha_c). \quad (7)$$

On top of the Ia populations, as described above, we also include a simplistic outlier population that also follows R15 (and therefore Kunz et al. 2007) as a Gaussian mixture; where the mean of the population is fixed to the Ia population, but the population width is set to a width of $\sigma^{\text{outl}} = 1$ in M_B, x_1 and c. With the spectroscopic DES sample, the contamination rate is expected to be far too low to actually fit contamination population, however in future works with photometric samples that will suffer from significantly more contamination it will be required that extra degrees of freedom are afforded the outlier population. Proof of concept simulation fits show that an acceptable parametrisation is to represent the typically brighter contaminant population as $\langle M_B^{\text{outl}}\rangle = \langle M_B\rangle - \delta_{M_B}^{\text{outl}}$, where $\delta_{M_B}^{\text{outl}}$ is constrained to be positive, or even to be greater than a small positive number to reduce degeneracy between the two populations. For the purposes of the DES spectroscopic sample, which will be dominated by confirmed Type Ia supernovae, $\delta_{M_B}^{\text{outl}} = 0$. We assume that supernovae fall into either population as determined by their observed classification probability $\hat{p}$.

3.3 Population Map

3.3.1 Cosmology

We formulate our model with three different cosmological parameterisations; Flat ΛCDM, Flat wCDM and standard ΛCDM. 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 $H_0 = 70\text{km s}^{-1}\,\text{Mpc}^{-1}$.

3.3.2 Standardisation Parameters

With increasingly large datasets and more nuanced analyses, the choice of how to handle α and β becomes an important consideration when constructing a model. R15 employs a broken linear relationship for both colour and stretch, where different values of α and β are adopted depending on whether x_1 and c are respectively positive or negative (although the cut could be placed at a location other than 0). Shariff et al. (2016) instead of employ a colour-dependent β, model β as redshift-dependent, testing two phenomenological models; $\beta(z) = \beta_0 + \beta_1 z$ and $\beta(z) = \beta_0 + \Delta\beta\,(0.5 + \arctan(100(z - z_t))/\pi)$, where the later effects a rapid but smooth change in β at a turnover redshift z_t.

We tested two models against simulated supernova sets; $\beta(c) = \beta_0 + \beta_1 c$ and $\beta(z) = \beta_0 + \beta_1 z$. See Section 4.2 for details on simulation generation. We found for both models that non-zero 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, and so in our final model 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 Equations (3) and (4).

3.3.3 *Host Galaxy Environment*

It is now well known that host galaxy environment has a significant effect on supernova properties. The latest sample of over 1300 spectroscopically confirmed Type Ia supernovae show $> 5\sigma$ 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) invoke a step function such that $\Delta M = 0.08\mathcal{H}(\log(M) - 10))$, where $\mathcal{H}$ is the Heaviside step function and M is the galaxy mass in solar masses. The scale of this step function varies from analysis to analysis, with the 0.08 value shown previously sourced from Sullivan et al. (2010) and used in Betoule et al. (2014). 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:

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

We also take flat priors on the parametrisation $\delta(0)$, $\delta(0)/\delta(\infty)$. With this correction, our calculation of absolute magnitude becomes

$$M_B = m_B - \mu(z) - \alpha x_1 + \beta c - \Delta M \cdot m. \tag{9}$$

3.3.4 *Uncertainty Propagation*

The chief difficulty with including systematic uncertainties in supernova analyses is that they generally occur during the observational pipeline, and have difficult-to-model effects on the output observations. As such, the normal treatment for systematics is to compute their effect on the supernova summary statistics – computing the numerical derivatives $\frac{\partial \hat{m_B}}{\partial \mathcal{Z}_i}$, $\frac{\partial \hat{x_1}}{\partial \mathcal{Z}_i}$, $\frac{\partial \hat{c}}{\partial \mathcal{Z}_i}$, where $\mathcal{Z}_i$ represents the i^{th} 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 original 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 adjusting the observed summary statistics is used throughout the traditional and BHM analyses, however it is normally constrained to band systematics. That is, each band for each survey has two systematics associated with it - the calibration uncertainty and the filter wavelength uncertainty. We include these in our approach, in addition to including HST Calspec calibration uncertainty, 10 SALT2 model systematic uncertainties, a dust systematic, a global redshift bias systematic and also the systematic peculiar velocity uncertainty. This gives thirteen global systematics shared by all surveys, plus two systematics per band in each survey. Denoting $\eta \equiv \{m_B, x_1, c\}$, our initial conditional likelihood for our observed summary statistics shown in Equation (6) becomes

$$P\left(\hat{\eta}, \frac{\partial \hat{\eta}}{\partial \mathcal{Z}_i}|\eta, \delta\mathcal{Z}_i, C\right) = \mathcal{N}\left(\hat{\eta} + \delta\mathcal{Z}_i \frac{\partial \hat{\eta}}{\partial \mathcal{Z}_i}|\eta, \mathcal{C}\right). \tag{10}$$

3.3.5 *Selection Effects*

Our treatment of selection effects is to incorporate selection efficiency into our model, rather than relying on simulation-driven data corrections. As such, 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 they happened, are subsequently successfully observed. 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

$$\mathcal{L} = P(\text{data}|\theta, S). \tag{11}$$

As the model so far described in previous sections describe components of a basic likelihood $P(D|\theta)$, 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

$$\mathcal{L} = \frac{P(S|\text{data}, \theta)P(\text{data}|\theta)}{\int P(S|D, \theta)P(D|\theta)\, dD}, \tag{12}$$

where the denominator represents an integral over all potential data. Full derivation of this can be found in Appendix A. As θ represents the vector of all model parameters, and D represents a vector of all observed variables, 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 HMC (a brief dismissal of months of struggle). We therefore simplify the integral and approximate the selection effects in apparent magnitude and redshift space independently, such that the denominator, denoted now w for simplicity, is given as

$$w = \int \left[\int P(S|m_B)P(m_B|z, \theta)\, dm_B\right] P(S|z)P(z|\theta)\, dz. \tag{13}$$

We apply two further approximations similar to those made in R15 – that the redshift distribution of the observed supernova reasonably well sampled the $P(S|z)P(z|\theta)$ distribution, and that the survey colour and stretch populations can be treated as Gaussian for the purposes of this integral. It was found that discarding the skewness entirely resulted in highly biased population recover, and so we instead characterise the skew normal colour distribution with a Gaussian that follows the mean and variance of a skew normal, with 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 removes the unintended bias when simply discarding skewness. More detail on this shift can be found in Appendix A3. The population $P(m_B|z, \theta)$ becomes $\mathcal{N}(m_B|m_B^*(z), \sigma_{m_B}^*)$, where

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

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

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

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

where Φ^c the complimentary CDF and $\mu_{\rm CDF}$ and $\sigma_{\rm CDF}$ 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 which suffer from saturation and thus rejection of low-z supernovae, or for groups of surveys treated together (as is common to do with low-redshift surveys), we find that a skew normal is a good analytic form, taking the form

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

The selection functions are fit to apparent magnitude efficiency ratios calculated from SNANA simulations, by calculating an efficiency ratio as a function of apparent magnitude. Uncertainty of the Malmquist bias (entering both through statistical uncertainty from finite sized simulations in the efficiency ratio and the discrepancy between the analytic approximation and non-analytic simulation results) is incorporated into the fitting for the analytic approximation. Uncertainty is uniformly added to the efficiency ratio 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.

With the well sampled approximation as specified previously, we can remove the redshift integral in Eq (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 respectively (denoted with a subscript 'Skew'), this correction becomes

$$w_{\rm CDF} = \Phi^c\left(\frac{m_B^* - \mu_{\rm CDF}}{\sqrt{\sigma_{m_B}^{*2} + \sigma_{\rm CDF}^2}}\right) \quad (18)$$

$$w_{\rm Skew} = 2\mathcal{N}\left(\frac{m_B^* - \mu_{\rm Skew}}{\sqrt{\sigma_{m_B}^{*2} + \sigma_{\rm Skew}^2}}\right) \times \Phi\left(\frac{\mathrm{sign}(\alpha_{\rm Skew})(m_B^* - \mu_{\rm Skew})}{\frac{\sigma_{m_B}^{*2}+\sigma_{\rm Skew}^2}{\sigma_{\rm Skew}^2}\sqrt{\frac{\sigma_{\rm Skew}^2}{\alpha_{\rm Skew}^2} + \frac{\sigma_{m_B}^{*2}\sigma_{\rm Skew}^2}{\sigma_{m_B}^{*2}+\sigma_{\rm Skew}^2}}}\right). \quad (19)$$

4 MODEL VERIFICATION

In order to verify our model we run it through several tests. First, we validate on toy models, verifying that there is not significant cosmological bias in highly constraining datasets. We then validate our model on SNANA simulations based on a collection of low redshift surveys and the DES three year spectroscopic sample.

4.1 Applied to Toy Spectroscopic Data

We generate simple toy data to validate the basic premise of the model. For both high-redshift and low-redshift (LowZ) data we draw from an underlying M_B, x_1, c population and translate into apparent magnitude space using $\Omega_m = 0.3$, $\alpha = 0.14$ and $\beta = 3.1$. Masses are randomly drawn from the interval 0 to 1, and a mass correction with $\delta(0) = 0.08$ and $\delta(0)/\delta(\infty) = 0.5$ included. Absolute magnitudes are drawn from $\mathcal{N}(-19.3, 0.1)$, true stretch values are drawn from $\mathcal{N}(0, 1)$ and true colour values are drawn from $\mathcal{N}^{\rm Skew}(0, 0.1, 2)$. Redshifts are drawn from a power law to increase the number of events as redshift increases.

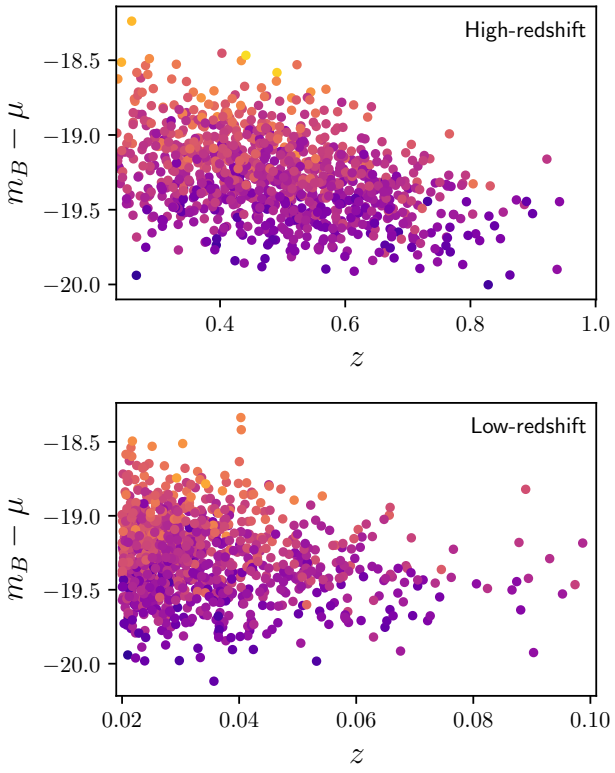

Figure 2. Population distributions shown in redshift and uncorrected absolute magnitude $m_B - \mu$ for 1000 supernovae in both high-redshift and low-redshift surveys. Selection effects are visible in both samples, where red supernovae are often cut as redshift increases.

Independent observational errors of 0.04, 0.2, 0.03 on m_B, x_1 and c (following the mean uncertainty for DES SNANA simulations) are added to create the observables. Extra dispersion of $0.02(1 + 3z)$ is added in quadrature to the c uncertainty to simulation unexplained colour dispersion from chromatic variation in supernovae. The selection functions (a skew normal for low-redshift and an error function for high-redshift) are given independent uncertainty of 0.01 on all parameters. The results are then passed through the selection effects, where each supernova is only selected based on $P(S|m_B)$, using a skew normal function for the LowZ supernovae and error function for the DES-like supernovae. We draw from each survey simulation until we have 1000 LowZ supernovae and 1000 DES-like supernovae, representing a statistical sample of greater power than the estimated 250 supernovae for the DES spectroscopic analysis. Sample data for 1000 high and low redshift supernovae are shown in Figure 2, confirming the presence of strong selection effects in both toy surveys.

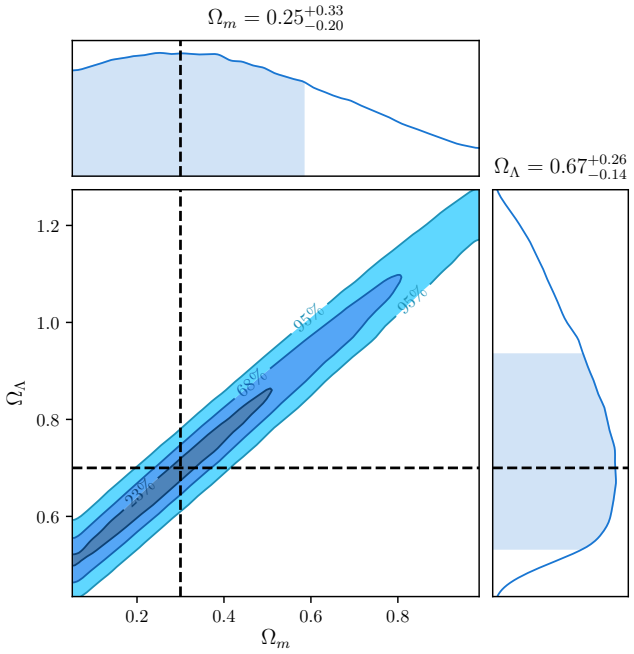

Figure 3. Posterior surfaces for 100 realisations of supernova data with the ΛCDM model. Even a large supernova sample when treated robustly is insufficient to provide tight constraints on either Ω_m and Ω_Λ due to the severe degeneracy between the parameters.

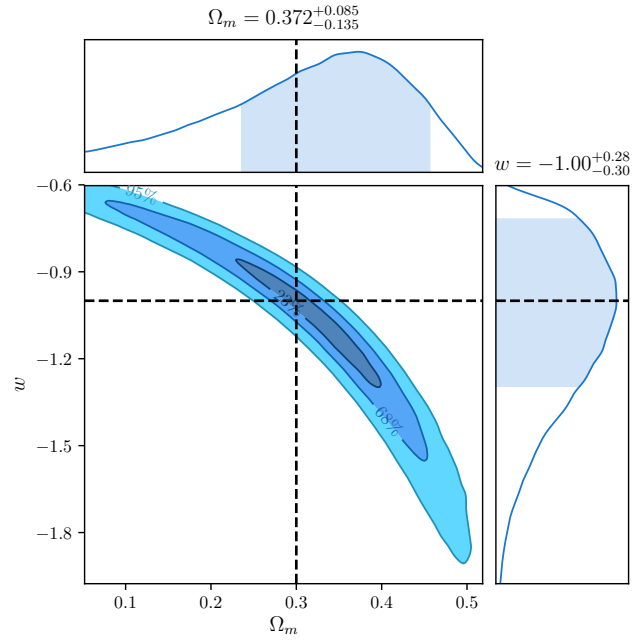

Figure 4. Posterior surfaces for 100 realisations of supernova data with the Flat wCDM model. The well known banana shaped contour is recovered, with the marginalised distributions in Ω_m and w providing incorrect statistics due to the highly non-Gaussian nature of the posterior surface.

Table 1. Maximum likelihood cosmological parameters determined from stacking the surfaces of 100 fits to independent realisations of toy supernova data. As described in the main text, each dataset comprised 1000 low-redshift supernovae and 1000 high-redshift supernovae. Model bias would appear as shifts away from the simulation values of $\Omega_m = 0.3$, $w = -1$. As our surfaces are stacked, a shift away from the simulated values of more than $\sigma/\sqrt{100} = 0.1\sigma$ would give a 1σ detection of bias. No significant bias is detected in either cosmological model.

Model	Ω_m	w
Flat ΛCDM	0.301 ± 0.016	–
Flat wCDM + Ω_m prior	$\left(300.5^{+10.5}_{-9.8}\right) \times 10^{-3}$	$-0.998^{+0.043}_{-0.047}$

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 realisations of supernovae datasets. Cosmological parameters are recovered without significant bias. Combined posterior surfaces of all 100 realisations for Flat *wCDM* fits are shown in Figure 4. By utilising the Stan framework and several efficient parametrisations (discussed further in Appendix B), fits to these simulations of 2000 supernovae take only on order of a single CPU-hour to run.

To enforce 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, and the results of the analysis are detailed in Table 1, and do not reveal evidence of systematic bias in our model.

4.2 DES SN data validation

Early analyses often treated intrinsic dispersion simply as scatter in the underlying absolute magnitude of the underlying population, but recent analyses require more a more sophisticated approach. In our development of this model and tests of intrinsic dispersion, we analyse the effects of two different scatter models. The first model is the Guy et al. (2010, hereafter denoted the G10 scatter model), which models intrinsic scatter with a 70% contribution from coherent variation and 30% from chromatic variation. The second model, denoted the C11 model is sourced from Chotard et al. (2011) and has variation with 25% contribution from coherent scatter and 75% from chromatic variation.

Simulations (using the SNANA package) follow the observational schedule and observing conditions for the DES and LowZ surveys. In addition to the improvements in the scatter models over the simple data, we also include peculiar velocities for the LowZ sample, and now also include our full treatment of systematics. Our simulated populations are sourced from Scolnic & Kessler (2016, hereafter SK16). The selection effects observed by comparing the generated supernovae to those that pass our simulated cuts and were successfully 'observed' are shown in Figure 5, and it is from this simulation that our analytic determination of the selection functions for the LowZ and DES survey are based. We note

Table 2. Tested population distributions, where the SK16 LowZ stretch distribution is formed as sum of two bifurcated Gaussians, with the mean and spread of each component given respectively.

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

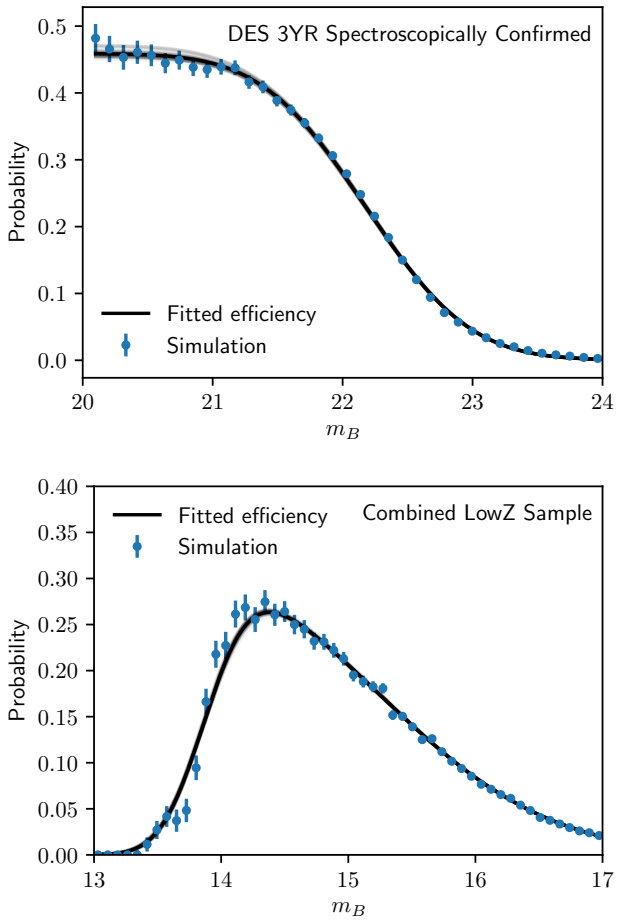

Figure 5. Fitting the selection function for both the DES 3YR spectroscopically confirmed supernova sample and the combined low-redshift sample. Blue errorbars represent the efficiency calculated by determining the ratio of discovered to generated supernovae in apparent magnitude bins for SNANA simulations. The black line represents the best fit analytic function for each sample, and the light grey lines surrounding the best fit value represent random realisations of analytic function taking into account uncertainty on the best fit value.

that there was no significant difference between using the G10 or C11 scatter model compared to the fit uncertainty.

Each realisation of cosmology fitted contains 137 LowZ supernovae, and 204 DES-like supernovae, such that the uncertainties found when combining chains is representative of the uncertainty in the final DES spectroscopic analysis. As our primary focus is upon Dark Energy, we now focus specifically on the Flat wCDM model with matter prior.

Combined posterior surfaces for 100 data realisations are shown in Figure 6. Note that we have combined the posteriors for 100 realisations, and so we should expect the size of the uncertainty to be representative of one realisation, but the statistical spread of the final surface should be $\sqrt{100} = 10$ times less than a single realisation. The parameter bounds are listed in Table 3.

Table 3. Investigating the combined 100 fits to G10 and C11 simulations, fitting with both statistics only and also when including systematics. Whilst the G10 models do not show significant evidence of bias, the C11 model does. With the current statistical size of the supernovae dataset, the bias is sub-dominant to the size of the uncertainty.

Model	w	Bias Detection Level
G10 Stat + Syst	-1.02 ± 0.13	1.5σ
C11 Stat + Syst	-0.97 ± 0.11	2.7σ
G10 Stat	-1.01 ± 0.11	0.9σ
C11 Stat	-0.95 ± 0.09	5.5σ

Cosmological bias is detected with $\approx 2.7\sigma$ significance for the C11 model and $\approx 1.5\sigma$ significance for the G10 model when fitting both systematics and statistics. For the statistics only fits, the respective bias detection levels are 0.9σ and 5.5σ. We investigate the source of the C11 bias and find its source to be bias in the observed summary statistics, in addition to incorrect reported uncertainty on those summary statistics. By replacing fully simulated observed m_B, x_1 and c with random numbers drawn from a true Gaussian centered on the simulated SALT2 m_B, x_1 and c values with covariance as reported by initial light curve fits, both the G10 and C11 fits recover $w = -1.00$ exactly. From this, the main challenge of improving our methodology is to handle the fact that observational uncertainty is incorrect, non-Gaussian and biased. Unfortunately, adding extra fit parameters to allow for shifting observables washes out cosmology, and applying a specific bias correction requires running a fiducial simulation (assuming cosmology, population and scatter model) and binning data, and is difficult to do whilst accounting 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 inherently. Colour bias corrections for the C11 scatter model are included in our cosmology model, the strength of which are controlled by a fit parameter, s_c, however neither the G10 nor C11 simulations constrained the fitting of this parameter, and thus the majority of bias in the C11 fits remain. Works such as Kessler & Scolnic (2017) show that bias corrections can be applied to supernovae datasets that can robustly handle multiple intrinsic scatter models, and future work will center on uniting these methodologies - incorporating better bias corrections without having to precompute standardisation parameters and populations.

For the sample size of the DES and LowZ supernova samples (of order 350 supernova), the bias from intrinsic scatter models are sub-dominant to the statistical uncertainty, representing at most a deviation of 0.27σ for the full statistics and systematics model fit for the C11 model, and as such we will leave more complicated treatment of them for future work.

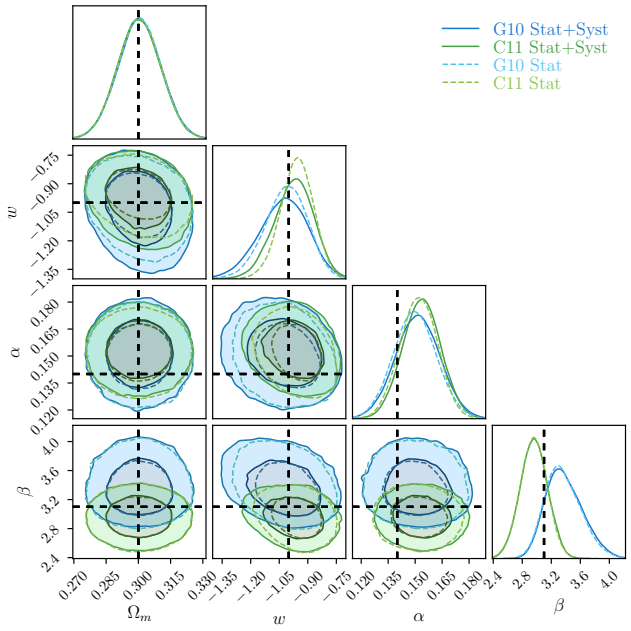

Figure 6. Posterior surfaces for 100 realisations of supernova data. Surfaces for Ω_m, w, α, β and $\langle M_B \rangle$ are shown, with the other fit parameters being marginalised over. The intrinsic scatter model has significant impact on the recovered β, which in itself effects cosmology, resulting in small biases in w.

5 CONCLUSIONS

Points to make:

1. Created more efficient BHM model with treatment of selection effects, populations, unexplained dispersion, systematics. 2. Main issue is still scatter model. Need to characterise this theoretically because adding full variance for them into the model results in weak cosmology. 3. Bias in observed summary statistics is difficult to model analytically, and future work should incorporate bias corrections calculated from simulations. However, even numerically calculated bias corrections run into issues in that the type Ia dispersion is not well characterised. 4. Model will be applied shortly to the DES 3yr spectroscopic sample.

ACKNOWLEDGEMENTS

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

REFERENCES

Abbott T., et al., 2016, Monthly Notices of the Royal Astronomical Society, 460, 1270
Alam S., et al., 2017, Monthly Notices of the Royal Astronomical Society, 470, 2617
Amanullah R., et al., 2010, The Astrophysical Journal, 716, 712
Astier P., et al., 2006, Astronomy and Astrophysics, 447, 31
Bailey S., et al., 2008, eprint arXiv:0810.3499
Balland C., et al., 2009, Astronomy and Astrophysics, 507, 85
Barbary K., et al., 2010, The Astrophysical Journal, 745, 27
Bernstein J. P., et al., 2012, The Astrophysical Journal, 753, 152
Betoule M., et al., 2014, Astronomy & Astrophysics, 568, 32
Carpenter B., et al., 2017, Journal of Statistical Software, 76, 1
Chotard N., et al., 2011, Astronomy & Astrophysics, 529, 6
Conley A., et al., 2011, The Astrophysical Journal Supplement Series, 192, 1
Contreras C., et al., 2010, The Astronomical Journal, 139, 519
D'Andrea C. B., et al., 2011, The Astrophysical Journal, 743, 172
Dilday B., et al., 2008, The Astrophysical Journal, 682, 262
Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, Publications of the Astronomical Society of Pacific, 125, 306
Freedman W. L., et al., 2009, The Astrophysical Journal, 704, 1036
Graur O., et al., 2013, The Astrophysical Journal, 783, 28
Gupta R. R., et al., 2011, ApJ, 740, 92
Guy J., et al., 2007, Astronomy and Astrophysics, 466, 11
Guy J., et al., 2010, Astronomy and Astrophysics, 523, 34
Hicken M., et al., 2009, The Astrophysical Journal, 700, 331
Hinshaw G., et al., 2013, The Astrophysical Journal Supplement Series, 208, 19
Hinton S., 2016, The Journal of Open Source Software, 1
Hlozek R., et al., 2012, The Astrophysical Journal, 752, 79
Ivezic Z., et al., 2008, eprint arXiv:0805.2366
Jennings E., Wolf R., Sako M., 2016, eprint arXiv:1611.03087, pp 1–22
Johansson J., et al., 2013, Monthly Notices of the Royal Astronomical Society, 435, 1680
Karpenka N. V., 2015, The supernova cosmology cookbook: Bayesian numerical recipes. (arXiv:1503.03844), http://arxiv.org/abs/1503.03844
Kelly P. L., Hicken M., Burke D. L., Mandel K. S., Kirshner R. P., 2010, The Astrophysical Journal, 715, 743
Kessler R., Scolnic D., 2017, The Astrophysical Journal, 836, 56
Kessler R., et al., 2009, Publications of the Astronomical Society of the Pacific, 121, 1028
Kessler R., et al., 2013, The Astrophysical Journal, 764, 48
Kessler R., et al., 2015, The Astronomical Journal, 150, 172
Kowalski M., et al., 2008, The Astrophysical Journal, 686, 749
Kunz M., Bassett B., Hlozek R., 2007, Physical Review D, 75, 1
LSST Science Collaboration et al., 2009, eprint arXiv:0912.0201
Lampeitl H., et al., 2010, The Astrophysical Journal, 722, 566
Ma C., Corasaniti P.-S., Bassett B. A., 2016, Monthly Notices of the Royal Astronomical Society, 463, 1651
Malmquist K. G. 1922, Lund Medd. Ser. I, 100, 1
Mandel K. S., Wood-Vasey W. M., Friedman A. S., Kirshner R. P., 2009, The Astrophysical Journal, 704, 629
Mandel K. S., Narayan G., Kirshner R. P., 2011, The Astrophysical Journal, 731, 120
March M. C., Trotta R., Berkes P., Starkman G. D., Vaudrevange P. M., 2011, Monthly Notices of the Royal Astronomical Society, 418, 2308
March M. C., Karpenka N. V., Feroz F., Hobson M. P., 2014, Monthly Notices of the Royal Astronomical Society, 437, 3298
Mosher J., et al., 2014, The Astrophysical Journal, 793, 16
Perlmutter S., et al., 1999, The Astrophysical Journal, 517, 565
Perrett K., et al., 2012, The Astronomical Journal, 144, 59
Phillips M. M., 1993, The Astrophysical Journal, 413, L105
Planck Collaboration et al., 2013, Astronomy & Astrophysics, 571, 66
Rest A., et al., 2014, The Astrophysical Journal, 795, 44
Riess A. G., et al., 1998, The Astronomical Journal, 116, 1009
Rigault M., et al., 2013, Astronomy & Astrophysics, 560, A66
Roberts E., Lochner M., Fonseca J., Bassett B. A., Lablanche P.-Y., Agarwal S., 2017, eprint arXiv:1704.07830
Rodney S. A., et al., 2014, The Astronomical Journal, 148, 13
Rubin D., et al., 2015, The Astrophysical Journal, 813, 15
Sako M., et al., 2014, eprint arXiv:1401.3317
Scolnic D., Kessler R., 2016, The Astrophysical Journal Letters, 822

Shariff H., Jiao X., Trotta R., van Dyk D. A., 2016, The Astrophysical Journal, 827, 1
Stan Development Team 2017, PyStan: the interface to Stan, http://mc-stan.org/
Sullivan M., et al., 2010, Monthly Notices of the Royal Astronomical Society, 406, 782
Suzuki N., et al., 2012, The Astrophysical Journal, 746, 85
Tripp R., 1998, A two-parameter luminosity correction for Type IA supernovae. Vol. 331, EDP Sciences [etc.], http://adsabs.harvard.edu/abs/1998A{%}26A...331..815T
Uddin S. A., Mould J., Lidman C., Ruhlmann-Kleider V., Zhang B. R., 2017, eprint arXiv:1709.05830
Weyant A., Schafer C., Wood-Vasey W. M., 2013, The Astrophysical Journal, 764, 116
Wood-Vasey W. M., et al., 2007, The Astrophysical Journal, 666, 694

APPENDIX A: SELECTION EFFECT DERIVATION

A1 General Selection Effects

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

$$P(\theta|\text{data}) \propto P(\text{data}|\theta)P(\theta), \tag{A1}$$

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 parametrised model, $\mathcal{L} \equiv P(\text{data}|\theta)$. When dealing with experiments which 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 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:

$$\mathcal{L} = P(\text{data}|\theta, S). \tag{A2}$$

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 characterise 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 utilising that $P(x, y, z) = P(x|y, z)P(y, z) = P(y|x, z)P(x, z)$:

$$P(\text{data}|S, \theta)P(S, \theta) = P(S|\text{data}, \theta)P(\text{data}, \theta) \tag{A3}$$

$$P(\text{data}|S, \theta) = \frac{P(S|\text{data}, \theta)P(\text{data}, \theta)}{P(S, \theta)} \tag{A4}$$

$$= \frac{P(S|\text{data}, \theta)P(\text{data}|\theta)P(\theta)}{P(S|\theta)P(\theta)} \tag{A5}$$

$$= \frac{P(S|\text{data}, \theta)P(\text{data}|\theta)}{P(S|\theta)} \tag{A6}$$

which is equal to the likelihood $\mathcal{L}$. Introducing an integral over all possible events D, so we can evaluate $P(S|\theta)$,

$$\mathcal{L} = \frac{P(S|\text{data}, \theta)P(\text{data}|\theta)}{\int P(S, D|\theta)\, dD} \tag{A7}$$

$$\mathcal{L} = \frac{P(S|\text{data}, \theta)P(\text{data}|\theta)}{\int P(S|D, \theta)P(D|\theta)\, dD}, \tag{A8}$$

where we define the denominator as w for simplicity in future derivations.

A2 Supernova Selection Effects

We assume that our selection effects can be reasonably well encapsulated by independent functions of (actual) apparent magnitude and redshift, such that $P(S|\text{data}, \theta) = P(S|z)P(S|m_B)$. Our denominator then becomes

$$w = \int d\hat{z}\, dm\hat{_B}\, dz\, dm_B P(S|z)P(S|m_B)P(\hat{z}|z)P(m\hat{_B}|m_B)P(z, m_B|\theta), \tag{A9}$$

where for simplicity we have not written out all the integrals which do not interact with the selection effects explicitly. Due to our assumed perfect measurement of redshift, $P(\hat{z}|z) = \delta(\hat{z} - z)$. $P(m\hat{_B}|m_B)$ is a Gaussian due to our Gaussian model of summary statistics m_B, x_1, c, and can be analytically integrated out, collapsing the integral over $m\hat{_B}$. Finally, 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 supernovae rates. We can thus state

$$w = \int \left[\int P(S|m_B)P(m_B|z, \theta)\, dm_B \right] P(S|z)P(z|\theta)\, dz. \tag{A10}$$

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

$$\int P(S|m_B)P(m_B|z, \theta)\, dm_B \tag{A11}$$

for each supernova in the dataset – i.e. Monte Carlo integration with assumed perfect importance sampling.

As stated in Section 3.3.5, 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

$$m^*_B(z) = \langle M_B \rangle + \mu(z) - \alpha\langle x_1(z)\rangle + \beta\left(\langle c(z)\rangle + \sqrt{\frac{2}{\pi}}\sigma_c\delta_c\right) \tag{A12}$$

$$\sigma^*_{m_B} = \sigma^2_{M_B} + (\alpha\sigma_{x_1})^2 + \left(\beta\sigma_c\sqrt{1 - \frac{2\delta_c^2}{\pi}}\right)^2. \tag{A13}$$

Then, modelling $P(S|m_B)$ as either a normal or a skew normal, we can analytically perform the integral and reach equations (18) and (19).

A3 Approximate Selection Effects

Equations (A12) and (A13) make the assumption that, for our colour 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 match the actual mean and standard deviation of skew normal. With $\delta \equiv \alpha/\sqrt{1+\alpha^2}$, the correct mean and standard deviation are

$$\mu_1 = \mu_0 + \sqrt{\frac{2}{\pi}}\delta\sigma_0 \tag{A14}$$

$$\sigma_1 = \sigma_0\sqrt{1 - \frac{2\delta^2}{\pi}}. \tag{A15}$$

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 did not produce stable posterior surfaces which Stan could fit when we included full covariance between m_B, x_1 and c in the underlying population, and so we simplified the model to have independent parameters. Stable surfaces with underlying population covariance were found when we 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. However, this resulted in several population parameters becoming biased, and so we do not fix σ_c. Comparing whether we shift our normal in the approximation or simply discard skewness, Figure A1 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 A2, and only the shifted normal approximation correctly recovers underlying population parameters.

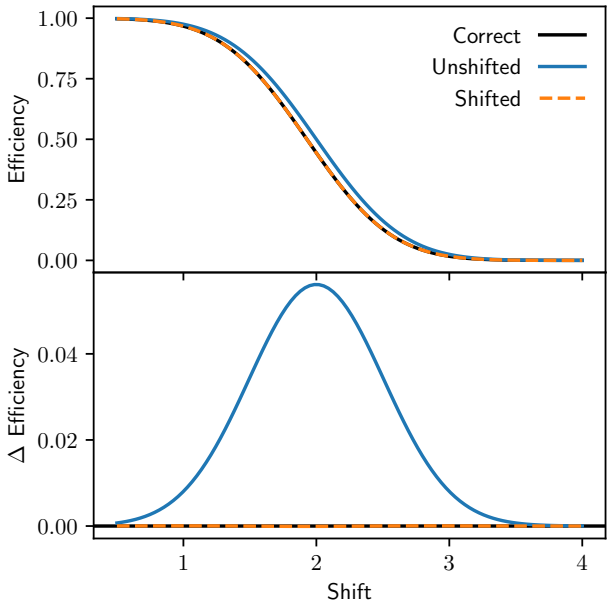

Figure A1. Testing the correctness of our normal approximation to the skewed colour distribution. The 'correct' line (shown in black) represents the exact integral $w = \int P(S|x)P(x)dx$ where $P(S|x)$ is an error function (following our high-redshift surveys) and $P(x) = \mathcal{N}^{\text{Skew}}(x, 0.1, 2)$, calculated numerically. The x-axis is analogous to m_B is cosmological context. As expected, all efficiencies drop towards zero as shift increases (as objects get fainter). The unshifted normal approximation shows significant discrepancy in the calculated efficiency as it transitions from 1 to 0, whilst the shifted normal approximation shows negligible error to the correct solution. From these plots, further refinement of the normal approximation (such as including kurtosis or higher powers) as unnecessary.

APPENDIX B: NUMERICAL OPTIMISATIONS

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 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 far exceeded.

The simplest and most commonly found optimisation we employed was to precompute as much as possible to reduce the complexity of the mathematical graph our model is translated into to compute the 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 results.

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 reparameterise it such that instead of sampling $\mathcal{N}(\vec{x}|\vec{\mu}, \Sigma)$, you sample $\mathcal{N}(\vec{\delta}|0, 1)$ where $\vec{x} = \vec{\mu} + L\vec{\delta}$ 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 parametrisation resulted in a computational increase of an order of magnitude, taking fits for a sample of approximately 500 supernovae from roughly four hours down to twenty minutes.

This parametrisation 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 parametrisation we could modify the matrix easily - for example when incorporating intrinsic dispersion we could simple add on a secondary matrix to create an updated covariance. Using cholesky decompositions, $\text{cholesky}(\Sigma_1 + \Sigma_2) \neq L_1 + L_2$, and so 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

$$L = \begin{pmatrix} a & 0 & 0 \\ b & c & 0 \\ d & e & f \end{pmatrix}, \tag{B1}$$

the original covariance matrix Σ is given by

$$\Sigma = \begin{pmatrix} a^2 & ab & ad \\ ab & b^2 + c^2 & bd + ce \\ ad & bd + ce & d^2 + e^2 + f^2 \end{pmatrix}. \tag{B2}$$

Now, the primary source of extra uncertainty in the in-

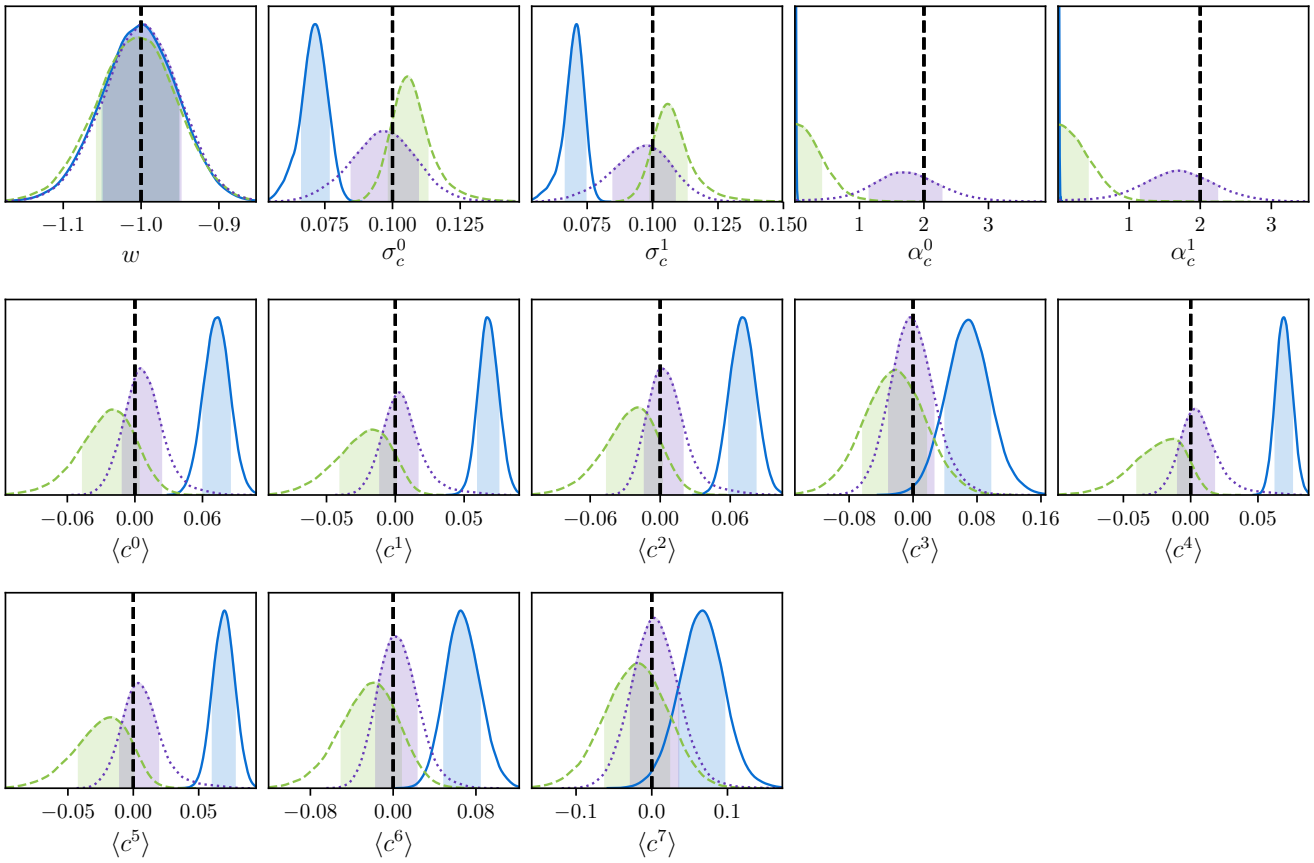

Figure A2. Marginalised probability distributions for 100 realisations of cosmology, fit to Flat wCDM with prior $\Omega_m \sim \mathcal{N}(0.3, 0.01)$, each containing 1000 simulated high-z and 1000 simulated low-z supernovae. The dashed green surfaces represent a fit to an underlying Gaussian colour population with the unshifted model. The blue solid surface represents fits to a skewed colour population with the unshifted model, and the purple dotted surface represents a fit to a skewed colour population with the shifted model. The superscript 0 and 1 denote the two different surveys (high-z and low-z respectively), and 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 colour $\langle c \rangle$ and width of the colour distribution σ_c. The unshifted model recovers the correct colour 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 fits to the colour population mean. Importantly, the unshifted model when run on skewed data (the solid blue) shows extreme bias in α_c, where it fits strongly around zero regardless, showing it to be a poor approximation.

trinsic dispersion models comes from chromatic smearing, which primarily influences the recovered colour 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 colour 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 colour an extra σ_e term, we get

$$C = \begin{pmatrix} \sigma_{m_B}^2 & \rho_{0,1}\sigma_{m_B}\sigma_{x_1} & \rho_{0,2}\sigma_{m_B}\sigma_c \\ \rho_{0,1}\sigma_{m_B}\sigma_{x_1} & \sigma_{x_1}^2 & \rho_{1,2}\sigma_{x_1}\sigma_c \\ \rho_{0,2}\sigma_{m_B}\sigma_c & \rho_{1,2}\sigma_{x_1}\sigma_c & \sigma_c^2 + \sigma_e^2 \end{pmatrix} \tag{B3}$$

. The cholesky decomposition of this is, in terms of the original cholesky decomposition, is

$$L = \begin{pmatrix} a & 0 & 0 \\ b & c & 0 \\ d & e & f+g \end{pmatrix}, \tag{B4}$$

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

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