MNRAS 000, 1–17 (2017) Preprint 24 September 2018 Compiled using MNRAS LATEX style file v3.0

Steve: 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

We present a hierarchical Bayesian model for use in supernova cosmology which takes into account a full range of analysis systematics and selection effects. This advances previous works with hierarchical Bayesian models by including additional sources of systematic uncertainty and improving treatment of Malmquist bias, whilst increasing numerical efficiency. Our model is optimised for efficient fitting and has been validated on a combined simulated dataset of more than 250 000 supernovae to rule out the presence of significant model biases in parameter recovery. We demonstrate its effectiveness by fitting simplified statistical simulations for ideal supernova cosmology, and sophisticated SNANA simulations configured to mimic the Dark Energy Survey (DES) 3-year spectroscopic supernova sample. SNANA simulations can recover the input cosmology between 0 and 0.5σ bias depending on the intrinsic scatter model for the DES 3-year spectroscopic sample, while taking into account an increased number of systematics and increasing simulation independent over previous analyses. Future development will focus on increasing model agnosticism towards intrinsic scatter model. Have left out explicit references for 'previous analyses' as I dont want to bloat the abstract. Its talked about in the paper, so is this alright?

Key words: keyword1 – keyword2 – keyword3

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 (SNIa) have 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. 2009a; 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; Sako et al. 2014). 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. 2017) have been combined with complementary probes of large scale structure and the CMB. For a recent review, see Huterer & Shafer (2018). Despite these prodigious efforts, the nature of dark energy remains an unsolved mystery.

In attempts to tease out the nature of dark energy, active and planned surveys are once again ramping up their statistical power. The Dark Energy Survey (DES, Bernstein et al. 2012; Abbott et al. 2016) has observed thousands of Type Ia supernova, attaining both spectroscopic and photometric selected samples. The Large Synoptic Survey Telescope (LSST, Ivezic et al. 2008; LSST Science Collaboration et al. 2009) will observe scores of thousands of photometrically classified supernovae. Such increased statistical power demands greater fidelity and flexibility in modelling the supernovae for cosmological purposes, as systematic uncertainties will prove to be the limiting factor in our analyses (Betoule et al. 2014; Scolnic et al. 2017).

As such, staggering effort is being put into 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 cosmological models. First used in ESSENCE analyses (Wood-Vasey et al. 2007), and then refined and improved in Kessler et al. (2009b), simulations are now a fundamental component of modern supernovae cosmology. Betoule et al. (2014) quantise and then correct observational bias using simulations, and more recently Scolnic & Kessler (2016) and Kessler & Scolnic (2017) explore simulations to quantify observational bias in cosmological parameters as a function of many factors to better remove fitting biases. Approximate Bayesian computation 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). 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), and either also use simulation corrections or attempt to find sufficient analytic approximations for complicated effects such as Malmquist bias.

In this paper, we lay out a new hierarchical model that builds

★ E-mail: samuelreay@gmail.com

off 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 supernova cosmology, and Section 3 outlines some of the common challenges faced by analysis methods. In Section 4 we outline our methodology. Model verification on simulated datasets is given in Section 5, along with details on potential areas of improvement. We summarise our methodology in Section 6.

2 REVIEW

Whilst supernova observations take the form of time-series photometric measurements of brightness in many photometric bands, 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 (Betoule et al. 2014; Scolnic et al. 2017). Ive just swapped out references here as this isnt a critical point for my analysis to drill down on. 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, $\mathcal{C}$. As all supernova 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. Removed confusing statement on intrinsic dispersion.

This represents an observed population, which – due to the presence of various selection effects – may not represent the true, underlying supernova population. Accurately characterising this underlying population, its evolution over redshift, and effects from environment, is one of the challenges of supernova cosmology. Given some modelled underlying supernova population that lives in the redshift-dependent space M_B (absolute magnitude), x_1 and c, the introduction of cosmology into the model is simple – it translates the underlying population in absolute magnitude space into the observed population living in apparent magnitude space. Specifically, for any given supernova our map between absolute magnitude and apparent magnitude may take the traditional form:

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

where M_B is the mean absolute magnitude for all SN Ia α is the stretch correction (Phillips 1993; Phillips et al. 1999), and β is the colour correction (Tripp 1998) that respectively encapsulate the empirical relation that broader (longer-lasting) and bluer supernovae are brighter. The cosmological term, $\mu(z)$ represents the distance modulus, and can be precisely calculated given cosmological parameters and a redshift, or can be 'observed' by rearranging equation (1) to isolate $\mu(z)$. The 'other corrections' term often includes corrections for host galaxy environment, as this has statistically significant correlations with 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). In traditional analyses, the other corections term often includes Malmquist bias corrections, which can take the form of a redshift-dependent function (Betoule et al. 2014).

2.1 Traditional Cosmology Analyses

Traditional χ^2 analyses such as that 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); Betoule et al. (2014), minimise the difference in distance modulus between the observed distance modulus attained from rearranging equation 1, and the cosmologically predicted values. To distinguish them, we denote the observed distance modulus μ_{obs} and the cosmologically calculated one μ_C. This function is given by

$$\chi^2 = (\mu_{\text{obs}} - \mu_C)^\dagger C^{-1} (\mu_{\text{obs}} - \mu_C), \quad (2)$$

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 cosmological μ_C is defined as

$$\mu_C = 5\log\left[\frac{(1+z)r}{10}\right], \quad (3)$$

$$r = \frac{c}{H_0}\int_0^z \frac{dz'}{\sqrt{\Omega_m(1+z')^3 + \Omega_k(1+z')^2 + \Omega_\Lambda(1+z')^{3(1+w)}}}, \quad (4)$$

where r is the comoving distance for redshift z given a specific cosmology, H_0 is the current value of Hubble's constant in $\mathrm{km\,s^{-1}\,Mpc^{-1}}$ 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. Firstly, formulating a χ^2 likelihood requires a loss of model flexibility by building into the model assumptions of Gaussian uncertainty. Secondly, the method of creating a global covariance matrix relies on computing partial derivatives and thus any uncertainty estimated from this method loses all information about correlation between sources of uncertainty. For example, the underlying supernova colour population's mean and skewness are highly correlated, however this correlation is lost when determining population uncertainty using numerical derivatives of population permutations. Thirdly, 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. A common solution to this computational cost problem is to bin the supernovae, which then produces a matrix of arbitrary size. Whilst binning data does result in some loss of information, recent works tested against simulationsshow that this loss of information does not create significant cosmological biases (Scolnic & Kessler 2016; Kessler & Scolnic 2017).

Selection efficiency, such as the well known Malmquist bias (Malmquist K. G. 1922) is accounted for by correcting data. Specifically, Malmquist bias is the result of losing the fainter tail of the supernova population at high redshift when supernovae increasingly fall below the detection threshold. An example of Malmquist bias is illustrated in Figure 1. 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. When using traditional fitting methods, these effects are not built into the likelihood and instead are formed by correcting data, which means their uncertainty is not captured fully in the χ^2 distribution,

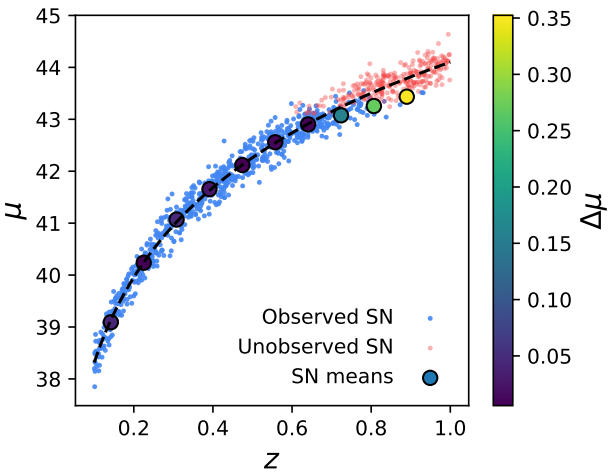

Figure 1. An example of the effects of Malmquist bias. Here are shown 1000 simulated supernovae redshifts and distance modulus given fiducial cosmology. The simulated survey is magnitude limited, and all supernovae brighter than magnitude 24 are successfully observed (shown as blue dots), and all dimmer than 24th magnitude are not successfully observed (shown as red dots). By binning the supernovae along 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 either be corrected by adjusting the data (such as subtracting the found bias) for by incorporating Malmquist bias explicitly in the cosmological model.

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 approaches, 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. 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 that plague supernova surveys. By using forward modelling directly from these simulations, 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 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, modelling 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) presents two metrics, which are used to measure the distance between the forward modelled population and observed population, and are minimised in fitting. The first metric compares forward modelled summary statistic populations (denoted the 'Tripp' metric) and the second utilises 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 Hierarchical Bayesian Models

Sitting between the traditional models simplicity and the complexity of forward modelling lies Bayesian hierarchical models (BHM). Hierarchical models utilise multiple layers of connected parameters in their models, with the layers linked via well defined and physical 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 easily incorporate different distributions, populations, and parameter inter-dependence which cannot be found in traditional analyses where uncertainty must be encapsulated in a covariance matrix.

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, 2011, 2017) construct a hierarchical model that they apply to supernova light curve fitting. March et al. (2011) derive a hierarchical model and simplify it by analytically marginalising over nuisance parameters to provide increased flexibility with reduced uncertainty over the traditional method, but do not incorporate bias correction. March et al. (2014); Karpenka (2015) improve upon this by incorporating redshift-dependent magnitude corrections from Perrett et al. (2010) to remove bias, and validate on 100 realisations of SNLS like simulations. The recent BAHAMAS model (Shariff et al. 2016) builds on this and reanalyses the JLA dataset (using redshift dependent bias corrections from Betoule et al. 2014), whilst 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 dataset 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 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 dataset (Suzuki et al. 2012). The simplification made by the UNITY analysis is that the bias is well described by an analytic function, and this function is fixed such that selection effects are assumed to be determined with perfect precision. Their model was validated to be free of significant biases using fits to thirty realisations of supernova datasets. 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). For the latter case, by inferring biases using Bayesian models, sophisticated corrections can be calculated and then applied to more traditional χ^2 models.

Whilst there are a large number of hierarchical models available, all so far mentioned, in the case where any validation has been done, have been done on either ΛCDM cosmology or Flat ΛCDM cosmology. None of them have undergone high-statistics simulation verification for wCDM cosmology to quantify each models'

respective bias, which is becoming critically important as precision supernovae cosmology comes into its own and focus shifts from determination of Ω_m to w. Dan, you mention underselling the prize here. Any suggestions on an appropriate sell, so to speak?

The flexibility afforded by hierarchical models allows for investigations into different treatments of underlying supernova magnitude, colour 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 whilst incorporating the flexibility of hierarchical models and increasing simulation independence to provide a valuable cross-check on analysis methodologies which are highly simulation-dependent.

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 when performing supernova cosmology analyses. In this Section, we outline several of the most prominent challenges.

Forefront among these challenges is our ignorance of the underlying Type Ia population dispersion. Ideally, analysis of the underlying population would make use of an ensemble of time-series spectroscopy to characterise the diversity of Type Ia supernovae, however this data is difficult to obtain, and recent efforts to quantify the population draw inference from photometric measurements. We utilise two dispersion models in this work. The first is the Guy et al. (2010, hereafter denoted G10) scatter model, which models intrinsic scatter with a 70% contribution from coherent variation in the spectral energy distribution and 30% from chromatic variation. The second, 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. 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 $\mathcal{C}$. These two scatter models are currently considered sufficient to span the possible range of scatter in the underlying supernova population. We have insufficient information to prefer one model over the other, and thus we have to account for both possible scatter models.

The underlying supernova population is further complicated by the presence of outliers. Non-Ia supernovae often trigger transient follow-up in surveys and can easily be mistaken for Type Ia supernovae. This contamination is not just a result of non-SNIa being observed, but can also arise from host galaxy misidentification causing incorrect redshifts. Depending on the cuts on the optimised ratio between purity and efficiency, this can result in between 3% to 9% misidentification (Gupta et al. 2016) and results 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 SNIa, datasets are commonly formed from the discoveries of multiple surveys. However, each different 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 which must all be modelled and accounted for.

4 OUR METHOD

We construct our hierarchical Bayesian model with several goals in mind: creation of a redshift-dependent underlying supernova population, increased treatment of systematics, and analytic correction of selection effects, including systematic uncertainty on those corrections. We also desire this to be more computationally efficient than prior works, such that cosmological results from thousands of supernovae are obtainable in the order of hours, rather than days. 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 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 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. In the following sections, we will describe these layers, mapping functions, and occurrences of these fatal pathologies. 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 flux measurement $\hat{m}_B$ (which is analogous to apparent magnitude), stretch $\hat{x}_1$ and colour $\hat{c}$, with uncertainty on those values encoded in the covariance matrix $\mathcal{C}$. 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. We will also have a probability of each supernovae 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 2.

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 $P(\hat{z}|z) = \delta(\hat{z} - z)$. Potential sources of redshift error (such as peculiar velocities) are taken into account not via uncertainty on redshift (which is technically challenging to implement) but instead uncertainty on distance modulus. This is discussed further in Section 4.3.4. 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 this model (and the R15 analysis) is that for future data sets where supernovae have been classified photometrically,

Table 1. Parameters defined in our model and a summary of their use. The parameters are broken into multiple sections, the top for parameters which are defined globally across all surveys, the second section for parameters which are defined for each survey, the third section are parameters which are defined for each supernova. The bottom section shows the observables (not parameters) that are the input data to the model.

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

and we expect some misclassification and misidentification of the host galaxies, those can naturally be modelled and taken into account by introducing secondary populations that supernovae have a non-zero probability of belonging to.

4.1.2 *Latent Variables for Observables*

The first layer of the hierarchy represents the parameters that describe each supernova. That is, observed parameters are denoted with a hat, whilst the true (latent) value is denoted without a hat. For example, c is the true colour of the supernova, whilst $\hat{c}$ is the colour we observe, which, as it has uncertainty, is different from c.

For the moment, let us consider a single supernova and it's classic summary statistics m_B, x_1, 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 curves and uncertainty. However, this is computational prohibitive, and as such we rely on initially fitting the light curve observations to produce a best fit $\hat{\eta}$ along with a 3×3 covariance matrix $\mathcal{C}$ describing the uncertainty on $\hat{\eta}$. When using this simplification, our latent variables are given by

$$p(\hat{\eta}|\eta) \sim \mathcal{N}(\hat{\eta}|\eta, \mathcal{C}). \quad (5)$$

As discussed in Section 3, this approximation is known to fail due to the SALT2 model not including intrinsic dispersion in the supernova model.

4.1.3 *Correcting biased summary statistics*

With the report summary statistics being biased and their uncertainty under-reported, we face a significant challenge 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 must also do this without assuming a specific dispersion model – either the G10 or C11 model.

We model the extra dispersion only in colour, and do so by adding extra independent uncertainty on the colour observation. We note that extra dispersion in magnitude (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 colour term, hence the requirement for modelling additional colour dispersion. Tests on incorporating extra dispersion on stretch as well show that stretch is less biased than colour, and cause negligible bias in cosmology.

As shown in (Kessler et al. 2013), the extra colour dispersion shows heavy redshift dependence, increasing with redshift. As the extra dispersion may arise from underlying supernova physics, rather than 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). The κ parameters are highly degenerate with the width of the intrinsic colour 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 $\mathcal{C}_{\rm tot} = \mathcal{C} + \text{DiagMatrix}\left[0, 0, (\kappa_0 + \kappa_1 z)^2\right]$.

Fully covariant extra dispersion on $\{m_B, x_1, c\}$ (rather than just dispersion on c) was also tested, by modelling the dispersion as a multivariate Gaussian, but it showed negligible improvement in recovering unbiased cosmology over just colour dispersion, and was far more computationally inefficient. We note here that we do model dispersion in magnitude, but this is done at the level of underlying populations, not observed populations. This extra dispersion is modelled with redshift independence. I'm not too sure what you mean about telling a story here. I've tried to change the wording to put it back in line with the method, is that sufficient? Also, I did not try redshift dependent extra scatter on m_B.

4.2 Underlying Population

4.2.1 *Type Ia population*

The underlying supernova population is characterised by underlying distributions in colour, stretch and absolute magnitude. We follow the prior work of R15 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 to

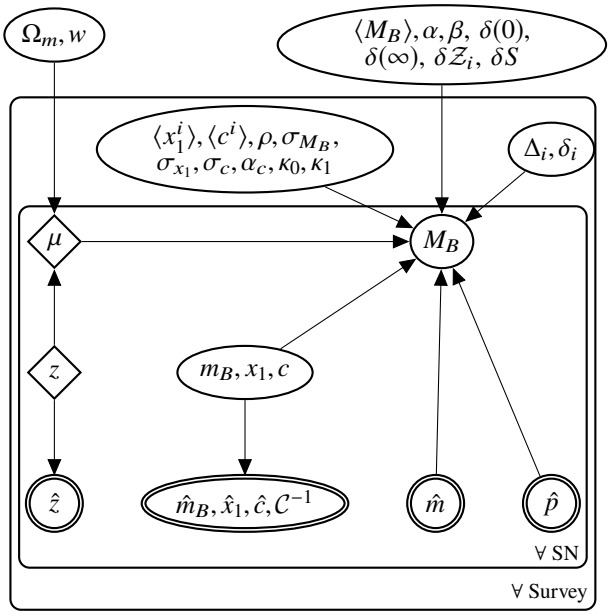

Figure 2. Probabilistic graphical model for our likelihood. Double-lined nodes represent observed variables, diamond nodes represent deterministic variables, single-lined ellipse nodes represent fit variables. The SN box represents observed variables and latent variables for each individual supernova, whilst 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.

determine the mean stretch and colour for a given redshift. Each set of nodes are anchored to equally spaced to span the minimum to maximum redshift of each survey. 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 parametrise the skewness α_c by sampling $\delta_c = \alpha_c/\sqrt{1+\alpha_c^2}$ which itself is given a uniform prior $\mathcal{U}(0, 0.98)$, which allows α_c to span positive values in the realm of physical plausibility as determined from constraints in Scolnic & Kessler (2016). 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 mean absolute magnitude $\langle M_B\rangle$, 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

$$P(M_B, x_1, c, z|\langle M_B\rangle, \langle x_1(z)\rangle, \langle c(z)\rangle, \sigma_{m_B}, \sigma_{x_1}, \sigma_c, \alpha_c) = \mathcal{N}(M_B|\langle M_B\rangle, \sigma_{M_B})\mathcal{N}(x_1|\langle x_1(z)\rangle, \sigma_{x_1})\mathcal{N}^{\text{skew}}(c|\langle c(z)\rangle, \sigma_c, \alpha_c). \quad (6)$$

4.2.2 *Outlier populations*

For future use with photometrically classified surveys, we include a simplistic outlier population model. We follow R15 (and therefore Kunz et al. 2007) by implementing a Gaussian 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 SNIa population in prior works. However, with 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 parametrisation 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 to be greater than a small positive number to reduce degeneracy between the two populations. We set the population width to $\sigma^{\text{outl}} = 1$ in M_B, x_1 and c. The probability of each supernovae falling into either population is determined by the observed Type Ia probability $\hat{p}$ discussed in the previous section. For the spectroscopic survey, we set this to unity. For the photometric proof of concepts we provide an accurate probability estimate. Further investigation on the effect of inaccurate estimates we leave for future improvements during the analysis of the DES photometric sample.

4.3 Mapping function

4.3.1 *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 $H_0 = 70\text{km s}^{-1}\text{ Mpc}^{-1}$. 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.

4.3.2 *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 model β as redshift-dependent, testing two phenomenological models; $\beta(z) = \beta_0 + \beta_1 z$ and a second model which 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 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. Removed comments about varying alpha and beta. Due to the recovery of non-zero 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 (4).

4.3.3 *Host Galaxy Environment*

It is now well known that host galaxy environment is significantly correlated with 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, M is the galaxy mass in solar masses and 0.08 represents the size of the magnitude step. 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:

$$\Delta M = \delta(0)\left[\frac{1.9\left(1-\frac{\delta(0)}{\delta(\infty)}\right)}{0.9+10^{0.95z}}+\frac{\delta(0)}{\delta(\infty)}\right], \tag{7}$$

where $\delta(0)$ represents the correction at redshift zero, and $\delta(\infty)$ a parameter allowing the behaviour to change with increasing redshift. We take flat priors on $\delta(0)$ and $\delta(0)/\delta(\infty)$. With this correction, our calculation of absolute magnitude becomes

$$M_B = m_B - \mu(z) - \alpha x_1 + \beta c - \Delta M \times \hat{m}. \tag{8}$$

4.3.4 *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. 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 HST Calspec calibration uncertainty, ten 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. With $\eta \equiv \{m_B, x_1, c\}$, our initial conditional likelihood for our observed summary statistics shown in Equation (5) becomes

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

4.3.5 *Selection Effects*

One large difference between traditional methods 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 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

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

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

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

where the denominator represents an integral over all potential data D. 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 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 (11), denoted now d for simplicity, is given as

$$d = \int \left[\int P(S|m_B)P(m_B|z, \theta)\, dm_B\right] P(S|z)P(z|\theta)\, dz, \tag{12}$$

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

We now apply two further approximations similar to those made in R15 – that the redshift distribution of the observed supernova reasonably well samples 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 evaluating $P(m_B|z, \theta)$. It was found that discarding the colour population skewness entirely resulted in highly biased population recovery (see Figure A1 to see the populations), 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 for colour completely removes the unintended bias when simply discarding skewness. This shift was not required for the stretch population, and so was left out for numerical reasons. The impact of this approximation on the calculated efficiency is shown in Figure 3, and more detail on this shift and resulting population recovery 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

$$m_B^*(z) = \langle M_B\rangle + \mu(z) - \alpha\langle x_1(z)\rangle + \beta\langle c(z)\rangle \tag{13}$$

$$\sigma_{m_B}^{*2} = \sigma_{M_B}^2 + (\alpha\sigma_{x_1})^2 + (\beta\sigma_c)^2. \tag{14}$$

What then remains is determining the functional form of

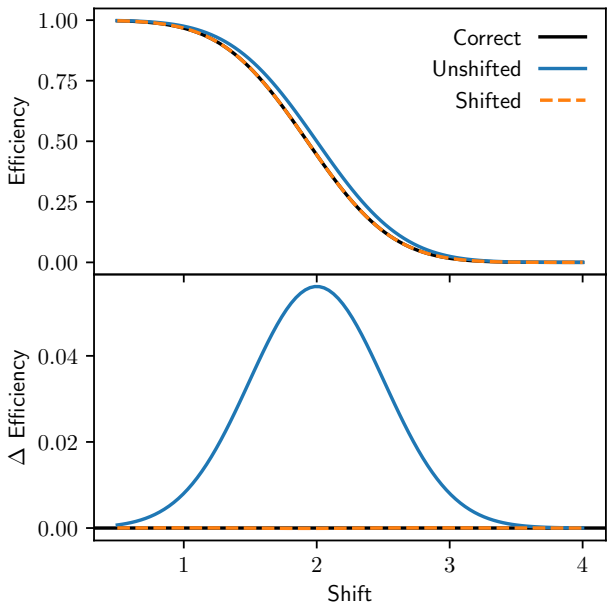

Figure 3. Testing the correctness of our normal approximation to the skewed colour distribution. The 'correct' line (shown in black) represents the exact integral $w = \int P(S|x)P(x)dx$ where $P(S|x)$ is an error function (following our high-redshift surveys) and $P(x) = \mathcal{N}^{\rm Skew}(x, 0.1, 2)$, 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.

$P(S|m_B)$. 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

$$P(S|m_B) = \Phi^c(m_B|\mu_{\rm CDF}, \sigma_{\rm CDF}), \tag{15}$$

where Φ^c the complimentary cumulative distribution function 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-redshift 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 better analytic form, taking the form

$$P(S|m_B) = \mathcal{N}^{\rm Skew}(m_B|\mu_{\rm Skew}, \sigma_{\rm Skew}, \alpha_{\rm Skew}). \tag{16}$$

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. That is, we calculate the probability 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 one, 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 can thus include into our model parametrised selection effects 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. Whilst this uncertainty encapsulates the potential error from the simulations not matching the analytic approximations, it does not cover 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 approximation we can remove the redshift integral in Eq (12) 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'), the correction *per SNIa* becomes

$$d_{\rm CDF} = \Phi^c\left(\frac{m_B^* - \mu_{\rm CDF}}{\sqrt{\sigma_{m_B}^{*2} + \sigma_{\rm CDF}^2}}\right) \tag{17}$$

$$d_{\rm Skew} = 2\mathcal{N}\left(\frac{m_B^* - \mu_{\rm Skew}}{\sqrt{\sigma_{m_B}^{*2} + \sigma_{\rm Skew}^2}}\right)\times \Phi\left(\frac{\mathrm{sign}(\alpha_{\rm Skew})(m_B^* - \mu_{\rm Skew})}{\frac{\sigma_{m_B}^{*2}+\sigma_{\rm Skew}^2}{\sigma_{\rm Skew}^2}\sqrt{\frac{\sigma_{\rm Skew}^2}{\alpha_{\rm Skew}^2} + \frac{\sigma_{m_B}^{*2}\sigma_{\rm Skew}^2}{\sigma_{m_B}^{*2}+\sigma_{\rm Skew}^2}}}\right), \tag{18}$$

and is incorporated into our likelihood. This is illustrated in Figure 4. Our corrections for the DES spectroscopic data utilise the CDF functional form, with the combined low redshift surveys being modelled with the skew normal efficiency. Further details on this choice are given in Section 5.2.

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 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.

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 $\mathcal{U}(0.008, 1.0)^{0.3}$.
2. Draw a true absolute magnitude, stretch and colour from the respective distributions $\mathcal{N}(-19.3, 0.1)$, $\mathcal{N}(0, 1)$, $\mathcal{N}(0, 0.1)$.

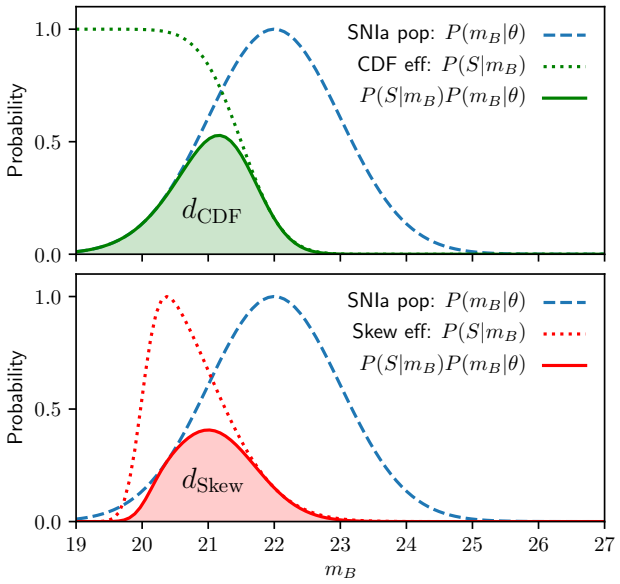

Figure 4. The efficiency for supernova discovery at an arbitrary redshift. Shown in both panels in dashed blue is the SNIa population distribution, which takes the form of a normal distribution. The top panel shows a CDF based survey efficiency (green dotted line), whilst the bottom panel shows a skew normal based survey efficiency (red dotted line), as functions of apparent magnitude. The survey efficiency, given the SNIa population, is shown as a solid line in both panels, and the probability of observing a SNIa is found by integrating over the population detection efficiency as described in equation (12), and has been shown by shading the area integrated. This area is what is analytically given by equations (17) and (18).

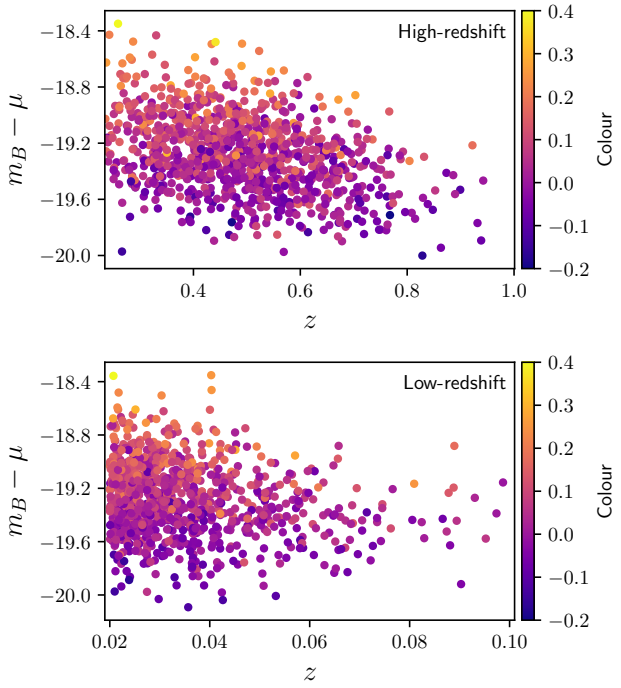

Figure 5. Population distributions shown in redshift and uncorrected absolute magnitude $m_B - \mu$ for 1000 supernovae in both high-redshift and low-redshift surveys. Selection effects are visible in both samples, where red supernovae are often cut as redshift increases. The colour of the data points is representative over the supernovae colour itself, a negative colour value showing bluer supernovae, with positive colour values representing redder supernovae.

Table 2. Cosmological parameters determined from 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. For each chain, we record the mean and standard deviation, and then show the average mean and average standard deviations in the table. The scatter introduced by simulation variance (the standard deviation of the 100 mean parameter values) is shown in brackets. Model bias would appear as shifts away from the simulation values of $\Omega_m = 0.3$, $w = -1$. As we are using 100 independent realisations, the precision of our determination of the mean simulation result is a tenth of the quoted standard deviation: $\sqrt{100} = 10$. As the deviation from truth values is below this threshold, no significant bias is detected in either the Flat ΛCDM model or the Flat wCDM model. For the Flat wCDM model, the value of w is reported with a prior on Ω_m of $\mathcal{N}(0.3, 0.01)$.

Model	Ω_m(scatter)	w(scatter)
Flat ΛCDM	$0.301 \pm 0.015(0.012)$	–
Flat wCDM	–	$-1.00 \pm 0.042(0.030)$

3. 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 (7).
4. Calculate $\mu(z)$ given the drawn redshift and cosmological parameters $\Omega_m = 0.3$, $w = -1$. Use this to determine the true apparent magnitude of the object m_B using equation (8).
5. Determine if the SNIa is detected using detection probability $P(S|m_B) = \mathcal{N}^{\rm skew}(13.72, 1.35, 5.87)$ for the low redshift survey (numeric values obtained by fitting to existing low redshift data).

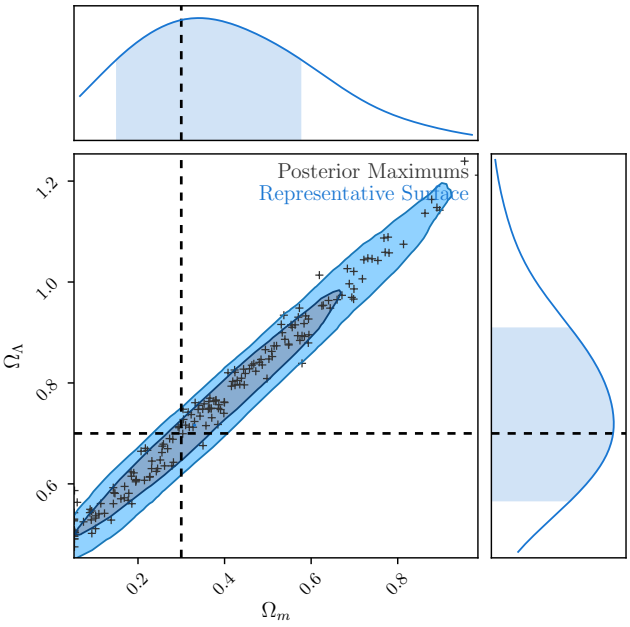

Figure 6. Maximal posterior points for 100 realisations of supernova data with the Flat ΛCDM model, with a representative contour from a single data realisation shown for context. Even a large supernova sample when treated robustly is insufficient to provide tight constraints on either Ω_m or Ω_Λ separately due to the severe degeneracy between the parameters.

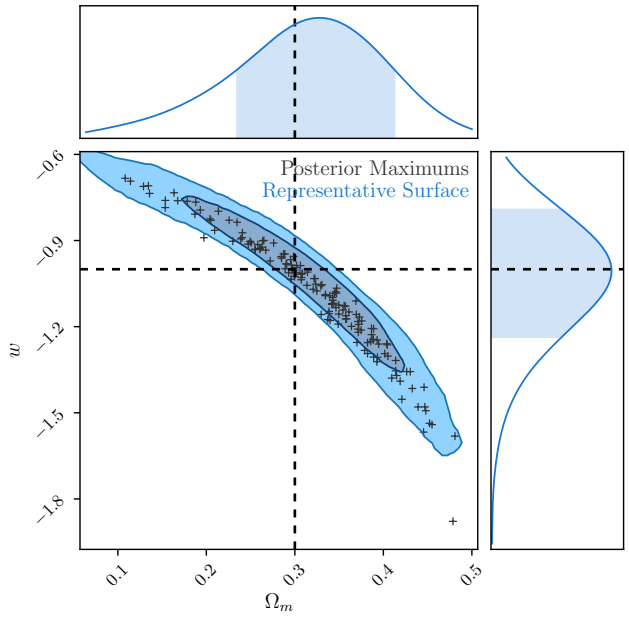

Figure 7. Maximal posterior points for 100 realisations of supernova data with the Flat wCDM model, with a representative contour from a single data realisation shown for context. The well known banana shaped contour is recovered, with the marginalised distributions in Ω_m and w providing misleading statistics due to the non-Gaussian nature of the posterior surface. The recovered posteior maximums show the same degeneracy direction as the representative surface, and scatter around the truth values input into the simulation, which are shown in dashed lines.

For the DES-like survey, accept with probability $P(S|m_B) = \Phi^C(23.14, 0.5)$. Repeat from step one until we have a supernova that passes.

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

For supernova-independent uncertainty, 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 (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 LowZ supernovae and 1000 DES-like supernovae, representing a statistical sample of greater power than the estimated 350 supernovae for the combined DES three-year and low-redshift spectroscopic analysis. 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 realisations of supernovae datasets. Cosmological parameters are recovered without significant bias. Combined posterior surfaces of all 100 realisations fits for ΛCDM are shown in Figure 6 and fits for Flat wCDM are shown in Figure 7. 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.

Table 3. 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	$^{+0.101}_{0.043}$

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 can appear to be large 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.

5.2 DES SN data validation

Early 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 analyse the effects of two different scatter models, the G10 and C11 models described in Section 3.

Simulations (using the SNANA package) follow the observational schedule and observing conditions for the DES and LowZ surveys, where the LowZ sample is comprised of CfA3 (Hicken et al. 2009a,b), CfA4 (Hicken et al. 2012) and CSP (Contreras et al. 2010; Folatelli et al. 2010; Stritzinger et al. 2011).

In addition to the improvements in the scatter models over the simple data, we also include peculiar velocities for the LowZ sample, and our full treatment of systematics. 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 colour and stretch populations centered on zero and with respective width 0.1 and 1, however cosmological parameters were not impacted by choice of the underlying population and we continue using only the SK16 population for computational efficiency. The selection effects were quantified by comparing all the generated supernovae to those that pass our simulated cuts, as shown in Figure 8. It is from this simulation that our analytic determination of the selection functions for the LowZ 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 selection effect of Malmquist bias between the two models.

Each realisation of simulated cosmology 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 Dark Energy, we now focus specifically on the Flat wCDM model with matter prior.

Points of maximum posterior for 100 data realisations are shown in Figure 9. The parameter bounds are listed in Table 4. Our

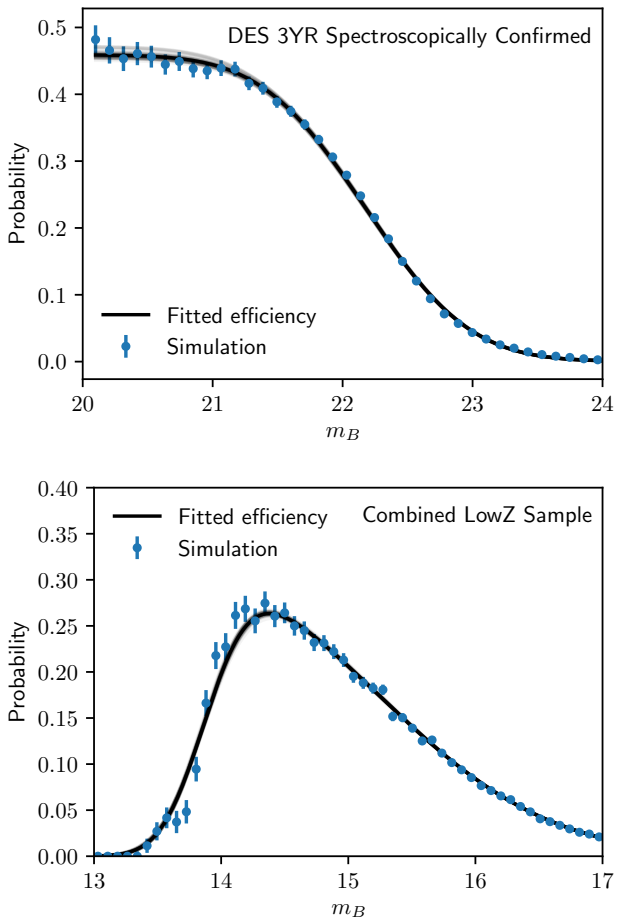

Figure 8. Fitting the selection function for both the DES 3YR spectroscopically confirmed supernova sample and the combined low-redshift sample. Blue errorbars represent the efficiency calculated by determining the ratio of discovered to generated supernovae in apparent magnitude bins for SNANA simulations. 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.

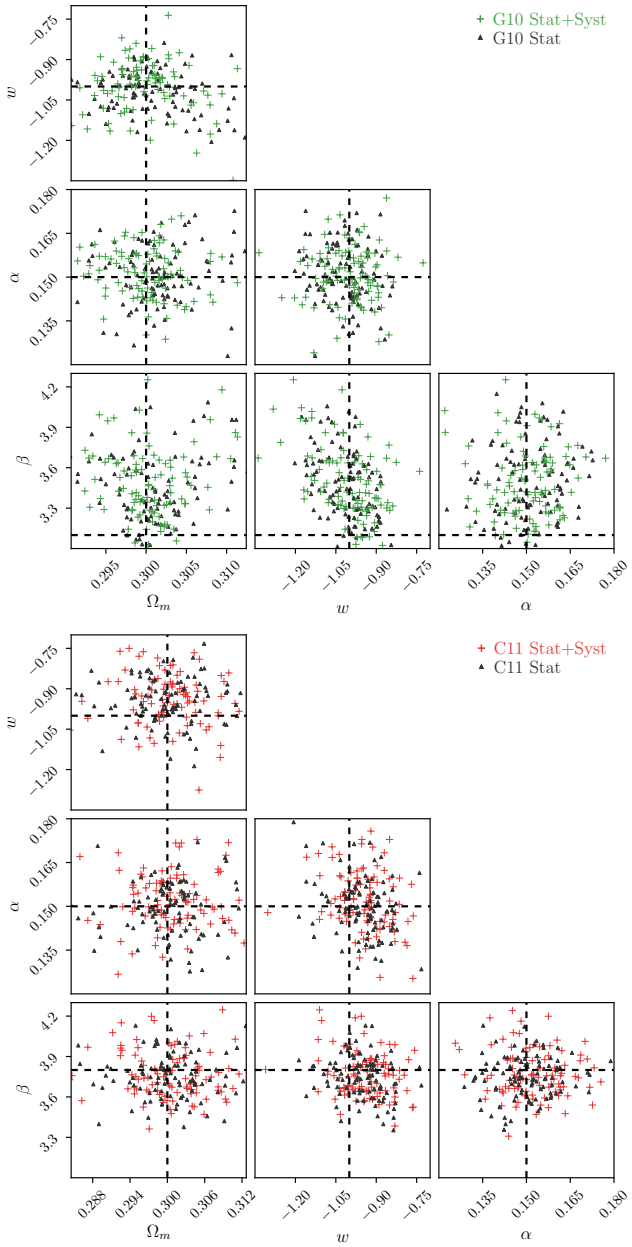

Figure 9. Maximum posterior points for 100 realisations of supernova data for two intrinsic dispersion models - the G10 model for the top panel and the C11 model for the bottom panel. Points are shown for parameters Ω_m, w, α, β and $\langle M_B \rangle$, with the other fit parameters being marginalised 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 effects cosmology, resulting in small biases in w.

recovered results are all within 0.5σ of the true values of $\Omega_m = 0.3$, $w = -1$.

We investigate the cosmological bias and find its source to be a bias in the observed summary statistics, in addition to incorrect reported uncertainty on the summary statistics. To confirm this, we run two tests. The first of which, we replace the fully simulated observed $\hat{m}_B$, $\hat{x}_1$ and $\hat{c}$ 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. With this test, both the G10 and C11 fits recover $w = -1.00$ exactly. Our second test, to allow measurement biases not sourced from intrinsic scatter through, we test a set of 100 simulations generated using only coherent magnitude scatter, and also find $w = -1.00$, showing that the source of the biases in summary statistics is the underlying intrinsic scatter model. From this, the main challenge of improving our methodology is to handle the fact that 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 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) and binning data. It is difficult to do whilst accounting for correlations with population and scatter model. This is compounded by the fact

Table 4. Investigating the combined 100 fits to G10 and C11 simulations, fitting with both statistics only and also when including systematics. The quoted value for w represents the average mean of the fits, with the uncertainty being the average standard deviation of the fits. In brackets, the simulation scatter (the standard deviation of the mean of 100 fits) is shown, and the bias significance represents our confidence that the deviation in the mean w away from -1 is not due to statistical fluctuation. With systematics enabled, both the G10 and C11 models show evidence of bias, scattering to either side of the simulated value of $w = -1$. 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 sub-dominant to both the size of the uncertainty for each fit, and the scatter induced by statistical variance in the simulations. We also note that the simulations do not vary cosmological parameters nor 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.

Model	w (scatter)	Bias
G10 Stat + Syst	-1.00 ± 0.10 (0.08)	0.0σ
C11 Stat + Syst	-0.95 ± 0.10 (0.06)	0.5σ
G10 Stat	-1.00 ± 0.08 (0.08)	0.0σ
C11 Stat	-0.95 ± 0.08 (0.05)	0.6σ

that bias corrections do not in general improve fits (increase the log posterior), and so are difficult to fit inherently. 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 that separate intrinsic scatter bias and non-Gaussian summary statistic bias from Malmquist bias, without having to precompute standardisation parameters and populations.

Table 5 lists the fit correlations between our model fit parameters (excluding the LowZ band systematics, and Malmquist bias uncertainty parameters which had negligible correlation), showing (in order) cosmological parameters, standardisation parameters, population width and skewness parameters, intrinsic dispersion parameters, mass-step parameters, population mean parameters, SALT2 model systematics, dust systematic, global HST calibration systematic, peculiar velocity systematic, global redshift systematic and DES band magnitude and wavelength systematics. Figure 10 show the full correlations between all non-systematic model parameters. As expected, w bias would be introduced primarily through incorrectly determined standardisation parameters and mean absolute magnitude, which itself is tied to the unexplained colour dispersion and biases in colour introduced by the unknown intrinsic scatter model. Other interesting correlations are shown and discussed in Figure 10. The band systematics for DES filters g, r and i also show significant correlation with w, highlighting the importance of minimising instrumental uncertainty.

For the sample size of the DES and LowZ supernova samples (of order 350 supernova), the bias from intrinsic scatter models is sub-dominant to the statistical uncertainty, as shown in Figure 9. For our full systematics model, the bias represents a deviation between 0σ to 0.5σ depending on scatter model, and as such we will leave more complicated treatment of them for future work.

5.3 Uncertainty Analysis

With the increased flexibility of Bayesian hierarchical models over traditional models, we expect to find an increased uncertainty on pa-

Table 5. Parameter Correlations with w for the combined 100 simulation fits. Correlations for the LowZ band systematics and the latent parameters representing selection function uncertainty are not shown but have negligible correlation.

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
$\sigma^0_{m_B}$	0.04	0.07
$\sigma^1_{m_B}$	0.23	0.18
$\sigma^0_{x_1}$	0.04	0.03
$\sigma^1_{x_1}$	0.05	0.01
σ^0_c	0.01	0.11
σ^1_c	0.08	0.04
α^0_c	-0.04	0.04
α^1_c	0.03	0.01
κ^0_{c0}	-0.10	-0.05
κ^1_{c0}	-0.20	-0.17
κ^0_{c1}	-0.05	-0.01
κ^1_{c1}	-0.01	0.01
$\delta(0)$	0.00	0.00
$\delta(\infty)/\delta(0)$	0.00	0.00
$\langle x^0_1 \rangle$	-0.01	-0.05
$\langle x^1_1 \rangle$	-0.02	0.02
$\langle x^2_1 \rangle$	-0.04	-0.04
$\langle x^3_1 \rangle$	-0.03	-0.06
$\langle x^4_1 \rangle$	-0.06	-0.06
$\langle x^5_1 \rangle$	0.04	0.02
$\langle x^6_1 \rangle$	0.04	0.04
$\langle x^7_1 \rangle$	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
δ[HSTCalib]	-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

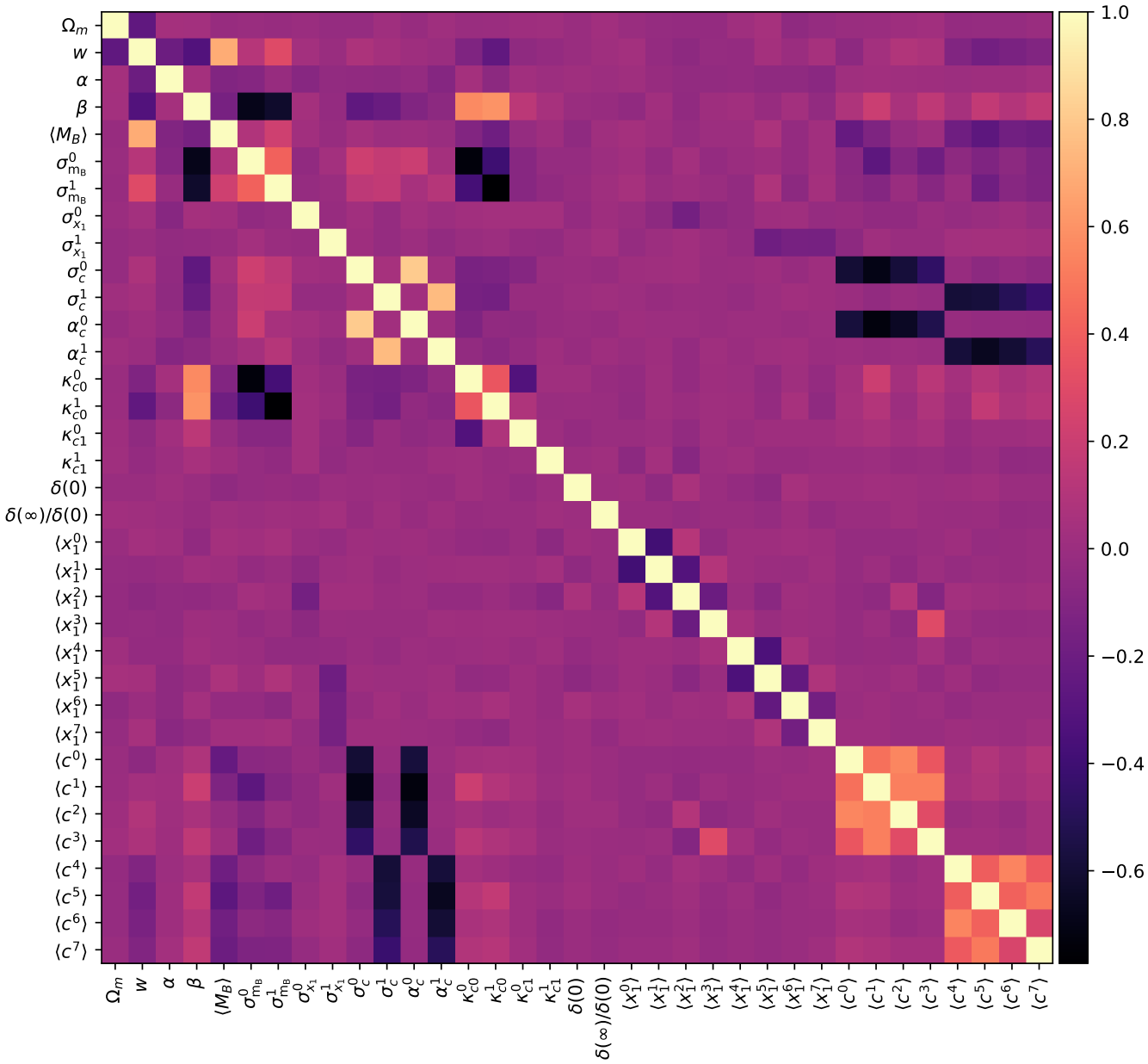

Figure 10. Parameter correlations for the combined fits to the 100 G10 scatter model simulations. We see that the primary correlations with w enter through α, β and $\langle M_B \rangle$, as shown in Table 5. Also visible in this figure are several other interesting relationships. β is strongly anti-correlated with intrinsic dispersion σ_{m_B} for both surveys (DES like and LowZ), with σ_{m_B} showing strong anti-correlation with κ^0_c. This relationship is indeed expected - as κ^0_c grows larger (more unexplained dispersion on the colour 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 colour causes the inherent magnitude smearing amount to decrease. And with extra freedom on the observed colour from κ^0_c, β shifts in response. The other striking feature in the plot are the strong correlation blocks in the bottom right and the anti-correlation stripes on the edges. These too are expected, for they show the relationship between the colour distributions mean value, its width and its skewness. As skewness or population width increases, the effective mean of the population shifts (see Appendix A3 for details), creating anti-correlation between skewness and the (Gaussian) mean colour population. Strong anti-correlation between κ^0_{c0} and κ^1_{c0} with σ_{m_B} reveals the strong population degeneracy, and – for the C11 simulation results – a constrained positive value shows that a finite non-zero extra colour dispersion is indeed preferred by our model.

rameter inference. To characterise the influence of the extra degrees of freedom in our model, we analyse the uncertainty on w averaged across 10 nominal simulations of the DES three-year spectroscopic survey with various model 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 6 shows that our uncertainty is still dominated by statistical error, as the total statistical uncertainty is on w is ±0.08. With the low number of supernovae in the DES three-year spectroscopic sample, this is expected. We note that the label 'Systematics' in Table 6 represents all numerically computed systematics (as discussed in Section 4.3.4) and systematic uncertainty on the selection function.

Table 6. The error budget on w, as determined from analysing uncertainty on simulation data whilst 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.

Feature	σ_w	Cumulative
Cosmology only	0.051	0.051
Standardisation parameters	0.046	0.068
Intrinsic scatter	0.020	0.071
Redshift-independent populations	0.022	0.074
Redshift-dependent populations	0.030	0.080
Systematics	0.054	0.096

5.4 Methodology Comparison

We compare the results of our model against those of the BBC method (Kessler & Scolnic 2017). BBC has been used in prior analyses, such as the Pantheon sample analysis of Scolnic et al. (2017). As a leading supernova cosmology method, it provides a good consistency as to the current levels of accuracy in recovering cosmological parameters.

To this end, we take the results of the BBC method also run on our set of 200 validation simulations and compare the recovered w values to those of our method. The results are detailed in Table 7, and a scatter plot of the simulation results is presented in Figure 11.

These results show that both BBC and BHM are sensitive to the intrinsic scatter model, finding differences of ~ 0.06 in w when varying the scatter model. The BBC method finds w biased low for G10 and w biased 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 (but not statistically significant) improvement in the insensitivity to the intrinsic scatter model, finding no bias for G10 but a w biased high for C11. 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 set 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 three year spectroscopic survey sample. When using Planck priors our uncertainty on w is reduced compared to using our simulation Gaussian prior on Ω_m, shrinking the average scatter from 0.06 to 0.04. After factoring this into our uncertainty, we expect our BHM method to, on average, recover $w_{\rm BHM} = w_{\rm BBC} + 0.04 \pm 0.04$.

Having established that our method exhibits similar shifts in their recovery of w compared to BBC, future work will focus on improving the parametrisation of intrinsic scatter model into our framework, with the goal of minimising 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 standardisation parameters, incorporates unexplained dispersion from intrinsic scatter colour 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 standardisation, population, mass-step and more, being explicitly parametrised in our model, are captured will covariance intact, an improvement on many previous methods. The model has been optimised to allow for hundreds of supernovae to be modelled fully with latent parameters and run in under an hour of CPU time and scales linearly with the number of supernovae, as opposed to 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 realisations cannot be overstated, however this validation is lacking in many earlier BHM models for supernova cosmology. We have validated this model against many realisations 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 which is traced back to light curve fits reporting biased observables and incorrect covariance. Allowing fully parametrised corrections on observed supernovae summary statis-

Table 7. Characterising the bias on w using the 100 simulations for the G10 scatter model and 100 simulations for C11 scatter model. 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. Finally, we also characterise the scatter between the methods in the final row. Averages are computed giving each simulation sample the same weight.

Model	G10	C11	Mean (G10 + C11)
Steve $\langle w \rangle$	−0.998	−0.945	−0.972
BBC $\langle w \rangle$	−1.044	−0.978	−1.010
(Steve - BBC) $\langle w \rangle$	+0.044	+0.033	+0.039
(Steve - BBC) σ_w	0.057	0.062	0.060

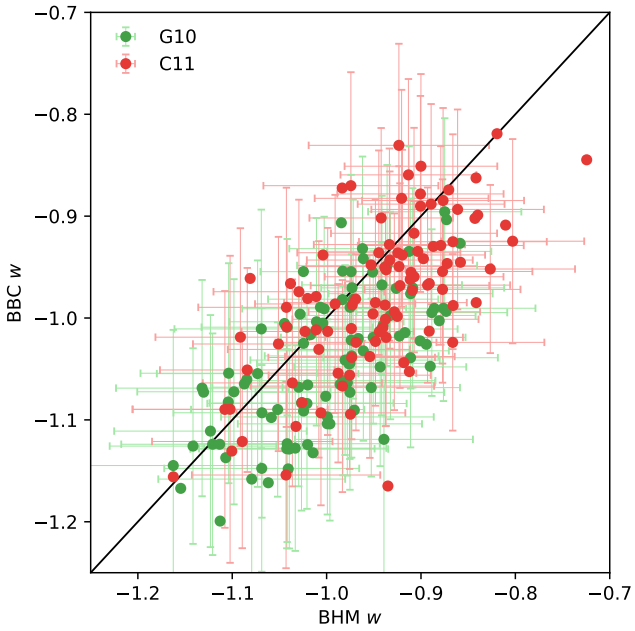

Figure 11. Recovered w for the 200 validation simulations with full treatment of statistical and systematic errors. Uncertainty on the recovered w value is shown for every second data point for visual clarity.

tics introduces too many degrees of freedom and is found to make cosmology fits too weak. Allowing simulation based 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 three-year spectroscopic survey, the determined biases should be sub-dominant to other sources of uncertainty, however this cannot be expected for future analyses with larger datasets. Stricter bias corrections calculated from simulations are required to reduce bias. Ideally, this would include further work on the calculation of intrinsic dispersion of the Type Ia supernovae population such that we can better characterise this bias.

With our model being validated against hundreds of simulation realisations, representing a combined dataset over more than 250 000 simulated supernovae, we have been able to accurately determine biases in our model and trace their origin. With the current biases being sub-dominant to the total uncertainty, we now prepare to analyse the DES three-year dataset.

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
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
Folatelli G., et al., 2010, The Astronomical Journal, 139, 120
Freedman W. L., et al., 2009, The Astrophysical Journal, 704, 1036
Frieman J. A., et al., 2008, AJ, 135, 338
Graur O., et al., 2013, The Astrophysical Journal, 783, 28
Gupta R. R., et al., 2011, ApJ, 740, 92
Gupta R. R., et al., 2016, AJ, 152, 154
Guy J., et al., 2007, Astronomy and Astrophysics, 466, 11
Guy J., et al., 2010, Astronomy and Astrophysics, 523, 34
Hicken M., et al., 2009a, The Astrophysical Journal, 700, 331
Hicken M., Wood-Vasey W. M., Blondin S., Challis P., Jha S., Kelly P. L., Rest A., Kirshner R. P., 2009b, Astrophysical Journal, 700, 1097
Hicken M., et al., 2012, Astrophysical Journal, Supplement Series, 200
Hinton S., 2016, The Journal of Open Source Software, 1
Hlozek R., et al., 2012, The Astrophysical Journal, 752, 79
Huterer D., Shafer D. L., 2018, Dark energy two decades after: Observables, probes, consistency tests (arXiv:1709.01091), doi:10.1088/1361-6633/aa997e, http://arxiv.org/abs/1709.01091http://dx.doi.org/10.1088/1361-6633/aa997e
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., 2009a, Publications of the Astronomical Society of the Pacific, 121, 1028
Kessler R., et al., 2009b, Astrophysical Journal, Supplement Series, 185, 32
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
Mandel K. S., Scolnic D., Shariff H., Foley R. J., Kirshner R. P., 2017, The Astrophysical Journal, 842, 26
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., 2010, Astronomical Journal, 140, 518
Perrett K., et al., 2012, The Astronomical Journal, 144, 59
Phillips M. M., 1993, The Astrophysical Journal, 413, L105
Phillips M. M., Lira P., Suntzeff N. B., Schommer R. A., Hamuy M., Maza J., 1999, The Astronomical Journal, 118, 1766
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
Scolnic D. M., et al., 2017, eprint arXiv:1710.00845
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/
Stritzinger M., et al., 2011, The Astronomical Journal, 142, 14
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 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 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 d for simplicity in future derivations.

A2 Supernova Selection Effects

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

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

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, $P(\hat{z}|z) = \delta(\hat{z} - z)$. $P(\hat{m}_B|m_B)$ is a Gaussian due to our Gaussian model of summary statistics, and m_B, x_1, c, and can be analytically integrated out, collapsing the integral over $\hat{m}_B$ (which is why they were not included in equation (A9)). 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

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

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 4.3.5, 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

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

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

Then, modelling $P(S|m_B)$ as either a normal or a skew normal, we can analytically perform the integral in equation (A11) and reach equations (17) and (18).

A3 Approximate Selection Effects

In this section, we investigate the effect of approximating the skew normal underlying colour distribution as a normal. Specifically, 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 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

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

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

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, and 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 A1, and only the shifted normal approximation correctly recovers underlying population parameters.

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 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 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 thirty 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. 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

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

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 intrinsic 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.

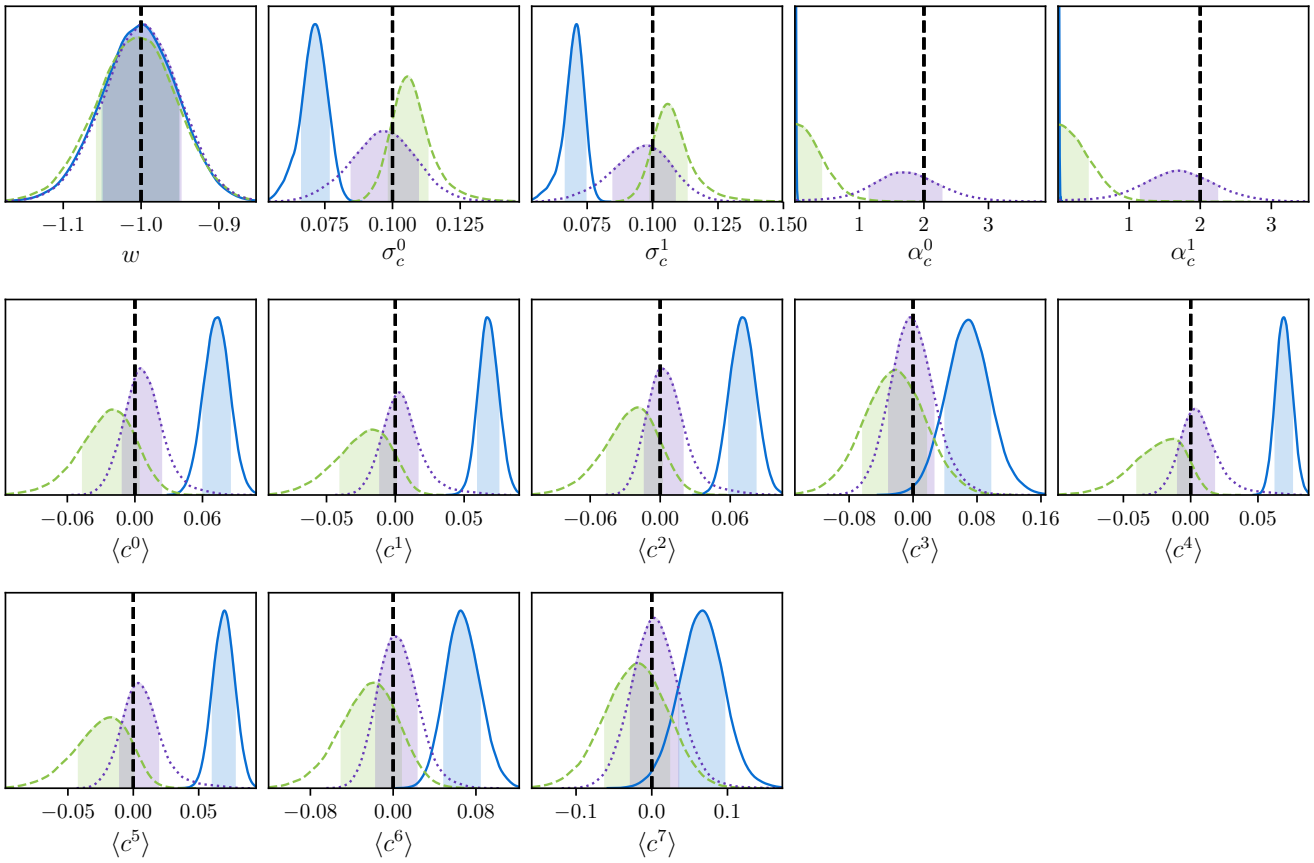

Figure A1. Marginalised probability distributions for 100 realisations of cosmology, fit to Flat wCDM with prior $\Omega_m \sim \mathcal{N}(0.3, 0.01)$, each containing 1000 simulated high-z and 1000 simulated low-z supernovae. The dashed green surfaces represent a fit to an underlying Gaussian colour 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. Based on these results and the fantastic performance in correctly recovering underlying populations of the shifted normal approximation, we adopt that in our model.