Draft version October 14, 2018
Preprint typeset using LaTeX style emulateapj v. 12/16/11

Steve: A HIERARCHICAL BAYESIAN MODEL FOR SUPERNOVA COSMOLOGY

S. R. Hinton[1,2], T. M. Davis[1,2], A. G. Kim[3], A. Möller[2,4], M. Sako[5], M. Smith[6],
[1] School of Mathematics and Physics, University of Queensland, Brisbane, QLD 4072, Australia
[2] ARC Centre of Excellence for All-sky Astrophysics (CAASTRO)
[3] Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA
[4] The Research School of Astronomy and Astrophysics, Australian National University, ACT 2601, Australia
[5] Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, PA 19104, USA and
[6] School of Physics and Astronomy, University of Southampton, Southampton, SO17 1BJ, UK
Draft version October 14, 2018

ABSTRACT

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

Subject headings: cosmology: supernovae

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 (SN Ia) has increased by more than an order of magnitude, with contributions from modern surveys at both low redshift (Bailey et al. 2008; Freedman et al. 2009; Hicken et al. 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; Chambers et al. 2016; Sako et al. 2018). Cosmological analyses of these supernova samples (Kowalski et al. 2008; Kessler et al. 2009b; Conley et al. 2011; Suzuki et al. 2012; Betoule et al. 2014; Rest et al. 2014; Scolnic et al. 2017) have been combined with complementary probes of large scale structure and the CMB. For a recent review, see Huterer & Shafer (2018). While these efforts have reduced the uncertainty on the equation-of-state of dark energy (w) by more than a factor of two, it is still consistent with a cosmological constant and the nature of dark energy remains an unsolved mystery.

In attempts to tease out the nature of dark energy, active and planned surveys are continually growing in size and scale. The Dark Energy Survey (DES, Bernstein et al. 2012; Abbott et al. 2016) has discovered thousands of type Ia supernovae, attaining both spectroscopically and photometrically identified samples. The Large Synoptic Survey Telescope (LSST, Ivezic et al. 2008; LSST Science Collaboration et al. 2009) will discover tens of thousands of photometrically classified supernovae. Such increased statistical power demands greater fidelity and flexibility in modelling supernovae for cosmological purposes, as we will require reduced systematic uncertainties to fully utilise these increased statistics (Betoule et al. 2014; Scolnic et al. 2017).

As such, considerable resources are aimed at developing more sophisticated supernova cosmology analyses. The role of simulations mimicking survey observations has become increasingly important in determining biases in cosmological constraints and validating specific supernova models. First used in SNLS (Astier et al. 2006) and ESSENCE analyses (Wood-Vasey et al. 2007), and then refined and improved for SDSS (Kessler et al. 2009b), simulations are a fundamental component of modern supernova cosmology. Betoule et al. (2014) quantise and correct observational bias using simulations, and more recently Scolnic & Kessler (2016) and Kessler & Scolnic (2017) explore simulations to quantify observational bias in SN Ia distances as a function of multiple factors to improve bias correction. 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). Bayesian Hierarchical 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 use simulation-determined distance-corrections to correct for biases, or attempt to find analytic approximations for effects such as Malmquist bias to model the biases inside the BHM itself.

In this paper, we lay out a new hierarchical model that builds off the past work of Rubin et al. (2015). We include additional sources of systematic uncertainty, including an analytic formulation of selection efficiency which incorporates parametric uncertainty. We also implement a different model of intrin-

sic dispersion to both incorporate redshift-dependent scatter and to increase numerical efficiency, allowing our model to perform rapid fits to supernovae datasets.

Section 2 is dedicated to a quick review of supernova cosmology analysis methods, and Section 3 outlines some of the common challenges faced by analysis methods. In Section 4 we outline our methodology. Model verification on simulated 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 photometric time-series brightness measurements in many bands and a redshift measurement of the supernova (or its assumed host), most analyses do not work from these measurements directly. Instead, most techniques fit an observed redshift and these photometric observations to a supernova model, with the most widely used being that of the empirical SALT2 model (Guy et al. 2007, 2010). This model is trained separately before fitting the supernova light curves for cosmology (Guy et al. 2010; Betoule et al. 2014). The resulting output from the model is, for each supernova, an amplitude x_0 (which can be converted into apparent magnitude, $m_B = -2.5\log(x_0)$), a stretch term x_1, and color term c, along with a covariance matrix describing the uncertainty on these summary statistics (C_η). As all supernovae are not identical, an ensemble of supernovae form a redshift-dependent, observed population of $\hat{m}_B$, $\hat{x}_1$ and $\hat{c}$, where the hat denotes an observed variable.

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

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

where M_B is the mean absolute magnitude for all SN Ia given $x_1 = c = 0$, α is the stretch correction (Phillips 1993; Phillips et al. 1999), and β is the color correction (Tripp 1998) that respectively encapsulate the empirical relation that broader (longer-lasting) and bluer supernovae are brighter. $\Delta M \cdot m$ refers to the host-galaxy mass correlation discussed in Section 4.4.3. The distance modulus $\mu_{\rm obs}$ is a product of our observations, however a distance modulus μ_C can be precisely calculated given cosmological parameters and a redshift. The 'other corrections' term often includes bias corrections for traditional χ^2 analyses. Bias corrections can take multiple forms, such as a redshift-dependent function (Betoule et al. 2014) or a 5D function of c, x_1, α, β and z (Kessler & Scolnic 2017; Scolnic et al. 2017).

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 equation 1, and the cosmologically predicted distance modulus. The χ^2 function minimised is

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

where C_μ^{-1} is an uncertainty matrix that combines statistical and systematic uncertainties (see Brout et al. (2018) for a review of these uncertainties for the DES supernova analysis). The predicted μ_C is defined as

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

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

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

where d_L is the luminosity distance for redshift z given a specific cosmology, H_0 is the current value of Hubble's constant in $\mathrm{km\,s^{-1}\,Mpc^{-1}}$ 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 results in a loss of model flexibility by assuming Gaussian uncertainty. Secondly, the method of creating a 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 color population's mean and skewness are highly correlated, however this correlation is ignored when determining population uncertainty using numerical derivatives of population permutations. Whilst correlations can be incorporated into a covariance matrix, it requires human awareness of the correlations and thus methods which can automatically capture correlated uncertainties are preferable. Thirdly, the computational efficiency is dependent on both creating the global covariance matrix, and then 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 recent solution to this computational cost problem is to bin the supernovae in redshift bins, which produces a matrix of manageable size and can allow fitting without matrix inversion on every step. Whilst binning data results in some loss of information, recent works tested against simulations show 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 the determined $\mu_{\rm obs}$ from the data, or equivalently, adding a distance bias to the μ_C prediction. Specifically, Malmquist bias is the result of losing the fainter tail of the supernova population at high redshift. An example of Malmquist bias is illustrated in Figure 1, which simulates supernovae according to equation (1). Simulations following survey observational

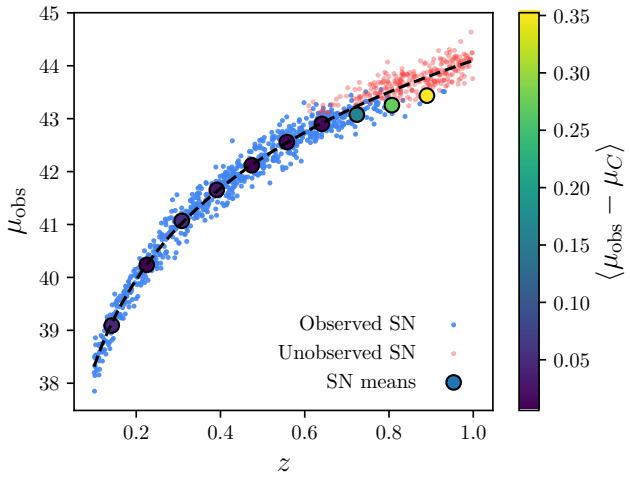

FIG. 1.— An example of the effects of Malmquist bias. Here are shown 1000 simulated supernovae redshifts and distance modulus given fiducial cosmology. The simulated survey is magnitude limited, and all supernovae brighter than magnitude 24 are successfully observed (shown as blue dots), and all dimmer 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.

strategies and geometry are used to calculate the expected bias in distance modulus, which is then subtracted from the observational data. When using traditional fitting methods such as that found in Betoule et al. (2014), these effects are not built into the likelihood and instead are formed by correcting data. This means that the bias uncertainty is not captured fully in the χ^2 distribution, and subtle correlations between cosmological or population parameters and the bias correction is lost. Recent developments such as the BBC method (Kessler & Scolnic 2017) incorporate corrections dependent on α and β, improving their capture of uncertainty on bias corrections in the χ^2 likelihood.

2.2. *Approximate Bayesian Computation*

To avoid 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. In both examples, the supernova simulation package SNANA (Kessler et al. 2009a) is used to forward model the data at each point in parameter space.

Simulations provide great flexibility and freedom in how to treat the systematic uncertainties and selection effects associated with supernova surveys. By using forward 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. *Bayesian Hierarchical 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, with the layers linked via well defined and physically motivated conditional probabilities. For example, an observation of a parameter from a population will be conditioned on the true value of the parameter, which itself will be conditioned on the population distribution of that parameter. We can thus incorporate different population distributions, and parameter inter-dependence which cannot be found in traditional analyses where uncertainty must be encapsulated in a covariance matrix. Unlike ABC methods, which can model arbitrary probability distributions, BHM methods are generally constrained to representing probabilities using analytic forms.

With the introduction of multiple layers in our model, we can add more flexibility than a traditional analysis 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) and 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 the methods ignore biases, however the UNITY framework given by Rubin et al. (2015) incorporates selection efficiency analytically in the model, and is applied to the Union 2.1 dataset (Suzuki et al. 2012). The assumption made by the UNITY analysis is that the bias in real data is perfectly described by an analytic function. They validate their model to be free of significant biases using fits to thirty realisations of supernova datasets that are constructed to mimic the UNITY framework. 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), most 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, none of them have undergone high-statistic tests using realistic simulations to verify each models' respective bias. Additionally, testing has generally been performed on supernovae simulations with either ΛCDM cosmology or Flat ΛCDM cosmology. However, quantifying the biases on wCDM cosmology simulations with realistic simulations is becoming critically important as precision supernovae cosmology comes into its own, and focus shifts from determination of Ω_m to both Ω_m and w.

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

3. CHALLENGES IN SUPERNOVA COSMOLOGY

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

Forefront among these challenges is our ignorance of the underlying type Ia intrinsic dispersion. Ideally, analysis of the underlying dispersion would make use of an ensemble of time-series spectroscopy to characterise the diversity of type Ia supernovae. However this data is difficult to obtain, and recent efforts to quantify the dispersion draw inference from photometric measurements. The underlying dispersion model is not a solved problem, and we therefore test against two dispersion models in this work. The first is based on the Guy et al. (2010, hereafter denoted G10) scatter model, the second is sourced from Chotard et al. (2011, hereafter denoted C11). As the SALT2 model does not include full treatment of intrinsic dispersion, each scatter model results in different biases in m_B, x_1, and c when fitting the SALT2 model to light curve observations, and results in increased uncertainty on the summary statistics that is not encapsulated in the reported covariance $\mathcal{C}_\eta$. These two scatter models are currently assumed to span the possible range of scatter in the underlying supernova population. We have insufficient information to prefer one model over the other, and thus we have to account for both possible scatter models.

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

Finally, one of the other persistent challenges facing supernova cosmology analyses are the high number of systematics. Because of the rarity of SN Ia events, datasets are commonly formed from the SN Ia discoveries of multiple surveys in order to increase the number of supernovae in a dataset. 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 *Steve* with several goals in mind: creation of a redshift-dependent underlying supernova population, treatment of an increased number of systematics, and analytic correction of selection effects, including systematic uncertainty on those corrections. We also desire *Steve* 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 a mapping from an underlying population onto an observed population, where cosmological parameters are encoded directly in the mapping function. The difficulty arises both in adequately describing the biases in the mapping function, and in adding sufficient, physically motivated flexibility in both the observed and underlying 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 brightness measurement $\hat{m}_B$ (which is analogous to apparent magnitude), stretch $\hat{x}_1$ and color $\hat{c}$, with uncertainty on those values encoded in the covariance matrix $\mathcal{C}_\eta$. Additionally, each supernova has an observed redshift $\hat{z}$ and a host-galaxy stellar mass associated with it, $\hat{m}$, where the mass measurement is converted into a probability of being above 10^{10} solar masses. We also have a probability of each supernova being a type Ia, $\hat{p}$. Our set of observables input into the

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	color 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
$\langle M_B \rangle$	Mean absolute magnitude
δS	Selection effect deviation
$\langle x_1^i \rangle$	Mean stretch nodes
$\langle c^i \rangle$	Mean color nodes
α_c	Skewness of color population
σ_{M_B}	Population magnitude scatter
σ_{x_1}	Population stretch scatter
σ_c	Population color scatter
κ_0	Extra color dispersion
κ_1	Redshift-dependence of extra color dispersion
m_B	True flux
x_1	True stretch
c	True color
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 color
$\boldsymbol{C}$	Covariance on flux, stretch and color
$\hat{z}$	Observed redshift
$\hat{m}$	Determined mass probability
$\hat{p}$	Determined Ia probability

Steve is therefore given as $\{\hat{m}_B, \hat{x}_1, \hat{c}, \hat{z}, \hat{p}, \hat{m}, \mathcal{C}_\eta\}$, 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. Similarly, we take the mass probability estimate $\hat{m}$ as correct, and do not model a latent variable to represent uncertainty on the probability estimate. One of the strengths of *Steve* (and the R15 analysis) is that for future data sets where supernovae have been classified photometrically, and we expect some misclassification and misidentification of the host galaxies, these misclassifications can naturally be modelled and taken into account by introducing additional populations that supernovae have a non-zero probability of belonging to.

4.1.2. *Latent Variables for Observables*

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

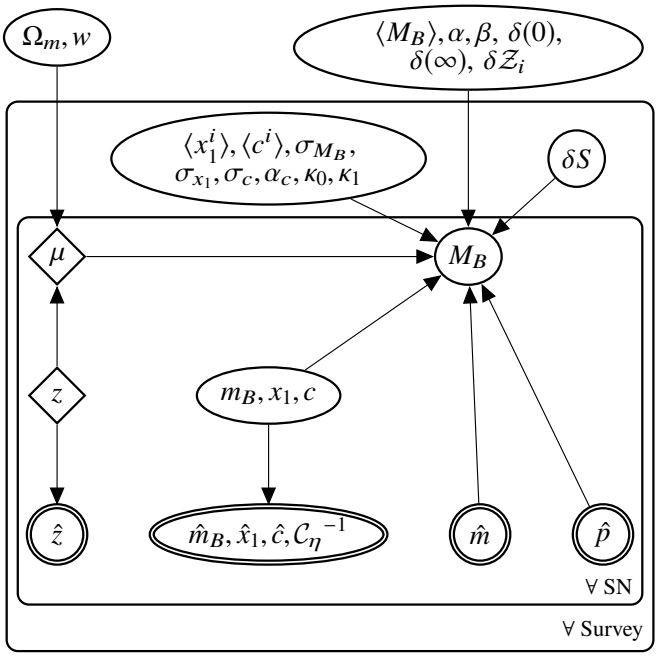

FIG. 2.— Probabilistic graphical model for our likelihood. Double-lined nodes represent observed variables, diamond nodes represent deterministic variables, single-lined ellipse nodes represent fit variables. The SN box represents observed variables and latent variables for each individual supernova,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

For the moment, let us consider a single supernova and its 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 fluxes and uncertainties. However, this is computationally prohibitive as it would require incorporating SALT2 light curve fitting inside our model fitting. Due to this computational expense, we rely on initially fitting the light curve observations to produce a best fit $\hat{\eta}$ along with a 3×3 covariance matrix $\mathcal{C}_\eta$ describing the uncertainty on $\hat{\eta}$. Using this simplification, our latent variables are given by

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

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

4.2. *Underlying Population*

4.2.1. *Type Ia population*

Unlike many previous formalisms which utilise M_B as a singular number and model magnitude scatter on the apparent magnitude m_B, we incorporate this scatter into the underlying rest-frame population by having a population in absolute magnitude space. To denote this difference, we refer to the mean of our absolute magnitude population with $\langle M_B \rangle$.

In addition to absolute magnitude, the underlying supernova population is also characterised by distributions in color and stretch. We follow the prior work of R15 and model the color population as an independent redshift-dependent skew normal distribution for each survey. For the stretch population, we adopt a redshift-dependent normal distribution, and magni-

tude dispersion is modelled as a normal distribution. We also tested a skew-normal approach for these parameters, reverting to the normal distributions as they are computationally easier to evaluate and we found no reduction in cosmological bias with the skew-normal distributions for stretch and magnitude. Following R15 we allow the mean color and stretch to vary over redshift, anchoring four equally spaced redshift nodes spanning the redshift range of each survey, linearly interpolating between the nodes to determine the mean stretch and color for a given redshift. These nodes are represented as $\langle x_1^i \rangle$ and $\langle c^i \rangle$. Both the color and stretch means are 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) to capture correlations between m_B and c, however pathological fitting complications required us to simplify our treatment. We parametrise the color skewness α_c by sampling $\delta_c = \alpha_c/\sqrt{1+\alpha_c^2}$ which itself is given a uniform prior $\mathcal{U}(0, 0.98)$ that allows α_c to span positive values in the realm of physical plausibility as determined from constraints in Scolnic & Kessler (2016). We sample δ_c in log space for efficiency in sampling close to zero. The width of the population, represented by the vector $\{\sigma_{M_B}, \sigma_{x_1}, \sigma_c\}$ is subject to Cauchy priors.

The only constant between survey populations is the absolute magnitude $\langle M_B \rangle$. We model the colour skewness and the redshift-dependent means and width of the colour and skew distributions 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

$$P(M_B, x_1, c, z | \langle M_B \rangle, \langle x_1(z) \rangle, \langle c(z) \rangle, \sigma_{m_B}, \sigma_{x_1}, \sigma_c, \alpha_c) = \mathcal{N}(M_B, x_1, c | \langle M_B \rangle, \langle x_1(z) \rangle, \langle c(z) \rangle, \mathbf{C}_{\rm tot}) \times \mathcal{N}^{\rm skew}(c | 0, \sigma_c, \alpha_c) \times \mathrm{DiagMatrix}\left[0, 0, \sigma_{\kappa_1} z\right]^2.$$

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 SN Ia population in prior works. However, with the increased number of contaminants expected in the DES photometric sample, we seek a more physically motivated outlier population. We find that an acceptable parametrisation is to model the outlier population with a mean absolute magnitude of $\langle M_B^{\rm outl} \rangle = \langle M_B \rangle + \delta_{M_B}^{\rm outl}$, where $\delta_{M_B}^{\rm outl}$ is constrained to be positive, or even to be greater than a small positive number to reduce degeneracy between the two populations. We note that this represents the mean brightness of outliers, and so outliers could both be brighter and dimmer than the mean SN Ia absolute magnitude. We set the population width to $\sigma^{\rm outl} = 1$ in M_B, x_1 and c. The probability of each supernova 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-concept we provide an accurate probability estimate. Further investigation on the effect of inaccurate estimates will be left for future improvements during the analysis of the DES photometric sample.

4.3. *Correcting biased summary statistics*

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

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

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

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 color dispersion, and was far more computationally inefficient. We note here that we model dispersion in magnitude, but this is done at the level of underlying populations, not observed populations. This magnitude dispersion is modelled with redshift independence.

4.4. *Mapping function*

4.4.1. *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 $H_0 = 70 \mathrm{km\,s^{-1}\,Mpc^{-1}}$. If the Hubble constant has a different value, the absolute magnitude is $M_B + 5\log(H_0/70 \mathrm{km\,s^{-1}\,Mpc^{-1}})$ with the other cosmological parameters unaffected.

4.4.2. *Supernova 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 color and stretch, where different values of α and β are adopted depending on whether x_1 and c are positive or negative (although the cut could be placed at a location other than 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. 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 (3).

4.4.3. *Host Galaxy Environment*

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

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

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 $\delta(0)/\delta(\infty)$.

4.4.4. *Uncertainty Propagation*

The chief difficulty with including systematic uncertainties in supernova analyses is that they have difficult-to-model effects on the output observations. As such, the traditional treatment for systematics is to compute their effect on the supernova summary statistics – computing the numerical derivatives $\frac{\mathrm{d}\hat{m}_B}{\mathrm{d}\mathcal{Z}_i}$, $\frac{\mathrm{d}\hat{x}_1}{\mathrm{d}\mathcal{Z}_i}$, $\frac{\mathrm{d}\hat{c}}{\mathrm{d}\mathcal{Z}_i}$, 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 values by constructing a $(3 \times N_{\text{sys}})$ matrix containing the numerical derivatives for the N_{sys} systematics and multiplying it with the row vector containing the offset for each systematic. By scaling the gradient matrix to represent the shift over 1σ of systematic uncertainty, we can simply enforce a unit normal prior on the systematic row vector to increase computational efficiency.

This method of creating a secondary covariance matrix using partial derivatives is used throughout the traditional and BHM analyses. For each survey and band, we have two systematics — the calibration uncertainty and the filter wavelength uncertainty. We include these in our approach, in addition to including HST Calspec calibration uncertainty, ten SALT2 model systematic uncertainties, a dust systematic, a global redshift bias systematic, and also the systematic peculiar velocity uncertainty. A comprehensive explanation of all systematics is given in Brout et al. (2018); see Table 4 for details. 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 (6) becomes

$$P\left(\hat{\eta}, \frac{\partial\hat{\eta}}{\partial\mathcal{Z}_i}|\eta, \delta\mathcal{Z}_i, C\right) = \mathcal{N}\left(\hat{\eta} + \delta\mathcal{Z}_i\frac{\partial\hat{\eta}}{\partial\mathcal{Z}_i}|\eta, \mathcal{C}_\eta\right). \tag{8}$$

4.4.5. *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 describe the probability that the events we observe are drawn from the distribution predicted by the underlying theoretical model *and* that those events, given they happened, are observed and pass cuts. To make this extra conditional explicit, we can write the likelihood of the data given an underlying model, θ, *and* that the data are included in our sample, denoted by S, as

$$\mathcal{L}(\theta; \text{data}) = P(\text{data}|\theta, S). \tag{9}$$

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

$$\mathcal{L}(\theta; \text{data}) = \frac{P(S|\text{data}, \theta)P(\text{data}|\theta)}{\int P(S|D, \theta)P(D|\theta)\, dD}, \tag{10}$$

where the denominator represents an integral over all potential data D. This is derived in Appendix A.1 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 (10), denoted now d for simplicity, is given as

$$d = \int\left[\int P(S|m_B)P(m_B|z, \theta)\, dm_B\right] P(S|z)P(z|\theta)\, dz, \tag{11}$$

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.2.

We now apply two further approximations similar to those made in R15 – that the redshift distribution of the observed

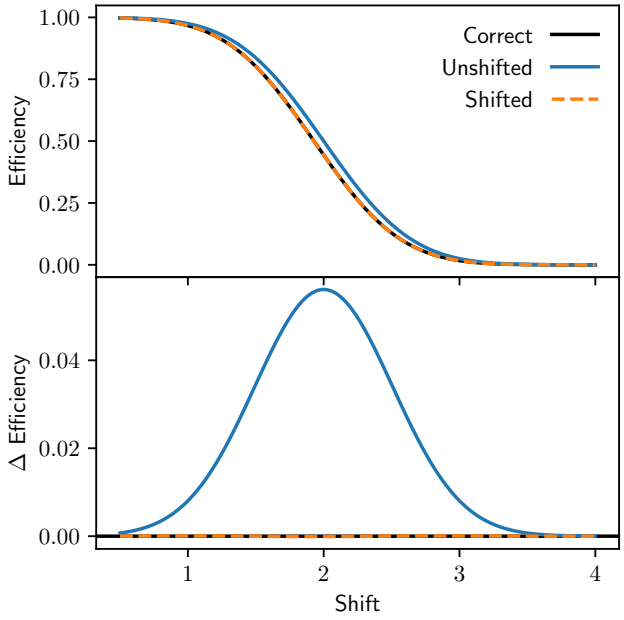

FIG. 3.— Testing the correctness of our normal approximation to the skewed color distribution. The 'correct' line (shown in black) represents the exact integral $w = \int P(S|x)P(x)dx$ where $P(S|x)$ is an error function (following our high-redshift surveys) and $P(x) = \mathcal{N}^{\rm Skew}(x, 0.1,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

supernovae reasonably well samples the $P(S|z)P(z|\theta)$ distribution, and that the survey color and stretch populations can be treated as Gaussian for the purposes of evaluating $P(m_B|z,\theta)$. We found that discarding the color population skewness entirely resulted in highly biased population recovery (see Figure 12 to see the populations), and so we instead characterise the skew normal color 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 color completely removes the unintended bias when simply discarding skewness. This shift was not required for the stretch population, and so was left out for the stretch population for numerical reasons. The impact of this approximation on the calculated efficiency is shown in Figure 3, and more detail on this shift and resulting population recovery can be found in Appendix A.3.

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

$$m_B^*(z) = \langle M_B\rangle + \mu(z) - \alpha\langle x_1(z)\rangle + \beta\langle c(z)\rangle \quad (12)$$

$$\sigma_{m_B}^{*2} = \sigma_{M_B}^2 + (\alpha\sigma_{x_1})^2 + (\beta\sigma_c)^2. \quad (13)$$

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

$$P(S|m_B) = \Phi^c(m_B|\mu_{\rm CDF},\sigma_{\rm CDF}), \quad (14)$$

where Φ^c is the complementary 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

$$P(S|m_B) = \mathcal{N}^{\rm Skew}(m_B|\mu_{\rm Skew},\sigma_{\rm Skew},\alpha_{\rm Skew}). \quad (15)$$

The selection functions are fit to apparent magnitude efficiency ratios calculated from SNANA simulations, by taking the ratio of supernovae that are observed and passed cuts over the total number of supernovae generated in that apparent magnitude bin. That is, we calculate the probability 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 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 redshift approximation we can remove the redshift integral in Eq (11) 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 SN Ia* becomes

$$d_{\rm CDF} = \Phi^c\left(\frac{m_B^* - \mu_{\rm CDF}}{\sqrt{\sigma_{m_B}^{*2} + \sigma_{\rm CDF}^2}}\right) \quad (16)$$

$$d_{\rm Skew} = 2\mathcal{N}\left(\frac{m_B^* - \mu_{\rm Skew}}{\sqrt{\sigma_{m_B}^{*2} + \sigma_{\rm Skew}^2}}\right) \times \Phi\left(\frac{\mathrm{sign}(\alpha_{\rm Skew})(m_B^* - \mu_{\rm Skew})}{\frac{\sigma_{m_B}^{*2} + \sigma_{\rm Skew}^2}{\sigma_{\rm Skew}^2}\sqrt{\frac{\sigma_{\rm Skew}^2}{\alpha_{\rm Skew}^2} + \frac{\sigma_{m_B}^{*2}\sigma_{\rm Skew}^2}{\sigma_{m_B}^{*2} + \sigma_{\rm Skew}^2}}}\right), \quad (17)$$

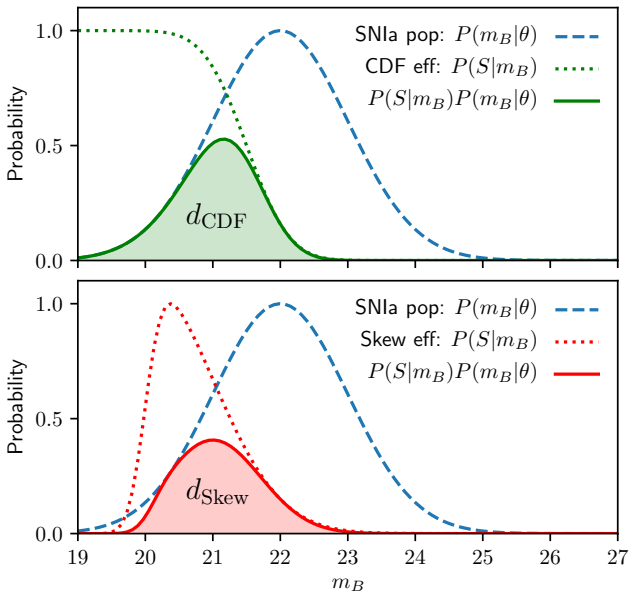

FIG. 4.— The efficiency for supernova discovery at an arbitrary redshift. Shown in both panels in dashed blue is the SN Ia population distribution, which takes the form of a normal distribution. The top panel shows a CDF based survey efficiency (green dotted line), whilst the bottom panel shows a skew normal based survey efficiency (red dotted line), as functions of apparent magnitude. The survey efficiency, given the SN Ia population, is shown as a solid line in both panels, and the probability of observing a SN Ia is found by integrating over the population detection efficiency as described in equation (11), and has been shown by shading the area integrated. This area is what is analytically given by equations (16) and (17).

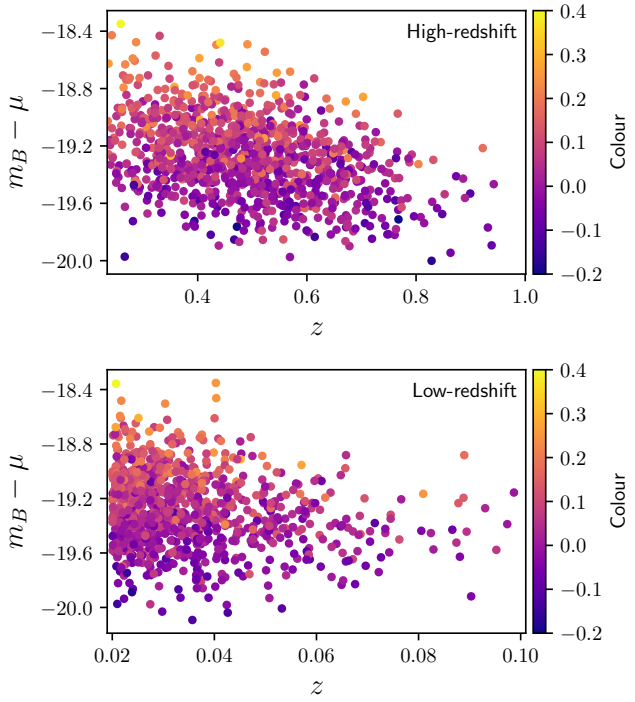

FIG. 5.— Population distributions shown in redshift and uncorrected absolute magnitude $m_B - \mu$ for 1000 supernovae in both high-redshift and low-redshift surveys. Selection effects are visible in both samples, where red supernovae are often cut as redshift increases, creating a skewed color population.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

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 when generating toy supernovae data constructed to satisfy the assumptions of the BHM construction. We then validate our model on SNANA simulations based on a collection of low redshift surveys and the DES three-year spectroscopic sample, termed the DES-SN3YR sample

5.1. *Applied to Toy Spectroscopic Data*

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

1. Draw a redshift from a power law distribution. For the low redshift survey this is $\mathcal{U}(0.0004, 0.01)^{0.5}$, and for the DES-like survey this is $\mathcal{U}(0.008, 1.0)^{0.3}$.

2. Draw a random mass probability from $\mathcal{U}(0, 1)$ and calculate the mass-brightness correction using $\delta(0) = 0.08$, $\delta(0)/\delta(\infty) = 0.5$, and equation (7).

3. Draw an absolute magnitude, stretch and color from the respective distributions $\mathcal{N}(-19.3, 0.1)$, $\mathcal{N}(0, 1)$, $\mathcal{N}(0, 0.1)$. Calculate the true absolute magnitude using Equation (1).

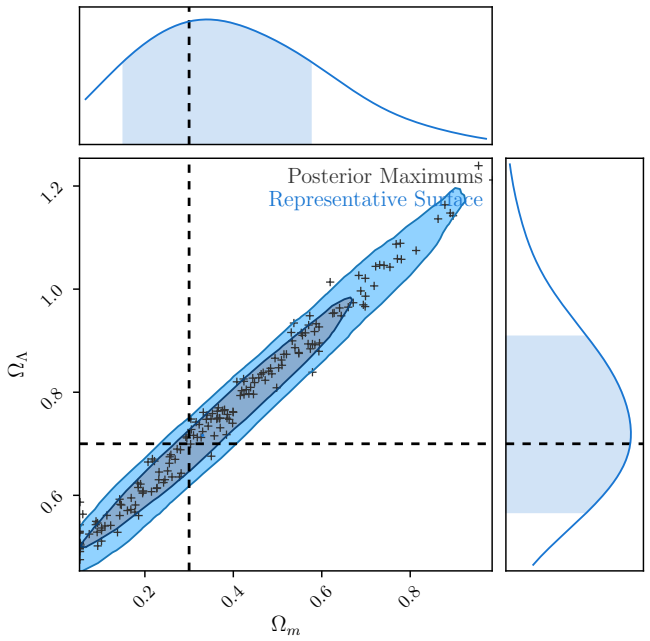

FIG. 6.— Maximal posterior points for 100 realisations of supernova data with the Flat ΛCDM model, with a representative contour from a single data realisation shown for context. Even a large supernova sample when treated robustly is insufficient to provide tight constraints on either Ω_m or Ω_Λ separately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

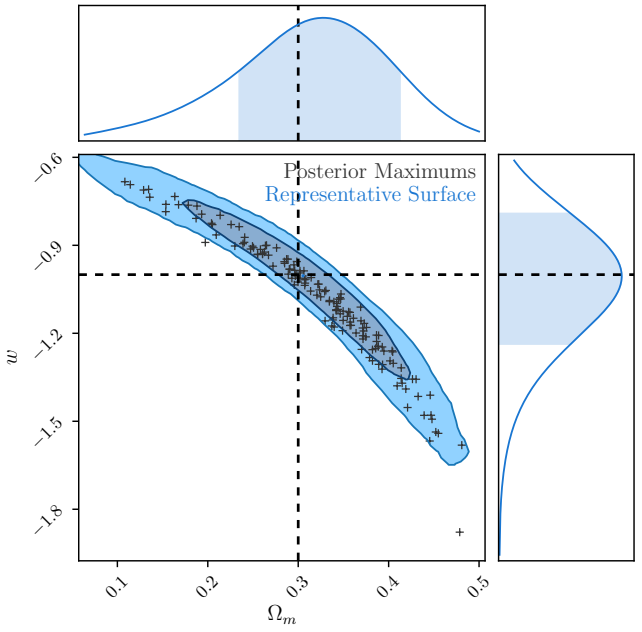

FIG. 7.— Maximal posterior points for 100 realisations of supernova data with the Flat wCDM model, with a representative contour from a single data realisation shown for context. The well known banana shaped contour is recovered, with the marginalised distributions in Ω_m and w providing misleading statistics 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.

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 DEVIATION 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 $\langle\mu\rangle$, $\langle\sigma\rangle$(scatter)	w $\langle\mu\rangle$, $\langle\sigma\rangle$(scatter)
Flat ΛCDM	0.301, 0.015 (0.012)	–
Flat wCDM	–	−1.00, 0.042 (0.030)

4. Calculate $\mu(z)$ given the drawn redshift and cosmological parameters $\Omega_m = 0.3$, $w = -1$ under Flat ΛCDM cosmology. Use this to determine the true apparent magnitude of the object m_B.

5. Determine if the SN Ia 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). 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. We use realistic values for the selection probability to ensure our model is numerically stable with highly skewed selection functions.

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 color uncertainty in quadrature of $\kappa_0 + \kappa_1 z$, where $\kappa_0 = \kappa_1 = 0.03$.

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

We test four models: Flat ΛCDM, Flat wCDM, ΛCDM, and Flat wCDM with a prior $\Omega_m \sim \mathcal{N}(0.3, 0.01)$, with the latter included to allow sensitive tests on bias for w. To achieve statistical precision, we fit 100 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.

To investigate biases in the model in fine detail, we look for systematic bias in Ω_m in the Flat ΛCDM cosmology test, and bias in w for the Flat wCDM test with strong prior $\Omega_m \sim \mathcal{N}(0.3, 0.01)$. This allows us to investigate biases without the investigative hindrances of non-Gaussian or truncated posterior surfaces. The strong prior on Ω_m cuts a slice through the traditional 'banana' posterior surface in the w-Ω_m plane of Figure 7. Without making such a slice, the variation in w is larger due to a shift along the degeneracy direction of the 'banana'. By focusing the slice at an almost fixed Ω_m, we can see the variation in the mean value of w approximately perpendicular to the lines of degeneracy, instead of along them. The results of the analysis are detailed in Table 2, and demonstrate the performance of our model in recovering the true cosmological parameters. With this simple data, we also correctly recover underlying supernova populations, which can be seen in Figure 12.

5.2. *DES SN data validation*

Many BHM methods have previously been validated on data constructed explicitly to validate the assumptions of the model. This is a useful consistency check that the model implementation is correct, efficient and free of obvious pathologies. However, the real test of a model is its application to realistic datasets that mimic expected observational data in as many possible ways. To this end, we test using simulations (using the SNANA package) that follow the observational schedule and observing conditions for the DES and low-z surveys, where the low-z sample is based on observations from CfA3 (Hicken et al. 2009a,b), CfA4 (Hicken et al. 2012) and CSP (Contreras et al. 2010; Folatelli et al. 2010; Stritzinger et al. 2011). Simulation specifics can be found in Kessler et al. (2018).

Additionally, prior analyses often treated intrinsic dispersion simply as scatter in the underlying absolute magnitude of the underlying population (Conley et al. 2011; Betoule et al. 2014), but recent analyses require a more sophisticated approach. In our development of this model and tests of intrinsic

TABLE 3
TESTED POPULATION DISTRIBUTIONS, WHERE THE SK16 LOW-z 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 low-z	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	$^{+0.101}_{0.043}$

dispersion, we analyse the effects of two different scatter models via simulations, the G10 and C11 models described in Section 3. The G10 models dispersion with 70% contribution from coherent variation and 30% from chromatic variation whilst the C11 model has 25% coherent scatter and 75% from chromatic variation. These two broadband scatter models are converted to spectral energy distribution models for use in simulations in Kessler et al. (2013a).

In addition to the improvements in testing multiple scatter models, we also include peculiar velocities for the low-z sample, and our full treatment of systematics as detailed in Brout et al. (2018). Our simulated populations are sourced from Scolnic & Kessler (2016, hereafter SK16) and shown in Table 3. Initial tests were also done with a second, Gaussian population with color and stretch populations centered on zero and 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 cuts, as shown in Figure 8. It is from this simulation that our analytic determination of the selection functions for the low-z and DES survey are based. We run two simulations to determine the efficiency using the G10 and C11 scatter models and find no difference in the functional form of the Malmquist bias between the two models.

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

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

Table 5 shows that the G10 model is consistent with $w = -1$, whilst the C11 model show evidence of bias on w, scattering high. However, their deviation from the truth value represents a shift of approximately 0.5σ when taking into account the uncertainty on fits to w. The bias is 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.

Table 4 shows a clear difference in both β and σ_{m_B} across the G10 and C11 simulations. As expected, the C11 simulations recover a far smaller intrinsic magnitude scatter, giving results of approximately 0.025 when compared to the result of 0.070 for the G10 simulations. The extra smearing of the C11 model does not result in a significantly biased β value compared to the average uncertainty on β, with recovery of $\beta \approx 3.76$ close

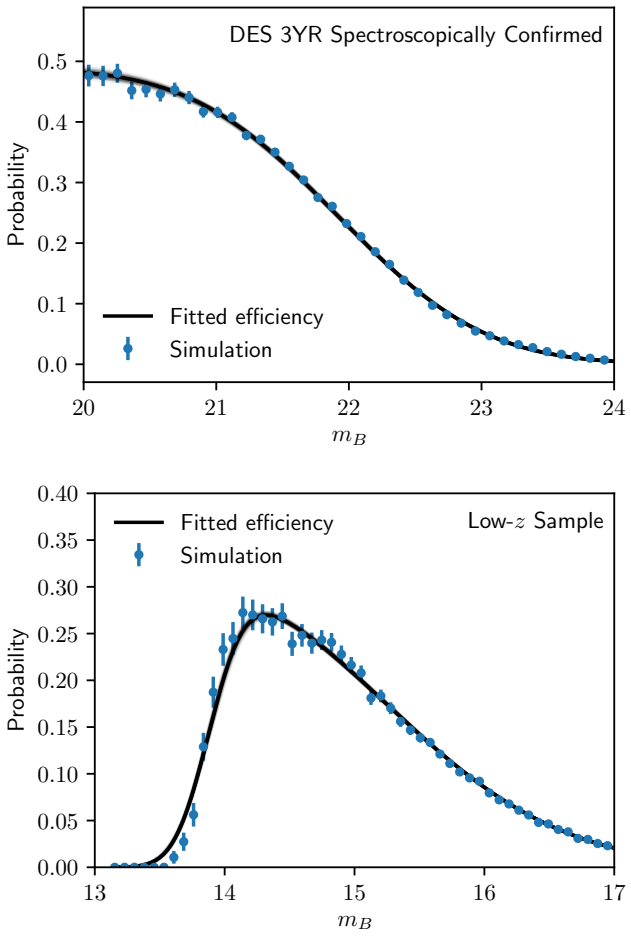

FIG. 8.— Fitting the selection function for both the DES 3YR spectroscopically confirmed supernova sample and the low-z sample. Blue errorbars represent the efficiency calculated by determining the ratio of discovered to generated supernovae in apparent magnitude bins for SNANA simulations. The black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

to the input truth value of 3.8, however the β recovery for the G10 simulations is biased high, finding $\beta \approx 3.44$ with an input of 3.1. Interestingly, w-bias is only found for the C11 simulations. A measure of the significance of the parameter bias can be calculated by comparing the bias to a tenth of the scatter (as our Monte-Carlo estimate uncertainty is $\sqrt{100}$ of the scatter). From this, we can see that most biases are detected with high statistical significance due to the large number of simulations tested against.

We investigate the cosmological bias and find its source to be a bias in the observed summary statistics (i.e. the $\hat{m}_B$, $\hat{x}_1$, $\hat{c}$ output from SALT2 light curve fitting), in addition to incorrect reported uncertainty on the summary statistics. To confirm this, we run two tests. The first of which, we replace the SALT2-fitted $\hat{m}_B$, $\hat{x}_1$ and $\hat{c}$ with random numbers drawn from a Gaussian centered on the true SALT2 m_B, x_1 and c values with covariance as reported by initial light curve fits. With this test, both the G10 and C11 fits recover $w = -1.00$ exactly. Our second test aims to test our model whilst allowing biases in the summary statistics not caused from intrinsic scatter through. To this end, we test a set of 100 simulations generated using an intrinsic dispersion model of only coherent magnitude scatter,

TABLE 4
STANDARDISATION PARAMETERS AND BASE INTRINSIC SCATTER PARAMETER RESULTS FOR THE 100 FITS TO G10 AND C11 SIMULATIONS. WE SHOW THE AVERAGE PARAMETER MEAN AND AVERAGE STANDARD DEVIATION RESPECTIVELY, WITH THE SIMULATION SCATTER SHOWN IN BRACKETS, SUCH THAT EACH CELL SHOWS $\langle\mu\rangle$ [$\langle\sigma\rangle$ (scatter)]. THE WIDTH OF THE INTRINSIC SCATTER (σ_{m_B}) DOES NOT HAVE AN INPUT TRUTH VALUE AS IT IS DETERMINED FROM THE SCATTER MODEL.

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

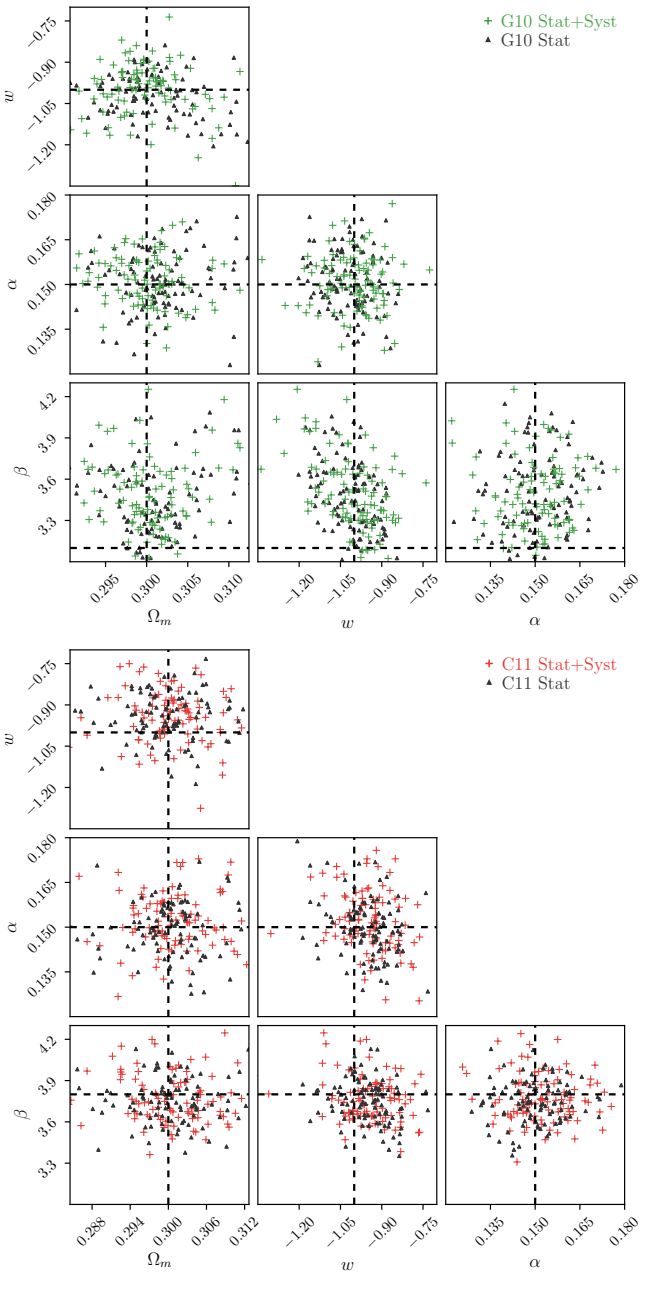

FIG. 9.— Maximum posterior points for 100 realisations of supernova data for two intrinsic dispersion models - the G10 model for the top panel and the C11 model for the bottom panel. Points are shown for parameters Ω_m, w, α and β, with the other fit parameters being marginalised over. As we are unable 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.

TABLE 5
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 AVERAGE UNCERTAINTY BEING SHOWN IN SQUARE BRACKETS AND THE SIMULATION SCATTER (THE STANDARD DEVIATION OF THE MEAN OF 100 FITS) SHOWN IN STANDARD BRACKETS. THE BIAS SIGNIFICANCE REPRESENTS OUR CONFIDENCE THAT THE DEVIATION IN THE MEAN w AWAY FROM -1 IS NOT DUE TO STATISTICAL FLUCTUATION.

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

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 by an unknown scatter model.

Unfortunately, adding extra fit parameters to allow for shifting observables washes out our ability to constrain cosmology, and applying a specific bias correction requires running a fiducial simulation (assuming cosmology, population and scatter model), which presents difficulties when trying to account for correlations with population and scatter model. This is compounded by the fact that bias corrections do not in general improve fits (increase the log posterior), and so are difficult to fit parametrically. Works such as Kessler & Scolnic (2017) show that bias corrections can be applied to 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.

Difficulty in providing an adequate parametrisation for realistic intrinsic dispersion, and the simplification of Malmquist bias to only apparent magnitude also leads to biases in the population parameters. Tests when fixing the population parameters to true does not resolve the cosmological bias observed in the C11 simulations. As the population parameters recovered using the simplistic toy supernova data in the previous section do not exhibit significant bias, future work will focus on intrinsic dispersion and Malmquist bias rather than alternate parametrisations of the underlying supernova population.

Table 6 lists the fit correlations between our model fit parameters (excluding the low-z band systematics, and Malmquist bias uncertainty parameters which had negligible correlation),

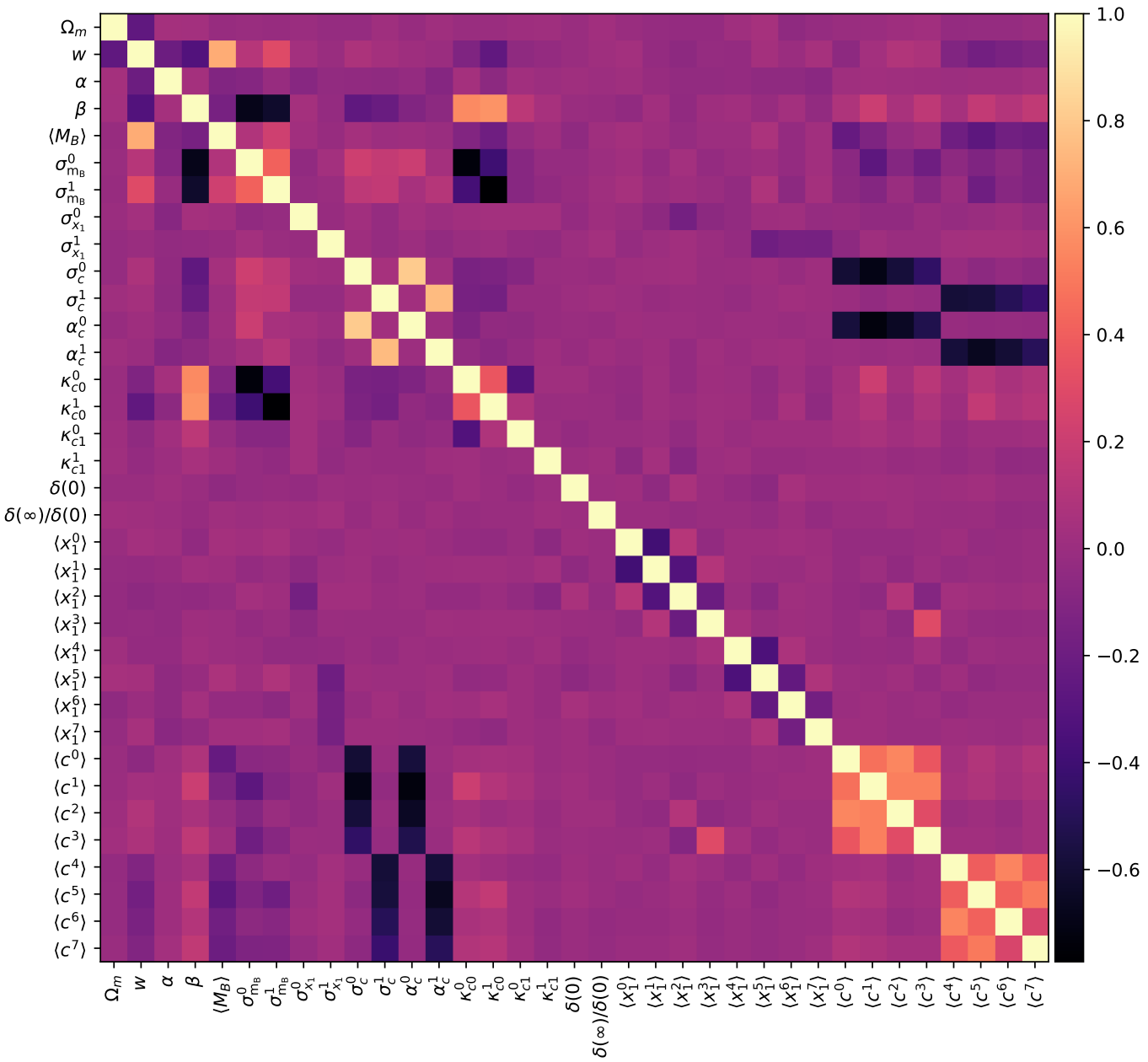

FIG. 10.— Parameter correlations for the combined fits to the 100 G10 scatter model simulations. We see that the primary correlations with w enter through α, β and $\langle M_B \rangle$, as shown in Table 6. Whilst $\langle M_B \rangle$ is generally thought to be a nuisance parameter, we find cosmological correlation. We note that, by fixing H_0 in our distance modulus calculation, $\langle M_B \rangle$ absorbs any cosmological uncertainty on this term. Additionally $\langle M_B \rangle$ also effects the selection efficiency, which was computed from simulations with a fixed M_B value, introducing a second plausible source of correlation. Also visible in this figure are several other interesting relationships. β is strongly anti-correlated with intrinsic dispersion σ_{m_B} for both surveys (DES-like and low-z), 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 color observation), the width of the supernova population in apparent magnitude space increases. As the fit prefers it to conform to the observed width of the distribution, the extra width in color causes the inherent magnitude smearing amount to decrease. And with extra freedom on the observed color from κ^0_c, β shifts in response. The other striking feature in the plot is 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 color distribution's mean value, its width and its skewness. As skewness or population width increases, the effective mean of the population shifts (see Appendix A.3 for details), creating anti-correlation between skewness and the (Gaussian) mean color 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 color dispersion is indeed preferred by our model.

TABLE 6
REDUCED PARAMETER CORRELATIONS WITH w FOR THE COMBINED 100 SIMULATION FITS. CORRELATIONS FOR THE LOW-z BAND SYSTEMATICS AND THE LATENT PARAMETERS REPRESENTING SELECTION FUNCTION UNCERTAINTY ARE NOT SHOWN BUT HAVE NEGLIGIBLE CORRELATION. ZERO SUPERSCRIPTS INDICATE THE DARK ENERGY SURVEY, AND A SUPERSCRIPT ONE REPRESENTS THE LOW-z SURVEY.

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
$\sigma^0_{\mathrm{m_B}}$	0.04	0.07
$\sigma^1_{\mathrm{m_B}}$	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

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. 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 + low-z supernova samples (332 supernova), the bias from intrinsic scatter models is subdominant to the statistical uncertainty, as shown in Figure 9. For our full systematics model, the bias represents a deviation between 0σ to 0.5σ depending on scatter model, and given that they remain sub-dominant, 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 parameter 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-SN3YR sample 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 7 shows that our uncertainty is dominated by statistical error, as the total statistical uncertainty is on w is ± 0.08. With the low number of supernovae in the DES-SN3YR sample, this is expected. We note that the label 'Systematics' in Table 7 represents all numerically computed systematics (as discussed in Section 4.4.4) and systematic uncertainty on the selection function.

5.4. *Methodology Comparison*

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

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

As shown in Brout et al. (2018), the BBC method recovers cosmological parameters without bias so long as the intrinsic scatter model is known. As we do not know the correct intrinsic scatter model, the BBC method averages the results

TABLE 7
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	Parameters	σ_w	Cumulative
Cosmology	Ω_m, w	0.051	0.051
Standardisation	$\alpha, \beta, \langle M_B \rangle, \delta(0), \delta(\infty)/\delta(0)$	0.046	0.068
Intrinsic scatter	κ_0, κ_1	0.020	0.071
Redshift-independent populations	$\sigma_{M_B}, \sigma_c, \sigma_{x_1}, \alpha_c$	0.022	0.074
Redshift-dependent populations	$\langle c_i \rangle, \langle x_{1,i} \rangle$	0.030	0.080
Systematics	$\delta\mathcal{Z}_i, \delta S$	0.054	0.096

TABLE 8
CHARACTERISING THE BIAS ON w USING THE 100 SIMULATIONS FOR THE G10 SCATTER MODEL AND 100 SIMULATIONS FOR C11 SCATTER MODEL. WE ALSO SHOW THE RESULTS WHEN COMBINING THE G10 AND C11 MODELS INTO A COMBINED SET OF 200 SIMULATIONS. THE MEAN w VALUE FOR OUR METHOD AND BBC ARE PRESENTED, ALONG WITH THE MEAN WHEN AVERAGING THE DIFFERENCE BETWEEN OUR METHOD AND BBC FOR EACH INDIVIDUAL SIMULATION. AVERAGES ARE COMPUTED GIVING EACH SIMULATION SAMPLE THE SAME WEIGHT. IN THE MODEL, Δ REPRESENTS STEVE - BBC. THE FINAL ROW SHOWS THE SCATTER BETWEEN STEVE AND BBC FOR THE DIFFERENT SIMULATIONS.

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

when using bias corrections from G10 and from C11. As such, we expect the BBC method to have a w-bias in one direction for G10 simulations and the other direction for C11 simulations. These results are consistent with those displayed in Table 8. Both BBC method and *Steve* are sensitive to the intrinsic scatter model, finding differences of ~ 0.066 and 0.053 respectively in w when varying the scatter model. The BBC method finds w biased low for G10 and 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 improvement in the insensitivity to the intrinsic scatter model (having a decrease in difference in w between the G10 and C11 models), finding no bias for G10 but a w biased high for C11. This decrease in error is not statistically significant as we have statistical uncertainty of ~ 0.01 for 100 simulation realisations. 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-SN3YR sample. When using Planck priors our uncertainty on w is reduced compared to using our simulation Gaussian prior on Ω_m, shrinking the average w-difference from 0.06 to 0.04. After factoring this into our uncertainty, we expect our BHM method to, on average, recover $w_{\rm BHM} = w_{\rm BBC} + 0.04 \pm 0.04$.

Having established that our method exhibits similar shifts in the 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.

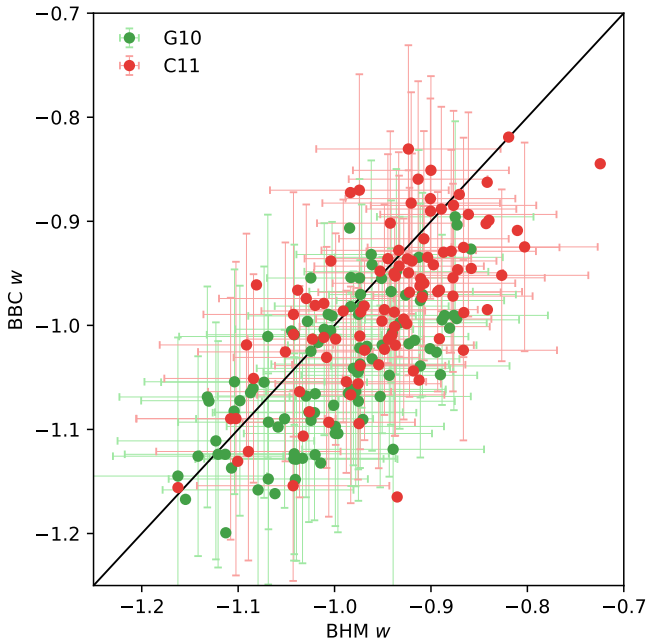

FIG. 11.— Recovered w for the 200 validation simulations with full treatment of statistical and systematic errors. Uncertainty on the recovered w value is shown for every second data point for visual clarity.

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 color smearing and incorporates uncertainty from peculiar velocities, survey calibration, HST calibration, dust, a potential global redshift offset, and SALT2 model uncertainty. Furthermore, our uncertainties in standardisation, population, mass-step and more, being explicitly parametrised in our model, are captured with 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. It runs 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.

[1] This value is computed with each simulation having the same weight. It disagrees with the value provided in Brout et al. (2018, Table 10, row 3) which uses inverse variance weighted averages. We do not utilise this weight because the variance is correlated with the value of w due to the Ω_m prior applied in the fitting process. We note that if inverse variance weighting is applied to both datasets, they both shift by $\Delta w \approx 0.005$, and thus the predicted difference between the BBC method and *Steve* remains the same.

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 statistics 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 supernova 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 60 000 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).

Funding for the DES Projects has been provided by the U.S. Department of Energy, the U.S. National Science Foundation, the Ministry of Science and Education of Spain, the Science and Technology Facilities Council of the United Kingdom, the Higher Education Funding Council for England, the National Center for Supercomputing Applications at the University of Illinois at Urbana-Champaign, the Kavli Institute of Cosmological Physics at the University of Chicago, the Center for Cosmology and Astro-Particle Physics at the Ohio State University, the Mitchell Institute for Fundamental Physics and Astronomy at Texas A&M University, Financiadora de Estudos e Projetos, Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro, Conselho Nacional de Desenvolvimento Científico e Tecnológico and the Ministério da Ciência, Tecnologia e Inovação, the Deutsche Forschungsgemeinschaft and the Collaborating Institutions in the Dark Energy Survey.

The Collaborating Institutions are Argonne National Laboratory, the University of California at Santa Cruz, the University of Cambridge, Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas-Madrid, the University of Chicago, University College London, the DES-Brazil Consortium, the University of Edinburgh, the Eidgenössische Technische Hochschule (ETH) Zürich, Fermi National Accelerator Laboratory, the University of Illinois at Urbana-Champaign, the Institut de Ciències de l'Espai (IEEC/CSIC), the Institut de Física d'Altes Energies, Lawrence Berkeley National Laboratory, the Ludwig-Maximilians Universität München and the associated Excellence Cluster Universe, the University of Michigan, the National Optical Astronomy Observatory, the University of Nottingham, The Ohio State University, the University of Pennsylvania, the University of Portsmouth, SLAC National Accelerator Laboratory, Stanford University, the University of Sussex, Texas A&M University, and the OzDES Membership Consortium.

Based in part on observations at Cerro Tololo Inter-American Observatory, National Optical Astronomy Observatory, which is operated by the Association of Universities for Research in Astronomy (AURA) under a cooperative agreement with the National Science Foundation.

The DES data management system is supported by the National Science Foundation under Grant Numbers AST-1138766 and AST-1536171. The DES participants from Spanish institutions are partially supported by MINECO under grants AYA2015-71825, ESP2015-66861, FPA2015-68048, SEV-2016-0588, SEV-2016-0597, and MDM-2015-0509, some of which include ERDF funds from the European Union. IFAE is partially funded by the CERCA program of the Generalitat de Catalunya. Research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Program (FP7/2007-2013) including ERC grant agreements 240672, 291329, and 306478. We acknowledge support from the Australian Research Council Centre of Excellence for All-sky Astrophysics (CAASTRO), through project number CE110001020, and the Brazilian Instituto Nacional de Ciência e Tecnologia (INCT) e-Universe (CNPq grant 465376/2014-2).

This manuscript has been authored by Fermi Research Alliance, LLC under Contract No. DE-AC02-07CH11359 with the U.S. Department of Energy, Office of Science, Office of High Energy Physics. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.

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
Brout D., Scolnic D., Kessler R., 2018, A&A, p. SYS
Carpenter B., et al., 2017, Journal of Statistical Software, 76, 1
Chambers K. C., et al., 2016, preprint, (`arXiv:1612.05560`)
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
Dark Energy Survey 2018, A&A, p. SYS
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/1709.01091 http://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., 2013a, ApJ, 764, 48
Kessler R., et al., 2013b, The Astrophysical Journal, 764, 48
Kessler R., Brout D., Crawforsd S., 2018, A&A, p. FILLHERE
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
Lewis A., Bridle S., 2002, Phys. Rev. D, 66, 103511
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
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
Sako M., et al., 2018, PASP, 130, 064002
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

SELECTION EFFECT DERIVATION

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.

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

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

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

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 (16) and (17).

Approximate Selection Effects

In this section, we investigate the effect of approximating the skew normal underlying color distribution as a normal. Specifically, equations (A12) and (A13) make the assumption that, for our color distribution, $\mathcal{N}^{\text{Skew}}(\mu,\sigma,\alpha)$ is well approximated by $\mathcal{N}(\mu,\sigma)$. We sought to improve on this approximation by adjusting the mean and standard deviation of the approximated normal to more accurately describe the actual mean and standard deviation of a skew normal. With $\delta \equiv \alpha/\sqrt{1+\alpha^2}$, the correct mean and standard deviation are

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

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

where we highlight that μ here represents the mean of the distribution, not distance modulus. We can then test the approximation $\mathcal{N}^{\text{Skew}}(\mu_0,\sigma_0,\alpha) \to \mathcal{N}(\mu_1,\sigma_1)$. 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 12, and only the shifted normal approximation correctly recovers underlying population parameters.

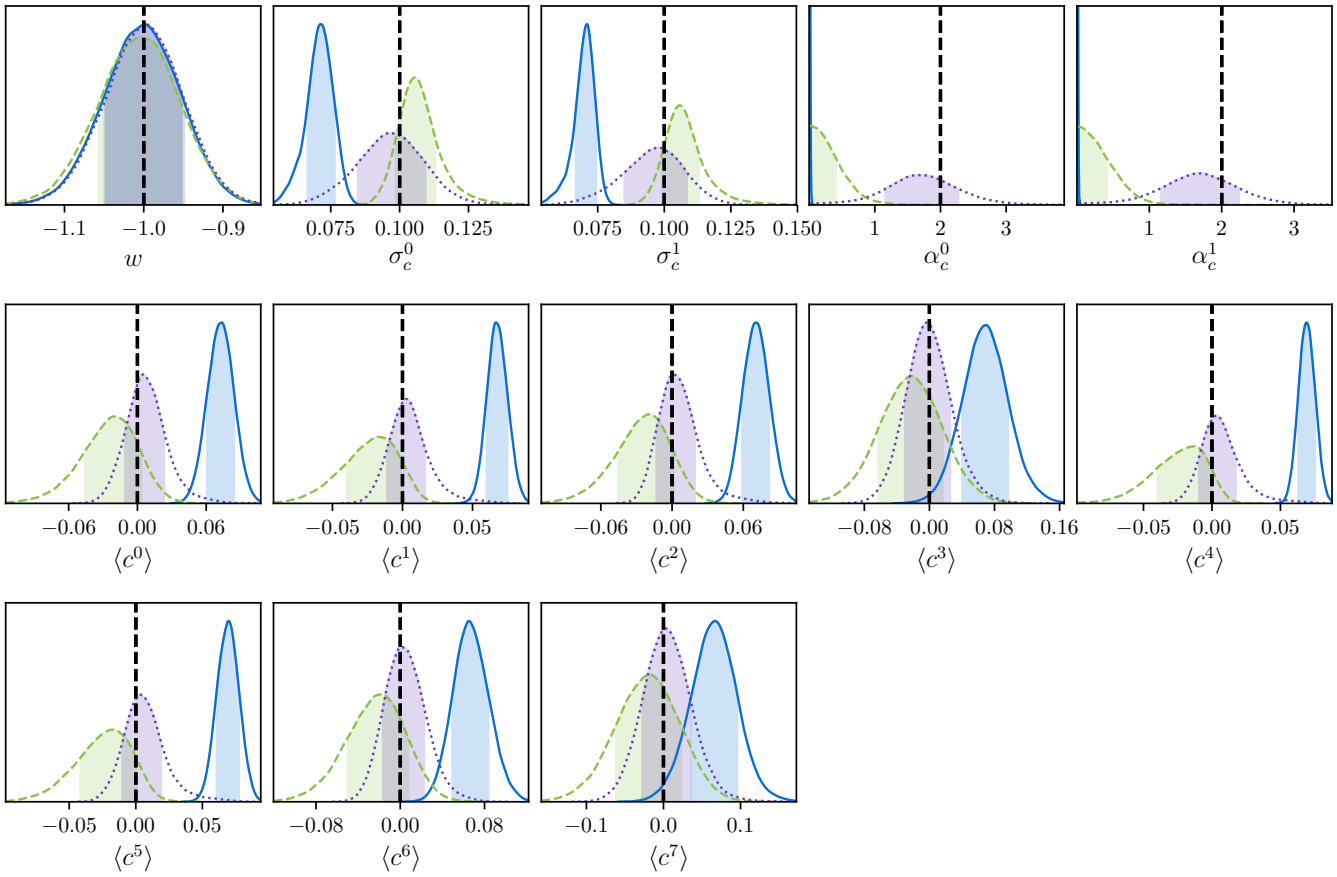

FIG. 12.— Marginalised probability distributions for 100 realisations of cosmology, fit to Flat wCDM with prior $\Omega_m \sim \mathcal{N}(0.3, 0.01)$, each containing 1000 simulated high-z and 1000 simulated low-z supernovae. The dashed green surfaces represent a fit to an underlying Gaussian color population with the unshifted model. The blue solid surface represents fits to a skewed color population with the unshifted model, and the purple dotted surface represents a fit to a skewed color population with the shifted model. The 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 color $\langle c \rangle$ and width of the color distribution σ_c. The unshifted model recovers the correct color mean and width if you approximate a skew normal as a normal: $\Delta\mu = \sqrt{2/\pi}\sigma_c\delta_c \approx 0.071$, which is approximately the deviation found in fits to the color 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 the shifted normal approximation in our model.

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 simply add on a secondary matrix to create an updated covariance. However as the cholesky decomposition of a sum of matrices is not equal to the sum of the cholesky decomposition of each individual matrix,

we would need to recompute the decomposition for each step, which discards most of the computational benefit just gained.

Considering a 3×3 matrix with cholesky decomposition

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

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 color parameter, which is placed as the last element in the observables vector $\{m_B, x_1, c\}$. We can now see that it is possible to add extra uncertainty to the color observation on the diagonal without having to recompute the cholesky decomposition - notice that f is unique in that it is the only element of L that appears in only one position in the covariance matrix. To take our covariance and add on the diagonal uncertainty for color an extra σ_e term, we get

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

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 color 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.