Improved constraints on cosmological parameters from Type Ia supernova data

M. C. March,[1]* R. Trotta,[1,2] P. Berkes,[3] G. D. Starkman[4] and P. M. Vaudrevange[4,5]

[1]*Imperial College London, Astrophysics Group, Blackett Laboratory, Prince Consort Road, London SW7 2AZ*
[2]*African Institute for Mathematical Sciences, 6 Melrose Road, Muizenberg, 7945 Cape Town, South Africa*
[3]*Brandeis University, Volen Center for Complex Systems, 415 South Street, Waltham, MA 02454-9110, USA*
[4]*CERCA, Department of Physics, Case Western Reserve University, 10900 Euclid Ave, Cleveland, OH 44106, USA*
[5]*DESY, Notkestrasse 85, 22607 Hamburg, Germany*

ABSTRACT

We present a new method based on a Bayesian hierarchical model to extract constraints on cosmological parameters from Type Ia supernova (SNIa) data obtained with the SALT-II light-curve fitter. We demonstrate with simulated data sets that our method delivers tighter statistical constraints on the cosmological parameters over 90 per cent of the time, that it reduces statistical bias typically by a factor of $\sim$2–3 and that it has better coverage properties than the usual χ^2 approach. As a further benefit, a full posterior probability distribution for the dispersion of the intrinsic magnitude of SNe is obtained. We apply this method to recent SNIa data, and by combining them with cosmic microwave background and baryonic acoustic oscillations data, we obtain $\Omega_m = 0.28 \pm 0.02$, $\Omega_\Lambda = 0.73 \pm 0.01$ (assuming $w = -1$) and $\Omega_m = 0.28 \pm 0.01$, $w = -0.90 \pm 0.05$ (assuming flatness; statistical uncertainties only). We constrain the intrinsic dispersion of the B-band magnitude of the SNIa population, obtaining $\sigma_\mu^{int} = 0.13 \pm 0.01$ mag. Applications to systematic uncertainties will be discussed in a forthcoming paper.

Key words: methods: statistical – supernovae: general – cosmological parameters – dark energy.

1 INTRODUCTION

Since the late 1990s when the Supernova Cosmology Project and the High-Z Supernova Search Team presented evidence that the expansion of the Universe is accelerating (Riess et al. 1998; Perlmutter et al. 1999), observations of Type Ia supernovae (SNeIa) have been seen as one of our most important tools for measuring the cosmic expansion as a function of time. Since a precise measurement of the evolution of the scalefactor is likely the key to characterizing the dark energy or establishing that general relativity must be modified on cosmological scales, the limited data that our universe affords us must be used to the greatest possible advantage. One important element of that task is the most careful possible statistical analysis of the data. Here, we report how an improved statistical approach to SNIa data, with, in particular, more consistent treatment of uncertainties, leads to significant improvements in both the precision and accuracy of inferred cosmological parameters.

The fundamental assumption underlying the past and proposed use of SNeIa to measure the expansion history is that they are 'standardizable candles'. SNeIa occur when material accreting on to a white dwarf from a companion drives the mass of the white dwarf above the maximum that can be supported by electron degeneracy pressure, the Chandrasekhar limit of about $1.4 \, M_\odot$. This triggers the collapse of the star and the explosive onset of carbon fusion, in turn powering the SN explosion. Because the collapse happens at a particular critical mass, all SNeIa are similar. Nevertheless, variability in several factors, including composition, rotation rate and accretion rate, can lead to measurable differences in SN observables as a function of time. Indeed, the intrinsic magnitude of the nearby SNeIa, the distances to which are known via independent means, exhibits a fairly large scatter. Fortunately, this scatter can be reduced by applying the so-called 'Phillips corrections' – phenomenological correlations between the intrinsic magnitude of SNeIa and their colour as well as between their intrinsic magnitudes and the time-scale for the decline of their luminosity (Phillips 1993; Phillips et al. 1999). Such corrections are derived from multiwavelength observations of the SNIa light curves (i.e. their apparent brightness as a function of time). Fortuitously, they make SNIa into standardizable candles – in other words, having measured the colour and light curve of an SNIa, one can infer its intrinsic magnitude with relatively low scatter, typically in the range 0.1–0.2 mag. Observations of SNIa at a range of redshifts can then be used to measure

*E-mail: marisa.march06@imperial.ac.uk

the evolution of luminosity distance as a function of redshift, and thence infer the evolution of the scalefactor, assuming that the intrinsic properties of SNIa do not themselves evolve (an assumption that has to be carefully checked).

The SNIa sample which is used to measure distances in the Universe has grown massively thanks to a world-wide observational effort (Astier et al. 2006; Wood-Vasey, Miknaitis & Stubbs 2007; Bailey et al. 2008; Kowalski et al. 2008; Balland et al. 2009; Freedman et al. 2009; Hicken et al. 2009; Kessler et al. 2009a; Amanullah et al. 2010; Contreras et al. 2010). Presently, several hundred SNeIa have been observed, a sample which is set to increase by an order of magnitude in the next 5 years or so. As observations have become more accurate and refined, discrepancies in their modelling have come into focus. Two main methods have emerged to perform the light-curve fit and derive cosmological parameter constraints. The multicolour light-curve shape (MLCS; Jha, Riess & Kirshner 2007) strategy is to simultaneously infer the Phillips corrections and the cosmological parameters of interest, applying a Bayesian prior to the parameter controlling extinction. The SALT and SALT-II (Guy et al. 2007) methodology splits the process in two steps. First, the Phillips corrections are derived from the light-curve data; the cosmological parameters are then constrained in a separate inference step. As the SN sample has grown and improved, the tension between the results of the two methods has increased.

Despite the past and anticipated improvements in the SN sample, the crucial inference step of deriving cosmological constraints from the SALT-II light-curve fits has remained largely unchanged. For details of how the cosmology fitting is currently done, see, for example, Astier et al. (2006), Kowalski et al. (2008), Amanullah et al. (2010) and Conley et al. (2011). As currently used, it suffers from several shortcomings, such as not allowing for rigorous model checking and not providing a rigorous framework for the evaluation of systematic uncertainties. The purpose of this paper is to introduce a statistically principled, rigorous, Bayesian hierarchical model (BHM) for cosmological parameter fitting to SNIa data from SALT-II light-curve fits. In particular, the method addresses identified shortcomings of standard χ^2 approaches – notably, it properly accounts for the dependence of the errors in the distance moduli of the SNe on fitted parameters. It also treats more carefully the uncertainty on the Phillips colour correction parameter, escaping the potential bias caused by the fact that the error is comparable to the width of the distribution of its value.

We will show that our new method delivers considerably tighter statistical constraints on the parameters of interest, while giving a framework for the full propagation of systematic uncertainties to the final inferences. (This will be explored in an upcoming work.) We also apply our BHM to current SNIa data, and derive new cosmological constraints from them. We derive the intrinsic scatter in the SNIa absolute magnitude and obtain a statistically sound uncertainty on its value.

This paper is organized as follows. In Section 2, we review the standard method used to perform cosmological fits from SALT-II light-curve results and we describe its limitations. We then present a new, fully Bayesian method, which we test on simulated data in Section 3, where detailed comparisons of the performance of our new method with the standard approach are presented. We apply our new method to current SN data in Section 4 and give our conclusions in Section 5.

2 COSMOLOGY FROM SALT-II LIGHT-CURVE FITS

2.1 Definition of the inference problem

Several methods are available to fit SNe light curves, including the MLCS method, the Δm_{15} method, CMAGIC (Wang et al. 2003; Conley et al. 2006), SALT, SALT-II and others. Recently, a sophisticated Bayesian hierarchical method to fit optical and infrared light-curve data has been proposed by Mandel et al. (2009) and Mandel, Narayan & Kirshner (2010). As mentioned above, MLCS fits the cosmological parameters at the same time as the parameters controlling the light-curve fits. The SALT and SALT-II methods, on the contrary, first fit to the SNe light curves three parameters controlling the SN magnitude, the stretch and colour corrections. From those fits, the cosmological parameters are then fitted in a second, separate step. In this paper, we will consider the SALT-II method (although our discussion is equally applicable to SALT), and focus on the second step in the procedure, namely the extraction of cosmological parameters from the fitted light curves. We briefly summarize below the light-curve fitting step, on which our method builds.

The rest-frame flux at wavelength λ and time t is fitted with the expression

$$\frac{\mathrm{d}F_{\mathrm{rest}}}{\mathrm{d}\lambda}(t, \lambda) = x_0 \left[M_0(t, \lambda) + x_1 M_1(t, \lambda)\right] \exp\left[c \cdot CL(\lambda)\right], \tag{1}$$

where M_0, M_1, CL are functions determined from a training process, while the fitted parameters are x_0 (which controls the overall flux normalization), x_1 (the stretch parameter) and c (the colour correction parameter). The B-band apparent magnitude m_B^* is related to x_0 by the expression

$$m_B^* = -2.5 \log\left[x_0 \int \mathrm{d}\lambda M_0(t = 0, \lambda) T^B(\lambda)\right], \tag{2}$$

where $T^B(\lambda)$ is the transmission curve for the observer's B band, and $t = 0$ is by convention the time of peak luminosity. After fitting the SNIa light-curve data with SALT-II algorithm, e.g. Kessler et al. (2009a) report the best-fitting values for m_B^*, x_1, c, the best-fitting redshift z of each SNIa and a covariance matrix $\hat{C}_i$ for each SN, describing the covariance between m_B^*, x_1, c from the fit, of the form

$$\hat{C}_i = \begin{pmatrix} \sigma_{m_B^* i}^2 & \sigma_{m_B^* i, x_1 i} & \sigma_{m_B^* i, ci} \\ \sigma_{m_B^* i, x_1 i} & \sigma_{x_1 i}^2 & \sigma_{x_1 i, ci} \\ \sigma_{m_B^* i, ci} & \sigma_{x_1 i, ci} & \sigma_{ci}^2 \end{pmatrix}. \tag{3}$$

Let us denote the result of the SALT-II light-curve fitting procedure for each SN as

$$D_i = \{\hat{z}_i, \hat{m}^*_{Bi}, \hat{x}_{1i}, \hat{c}_i, \hat{C}_i\} \qquad (4)$$

(where i runs through the n SNe in the sample, and measured quantities are denoted by a hat). We assume (as it is implicitly done in the literature) that the distribution of $\hat{m}^*_{Bi}, \hat{x}_{1i}, \hat{c}_i$ is a multinormal Gaussian with covariance matrix $\hat{C}_i$.

The distance modulus μ_i for each SN (i.e. the difference between its apparent B-band magnitude and its absolute magnitude) is modelled as

$$\mu_i = m^*_{Bi} - M_i + \alpha \cdot x_{1i} - \beta \cdot c_i, \qquad (5)$$

where M_i is the (unknown) B-band absolute magnitude of the SN, while α, β are nuisance parameters controlling the stretch and colour correction (so-called 'Phillips corrections'), respectively, which have to be determined from the data at the same time as the parameters of interest. The purpose of applying the Phillips corrections is to reduce the scatter in the distance modulus of the SNe, so they can be used as almost standard candles. However, even after applying the corrections, some intrinsic dispersion in magnitude is expected to remain. Such intrinsic dispersion can have physical origin [e.g. host galaxies properties such as mass (Kelly et al. 2010; Sullivan et al. 2011) and star formation rate (Sullivan et al. 2006), host galaxy reddening (Mandel et al. 2010), possible SNeIa evolution (Gonzalez-Gaitan et al. 2011)] or be associated with undetected or underestimated systematic errors in the surveys. Below, we show how to include the intrinsic dispersion explicitly in the statistical model.

Turning now to the theoretical predictions, the cosmological parameters we are interested in constraining are

$$\mathscr{C} = \{\Omega_m, \Omega_\Lambda \text{ or } w, h\}, \qquad (6)$$

where Ω_m is the matter density (in units of the critical energy density), Ω_Λ is the dark energy density, w is the dark energy equation of state (EOS; taken to be redshift-independent, although this assumption can easily be generalized) and h is defined as $H_0 = 100\, h\, \text{km s}^{-1}\, \text{Mpc}^{-1}$, where H_0 is the value of the Hubble constant today.[1] The curvature parameter Ω_κ is related to the matter and dark energy densities by the constraint equation

$$\Omega_\kappa = 1 - \Omega_m - \Omega_\Lambda. \qquad (7)$$

In the following, we shall consider either a Universe with non-zero curvature (with $\Omega_\kappa \neq 0$ and an appropriate prior) but where the dark energy is assumed to be a cosmological constant, i.e. with $w = -1$ [the Λ cold dark matter (ΛCDM) model], or a flat Universe where the effective EOS parameter is allowed to depart from the cosmological constant value, i.e. $\Omega_\kappa = 0$, $w \neq -1$ (the wCDM model).

In a Friedmann–Robertson–Walker cosmology defined by the parameters $\mathscr{C}$, the distance modulus to an SN at redshift z_i is given by

$$\mu_i = \mu(z_i, \mathscr{C}) = 5 \log \left[\frac{D_L(z_i, \mathscr{C})}{\text{Mpc}} \right] + 25, \qquad (8)$$

where D_L denotes the luminosity distance to the SN. This can be rewritten as

$$\mu_i = \eta + 5 \log d_L(z_i, \Omega_m, \Omega_\Lambda, w), \qquad (9)$$

where

$$\eta = -5 \log \frac{100h}{c} + 25 \qquad (10)$$

and c is the speed of light in km s^{-1}. We have defined the dimensionless luminosity distance (with $D_L = c/H_0 d_L$, where c is the speed of light)

$$d_L(z, \Omega_m, \Omega_\Lambda, w) = \frac{(1+z)}{\sqrt{|\Omega_\kappa|}} \text{sinn} \left\{ \sqrt{|\Omega_\kappa|} \int_0^z dz' [(1+z')^3 \Omega_m + \Omega_{de}(z') + (1+z')^2 \Omega_\kappa]^{-1/2} \right\}, \qquad (11)$$

with the dark energy density parameter

$$\Omega_{de}(z) = \Omega_\Lambda \exp \left[3 \int_0^z \frac{1 + w(x)}{1 + x} dx \right]. \qquad (12)$$

In the above equation, we have been completely general about the functional form of the dark energy EOS, $w(z)$. In the rest of this work, however, we will make the further assumption that w is constant with redshift, i.e. $w(z) = w$. We have defined the function $\text{sinn}(x) = x$, $\sin(x)$ or $\sinh(x)$ for a flat Universe ($\Omega_\kappa = 0$), a closed Universe ($\Omega_\kappa < 0$) or an open Universe, respectively.

The problem now is to infer, given data D in equation (4), the values (and uncertainties) of the cosmological parameters $\mathscr{C}$, as well as the nuisance parameters $\{\alpha, \beta\}$, appearing in equation (5) and any further parameter describing the SNe population and its intrinsic scatter. Before building a full BHM to solve this problem, we briefly describe the usual approach and its shortcomings.

2.2 Shortcomings of the usual χ^2 method

The usual analysis (e.g. Astier et al. 2006; Kowalski et al. 2008; Kessler et al. 2009a; Conley et al. 2011) defines a χ^2 statistics as follows:

$$\chi^2_\mu = \sum_i \frac{(\mu_i - \mu_i^{\text{obs}})^2}{\sigma^2_{\mu i}}, \qquad (13)$$

---

[1] The Hubble parameter h actually plays the role of a nuisance parameter, as it cannot be constrained by distance modulus measurements independently unless the absolute magnitude of the SNe is known, for the two quantities are perfectly degenerate.

where μ_i is given by equation (9) as a function of the cosmological parameters, and the 'observed' distance modulus μ_i^{obs} is obtained by replacing in equation (5) the best-fitting values for the colour and stretch correction and B-band magnitude from the SALT-II output (denoted by hats). Furthermore, the intrinsic magnitude for each SN, M_i, is replaced by a global parameter M, which represents the mean intrinsic magnitude of all SNe in the sample:

$$\mu_i^{\text{obs}} = \hat{m}_{Bi}^* - M + \alpha \cdot \hat{x}_{1i} - \beta \cdot \hat{c}_i, \tag{14}$$

where the mean intrinsic magnitude M is unknown. The variance $\sigma_{\mu i}^2$ comprises several sources of uncertainty, which are added in quadrature:

$$\sigma_{\mu i}^2 = \left(\sigma_{\mu i}^{\text{fit}}\right)^2 + \left(\sigma_{\mu i}^{z}\right)^2 + \left(\sigma_{\mu}^{\text{int}}\right)^2, \tag{15}$$

where $\sigma_{\mu i}^{\text{fit}}$ is the statistical uncertainty from the SALT-II light-curve fit:

$$\left(\sigma_{\mu i}^{\text{fit}}\right)^2 = \underline{\Psi}^{\text{T}} \hat{C}_i \underline{\Psi}, \tag{16}$$

where $\Psi = (1, \alpha, -\beta)$ and $\hat{C}_i$ is the covariance matrix given in equation (3). $\sigma_{\mu i}^{z}$ is the uncertainty on the SN redshift from spectroscopic measurements and peculiar velocities of and within the host galaxy; finally, $\sigma_{\mu}^{\text{int}}$ is an unknown parameter describing the SN intrinsic dispersion. For further discussions of the unknown $\sigma_{\mu}^{\text{int}}$ estimation problem, see Blondin, Mandel & Kirshner (2011), Kim (2011) and Gopal Vishwakarma & Narlikar (2011). As mentioned above, ideally $\sigma_{\mu}^{\text{int}}$ is a single quantity that encapsulates the remaining intrinsic dispersion in the SNIa sample, folding in all of the residual scatter due to physical effects not well captured by the Phillips corrections. However, observational uncertainties such as the estimation of photometric errors can lead to a variation of $\sigma_{\mu}^{\text{int}}$ sample by sample (for which there is a growing body of evidence). While we do not consider the latter scenario in this paper, it is important to keep in mind that describing the whole SN population with a single scatter parameter $\sigma_{\mu}^{\text{int}}$ is likely to be an oversimplification.

Further error terms are added in quadrature to the total variance, describing uncertainties arising from dispersion due to lensing, Milky Way dust extinction, photometric zero-point calibration, etc. In this work, we do not deal with such systematic uncertainties, though they can be included in our method and we comment further on this below.

The cosmological parameter fit proceeds by minimizing the χ^2 in equation (13), simultaneously fitting the cosmological parameters, α, β and the mean intrinsic SN magnitude M. The value of $\sigma_{\mu}^{\text{int}}$ is adjusted to obtain $\chi_{\mu}^2/\text{degreesoffreedom(d.o.f.)} \sim 1$ (usually on a sample-by-sample basis), often rejecting individual SNe with a residual pull larger than some arbitrarily chosen cut-off. It was recognized early that fitting the numerator and denominator of equation (13) iteratively leads to a large 'bias' in the recovered value of β (Astier et al. 2006; Wang et al. 2006; Kowalski et al. 2008). This has been traced back to the fact that the error on the colour correction parameter c_i is as large as or larger than the width of the distribution of values of c_i, especially for high-redshift SNe. This is a crucial observation, which constitutes the cornerstone of our BHM, as explained below. We demonstrate that an appropriate modelling of the distribution of values of c_i leads to an effective likelihood that replaces the χ^2 of equation (13). With this effective likelihood and appropriate Bayesian priors, all parameters can be recovered without bias. If instead one adopts a properly normalized likelihood function, i.e. replacing the χ^2 of equation (13) with

$$\mathcal{L} = L_0 \exp\left(-\frac{1}{2}\chi_{\mu}^2\right) \tag{17}$$

(with the pre-factor L_0 chosen so that the likelihood function integrates to unity in data space), marginalization over α, β leads to catastrophic biases in the recovered values (up to $\sim 6\sigma$ in some numerical tests we performed). This is a strong hint that the naive form of the likelihood function above is incorrect for this problem. The effective likelihood we derive below solves this problem.

The standard approach to cosmological parameters fitting outlined above adopted in most of the literature to date has several shortcomings, which can be summarized as follows.

(i) The expression for the χ^2, equation (13), has no fundamental statistical justification, but is based on a heuristic derivation. The fundamental problem with equation (13) is that some of the parameters being fitted (namely α, β) control both the location and the dispersion of the χ^2 expression, as they appear both in the numerator and the denominator, via the $(\sigma_{\mu i}^{\text{fit}})^2$ term. Therefore, the statistical problem is one of jointly estimating both the location and the variance. We show below how this can be tackled using a principled statistical approach.

(ii) Adjusting $\sigma_{\mu}^{\text{int}}$ to obtain the desired goodness of fit is problematic, as it does not allow one to carry out any further goodness-of-fit test on the model itself, for obviously the variance has been adjusted to achieve a good fit by construction. This means that model checking is by construction not possible with this form of the likelihood function.

(iii) It would be interesting to obtain not just a point estimate for $\sigma_{\mu}^{\text{int}}$, but an actual probability distribution for it, as well. This would allow consistency checks, e.g., among different surveys, to verify whether the recovered intrinsic dispersions are mutually compatible (within error bars). This is currently not possible given the standard χ^2 method. A more easily generalizable approach is desirable, that would allow one to test the hypothesis of multiple SNe populations with different values of intrinsic dispersion, for example as a consequence of evolution with redshift or correlated with host galaxy properties. Current practice is to split the full SN sample in subsamples (e.g. low redshift and high redshift, or for different values of the colour correction) and check for the consistency of the recovered values from each of the subsamples. Our method allows for a more systematic approach to this kind of important model checking procedure.

(iv) It is common in the literature to obtain inferences on the parameters of interest by minimizing (i.e. profiling) over nuisance parameters entering in equation (13). This is in general much more computationally costly than marginalization from, e.g., Markov chain Monte Carlo (MCMC) samples (Feroz et al. 2011a). There are also examples where some nuisance parameters are marginalized over, while others are maximized (Astier et al. 2006), which is statistically inconsistent and should best be avoided. It should also be noted that maximization
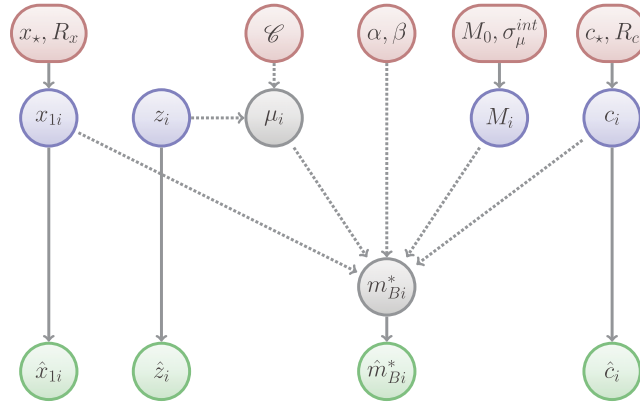
Figure 1. Graphical network showing the deterministic (dashed) and probabilistic (solid) connections between variables in our BHM. Variables of interest are in red, latent (unobserved) variables are in blue and observed data (denoted by hats) are in green.

and marginalization do not in general yield the same errors on the parameters of interest when the distribution is non-Gaussian. From a computational perspective, it would be advantageous to adopt a fully Bayesian method, which can be used in conjunction with fast and efficient MCMC and nested sampling techniques for the exploration of the parameter space. This would also allow one to adopt Bayesian model selection methods (which cannot currently be used with the standard χ^2 approach as they require the full marginalization of parameters to compute the Bayesian evidence).

(v) The treatment of systematic errors is being given great attention in the recent literature (see e.g. Nordin, Goobar & Jonsson 2008), but the impact of various systematics on the final inference for the interesting parameters, $\mathscr{C}$, has often been propagated in an approximate way (e.g. Kessler et al. 2009a, appendix F). As we are entering an epoch when SN cosmology is likely to be increasingly dominated by systematic errors, it would be desirable to have a consistent way to include sources of systematic uncertainties in the cosmological fit and to propagate the associated error consistently on the cosmological parameters. The inclusion in the analysis pipeline of systematic error parameters has been hampered so far by the fact that this increases the number of parameters being fitted above the limit of what current methods can handle. However, if a fully Bayesian expression for the likelihood function was available, one could then draw on the considerable power of Bayesian methods (such as MCMC or nested sampling) which can efficiently handle larger parameter spaces.

Motivated by the above problems and limitations of the current standard method, we now proceed to develop in the next section a fully Bayesian formalism from first principles, leading to the formulation of a new effective likelihood which will overcome, as will be shown below, the above problems. A more intuitive understanding of our procedure can be acquired from the simpler toy problem described in Appendix B.

2.3 The Bayesian hierarchical model

We now turn to developing a BHM for the SNe data from SALT-II light-curve fits. The same general linear regression problem with unknown variance has been addressed by Kelly (2007), and applied in that paper to X-ray spectral slope fitting. The gist of our method is shown in the graphical network of Fig. 1, which displays the probabilistic and deterministic connection between variables. The fundamental idea is that we introduce a new layer of so-called 'latent' variables – that is, quantities which describe the 'true' value of the corresponding variables, and which are obviously unobserved (represented in blue in Fig. 1). In particular, we associate with each SN a set of latent variables (z_i, c_i, x_{1i}, M_i), which give the true value of the redshift, colour correction, stretch correction and intrinsic magnitude for that SN. For the stretch and colour correction, we have noisy estimates in the form of data $\hat{c}_i$, $\hat{x}_{1i}$ (with an associated covariance matrix) resulting from the SALT-II fits (solid vertical lines). For the redshift z_i, we have noisy estimates given by the measured values $\hat{z}_i$. As above, all observed quantities are denoted by a hat.

Each of the latent variables (c_i, x_{1i}, M_i) is modelled as an uncorrelated random variable drawn from an underlying parent distribution, described by a Gaussian with two parameters, a population mean and a variance.[2] The latent intrinsic magnitudes M_i are thought of as realizations from an SN population with overall mean intrinsic magnitude M_0 and intrinsic dispersion σ_μ^{int} (both to be estimated from the data):

$$M_i \sim \mathcal{N}(M_0, (\sigma_\mu^{\text{int}})^2). \tag{19}$$

---

[2] In this paper, we use the notation

$$x \sim \mathcal{N}(m, \sigma^2) \tag{18}$$

to denote a random variable x being drawn from an underlying Gaussian distribution with mean m and variance σ^2. In vector notation, m is replaced by a vector $\underline{m}$, while σ^2 is replaced by the covariance matrix Σ and the distribution is understood to be the appropriate multidimensional normal (see Appendix A).
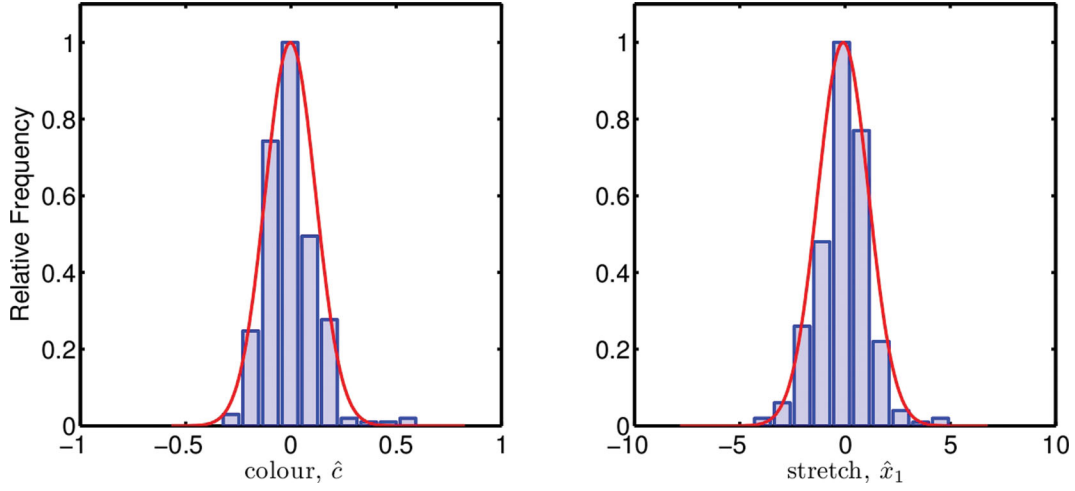
Figure 2. Histogram of observed stretch parameters $\hat{x}_{1i}$ and observed colour parameters $\hat{c}_i$ from the 288 SNeIa from Kessler et al. (2009a), compared with a Gaussian fit (red curve).

This is represented in Fig. 1 by the solid arrow connecting the parameters M_0, $(\sigma_\mu^{\text{int}})^2$ with the latent variables M_i. This represents statistically the physical statement that an intrinsic dispersion $(\sigma_\mu^{\text{int}})^2$ is expected to remain in the distribution of SNe magnitudes, even after applying the Phillips corrections to the observed magnitudes. The simple *ansatz* of equation (19) can be replaced by a more refined modelling, which describes the existence of multiple SN populations, for example with intrinsic magnitude correlated with host galaxy properties, for which there is a growing body of evidence (Sullivan et al. 2006, 2011; Mandel et al. 2010). We shall investigate this possibility in a forthcoming work.

The parent distribution of the true colour and stretch corrections is similarly represented by Gaussians, again parametrized each by a mean $(c_\star, x_\star)$ and a variance (R_c^2, R_x^2) as

$$c_i \sim \mathcal{N}\left(c_\star, R_c^2\right), \quad x_{1i} \sim \mathcal{N}\left(x_\star, R_x^2\right). \tag{20}$$

As above, the quantities $c_\star, R_c^2, x_\star, R_x^2$ have to be estimated from the data. The choice of a Gaussian distribution for the latent variables $\underline{c}$ and $\underline{x}_1$ is justified by the fact that the observed distribution of $\underline{\hat{c}}$ and $\underline{\hat{x}}_1$, shown in Fig. 2 for the actual SNIa sample described in Section 3.1 below, is fairly well described by a Gaussian. As shown in Fig. 2, there might be a hint for a heavier tail for positive values of $\underline{\hat{c}}$, but this does not fundamentally invalidate our Gaussian approximation. It would be easy to expand our method to consider other distributions, for example mixture models of Gaussians to describe a more complex population or a distribution with heavier tails, if non-Gaussianities in the observed distribution should make such modelling necessary. In this paper, we consider the simple unimodal Gaussians given by equation (20).

While we are interested in the value (and distribution) of the SNe population parameters M_0, $(\sigma_\mu^{\text{int}})^2$, the parameters entering equations (20) are not physically interesting. Therefore, they will be marginalized out at the end. It is, however, crucial to introduce them explicitly in this step: it is precisely the lack of modelling in the distribution of c_i that usually leads to the observed 'biases' in the reconstruction of β (see Appendix B for an explicit example with a toy model). For ease of notation, we rewrite equations (19) and (20) in matrix notation as

$$\underline{M} \sim \mathcal{N}(\underline{M}_0, \Sigma_\Delta), \tag{21}$$

$$\underline{c} \sim \mathcal{N}\left(c_\star \cdot \mathbf{1}_n, \text{diag}\left(R_c^2 \cdot \mathbf{1}_n\right)\right) = p(\underline{c} \,|\, c_\star, R_c), \tag{22}$$

$$\underline{x}_1 \sim \mathcal{N}\left(x_\star \cdot \mathbf{1}_n, \text{diag}\left(R_x^2 \cdot \mathbf{1}_n\right)\right) = p(\underline{x}_1 \,|\, x_\star, R_x), \tag{23}$$

where

$$\underline{M} = (M_1, \ldots, M_n) \in \mathbb{R}^n, \tag{24}$$

$$\underline{M}_0 = M_0 \cdot \mathbf{1}_n \in \mathbb{R}^n, \tag{25}$$

$$\Sigma_\Delta = \text{diag}\left(\left(\sigma_\mu^{\text{int}}\right)^2 \cdot \mathbf{1}_n\right) \in \mathbb{R}^{n \times n}. \tag{26}$$

Having introduced $3n$ latent (unobserved) variables $(\underline{c}, \underline{x}_1, \underline{M})$, where n is the number of SNe in the sample, the fundamental strategy of our method is to link them to underlying population parameters via equations (19) and (20), then to use the observed noisy estimates to infer constraints on the population parameters of interest (alongside the cosmological parameters), while marginalizing out the unobserved latent variables.

The intrinsic magnitude M_i is related to the observed B-band magnitude $\hat{m}_B^*$ and the distance modulus μ by equation (5), which can be rewritten in vector notation as

$$\underline{m}_B^* = \underline{\mu} + \underline{M} - \alpha \underline{x}_1 + \beta \underline{c}. \tag{27}$$

Note that the above relation is *exact*, i.e. $\underline{M}, \underline{x}_1, \underline{c}$ are here the latent variables (not the observed quantities), while $\underline{m}_B^*$ is the true value of the B-band magnitude (also unobserved). This is represented by the dotted (deterministic) arrows connecting the variables in Fig. 1.

We seek to determine the posterior probability density function (pdf) for the parameters of interest $\Theta = \{\mathscr{C}, \alpha, \beta, \sigma_\mu^{\text{int}}\}$, while marginalizing over the unknown population mean intrinsic magnitude, M_0. From Bayes' theorem, the marginal posterior for Θ is given by (see e.g. Trotta 2008; Hobson et al. 2010, for applications of Bayesian methods to cosmology)

$$p(\Theta|D) = \int dM_0 \, p(\Theta, M_0|D) = \int dM_0 \frac{p(D \mid \Theta, M_0) p(\Theta, M_0)}{p(D)}, \tag{28}$$

where $p(D)$ is the Bayesian evidence (a normalizing constant) and the prior $p(\Theta, M_0)$ can be written as

$$p(\Theta, M_0) = p(\mathscr{C}, \alpha, \beta) p\left(M_0, \sigma_\mu^{\text{int}}\right) = p(\mathscr{C}, \alpha, \beta) p\left(M_0|\sigma_\mu^{\text{int}}\right) p\left(\sigma_\mu^{\text{int}}\right). \tag{29}$$

We take a uniform prior on the variables $\mathscr{C}, \alpha, \beta$ (on a sufficiently large range so as to encompass the support of the likelihood), as well as a Gaussian prior for $p(M_0|\sigma_\mu^{\text{int}})$, since M_0 is a location parameter of a Gaussian (conditional on σ_μ^{int}). Thus, we take

$$p(M_0 \mid \sigma_\mu^{\text{int}}) = \mathcal{N}_{M_0}\left(M_{\text{m}}, \sigma_{M_0}^2\right), \tag{30}$$

where the mean of the prior ($M_{\text{m}} = -19.3\,\text{mag}$) is taken to be a plausible value in the correct ballpark, and the variance ($\sigma_{M_0} = 2.0\,\text{mag}$) is sufficiently large so that the prior is very diffuse and non-informative (the precise choice of mean and variance for this prior does not impact on our numerical results). Finally, the appropriate prior for σ_μ^{int} is a Jeffreys' prior, i.e. uniform in $\log \sigma_\mu^{\text{int}}$, as σ_μ^{int} is a scale parameter (see e.g. Box & Tiao 1992). Although the intrinsic magnitude M_0 and the Hubble constant H_0 are perfectly degenerate as far as SNIa data are concerned, we do not bundle them together in a single parameter but treat them separately with distinct priors, as we are interested in separating out the variability due to the distribution of the SNIa intrinsic magnitude.

We now proceed to manipulate further the likelihood, $p(D|\Theta, M_0) = p(\underline{\hat{c}}, \underline{\hat{x}}_1, \underline{\hat{m}}_B^*|\Theta, M_0)$:

$$p(\underline{\hat{c}}, \underline{\hat{x}}_1, \underline{\hat{m}}_B^*|\Theta, M_0) = \int d\underline{c} \, d\underline{x}_1 \, d\underline{M} \; p(\underline{\hat{c}}, \underline{\hat{x}}_1, \underline{\hat{m}}_B^* \mid \underline{c}, \underline{x}_1, \underline{M}, \Theta, M_0) p(\underline{c}, \underline{x}_1, \underline{M}|\Theta, M_0) \tag{31}$$

$$= \int d\underline{c} \, d\underline{x}_1 \, d\underline{M} \; p(\underline{\hat{c}}, \underline{\hat{x}}_1, \underline{\hat{m}}_B^* \mid \underline{c}, \underline{x}_1, \underline{M}, \Theta)$$
$$\times \int dR_c \, dR_x \, dc_\star \, dx_\star \; p(\underline{c} \mid c_\star, R_c) p(\underline{x}_1|x_\star, R_x) p(\underline{M} \mid M_0, \sigma_\mu^{\text{int}}) p(R_c) p(R_x) p(c_\star) p(x_\star). \tag{32}$$

In the first line, we have introduced a set of $3n$ latent variables, $\{\underline{c}, \underline{x}_1, \underline{M}\}$, which describe the *true* value of the colour, stretch and intrinsic magnitude for each SNIa. Clearly, as those variables are unobserved, we need to marginalize over them. In the second line, we have replaced $p(\underline{c}, \underline{x}_1, \underline{M}|\Theta, M_0)$ by the distributions of the latent $\{\underline{c}, \underline{x}_1, \underline{M}\}$ given by the probabilistic relationships of equation (21) and equations (22) and (23), obtained by marginalizing out the population parameters $\{R_c, R_x, c_\star, x_\star\}$:

$$p(\underline{c}, \underline{x}_1, \underline{M}|\Theta, M_0) = \int dR_c \, dR_x \, dc_\star \, dx_\star \; p(\underline{c} \mid c_\star, R_c) p(\underline{x}_1 \mid x_\star, R_x) p\left(\underline{M}|M_0, \sigma_\mu^{\text{int}}\right) p(R_c) p(R_x) p(c_\star) p(x_\star) \tag{33}$$

(we have also dropped M_0 from the likelihood, as conditioning on M_0 is irrelevant if the latent $\underline{M}$ are given). If we further marginalize over M_0 [as in equation (28), including the prior on M_0], the expression for the effective likelihood, equation (32), then becomes

$$p(\underline{\hat{c}}, \underline{\hat{x}}_1, \underline{\hat{m}}_B^*|\Theta) = \int d\underline{c} \, d\underline{x}_1 \, d\underline{M} \; p(\underline{\hat{c}}, \underline{\hat{x}}_1, \underline{\hat{m}}_B^* \mid \underline{c}, \underline{x}_1, \underline{M}, \Theta)$$
$$\times \int dR_c \, dR_x \, dc_\star \, dx_\star \, dM_0 \; p(\underline{c} \mid c_\star, R_c) p(\underline{x}_1 \mid x_\star, R_x) p\left(\underline{M}|M_0, \sigma_\mu^{\text{int}}\right) p(R_c) p(R_x) p(c_\star) p(x_\star) p(M_0 \mid \sigma_\mu^{\text{int}}). \tag{34}$$

The term $p(\underline{\hat{c}}, \underline{\hat{x}}_1, \underline{\hat{m}}_B^*|\underline{c}, \underline{x}_1, \underline{M}, \Theta)$ is the conditional probability of observing values $\{\underline{\hat{c}}, \underline{\hat{x}}_1, \underline{\hat{m}}_B^*\}$ if the latent (true) values of $\underline{c}, \underline{x}_1, \underline{M}$ and of the other cosmological parameters were known. From Fig. 1, $\underline{m}_B^*$ is connected only deterministically to all other variables and parameters, via equation (27). Thus, we can replace $\underline{m}_B^* = \underline{\mu} + \underline{M} - \alpha \cdot \underline{x}_1 + \beta \cdot \underline{c}$ and write

$$p(\underline{\hat{c}}, \underline{\hat{x}}_1, \underline{\hat{m}}_B^*|\underline{c}, \underline{x}_1, \underline{M}, \Theta) = \prod_{i=1}^{n} \mathcal{N}(\mu_i + M_i - \alpha \cdot x_{1i} + \beta \cdot c_i, \hat{C}_i) \tag{35}$$

$$= |2\pi \, \Sigma_C|^{-1/2} \exp\left\{-\frac{1}{2}\left[(X - X_0)^{\text{T}} \Sigma_C^{-1}(X - X_0)\right]\right\}, \tag{36}$$

where $\mu_i \equiv \mu_i(z_i, \Theta)$, and we have defined

$$X = \{X_1, \ldots, X_n\} \in \mathbb{R}^{3n}, \quad X_0 = \{X_{0,1}, \ldots, X_{0,n}\} \in \mathbb{R}^{3n}, \tag{37}$$

$$X_i = \{c_i, x_{1,i}, (M_i - \alpha x_{1,i} + \beta c_i)\} \in \mathbb{R}^3, \quad X_{0,i} = \{c_i, x_{1,i}, \hat{m}_{Bi}^* - \mu_i\} \in \mathbb{R}^3, \tag{38}$$

as well as the $3n \times 3n$ block covariance matrix[3]

$$
\Sigma_C = \begin{pmatrix} \hat{C}_1 & 0 & 0 & 0 \\ 0 & \hat{C}_2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \hat{C}_n \end{pmatrix}. \tag{39}
$$

Finally, we explicitly include redshift uncertainties in our formalism. The observed apparent magnitude, $\underline{\hat{m}}_B^*$, on the left-hand side of equation (35), is the value at the observed redshift, $\underline{\hat{z}}$. However, μ in equation (35) should be evaluated at the true (unknown) redshift, $\underline{z}$. As above, the redshift uncertainty is included by introducing the latent variables $\underline{z}$ and integrating over them:

$$
p(\underline{c}, \underline{x}_1, \underline{M} \,|\, \underline{c}, \underline{x}_1, \underline{M}, \Theta) = \int d\underline{z} \;\; p(\underline{c}, \underline{x}_1, \underline{M} \,|\, \underline{c}, \underline{x}_1, \underline{M}, \underline{z}, \Theta) p(\underline{z} \,|\, \underline{\hat{z}}), \tag{40}
$$

where we model the redshift errors $p(\underline{z}|\underline{\hat{z}})$ as Gaussians:

$$
\underline{z} \sim \mathcal{N}(\underline{\hat{z}}, \Sigma_z), \tag{41}
$$

with an $n \times n$ covariance matrix:

$$
\Sigma_z = \mathrm{diag}\left(\sigma_{z_1}^2, \dots, \sigma_{z_n}^2\right). \tag{42}
$$

It is now necessary to integrate out all latent variables and nuisance parameters from the expression for the likelihood, equation (34). This can be done analytically, as all necessary integrals are Gaussian. The detailed steps are described in Appendix C. In summary, the procedure consists of the following:

(i) marginalization over the intrinsic redshifts (see equation C2),
(ii) marginalization over the latent variables $\{\underline{M}, \underline{c}, \underline{x}_1\}$ (see equation C24) and
(iii) marginalization over the nuisance parameters $\{x_\star, c_\star\}$ (see equation C28).

This leads to the final expression for the effective likelihood of equation (C32):

$$
\begin{aligned}
p(\underline{\hat{c}}, \underline{\hat{x}}_1, \underline{\hat{m}}_B^* | \Theta) = \int & d\log R_c \; d\log R_x \; |2\pi\Sigma_C|^{-1/2} |2\pi\Sigma_P|^{-1/2} |2\pi\Sigma_A|^{1/2} |2\pi\Sigma_0|^{-1/2} |2\pi K|^{1/2} \\
& \times \exp\left[-\frac{1}{2} \left(X_0^{\mathrm{T}} \Sigma_C^{-1} X_0 - \Delta^{\mathrm{T}} \Sigma_A \Delta - k_0^{\mathrm{T}} K^{-1} k_0 + \underline{b}_m^{\mathrm{T}} \Sigma_0^{-1} \underline{b}_m \right) \right],
\end{aligned}
$$

where the various vectors and matrices appearing in the above expression are defined in Appendix C. This equation is the main result of our paper. The two remaining nuisance parameters R_c, R_x cannot be integrated out analytically, so they need to be marginalized numerically. Hence, they are added to our parameters of interest and are sampled over numerically, and then marginalized out from the joint posterior.

3 NUMERICAL TESTS OF THE METHOD

3.1 Generation of simulated SNe data

We now proceed to test the performance of our method on simulated data sets, and to compare it with the usual χ^2 approach. We take as starting point the recent compilation of 288 SNeIa from Kessler et al. (2009a), the data set on which we apply our BHM in the next section. Kessler et al. (2009a) refitted the light curves from five different surveys:

(i) Sloan Digital Sky Survey (SDSS): 103 SNe (Kessler et al. 2009a),
(ii) ESSENCE: 56 SNe (Miknaitis et al. 2007; Wood-Vasey et al. 2007),
(iii) SNLS: 62 SNe (Astier et al. 2006),
(iv) nearby sample: 33 SNe (Jha et al. 2007) and
(v) *Hubble Space Telescope* (*HST*): 34 SNe (Garnavich, Kirshner & Challis 1998; Knop et al. 2003; Riess & Strolger 2004, 2007),

using both the SALT-II method and the MLCS fitter. In the following, we are exclusively employing the results of their SALT-II fits and use those as the observed data set for the purposes of our current work, as described in the previous section. More refined procedures could be adopted, for example, by simulating light curves from scratch, using, e.g., the publicly available package SNANA (Kessler et al. 2009b). In this paper, we chose a simpler approach, consisting of simulating SALT-II fit results in such a way to broadly match the distributions and characteristics of the real data set used in Kessler et al. (2009a).

---

[3] Note that we neglect correlations between different SNeIa, which is reflected in the fact that Σ_C takes a block-diagonal form. It would be, however, very easy to add arbitrary cross-correlations to our formalism (e.g. coming from correlated systematic within survey, for example zero-point calibration) by adding such non-block-diagonal correlations to equation (39).

Table 1. Input parameter values used in the generation of the simulated SNe SALT-II fits.

Parameter	Symbol	True value
Matter energy density parameter	Ω_m	0.3
Dark energy density parameter	Ω_Λ	0.7
Dark energy EOS	w	-1
Spatial curvature	Ω_κ	0.0
Hubble expansion rate	H_0 (km s^{-1} Mpc^{-1})	72.0
Mean absolute magnitude of SNe	M_0 (mag)	-19.3
Intrinsic dispersion of SNe magnitude	σ_μ^{int} (mag)	0.1
Stretch correction	α	0.13
Colour correction	β	2.56
Mean of distribution of $\underline{x}_1$	$x_\star$	0.0
Mean of distribution of $\underline{c}$	$c_\star$	0.0
s.d. of distribution of $\underline{x}_1$	R_x	1.0
s.d. of distribution of $\underline{c}$	R_c	0.1
Observational noise on m_B^*	$\sigma_{m_{Bi}^*}$	Depending on survey
Observational noise on $\underline{x}_1$	$\sigma_{x_1 i}$	Depending on survey
Observational noise on $\underline{c}$	σ_{ci}	Depending on survey
Correlation between $\underline{x}_1$ and $\underline{c}$	$\sigma_{x_1 i, ci}$	0.0

The numerical values of the parameters used in the simulation are shown in Table 1. We adopt a vanilla, flat ΛCDM cosmological model as fiducial cosmology. The Phillips correction parameters are chosen to match the best-fitting values reported in Kessler et al. (2009a), while the distributional properties of the colour and stretch correction match the observed distribution of their total SN sample. For each survey, we generate a number of SNe matching the observed sample, and we model their redshift distribution as a Gaussian, with mean and variance estimated from the observed distribution within each survey. The observational error of $m_B^*, \underline{c}, \underline{x}_1$ is again drawn from a Gaussian distribution whose mean and variance have been matched to the observed ones for each survey. Finally, the pseudo-data (i.e. the simulated SALT-II fits results) are generated by drawing from the appropriate distributions centred around the latent variables. For simplicity, we have set to 0 the off-diagonal elements in the correlation matrix (39) in our simulated data, and neglected redshift errors. None of these assumptions has a significant impact on our results. In summary, our procedure for each survey is as follows.

(i) Draw a value for the latent redshift z_i from a normal distribution with mean and variance matching the observed ones. As we neglect redshift errors in the pseudo-data for simplicity (since the uncertainty in z is subdominant in the overall error budget), we set $\hat{z}_i = z_i$.

(ii) Compute μ_i using the fiducial values for the cosmological parameters $\mathscr{C}$ and the above z_i from equation (8).

(iii) Draw the latent parameters x_{1i}, c_i, M_i from their respective distributions (in particular, including an intrinsic scatter $\sigma_\mu^{int} = 0.1$ mag in the generation of M_i).

(iv) Compute m_{Bi}^* using x_{1i}, c_i, M_i and the Phillips relation equation (5).

(v) Draw the value of the standard deviations $\sigma_{x_1 i}, \sigma_{c_i}, \sigma_{m_i}$, from the appropriate normal distributions for each survey type. A small, z_i-dependent stochastic linear addition is also made to $\sigma_{x_1 i}, \sigma_{c_i}, \sigma_{m_i}$, to mimic the observed correlation between redshift and error.

(vi) Draw the SALT-II fit results from $\hat{x}_{1i} \sim \mathcal{N}(x_{1i}, \sigma_{x_1 i})$, $\hat{c}_i \sim \mathcal{N}(c_i, \sigma_{c_i})$ and $\hat{m}_{Bi}^* \sim \mathcal{N}(m_{Bi}^*, \sigma_{m_i})$.

As shown in Fig. 3, the simulated data from our procedure have broadly similar distributions to the real ones. The two notable exceptions are the overall vertical shift observed in the distance modulus plot, and the fact that our synthetic data cannot reproduce the few outliers with large values of the variances (bottom panels). The former is a consequence of the different intrinsic magnitude used in our simulated data (as the true one is unknown). However, the intrinsic magnitude is marginalized over at the end, so this difference has no impact on our inferences. The absence of outliers is a consequence of the fact that our simulation is a pure phenomenological description of the data, hence it cannot encapsulate such fine details. While in principle we could perform outlier detection with dedicated Bayesian procedures, we do not pursue this issue further in this paper. We stress once more that the purpose of our simulations is not to obtain realistic SNIa data. Instead, they should only provide us with useful mock data sets coming from a known model so that we can test our procedure. More sophisticated tests based on more realistically generated data (e.g. from SNANA) are left for future work.

3.2 Numerical sampling

After analytical marginalization of the latent variables, we are left with the following eight parameters entering the effective likelihood of equation (43):

$$\{\Omega_m, \Omega_\kappa \text{ or } w, H_0, \sigma_\mu^{int}, \alpha, \beta, R_c, R_x\}. \tag{43}$$

As mentioned above, in keeping with the literature we only consider either flat Universes with a possible $w \neq -1$ (the ΛCDM model) or curved Universes with a cosmological constant ($w = -1$, the wCDM model). Of course, it is possible to relax those assumptions and consider
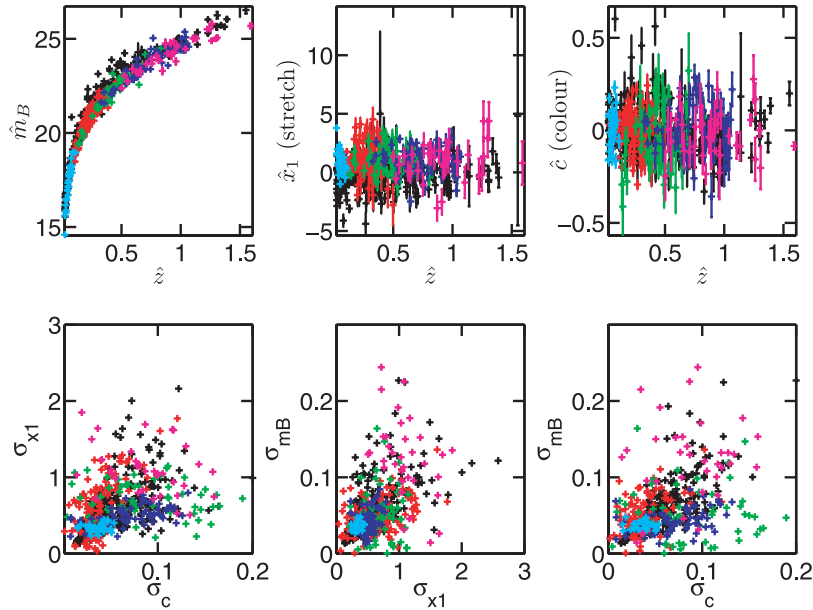
Figure 3. An example realization of our simulated data sets (coloured according to survey), superimposed on real data (black). Colour code for simulated data survey: nearby sample (cyan), ESSENCE (green), SDSS (red), SNLS (blue) and *HST* (magenta).

Table 2. Priors on our model's parameters used when evaluating the posterior distribution. Ranges for the uniform priors have been chosen so as to generously bracket plausible values of the corresponding quantities.

Parameter	ΛCDM	wCDM
Ω_m	Uniform: $\mathcal{U}(0.0, 1.0)$	Uniform: $\mathcal{U}(0.0, 1.0)$
Ω_κ	Uniform: $\mathcal{U}(-1.0, 1.0)$	Fixed: 0
w	Fixed: -1	Uniform: $\mathcal{U}(-4, 0)$
H_0 (km s^{-1} Mpc^{-1})	$\mathcal{N}(72, 8^2)$	$\mathcal{N}(72, 8^2)$
	Common priors	
σ_μ^{int} (mag)	Uniform on $\log \sigma_\mu^{\text{int}}$: $\mathcal{U}(-3.0, 0.0)$	
M_0 (mag)	Uniform: $\mathcal{U}(-20.3, -18.3)$	
α	Uniform: $\mathcal{U}(0.0, 1.0)$	
β	Uniform: $\mathcal{U}(0.0, 4.0)$	
R_c	Uniform on $\log R_c$: $\mathcal{U}(-5.0, 2.0)$	
R_x	Uniform on $\log R_c$: $\mathcal{U}(-5.0, 2.0)$	

more complicated cosmologies with a larger number of free parameters if one so wishes (notably including evolution in the dark energy EOS).

Of the parameters listed in equation (43), the quantities R_c, R_x are of no interest and will be marginalized over. As for the remaining parameters, we are interested in building their marginal one- and two-dimensional posterior distributions. This is done by plugging the likelihood (43) into the posterior of equation (28), with priors on the parameters chosen according to Table 2. We use a Gaussian prior on the Hubble parameter $H_0 = 72 \pm 8$ km s^{-1} Mpc^{-1} from local determinations of the Hubble constant (Freedman et al. 2001). However, as H_0 is degenerate with the intrinsic population absolute magnitude M_0 (which is marginalized over at the end), replacing this Gaussian prior with a less informative prior H_0 (km s^{-1} Mpc^{-1}) $\sim \mathcal{U}(20, 100)$ has no influence on our results.

Numerical sampling of the posterior is carried out via a nested sampling algorithm (Skilling 2004, 2006; Feroz & Hobson 2008; Feroz, Hobson & Bridges 2009). Although the original motivation for nested sampling was to compute the Bayesian evidence, the recent development of the MultiNest algorithm (Feroz & Hobson 2008; Feroz et al. 2009) has delivered an extremely powerful and versatile algorithm that has been demonstrated to be able to deal with extremely complex likelihood surfaces in hundreds of dimensions exhibiting multiple peaks. As samples from the posterior are generated as a by-product of the evidence computation, nested sampling can also be used to obtain parameter constraints in the same run as computing the Bayesian evidence. In this paper, we adopt the publicly available MultiNest algorithm (Feroz & Hobson 2008) to obtain samples from the posterior distribution of equation (28). We use 4000 live points and a tolerance parameter 0.1, resulting in about 8×10^5 likelihood evaluations.[4]

---

[4] A FORTRAN code implementing our method is available from the authors upon request.

© 2011 The Authors, MNRAS **418**, 2308–2329
Monthly Notices of the Royal Astronomical Society © 2011 RAS

We also wish to compare the performance of our BHM with the usually adopted χ^2 minimization procedure. To this end, we fit the pseudo-data using the χ^2 expression of equation (13). In order to mimic what is done in the literature as closely as possible, we first fix a value of σ_μ^{int}. Then, we simultaneously minimize the χ^2 with respect to the fit parameters $\vartheta = \{\Omega_m, \Omega_\kappa \text{ or } w, H_0, M_0, \alpha, \beta\}$, as described below. We then evaluate the χ^2/d.o.f. from the resulting best-fitting point, and we adjust σ_μ^{int} to obtain χ^2/d.o.f. = 1. We then repeat the above minimization over ϑ for this new value of σ_μ^{int}, and iterate the procedure. Once we have obtained the global best-fitting point, we derive one- and two-dimensional confidence intervals on the parameters by profiling (i.e. maximizing over the other parameters) over the likelihood

$$L(\vartheta) = \exp\left(-\frac{1}{2}\chi(\vartheta)^2\right),\tag{44}$$

with χ^2 given by equation (13). According to Wilks' theorem, approximate confidence intervals are obtained from the profile likelihood as the regions where the χ^2 increases by $\Delta\chi^2$ from its minimum value, where $\Delta\chi^2$ can be computed from the χ^2 distribution with the number of d.o.f. corresponding to the number of parameters of interest and is given in standard look-up tables.

Obtaining reliable estimates of the profile likelihood using Bayesian algorithms (such as MultiNest) is a considerably harder numerical task than mapping out the Bayesian posterior. However, it has been shown that MultiNest can be successfully used for this task even in highly challenging situations (Feroz et al. 2011b), provided the number of live points and tolerance value used are adjusted appropriately. For our χ^2 scan, we adopt 10^4 live points and a tolerance of 0.1. We have found that those values give accurate estimates of the profile likelihood more than 2σ into the tails of the distribution for an eight-dimensional Gaussian toy model (whose dimensionality matches the case of interest here). With these MultiNest settings, we gather 1.5×10^5 samples, from which the profile likelihood is derived.

Our implementation of the χ^2 method is designed to match the main features of the fitting procedure usually adopted in the literature (namely, maximization of the likelihood rather than marginalization of the posterior, and iterative determination of the intrinsic dispersion), although we do not expect that it exactly reproduces the results obtained by any specific implementation. Its main purpose is to offer a useful benchmark against which to compare the performance of our new Bayesian methodology.

3.3 Parameter reconstruction

We compare the cosmological parameters reconstructed from the standard χ^2 method and our Bayesian approach in Fig. 4 for a typical data realization. The left-hand panel shows constraints in the Ω_m–Ω_Λ plane for the ΛCDM model, both from our Bayesian method (filled regions, marginalized 68 and 95 per cent posterior) and from the standard χ^2 method (red contours, 68 and 95 per cent confidence regions from the profile likelihood). In the right-hand panel, constraints are shown in the w–Ω_m plane instead (this fit assumes a flat Universe). In a typical reconstruction, our Bayesian method produced considerably tighter constraints on the cosmological parameters of interest than the usual χ^2 approach. Our constraints are also less biased with respect to the true value of the parameters, an important advantage that we further characterize below.

Our BHM further produces marginalized posterior distributions for all the other parameters of the fit, including the stretch and colour corrections and the intrinsic dispersion of the SNe. The one-dimensional marginal posteriors for those quantities are shown in Fig. 5. The recovered posterior means lie within 1σ of the true values. Note that we do not expect the posterior mean to match exactly the true value, because of realization noise in the pseudo-data. However, as shown below, our method delivers less biased estimates of the parameters, and a reduced mean squared error (MSE) compared with the standard χ^2 approach. The stretch correction α is determined with 8 per cent accuracy, while the colour correction parameter β is constrained with an accuracy better than 3 per cent. A new feature of our method is that it produces
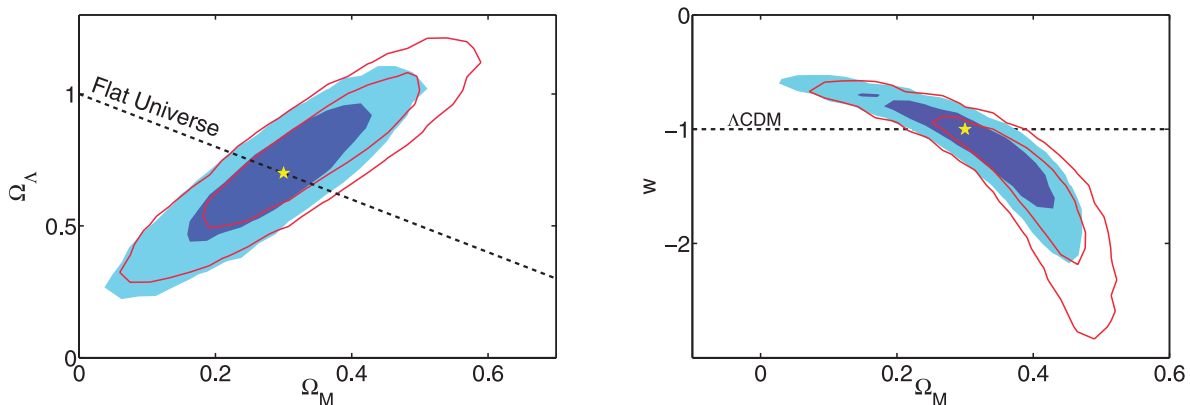

Figure 4. Reconstruction of cosmological parameters from a simulated data set encompassing 288 SNeIa, with characteristics matching presently available surveys (including realization noise). Blue regions contain 95 and 68 per cent of the posterior probability (other parameters marginalized over) from our BHM method, the red contours delimit 95 and 68 per cent confidence intervals from the standard χ^2 method (other parameters maximized). The yellow star indicates the true value of the parameters. The left-hand panel assumes $w = -1$, while the right-hand panel assumes $\Omega_\kappa = 0$. Note how out method produced considerably less biased constraints on the parameters.
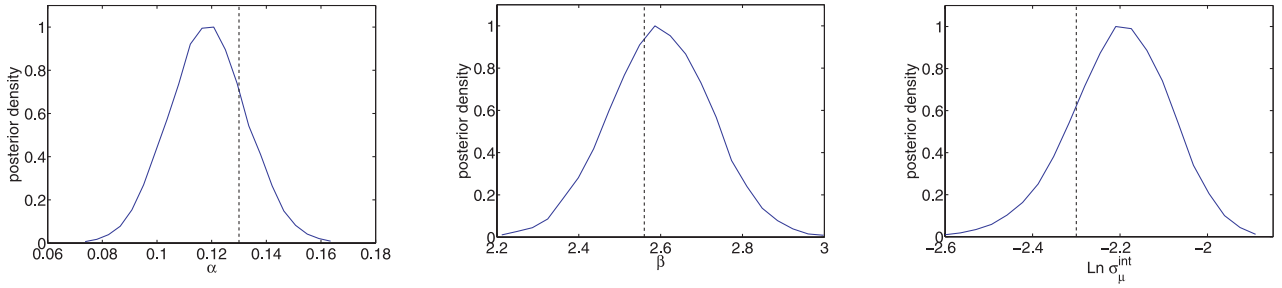
Figure 5. Marginalized posterior for the stretch correction α, colour correction β parameter and logarithm of the intrinsic dispersion of SNe, $\log \sigma_\mu^{\rm int}$, from a simulated data set from our Bayesian method. The vertical, dashed line gives the true value for each quantity.

a posterior distribution for the SN population intrinsic dispersion, $\sigma_\mu^{\rm int}$ (right-hand panel of Fig 5). This allows one to determined the intrinsic dispersion of the SNIa population to typically about 10 per cent accuracy.

3.4 Comparison of long-term performance of the two methods

Encouraged by the good performance of our BHM on single instances of simulated data, we now wish to compare the long-term performance of the two methods for a series of simulated data realizations. We are interested in comparing the average ability of both methods to recover parameter values that are as much as possible unbiased with respect to their true values, as well as to establish the coverage properties of the credible and confidence intervals.

Coverage is defined as the probability that an interval contains (covers) the true value of a parameter, in a long series of repeated measurements. The defining property of a, e.g., 95 per cent frequentist confidence interval is that it should cover the true value 95 per cent of the time; thus, it is reasonable to check if the intervals have the properties they claim. Coverage is a frequentist concept: intervals based on Bayesian techniques are meant to contain a given amount of posterior probability for a *single measurement* (with no reference to repeated measurements) and are referred to as credible intervals to emphasize the difference in concept. While Bayesian techniques are not designed with coverage as a goal, it is still meaningful to investigate their coverage properties. To our knowledge, the coverage properties of even the standard χ^2 method (which, being a frequentist method, would ideally be expected to exhibit exact coverage) have never been investigated in the SN literature.

We generate 100 realizations of the pseudo-data from the fiducial model of Table 1 as described in Section 3.1, and we analyse them using our BHM method and the standard χ^2 approach, using the same priors as above, given in Table 2. For each parameter of interest θ, we begin by considering the relative size of the posterior 68.3 per cent range from out method, $\sigma_\theta^{\rm BHM}$, compared with the 68.3 per cent confidence interval from the χ^2 method, $\sigma_\theta^{\chi^2}$, which is summarized by the quantity $\mathcal{S}_\theta$, which shows the percentage change in error bar size with respect to the error bar derived using the χ^2 method:

$$\mathcal{S}_\theta \equiv \left(\frac{\sigma_\theta^{\rm BHM}}{\sigma_\theta^{\chi^2}} - 1 \right) \times 100. \tag{45}$$

A value $\mathcal{S}_\theta < 1$ means that our method delivers tighter error bars on the parameter θ. A histogram of this quantity for the variables of interest is shown in Fig. 6, from which we conclude that our method gives smaller error bars on $\Omega_{\rm m}$, Ω_Λ and w in almost all cases. The flip side of
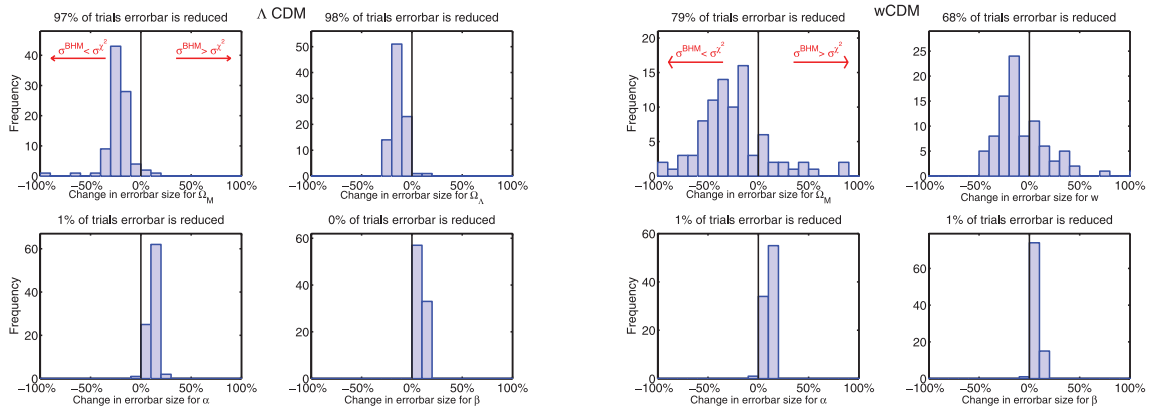

Figure 6. Histograms of the quantity defined in equation (45), comparing of the error bars on each parameter from our method and from the standard χ^2 approach for 100 realization, for the ΛCDM model (left) and the wCDM model (right). A change in error bar size of -10 per cent indicates BHM error bars are 10 per cent smaller than χ^2 error bars. A change in error bar size of $+10$ per cent indicates BHM error bars are 10 per cent larger than χ^2 error bars. Our BHM method generally delivers smaller errors on the cosmological parameters (top row) but larger errors on the Phillips correction parameters (bottom row).
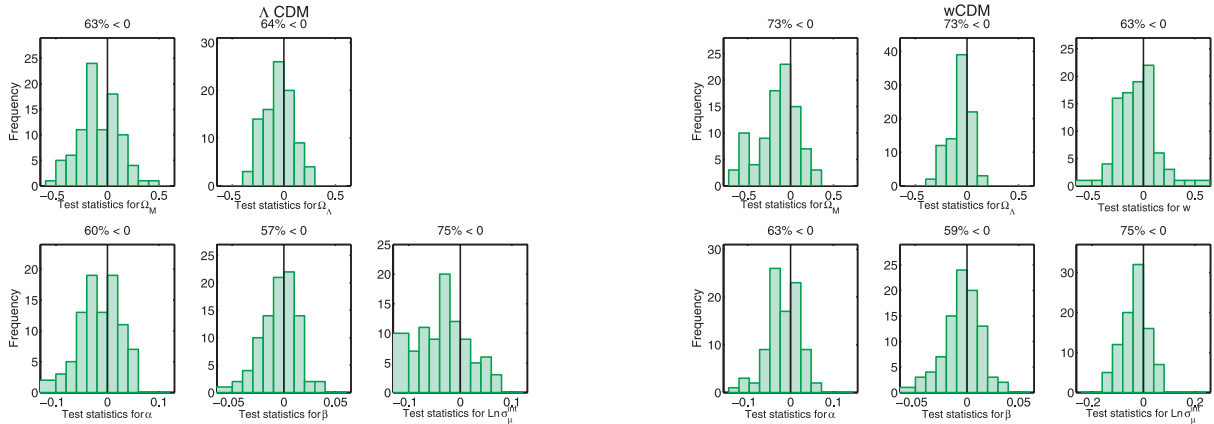
Figure 7. Histograms of the test statistics defined in equation (46), comparing the long-term performance of the two methods for the parameters of interest in the ΛCDM model (left) and the wCDM model (right). A predominantly negative value of the test statistics means that our method gives a parameter reconstruction that is closer to the true value than the usual χ^2, i.e. less biased. For the cosmological parameters (top row), our method outperforms χ^2 about two times out of three.

this result is that the uncertainty on α, β is larger from our method than from the χ^2 approach in essentially all data realization. This is a consequence of the large number of latent variables our method marginalizes over.

Tight error bars are good, but not if they come at the expense of a biased reconstruction. To investigate this aspect, we build the following test statistics from each reconstruction:

$$\mathcal{T}_\theta \equiv |\overline{\theta}_{\text{BHM}}/\theta_{\text{true}} - 1| - |\theta_{\chi^2}^{\text{bf}}/\theta_{\text{true}} - 1|, \qquad (46)$$

where $\overline{\theta}_{\text{BHM}}$ is the posterior mean recovered using our BHM method, $\theta_{\chi^2}^{\text{bf}}$ is the best-fitting value for the parameter recovered using the standard χ^2 approach and θ_{true} is the true value for that parameter. The meaning of $\mathcal{T}_\theta$ is the following: for a given data realization, if the reconstructed posterior mean from our BHM is closer to the true parameter value than the best-fitting χ^2, then $\mathcal{T}_\theta < 0$, which means that our method is less biased than χ^2. A histogram of the distribution of $\mathcal{T}_\theta$ across the 100 realizations, shown in Fig. 7, can be used to compare the two methods: a negative average in the histogram means that the BHM outperforms the usual χ^2. For all of the parameters considered, our method is clearly much less biased than the χ^2 method, outperforming χ^2 about two-thirds of the time. Furthermore, the reconstruction of the intrinsic dispersion is better with our Bayesian method almost three times out of four. We emphasize once more that our methodology also provides an estimate of the uncertainty in the intrinsic dispersion, not just a best-fitting value as the χ^2 approach.

We can further quantify the improvement in the statistical reconstruction by looking at the bias and MSE for each parameter, defined as

$$\text{bias} = \langle \hat{\theta} - \theta_{\text{true}} \rangle \qquad (47)$$

$$\text{MSE} = \text{bias}^2 + \text{Var}, \qquad (48)$$

respectively, where the expectation is taken by averaging over the observed values in our 100 simulated trials, $\hat{\theta} = \overline{\theta}_{\text{BHM}}$ ($\hat{\theta} = \theta_{\chi^2}^{\text{bf}}$) for the BHM (for the χ^2 approach) and Var is the observed parameter variance. The bias is the expectation value of the difference between estimator and true value, while the MSE measures the average of the squares of the errors, i.e. the amount by which the estimator differs from the true value for each parameter. Obviously, a smaller bias and a smaller MSE imply a better performance of the method. The results for the two methods are summarized in Table 3, which shows how our method reduces the bias by a factor of ~2–3 for most parameters, while reducing the MSE by a factor of ~2. The only notable exception is the bias of the EOS parameter w, which is larger in our method than in the χ^2 approach.

Finally, in Fig. 8 we plot the coverage of each method for 68 and 95 per cent intervals. Error bars give an estimate of the uncertainty of the coverage result, by giving the binomial sampling error from the finite number of realizations considered, evaluated from the binomial variance as $Np(1 - p)$, where $N = 100$ is the number of trials and p is the observed fractional coverage. Both method slightly undercover, i.e. the credible region and confidence intervals are too short, although the lack of coverage is not dramatic: e.g., the typical coverage of the 1σ (2σ) intervals from our method is ~60 per cent (90 per cent). Our method shows slightly better coverage properties than the χ^2 method, while producing considerably tighter and less biased constraints (as demonstrated above). This further proves that the tighter intervals recovered by our method do not suffer from bias with respect to the true values.

4 COSMOLOGICAL CONSTRAINTS FROM CURRENT SNIa DATA

We now apply our BHM to fitting real SN data. We use the SALT-II fits result for 288 SNeIa from Kessler et al. (2009a), which have been derived from five different surveys. Our method only includes statistical errors according to the procedure described in Section 2.3, coming from redshift uncertainties (arising from spectroscopic errors and peculiar velocities), intrinsic dispersion (which is determined from the data) and full error propagation of the SALT-II fit results. Systematic uncertainties play an important role in SNIa cosmology fitting, and (although

Table 3. Comparison of the bias and MSE for our Bayesian method and the usual χ^2 approach. The columns labelled 'Improvement' give the factor by which our Bayesian method reduces the bias and the MSE with respect to the χ^2 approach.

	Parameter	Bias			MSE		
		Bayesian	χ^2	Improvement	Bayesian	χ^2	Improvement
ΛCDM	Ω_m	-0.0188	-0.0183	1.0	0.0082	0.0147	1.8
	Ω_Λ	-0.0328	-0.0223	0.7	0.0307	0.0458	1.5
	α	0.0012	0.0032	2.6	0.0001	0.0002	1.4
	β	0.0202	0.0482	2.4	0.0118	0.0163	1.4
	σ_μ^{int}	-0.0515	-0.1636	3.1	0.0261	0.0678	2.6
wCDM	Ω_m	-0.0177	-0.0494	2.8	0.0072	0.0207	2.9
	Ω_Λ	0.0177	0.0494	2.8	0.0072	0.0207	2.9
	w	-0.0852	-0.0111	0.1	0.0884	0.1420	1.6
	α	0.0013	0.0032	2.5	0.0001	0.0002	1.5
	β	0.0198	0.0464	2.3	0.0118	0.0161	1.4
	σ_μ^{int}	-0.0514	-0.1632	3.2	0.0262	0.0676	2.6
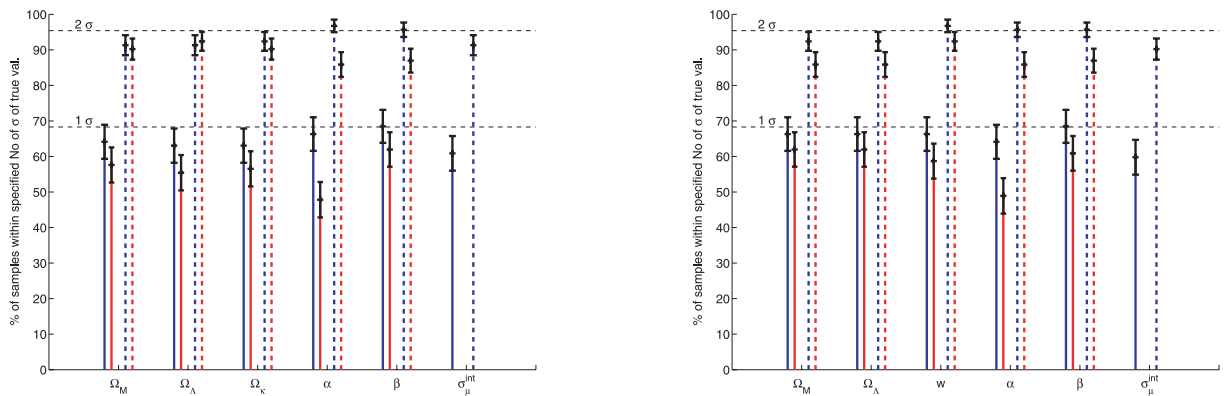
Figure 8. Coverage of our method (blue) and standard χ^2 (red) for 68 per cent (solid) and 95 per cent (dashed) intervals, from 100 realizations of pseudo-data for the ΛCDM model (left) and the wCDM model (right). While both methods show significant undercoverage for all parameters, our method has a comparable coverage to the standard χ^2, except for w. Coverage values for the intrinsic dispersion σ_μ^{int} are not available from the χ^2 method, as it does not produce an error estimate for this quantity.
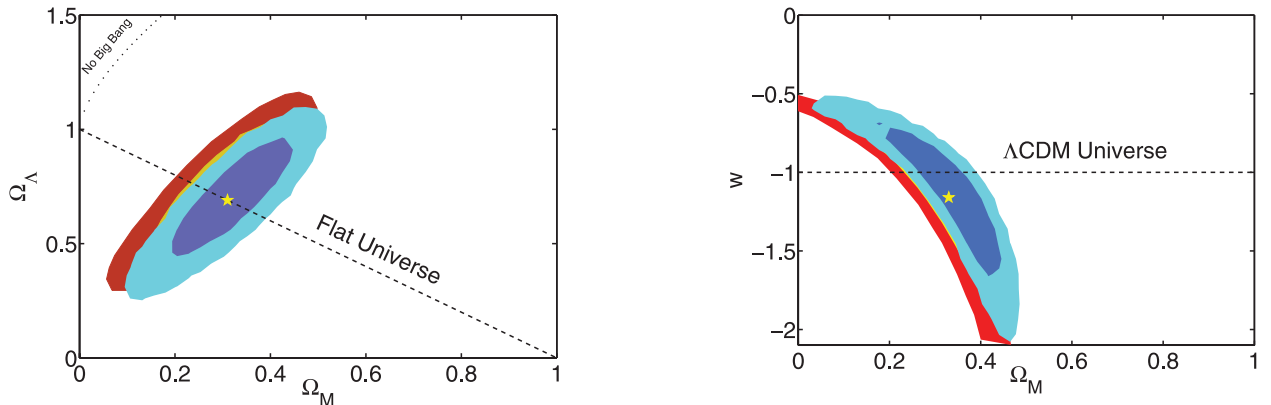

Figure 9. Constraints on the cosmological parameters Ω_m, Ω_Λ (left-hand panel, assuming $w = -1$) and w, Ω_m (right-hand panel, assuming $\Omega_\kappa = 0$) from our Bayesian method (light/dark blue regions, 68 and 95 per cent marginalized posterior), compared with the statistical errors from the usual χ^2 approach (yellow/red regions, same significance level; from Kessler at al. 2009a). The yellow star gives the posterior mean from our analysis.

not included in this study) can also be treated in our formalism in a fully consistent way. We comment on this aspect further below, though we leave a complete exploration of systematics with our BHM to a future, dedicated work (M. C. March, R. Trotta, P. Berkes, G. D. Starkman, P. M. Vaudrevange, in preparation).

We show in Fig. 9 the constraints on the cosmological parameters Ω_m–Ω_Λ (left-hand panel, assuming $w = -1$) and w–Ω_m (right-hand panel, assuming $\Omega_\kappa = 0$) obtained with our method. All other parameters have been marginalized over. In order to be consistent with the literature, we have taken a non-informative prior on H_0, uniform in the range $[20, 100]\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$. The figure also compares our results

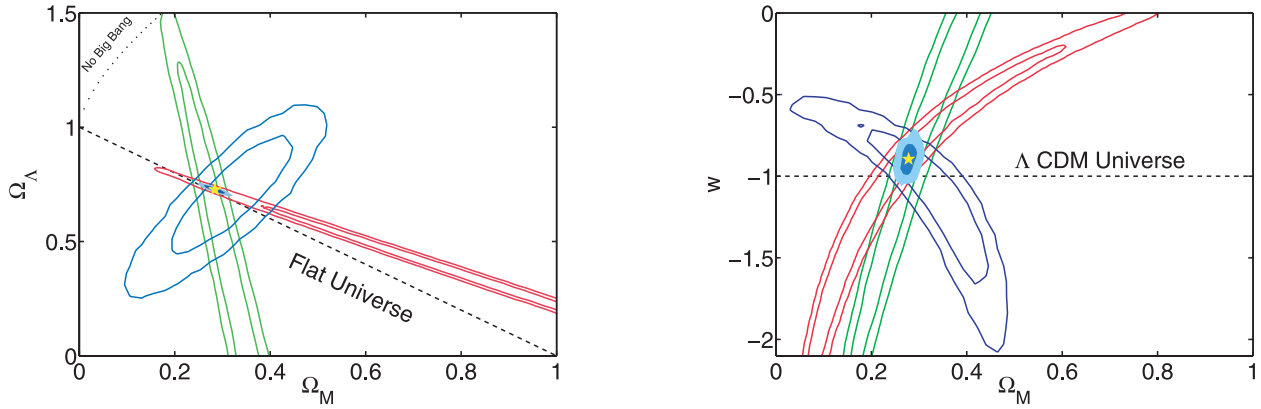**Figure 10.** Combined constraints on the cosmological parameters Ω_m, Ω_Λ (left-hand panel, assuming $w = -1$) and w, Ω_m (right-hand panel, assuming $\Omega_\kappa = 0$) from SNIa, CMB and BAO data. Red contours give 68 and 95 per cent regions from CMB alone, green contours from BAO alone, blue contours from SNIa alone from our Bayesian method. The filled regions delimit 68 and 95 per cent combined constraints, with the yellow star indicating the posterior mean.
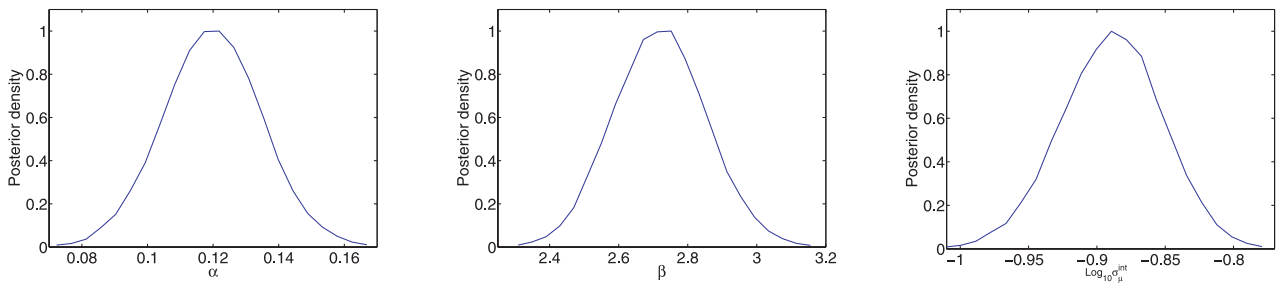

Figure 11. Marginalized posterior for the stretch correction α, colour correction β parameter and logarithm of the intrinsic dispersion of SNe, $\log \sigma_\mu^{\text{int}}$, from current SNIa data.

with the statistical contours from Kessler et al. (2009a), obtained using the χ^2 method. (Note that we compare with the contours including only statistical uncertainties for consistency.) In Fig. 10, we combine our SNIa constraints with cosmic microwave background (CMB) data from 5-year *Wilkinson Microwave Anisotropy Probe* measurements (Komatsu et al. 2009) and baryonic acoustic oscillations (BAO) constraints from the SDSS luminous red galaxy (LRG) sample (Eisenstein et al. 2005), using the same method as Kessler et al. (2009a). The combined SNIa, CMB and BAO statistical constraints result in $\Omega_m = 0.28 \pm 0.02$, $\Omega_\Lambda = 0.73 \pm 0.01$ (for the ΛCDM model) and $\Omega_m = 0.28 \pm 0.01$, $w = -0.90 \pm 0.05$ (68.3 per cent credible intervals) for the wCDM model. Although the statistical uncertainties are comparable to the results by Kessler et al. (2009a) from the same sample, our posterior mean values present shifts of up to $\sim 2\sigma$ compared to the results obtained using the standard χ^2 approach. This is a fairly significant shift, which can be attributed to our improved statistical method, which exhibits a reduced bias with respect to the χ^2 approach.

Fig. 11 shows the one-dimensional marginalized posterior distributions for the Phillips correction parameters and for the intrinsic dispersion. All parameters are well constrained by the posterior, and we find $\alpha = 0.12 \pm 0.02$, $\beta = 2.7 \pm 0.1$ and a value of the intrinsic dispersion (for the whole sample) $\sigma_\mu^{\text{int}} = 0.13 \pm 0.01$ mag.

Kessler et al. (2009a) find values for the intrinsic dispersion ranging from 0.08 (for SDSS-II) to 0.23 (for the *HST* sample), but their χ^2 method does not allow them to derive an error on those determinations. With our method, it would be easy to derive constraints on the intrinsic dispersion of each survey – all one needs to do is to replace equation (19) with a corresponding expression for each survey. This introduces one pair of population parameters (M_0, σ_μ^{int}) for each survey. In the same way, one could study whether the intrinsic dispersion evolves with redshift. We leave a detailed study of these issues to a future work.

The value of α found in Kessler et al. (2009a) is in the range 0.10–0.12, depending on the details of the assumptions made, with a typical statistical uncertainty of the order of ~ 0.015. These results are comparable with our own. As for the colour correction parameter β, constraints from Kessler et al. (2009a) vary in the range 2.46–2.66, with a statistical uncertainty of the order of 0.1–0.2. This stronger dependence on the details of the analysis seems to point to a larger impact of systematic uncertainties for β, which is confirmed by evidence of evolution with redshift of the value of β (Kessler et al. 2009a, fig. 39). Our method can be employed to carry out a rigorous assessment of the evolution with redshift of colour corrections. A possible strategy would be to replace β with a vector of parameters β_1, β_2,..., with each element describing the colour correction in a different redshift bin. The analysis proceeds then as above, and it produces posterior distributions for the components of β, which allows one to check the hypothesis of evolution. Finally, in such an analysis the marginalized constraints on all other parameters (including the cosmological parameters of interest) would automatically include the full uncertainty propagation from the colour correction evolution, without the need for further ad hoc inflation of the error bars. These kind of tests will be pursued in a forthcoming publication (March et al., in preparation).

5 CONCLUSIONS

We have presented a statistically principled approach for the rigorous analysis of SALT-II SNIa light-curve fits, based on a BHM. The main novelty of our method is that it produces an effective likelihood that propagates uncertainties in a fully consistent way. We have introduced an explicit statistical modelling of the intrinsic magnitude distribution of the SNIa population, which for the first time allows one to derive a full posterior distribution of the SNIa intrinsic dispersion.

We have tested our method using simulated data sets and found that it compares favourably with the standard χ^2 approach, both on individual data realizations and in the long-term performance. Statistical constraints on cosmological parameters are significantly improved, while in a series of 100 simulated data sets our method outperforms the χ^2 approach at least two times out of three for the parameters of interest. We have also demonstrated that our method is less biased and has better coverage properties than the usual approach.

We applied our methodology to a sample of 288 SNeIa from multiple surveys. We find that the flat ΛCDM model is still in good agreement with the data, even under our improved analysis. However, the posterior mean for the cosmological parameters exhibits up to 2σ shifts with respect to results obtained with the conventional χ^2 approach. This is a consequence of our improved statistical analysis, which benefits from a reduced bias in estimating the parameters.

While in this paper we have only discussed statistical constraints, our method offers a new, fully consistent way of including systematic uncertainties in the fit. As our method is fully Bayesian, it can be used in conjunction with fast and efficient Bayesian sampling algorithms, such as MCMC and nested sampling. This will allow one to enlarge the number of parameters controlling systematic effects that can be included in the analysis, thus taking SNIa cosmological parameter fitting to a new level of statistical sophistication. The power of our method as applied to systematic errors analysis will be presented in a forthcoming, dedicated work.

At a time when SNIa constraints are entering a new level of precision, and with a tenfold increase in the sample size expected over the next few years, we believe it is timely to upgrade the cosmological data analysis pipeline in order to extract the most information from present and upcoming SNIa data. This paper represents a first step in this direction.

REFERENCES

Amanullah R. et al., 2010, ApJ, 716, 712
Andreon S., 2006, MNRAS, 369, 969
Andreon S., Hurn M. A., 2010, MNRAS, 404, 1922
Astier P. et al., 2006, A&A, 447, 31
Bailey S., Aragon C., Romano R., Thomas R. C., Weaver B. A., Wong D., 2008, preprint (arXiv:0810.3499)
Balland C. et al., 2009, A&A, 507, 85
Blondin S., Mandel K. S., Kirshner R. P., 2011, A&A, 526, 81
Box G. E. P., Tiao G. C., 1992, Bayesian Inference in Statistical Analysis. John Wiley & Sons, Chicester, UK
Conley A., 2006, ApJ, 644, 1
Conley A. et al., 2011, ApJS, 192, 1
Contreras C. et al., 2010, ApJ, 139, 519
D'Agostini G., 2005, preprint (arXiv:physics/0511182v1)
Eisenstein D. J. et al., 2005, ApJ, 633, 560
Feroz F., Hobson M. P., 2008, MNRAS, 384, 449
Feroz F., Hobson M. P., Bridges M., 2009, MNRAS, 398, 1601
Feroz F., Cranmer K., Hobson M., de Austri R. R., Trotta R., 2011a, preprint (arXiv:1101.3296)
Feroz F., Cranmer K., Hobson M., Ruiz de Austri R., Trotta R., 2011b, preprint (arXiv:1101.3296)
Freedman W. L. et al., 2001, ApJ, 553, 47
Freedman W. L. et al., 2009, ApJ, 704, 1036
Garnavich P. M., Kirshner R. P., Challis P., 1998, ApJ, 493, L53
Gonzalez-Gaitan S. et al., 2011, ApJ, 727, 107
Gopal Vishwakarma R., Narlikar J. V., 2011, Res. Astron. Astrophys., 10, 1195
Gull S., 1989, in Skilling J., ed., Maximum Entropy and Bayesian Methods. Kluwer Academic, Dordrecht, p. 511
Guy J. et al., 2007, A&A, 466, 11
Hicken M. et al., 2009, ApJ, 700, 331
Hobson M., Jaffe A., Liddle A., Mukherjee P., Parkinson D. R., eds, 2010, Bayesian Methods in Cosmology. Cambridge Univ. Press, Cambridge

Hogg D. W., Bovy J., Lang D., 2010, preprint (arXiv:1008.4686)
Jha S., Riess A. G., Kirshner R. P., 2007, ApJ, 659, 122
Kelly B. C., 2007, ApJ, 665, 1489
Kelly P. L., Hicken M., Burke D. L., Mandel K. S., Kirshner R. P., 2010, ApJ, 715, 743
Kessler R. et al., 2009a, ApJS, 185, 32
Kessler R. et al., 2009b, preprint (arXiv:0908.4280)
Kim A. G., 2011, PASP, 123, 230
Knop R. A., Aldering G., Amanullah R., Astier P., 2003, ApJ, 598, 102
Komatsu E. et al., 2009, ApJS, 180, 330
Kowalski M. et al., 2008, ApJ, 686, 749
Mandel K. S., Wood-Vasey W., Friedman A. S., Kirshner R. P., 2009, ApJ, 704, 629
Mandel K. S., Narayan G., Kirshner R. P., 2010, preprint (arXiv:1011.5910)
Miknaitis G. et al., 2007, ApJ, 666, 674
Nordin J., Goobar A., Jonsson J., 2008, J. Cosmol. Astropart. Phys., 2008, 02, 008
Perlmutter S. et al., 1999, ApJ, 517, 565
Phillips M., 1993, ApJ, 413, L105
Phillips M. M., Lira P., Suntzeff N. B., Schommer R. A., Hamuy M., Maza J., 1999, AJ, 118, 1766
Riess A. G., Strolger L., 2004, ApJ, 607, 665
Riess A. G., Strolger L., 2007, ApJ, 659, 98
Riess A. G. et al., 1998, AJ, 116, 1009
Skilling J., 2004, in Fischer R., Preuss R., Toussaint U. V., eds, AIP Conf. Ser. Vol. 735, Nested Sampling. Am. Inst. Phys., New York, p. 395
Skilling J., 2006, Bayesian Analysis, 1, 833
Sullivan M. et al., 2006, ApJ, 648, 868
Sullivan M. et al., 2011, preprint (arXiv:1104.1444)
Trotta R., 2008, Contemporary Phys., 49, 71
Wang L., Goldhaber G., Aldering G., Perlmutter S., 2003, ApJ, 590, 944
Wang L., Strovink M., Conley A., Goldhaber G., Kowalski M., Pelmutter S., Siegrist J., 2006, ApJ, 641, 50
Wood-Vasey W. M., Miknaitis G., Stubbs C. W., 2007, ApJ, 666, 694

APPENDIX A: NOTATION

For reference, we collect here a few useful formulae relating to Gaussian integrals. Use the notation $\underline{x} \sim \mathcal{N}_{\underline{x}}(\underline{\mu}, \Sigma)$ to denote that the random variable $\underline{x}$ is drawn from a normal distribution of mean $\underline{\mu}$ and inverse covariance matrix Σ, given by

$$p(\underline{x}) = \frac{1}{|2\pi\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(\underline{x} - \underline{\mu})^{\mathrm{T}}\Sigma^{-1}(\underline{x} - \underline{\mu})\right]. \tag{A1}$$

In performing Gaussian integrals, it is also convenient to recall that

$$\mathcal{N}_{\underline{x}}(\underline{\mu}_1, \Sigma_1) \cdot \mathcal{N}_{\underline{x}}(\underline{\mu}_2, \Sigma_2) = f_0 \cdot \mathcal{N}_{\underline{x}}(\underline{\mu}_f, \Sigma_f) \tag{A2}$$

where

$$f_0 = \mathcal{N}_{\underline{\mu}_1}(\underline{\mu}_2, (\Sigma_1 + \Sigma_2)) \tag{A3}$$

$$\underline{\mu}_f = (\Sigma_1^{-1} + \Sigma_2^{-1})^{-1}\left(\Sigma_1^{-1}\underline{\mu}_1 + \Sigma_2^{-1}\underline{\mu}_2\right) \tag{A4}$$

$$\Sigma_f = \left(\Sigma_1^{-1} + \Sigma_2^{-1}\right)^{-1}. \tag{A5}$$

Finally, the integral of the multidimensional Gaussian of equation (A1) over all space is unity.

APPENDIX B: TOY MODEL FOR GENERAL LINEAR FITTING

We can get an intuitive understanding of the central ingredient in our Bayesian hierarchical method by considering a simpler toy model, which highlights the salient features of the problem.

As mentioned in Section 2.2, several authors have observed in the past that the spread of fit values for the colour correction, $\hat{c}$, is of the same order as the size of the statistical uncertainty on the values themselves. This leads to a bias in the best-fitting value of β obtained from minimizing the χ^2 in equation (13). The quantity β gives the slope of the linear relationship between $\underline{c}$ and $\underline{\mu}$ (see equation 27). In the usual χ^2 approach, the latent (true) $\underline{c}$ is replaced by the observed value, $\hat{c}$, as in equation (14). If the statistical uncertainty in the independent variable, $\hat{c}$, is as large as the spread of its values, the best-fitting slope obtained from minimizing the χ^2 may be biased, as the large uncertainty in the actual location of the independent variable leads to confusion as to the value of the slope. The (Bayesian) solution to this problem is to determine the spread of the independent variable directly from the data, and to marginalize over it with an appropriate prior. This gives the most general solution to the problem of linear fitting with errors in both the independent and dependent variable, as shown by Gull (1989). Recent literature in cosmology and astronomy (D'Agostini 2005; Hogg, Bovy & Lang 2010) addresses linear fitting, but not this general case, which has been treated before in Kelly (2007) (see also Andreon 2006; Andreon & Hurn 2010, for related examples). In this short appendix, we will analyse the toy model of fitting a linear function, and compare the performance of a BHM and the χ^2 approach.
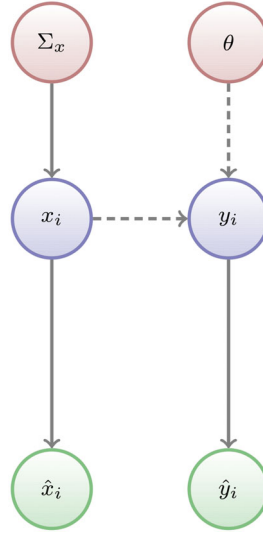
Figure B1. Bayesian network showing the deterministic and probabilistic connections in the toy linear model. Solid lines indicate probabilistic connections, dashed lines represent deterministic connections. Parameters to be constrained are in red, latent variables in blue and data in green (denoted by hats). θ denotes the parameters a, b, i.e. the intercept and slope of the linear relation of equation (B1).

B1 Bayesian linear fitting in the presence of large x and y uncertainties

In this subsection, we give a short review of the results of Gull (1989). The simplest toy model which illustrates the methodology we adopt in our paper is shown by the graphical network of Fig. B1. A linear relationship is assumed between the latent variables x_i and y_i, described by a slope a and intercept b (which are collectively denoted as a parameter vector θ):

$$y_i = ax_i + b. \tag{B1}$$

The observed values for the dependent and independent variables are denoted by hats ($\hat{x}_i$, $\hat{y}_i$, $i = 1, \ldots, N$), and they are obtained from the latent values under the assumption of Gaussian noise (with known variances σ_x^2, σ_y^2):

$$\hat{x}_i \sim \mathcal{N}(x_i, \sigma_x^2) \quad \text{and} \quad \hat{y}_i \sim \mathcal{N}\left(y_i, \sigma_y^2\right). \tag{B2}$$

This probabilistic relationship is depicted in Fig. B1 by the solid arrows connecting the latent variables to the observed quantities. Assuming that errors are uncorrelated, the joint likelihood is given by[5]

$$p(\hat{x}, \hat{y}|x, y, \sigma_x, \sigma_y, \theta) = \left(4\pi^2\sigma_x^2\sigma_y^2\right)^{-N/2} \exp\left\{-\frac{1}{2}\left[\frac{\sum_i(\hat{x}_i - x_i)^2}{\sigma_x^2} + \frac{\sum_i(\hat{y}_i - y_i)^2}{\sigma_y^2}\right]\right\}. \tag{B3}$$

The problem can be made more symmetric by defining rescaled versions of the data:

$$\hat{X} = \frac{\hat{x} - x_0}{R_x} \quad \text{and} \quad \hat{Y} = \frac{\hat{y} - y_0}{R_y}, \tag{B4}$$

where the variables x_0, y_0 describe the mean value of $\hat{x}$, $\hat{y}$, while R_x, R_y describe their spread. These new variables are related to the old ones (a, b) by

$$b = y_0 - ax_0 \quad \text{and} \quad a = R_y/R_x. \tag{B5}$$

Neglecting the normalization constant, the joint posterior for x, y, x_0, y_0, R_x, R_y can be written as

$$p(x, y, x_0, y_0, R_x, R_y \,|\, \hat{x}, \hat{y}, \sigma_x, \sigma_y) \propto p(\hat{x}, \hat{y} \,|\, x, x_0, y_0, R_x, R_y) p(x|x_0, y_0, R_x, R_y) p(x_0, y_0, R_x, R_y), \tag{B6}$$

where the first term on the right-hand side is the likelihood of equation (B3). The key step is to recognize that the appropriate conditional distribution for the latent x is

$$p(x|x_0, y_0, R_x, R_y) = \left(4\pi^2 R_x^2\right)^{-1/2} \exp\left[-\frac{1}{2}\frac{\sum_i(x_i - x_0)^2}{R_x^2}\right]. \tag{B7}$$

This describes a prior over x centred around the hyperparameter x_0 and with standard deviation R_x. Crucially, both x_0 and R_x are unknown, and are explicitly determined from the data in the joint posterior, before being marginalized out at the end. Finally, for the prior $p(x_0, y_0, R_x, R_y)$ appearing in equation (B6) we adopt a uniform prior on x_0, y_0 (as those are location variables) and a prior uniform in $\log R_x$, $\log R_y$ [those being scale variables, as apparent from (B7)].

---

[5] For ease of notation, quantities without subscript i denote in the following N-dimensional vectors, e.g. $x = \{x_1, \ldots, x_N\}$.

From here, one can eliminate y from equation (B6) using equation (B1), then trade b for R_y using equation (B5), to obtain

$$p(x, x_0, y_0, a, \log R_x, \log R_y | \hat{x}, \hat{y}, \sigma_x, \sigma_y) \propto \left(8\pi^3 \sigma_x^2 \sigma_y^2 R_x^2\right)^{-N/2} \exp\left\{-\frac{1}{2}\left[\frac{\sum_i (\hat{x}_i - x_i)^2}{\sigma_x^2} + \frac{\sum_i (\hat{y}_i - ax_i - y_0 + ax_0)^2}{\sigma_y^2} + \frac{\sum_i (x_i - x_0)^2}{R_x^2}\right]\right\}.$$

(B8)

From this expression, the latent x can be marginalized out analytically, as well as the nuisance parameters x_0, y_0, by using appropriate completions of the square in the Gaussian. After some algebra, one finally obtains

$$p(a, \log R | \hat{x}, \hat{y}, \sigma_x, \sigma_y) \propto \left(a^2 \sigma_x^2 R_x^2 + \sigma_x^2 \sigma_y^2 + \sigma_y^2 R_x^2\right)^{-(N-1)/2} \exp\left[-\frac{1}{2} \frac{V_{xx}\left(a^2 R_x^2 + \sigma_x^2\right) - 2V_{xy} a R_x^2 + V_{yy}\left(R_x^2 + \sigma_x^2\right)}{a^2 \sigma_x^2 R_x^2 + \sigma_x^2 \sigma_y^2 + \sigma_y^2 R_x^2}\right],$$

(B9)

where

$$V_{xx}^2 = \sum_i (\hat{x}_i - \bar{x})^2, \quad V_{xy}^2 = \sum_i (\hat{x}_i - \bar{x})(\hat{y}_i - \bar{y}), \quad V_{yy}^2 = \sum_i (\hat{y}_i - \bar{y})^2, \quad R = (R_x R_y)^{1/2}.$$

(B10)

In the above, $\bar{x} = \sum_i \hat{x}_i / N$, and similarly for $\bar{y}$. The marginal posterior for the slope a is obtained by numerical marginalization of $\log R$ from the above expression.

B2 Comparison with the χ^2 approach

If, instead of the statistically principled solution found above, one writes down a simple χ^2 expression for the likelihood, including error propagation from the linear relationship (B1), one would obtain (D'Agostini 2005)

$$L(a, b) \propto \exp\left[-\frac{1}{2} \sum_i \frac{(\hat{y}_i - a\hat{x}_i - b)^2}{\sigma_y^2 + a^2 \sigma_x^2}\right],$$

(B11)

from where the intercept b is eliminated by maximizing over it (profiling). We now compare the reconstruction of the slope parameter a from equation (B11) (with b eliminated via profiling) with the result obtained using the Bayesian expression, (B9), marginalized numerically over $\log R$ (with a uniform prior on $\log R$) with the help of simulated data.

Figs B2 and B3 show in the upper left-hand panel the simulated data points ($N = 300$), with the error bars giving the size of the standard deviation in the x and y directions for each datum (i.e. the value of σ_x, σ_y). The contour plots depict joint posterior regions for $(\log a, \log R)$ from the Bayesian expression of equation (B9), and confidence regions from the χ^2 likelihood (B11) in the a, b plane. The lower right-hand panel shows the one-dimensional marginalized posterior distribution for $\log a$ (blue) and the profile likelihood from the χ^2 approach (black). Fig. B2 shows that the two results are essentially identical when the size of the statistical error is smaller than the spread of values of the data points (in this particular example, by a factor of 2). However, when the statistical uncertainty in the x direction is as large as or larger than the spread of the points (Fig. B3), the χ^2 approach gives a biased result for the slope parameter a. The Bayesian posterior, on the contrary, is closer to the true value, although it (correctly) shows a larger uncertainty.
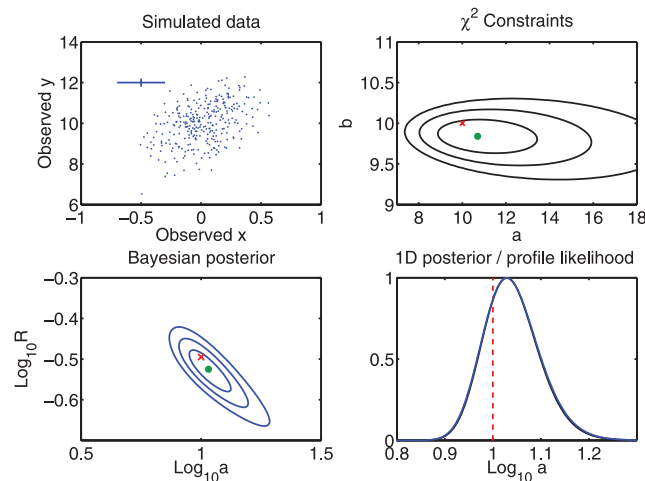

Figure B2. Numerical comparison between the χ^2 approach and the Bayesian method for linear fitting. Upper left-hand panel: data set of $N = 300$ observations with errors both in the x and y directions, given by the error bar. Upper right-hand panel: reconstruction of the slope a and intercept b using a χ^2 likelihood (red cross is the true value, green circle the maximum likelihood value). Lower left-hand panel: Bayesian posterior (equation B9) in the $\log a$, $\log R$ plane, with green circle showing posterior mean. In both panels, contours enclose 1σ, 2σ and 3σ regions. Lower right-hand panel: marginalized Bayesian posterior (blue) and profiled χ^2 likelihood (black, lying exactly on top of the blue curve), with the dashed line showing the true value. The two methods give essentially identical results in this case.
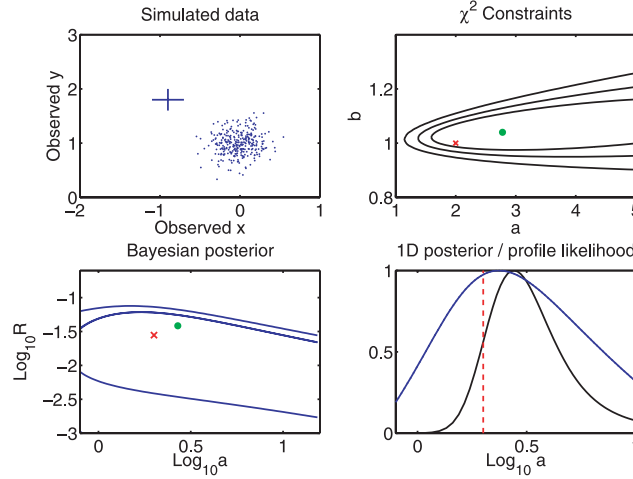
Figure B3. As in Fig. B2, but now with a larger statistical uncertainty in the data compared to their spread. The Bayesian marginal posterior for the slope (blue, lower right-hand panel) peaks much closer to the true value than the χ^2 expression (black). In the lower left-hand panel, the 3σ contour from the Bayesian method lies outside the range of the plot.

APPENDIX C: DERIVATION OF THE EFFECTIVE LIKELIHOOD

C1 Integration over intrinsic redshifts

In order to perform the multidimensional integral over $\underline{z}$, we Taylor expand $\underline{\mu}$ around $\hat{\underline{z}}$ (as justified by the fact that redshift errors are typically small: the error from $300\,\mathrm{km\,s^{-1}}$ peculiar velocity is $\sigma_{z_i} = 0.0012$, while the error from spectroscopic redshifts from SNe themselves is $\sigma_{z_i} = 0.005$; see Kessler et al. 2009a) :

$$\mu_j = \mu(z_j) = 5\log_{10}\left(\frac{D_\mathrm{L}(z_j)}{\mathrm{Mpc}}\right) + 25 \approx \mu(\hat{z}_j) + 5(\log_{10} e)\left.\frac{\partial_{z_j} D_\mathrm{L}(z_j)}{D_\mathrm{L}(z_j)}\right|_{\hat{z}_j}(z_j - \hat{z}_j). \tag{C1}$$

With this approximation, we can now carry out the multidimensional integral of equation (40), obtaining

$$p(\hat{\underline{m}}_B^* | \underline{c}, \underline{x}_1, \underline{M}, \Theta) = |2\pi\Sigma_m|^{-1/2}$$
$$\times \exp\left\{-\frac{1}{2}[\hat{\underline{m}}_B^* - (\underline{\mu} + \underline{M} - \alpha \cdot \underline{x}_1 + \beta \cdot \underline{c})]^\mathrm{T}\Sigma_m^{-1}[\hat{\underline{m}}_B^* - (\underline{\mu} + \underline{M} - \alpha \cdot \underline{x}_1 + \beta \cdot \underline{c})]\right\}, \tag{C2}$$

where, from now on, $\underline{\mu} = \underline{\mu}(\hat{\underline{z}})$ and

$$\Sigma_m = \Sigma_C + f\Sigma_z f^\mathrm{T}, \tag{C3}$$

$$f = \mathrm{diag}(f_1, \dots, f_n), \tag{C4}$$

$$f_i = 5\log_{10}(e)\left.\frac{D_\mathrm{L}'(z_i)}{D_\mathrm{L}(z_i)}\right|_{\hat{z}_i} \tag{C5}$$

$$= \frac{5\log_{10}(e)}{D_\mathrm{L}(\hat{z}_i)}\left[\frac{D_\mathrm{L}(\hat{z}_i)}{1 + z_i} + \frac{c}{H_0}(1 + \hat{z}_i) \times \mathrm{cosn}\left\{\sqrt{|\Omega_\kappa|}\int_0^{\hat{z}} \mathrm{d}z'[(1 + z')^3\Omega_\mathrm{m} + \Omega_\mathrm{de}(z) + (1 + z)^2\Omega_\kappa]^{-1/2}\right\} \tag{C6}$$

$$\times \left[(1 + z')^3\Omega_\mathrm{m} + \Omega_\mathrm{de}(z) + (1 + z)^2\Omega_\kappa\right]^{-1/2}\Bigg]. \tag{C7}$$

Strictly speaking, one should integrate over redshift in the range $0 \le z_i < \infty$, not $-\infty < z_i < \infty$, which would result in the appearance of Gamma functions in the final result. However, as long as $\frac{\sigma_{z_i}}{z_i} \ll 1$ (as is the case here), this approximation is expected to be excellent.

C2 Integration over latent $\{\underline{c}, \underline{x}_1, \underline{M}\}$

From equation (32) and using the expression in equation (35), we wish to integrate out the latent variables

$$Y = \{Y_1, \dots, Y_n\} \in \mathbb{R}^{3n}, \tag{C8}$$

$$Y_i = \{c_i, x_{1,i}, M_i\} \in \mathbb{R}^3. \tag{C9}$$

$$\tag{C10}$$

We therefore recast expression (35) as

$$p(\hat{\underline{c}}, \hat{\underline{x}}_1, \hat{\underline{m}}_B^* | \underline{c}, \underline{x}_1, \underline{M}, \Theta) = |\Sigma_C|^{-1/2} \exp\left(-\frac{1}{2}\left[(AY - X_0)^{\mathrm{T}} \Sigma_C^{-1}(AY - X_0)\right]\right), \tag{C11}$$

where we have defined the block-diagonal matrix

$$A = \mathrm{diag}(T, T, \ldots, T) \in \mathbb{R}^{3n \times 3n}, \tag{C12}$$

with

$$T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \beta & -\alpha & 1 \end{bmatrix} \begin{bmatrix} c_i \\ x_i \\ M_i \end{bmatrix}. \tag{C13}$$

The prior terms appearing in equation (34), namely $p(\underline{c}|c_\star, R_c)p(\underline{x}_1|x_\star, R_x)p(\underline{M}|M_0, \sigma_\mu^{\mathrm{int}})$, may be written as

$$p(\underline{c}|c_\star, R_c)p(\underline{x}_1|x_\star, R_x)p\left(\underline{M}|M_0, \sigma_\mu^{\mathrm{int}}\right) = |2\pi\Sigma_P|^{-1/2} \exp\left(-\frac{1}{2}\left[(Y - Y_*)^{\mathrm{T}} \Sigma_P^{-1}(Y - Y_*)\right]\right), \tag{C14}$$

where

$$S^{-1} = \mathrm{diag}\left(R_c^{-2}, R_x^{-2}, \left(\sigma_\mu^{\mathrm{int}}\right)^{-2}\right) \in \mathbb{R}^{3\times 3}, \quad \Sigma_P^{-1} = \mathrm{diag}\left(S^{-1}, S^{-1}, \ldots, S^{-1}\right) \in \mathbb{R}^{3n \times 3n}, \tag{C15}$$

$$Y_* = \underline{J} \cdot \underline{b} \in \mathbb{R}^{3n \times 1}, \tag{C16}$$

$$\underline{J} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \in \mathbb{R}^{3n \times 3}, \tag{C17}$$

$$\underline{b} = \begin{bmatrix} c_* \\ x_* \\ M_0 \end{bmatrix} \in \mathbb{R}^{3\times 1}. \tag{C18}$$

Now the integral over $\mathrm{d}Y = \mathrm{d}\underline{c}\,\mathrm{d}\underline{x}_1\,\mathrm{d}\underline{M}$ in equation (34) can be performed, giving

$$\int \mathrm{d}Y\; p(\hat{\underline{c}}, \hat{\underline{x}}_1, \hat{\underline{m}}_B^* | \underline{c}, \underline{x}_1, \underline{M}, \Theta)p(\underline{c}\,|\,c_\star, R_c)p(\underline{x}_1\,|\,x_\star, R_x)p\left(\underline{M}\,|\,M_0, \sigma_\mu^{\mathrm{int}}\right)$$

$$= |2\pi\Sigma_C|^{-1/2}|2\pi\Sigma_P|^{-1/2}|2\pi\Sigma_A|^{1/2} \exp\left(-\frac{1}{2}\left[X_0^{\mathrm{T}}\Sigma_C^{-1}X_0 - Y_0^{\mathrm{T}}\Sigma_A^{-1}Y_0 + Y_*^{\mathrm{T}}\Sigma_P^{-1}Y_*\right]\right), \tag{C19}$$

where

$$\Sigma_A^{-1} = A^{\mathrm{T}}\Sigma_C^{-1}A + \Sigma_P^{-1} \in \mathbb{R}^{3n \times 3n}, \tag{C20}$$

$$\Sigma_A^{-1}Y_0 = A^{\mathrm{T}}\Sigma_C^{-1}X_0 + \Sigma_P^{-1}Y_*, \tag{C21}$$

$$Y_0 = \Sigma_A\left(A^{\mathrm{T}}\Sigma_C^{-1}X_0 + \Sigma_P^{-1}Y_*\right)\Sigma_A\left(\Delta + \Sigma_P^{-1}Y_*\right), \tag{C22}$$

$$\Delta = A^{\mathrm{T}}\Sigma_C^{-1}X_0 \in \mathbb{R}^{3n \times 1}. \tag{C23}$$

Substituting equation (C19) back into equation (34) gives

$$p(\hat{\underline{c}}, \hat{\underline{x}}_1, \hat{\underline{m}}_B^* | \Theta) = \int \mathrm{d}R_c\,\mathrm{d}R_x\,\mathrm{d}c_\star\,\mathrm{d}x_\star\, |2\pi\Sigma_C|^{-1/2}|2\pi\Sigma_P|^{-1/2}|2\pi\Sigma_A|^{1/2}$$

$$\times \exp\left[-\frac{1}{2}\left(X_0^{\mathrm{T}}\Sigma_C^{-1}X_0 - Y_0^{\mathrm{T}}\Sigma_A^{-1}Y_0 + Y_*^{\mathrm{T}}\Sigma_P^{-1}Y_*\right)\right]$$

$$\times p(R_c)p(R_x)p(c_\star)p(x_\star)p\left(M_0\,|\,\sigma_\mu^{\mathrm{int}}\right). \tag{C24}$$

C3 Integration over population variables $\{c_\star, x_\star, M_0\}$

The priors on the population variables $\underline{b} = \{c_\star, x_\star, M_0\}$ in equation (C24) can be written as

$$p(\underline{b}) = p(c_\star)p(x_\star)p(M_0|\sigma_\mu^{\text{int}}) = |2\pi\Sigma_0|^{-1/2}\exp\left(-\frac{1}{2}(\underline{b}-\underline{b}_m)^{\text{T}}\Sigma_0^{-1}(\underline{b}-\underline{b}_m)\right), \tag{C25}$$

where

$$\Sigma_0^{-1} = \begin{bmatrix} 1/\sigma_{c_*}^2 & 0 & 0 \\ 0 & 1/\sigma_{x_*}^2 & 0 \\ 0 & 0 & 1/\sigma_{M_0}^2 \end{bmatrix} \tag{C26}$$

and

$$\underline{b}_m = \begin{bmatrix} 0 \\ 0 \\ M_{\text{m}} \end{bmatrix} \in \mathbb{R}^{3\times 1}. \tag{C27}$$

Thus, equation (C24) can be written as

$$p(\hat{\underline{c}}, \hat{\underline{x}}_1, \hat{\underline{m}}_B^*|\Theta) = \int dR_c\,dR_x\,d\underline{b}\,|2\pi\Sigma_C|^{-1/2}|2\pi\Sigma_P|^{-1/2}|2\pi\Sigma_A|^{1/2}|2\pi\Sigma_0|^{-1/2}p(R_c)p(R_x)$$

$$\times\exp\left(-\frac{1}{2}\left\{X_0^{\text{T}}\Sigma_C^{-1}X_0 - \left[\Sigma_A\left(\Delta+\Sigma_P^{-1}\underline{J}\cdot\underline{b}\right)\right]^{\text{T}}\Sigma_A^{-1}\left[\Sigma_A\left(\Delta+\Sigma_P^{-1}\underline{J}\cdot\underline{b}\right)\right]\right.\right.$$

$$\left.\left.+\underline{b}^{\text{T}}\underline{J}^{\text{T}}\Sigma_P^{-1}\underline{J}b + (\underline{b}-\underline{b}_m)^{\text{T}}\Sigma_0^{-1}(\underline{b}-\underline{b}_m)\right\}\right)$$

$$= \int dR_c\,dR_x\,|2\pi\Sigma_C|^{-1/2}|2\pi\Sigma_P|^{-1/2}|2\pi\Sigma_A|^{-1/2}|2\pi\Sigma_0|^{-1/2}p(R_c)p(R_x)$$

$$\times\exp\left[-\frac{1}{2}\left(X_0^{\text{T}}\Sigma_C^{-1}X_0 - \Delta^{\text{T}}\Sigma_A\Delta - k_0^{\text{T}}K^{-1}k_0 + \underline{b}_m^{\text{T}}\Sigma_0^{-1}\underline{b}_m\right)\right]\int d\underline{b}\,\exp\left\{-\frac{1}{2}\left[(\underline{b}-k_0)^{\text{T}}K^{-1}(\underline{b}-k_0)\right]\right\}, \tag{C28}$$

where

$$K^{-1} = -\underline{J}^{\text{T}}\Sigma_P^{-1}\Sigma_A\Sigma_P^{-1}\underline{J} + \underline{J}^{\text{T}}\Sigma_P^{-1}\underline{J} + \Sigma_0^{-1} \in \mathbb{R}^{3\times 3}, \tag{C29}$$

$$K^{-1}k_0 = \underline{J}^{\text{T}}\Sigma_P^{-1}\Sigma_A\Delta + \Sigma_0^{-1}\underline{b}_m \in \mathbb{R}^{3\times 1}, \tag{C30}$$

$$k_0 = K\left(\underline{J}^{\text{T}}\Sigma_P^{-1}\Sigma_A\Delta + \Sigma_0^{-1}\underline{b}_m\right). \tag{C31}$$

We can now carry out the Gaussian integral over $\underline{b}$ in equation (C24), obtaining our final expression for the effective likelihood:

$$p(\hat{\underline{c}}, \hat{\underline{x}}_1, \hat{\underline{m}}_B^*|\Theta) = \int d\log R_c\,d\log R_x\,|2\pi\Sigma_C|^{-1/2}|2\pi\Sigma_P|^{-1/2}|2\pi\Sigma_A|^{1/2}|2\pi\Sigma_0|^{-1/2}|2\pi K|^{1/2}$$

$$\times\exp\left[-\frac{1}{2}\left(X_0^{\text{T}}\Sigma_C^{-1}X_0 - \Delta^{\text{T}}\Sigma_A\Delta - k_0^{\text{T}}K^{-1}k_0 + \underline{b}_m^{\text{T}}\Sigma_0^{-1}\underline{b}_m\right)\right], \tag{C32}$$

where we have chosen an improper Jeffreys' prior for the scale variables R_c, R_x:

$$p(R_c) \propto R_c^{-1} \Rightarrow p(R_c)dR_c \propto d\log R_c, \tag{C33}$$

and analogously for R_x. These two remaining nuisance parameters cannot be integrated out analytically, so they need to be marginalized numerically. Hence, R_c, R_x are added to our parameters of interest and are sampled over numerically, and then marginalized out from the joint posterior.

This paper has been typeset from a TeX/LaTeX file prepared by the author.