

Improving Standardisable Candles for Precision Cosmology

Sam M. Ward

Supervisor: Dr K. Mandel

June 29, 2021

Abstract

Type Ia Supernovae (SNe Ia) are stellar explosions used as standardisable candles to measure luminosity distances in the late-time Universe ($z \lesssim 1$), and thereby place constraints on cosmological parameters. SN Ia absolute magnitudes are conventionally standardised by correcting for correlations of optical luminosity with light curve shape and apparent colour. Recent work in the field is focused on reducing standardised absolute magnitude dispersion even further, by identifying more precise standardisation corrections ($\lesssim 0.1\text{mag}$), for improved distance estimates, and hence constraints on e.g. the Hubble constant, H_0 , and the dark energy equation-of-state parameter, w . However, many of these analyses are plagued by the confounding of SN host galaxy dust effects with variations in intrinsic SN colour, which can bias results.

This work leverages *BayeSN*, a next-generation hierarchical Bayesian model of Type Ia SN optical-near-infrared (NIR) spectral energy distributions (SEDs), that is shown to determine SN distance constraints with unprecedented accuracy and precision. An independent empirical validation of light curve features and correlations learned by this state-of-the-art model is warranted, and performed in § 2 using Gaussian processes (GPs) fitted to the Foundation DR1 sample - a nearby ($0.015 \lesssim z \lesssim 0.08$) set of SN Ia optical light curves. I construct an algorithm to build a reliable SN sub-sample based on the quality of the fit. Results show that the independent GP fits - which do not allow for sharing of information among individual SNe in a population - biases feature estimates of the quickly varying z -band photometry. Factoring in this limitation, *BayeSN* and GP fits are highly consistent.

We investigate correlations of SN Ia luminosity with velocity of the SN explosion ejecta (as estimated by blue-shift measurements of the silicon-II absorption line; § 3). We find no evidence for the ‘redder-faster’ relation in $(B - V)$ optical colours reported in the literature, which is likely a product of confounding intrinsic colour with dust. Instead, there is evidence for a ‘redder-faster’ relation in the intrinsic $(B - H)$ optical-NIR colours, which indicates velocity information could be used to improve distance estimates. In future (§ 5.4), we will compute a distribution of p -values using samples from the MCMC chains, examine host-galaxy correlations with velocity, and determine the improvement in distance accuracy on incorporating velocity measurements into *BayeSN*. Beyond this, we will improve the *BayeSN* SED model by training on optical and NIR spectroscopic time series, enabling investigations of spectral correlations with intrinsic absolute magnitude and colours across λ (§ 5.4.2).

A new SNe Ia triplet in NGC 3147, comprising SN 2021hpr (21hpr, discovered 08/04/21), SN 2008fv, and SN 1997bq, provides another opportunity to test the consistency of SN Ia distance estimates (§ 5.1). I am leading the analysis of the proprietary Pan-STARRS 21hpr data, within the Young Supernova Experiment (YSE). We fit *BayeSN* to each of the SN datasets, and compare distance modulus estimates. Given these SNe are all at the same distance, this provides for a within-model consistency check. We will perform a between-model comparison across *BayeSN*, *SNoPy* and *SALT2/3*. Following this, a joint

model for a single shared distance using all the datasets will be implemented in *BayeSN*, and the shifts in distance modulus, extinction, and light curve shape parameter will be evaluated. Contrasting the triplet diversity in extinction/light curve shape with a population distribution will be informative in evaluating the strength of host-galaxy correlations. With a forthcoming Cepheid distance modulus, and in conjunction with a Hubble flow SN set, this work will yield the first *BayeSN* measurement of H_0 .

As part of the new DEHVILS survey (§ 5.2), we will analyse the first year of optical-NIR data of ~ 125 nearby SNe Ia ($0.01 \lesssim z \lesssim 0.08$) with *BayeSN*. We will investigate SED correlations with host galaxy properties, taking advantage of dust-insensitive YJH bands to improve the use of SNe Ia as standard candles. As members of YSE, we will analyse the premier low- z SNe Ia sample for anchoring high-redshift samples, investigate the systematic uncertainties due to dust, and contribute to YSE cosmological analyses to constrain w (§ 5.3).

Finally, I investigate the hypothesised crowding bias mis-estimate in the Wesenheit magnitudes of Cepheid variable stars, scrutinising the assumptions of flux light curve symmetry about the mean, and the method of artificial star injection (§ 4). Future work can adapt the Bayesian functional principal component architecture in *BayeSN* to model Milky-Way Cepheids, and hierarchically infer the crowding bias, informed by the reduction in light curve fractional amplitude, in extragalactic Cepheids (§ 5.6).

Contents

1	Introduction	4
1.1	Cosmological Tensions	4
1.2	SNe Ia as Probes of Cosmological Parameters	6
1.3	Late Time Systematics	7
1.3.1	<i>Standardisable Candles</i>	7
1.3.2	<i>SN Ia Colour and Host Galaxy Dust Systematics</i>	8
1.3.3	<i>SN-Host Galaxy Correlations</i>	9
1.4	This Work	9
2	Empirical Validation of <i>BayeSN</i> Predictions using Gaussian Processes	11
2.1	<i>BayeSN</i>	11
2.2	Gaussian Processes	12
2.2.1	<i>Motivation</i>	12
2.2.2	<i>Covariance Between Variates</i>	13
2.3	Empirical Validation	14
2.3.1	<i>Rejection Conditions</i>	14
2.3.2	<i>Multiple Linear Regression</i>	17
2.3.3	<i>Feature Comparison & Hyper-parameter Dependence</i>	20
2.3.4	<i>Dust Laws across the Mass-Step</i>	26
3	Velocity-SED Correlations	28
3.1	The Velocity-Colour Relation	28
3.2	<i>BayeSN</i> Methodology	29
3.3	Preliminary Results	30
4	Cepheid Amplitudes Encode Crowding Bias	34
4.1	Crowding Setup	34
4.2	A Compromising Approximation	35
4.3	Parameter Space Evaluation	35
4.4	Interpretation	39
4.5	Artificial Star Injection	39
5	Ongoing and Future Work	44
5.1	The Supernova Triplet in NGC 3147 [\sim Dec 2021]	45
5.2	DEHVILS & Dust-Laws Across the Mass-Step [\sim Dec 2022]	48
5.3	Pan-STARRS + DECam & Constraints on w [\sim Dec 2022]	48
5.4	Velocity-SED Correlations [\sim Dec 2023]	49
5.4.1	Silicon-II Velocity-SED Correlations	49
5.4.2	Kaepora Spectra & Spectroscopic Training	50
5.5	Adjacent Empirical Analyses	51
5.6	Bayesian Crowding	51
A	Intro Cosmology	56
B	Gaussian Cheat Sheet	57
C	NIR-Times Correlate with the First Functional Principal Component	58

1 Introduction

Type Ia Supernovae (SNe Ia) are employed as ‘standard candles’, objects of known brightness, which probe the distance to the source, and hence the underlying cosmological parameters which describe our expanding Universe. In the late 1990’s, high redshift ($z \sim 1$) SN Ia analyses concluded the expansion of the Universe was accelerating (Riess et al., 1998; Perlmutter et al., 1999), providing the first direct evidence of dark energy.

1.1 Cosmological Tensions

Fast forward two decades, a number of observations point to tensions within the working Λ CDM model of the Universe (for recent reviews, see Di Valentino et al. (2021); Perivolaropoulos & Skara (2021)). The ‘base’ Λ CDM model is rooted in the ‘Cosmological Principle’, that the Universe is statistically homogeneous and isotropic (i.e. ‘uniform’) on the largest scales ($\gtrsim 100\text{Mpc}$). Subject to this constraint, the FLRW metric is employed to characterise length scales - in particular luminosity and angular distances - as it is the most general metric that permits an expanding space while satisfying uniformity. Necessarily described by a *constant* spatial curvature, $\Omega_K = 0$ is adopted in base Λ CDM (supported in e.g. Planck Collaboration et al. (2020b): $\Omega_K = 0.0007 \pm 0.0019$).

The most notable tension is a 4.2σ discrepancy between the values of the present day expansion rate of the Universe, the Hubble constant, H_0 , as estimated from the ‘early-time’ 6-parameter base Λ CDM analysis of the Cosmic Microwave Background (CMB: radiation from baryon-photon decoupling at $z_{rec} \sim 10^3$), and from 3-rung distance ladder ‘late-time’ methodologies using Parallax-Cepheid-SN observations. The most recent Planck analysis gives (Planck Collaboration et al., 2020b),

$$H_{0;early} = 67.4 \pm 0.5 \text{ km s}^{-1} \text{ Mpc}^{-1}, \quad (1)$$

whereas the SH0ES collaboration estimate is (Riess et al., 2021),

$$H_{0;late} = 73.2 \pm 1.3 \text{ km s}^{-1} \text{ Mpc}^{-1}. \quad (2)$$

Such discrepancies have motivated a renewed effort to identify and control underlying systematic uncertainties, and have inspired a multitude of independent methods to determine H_0 . Fig. 1 from a recent review, Di Valentino et al. (2021), provides a flavour for the vastness of these studies.

Tensions are not limited to H_0 . For example, the Planck clustering amplitude, $S_8 \equiv \sigma_8 \sqrt{\frac{\Omega_{m,0}}{0.3}} = 0.832 \pm 0.013$ (Planck Collaboration et al., 2020b; Perivolaropoulos & Skara, 2021), is in $\sim 2.6\sigma$ tension with the most recent DES Y-3 results: $S_8 = 0.776 \pm 0.017$ (Abbott, 2021). Whilst uniformity is strongly supported by the isotropy of the CMB, with $\Delta T/T \sim 10^{-5}$, and the broad agreement with Λ CDM from all early/late probes (e.g. Fig. 1), deviations from Λ CDM are also implied by various dipole measurements. Excess magnitudes (relative to Λ CDM expectations) of both local bulk flows (estimates ranging from $\sim 400 \text{ km s}^{-1}$ at $z \sim 0.03$, to 1000 km s^{-1} at $z \sim 0.2$) and quasar density dipoles, and also a variation of the fine structure constant α across the sky are hard to reconcile with Λ CDM (although these findings are highly debated, see Section III.3 Perivolaropoulos & Skara (2021), Perivolaropoulos (2014); Nadolny et al. (2021) and references therein). A violation of the Cosmological Principle may also be evidenced by an apparently direction dependent H_0 inferred from SN analyses, and quasar strong lensing H_0 estimates, that appear to increase with closer alignment with the CMB dipole (Krishnan et al., 2021).

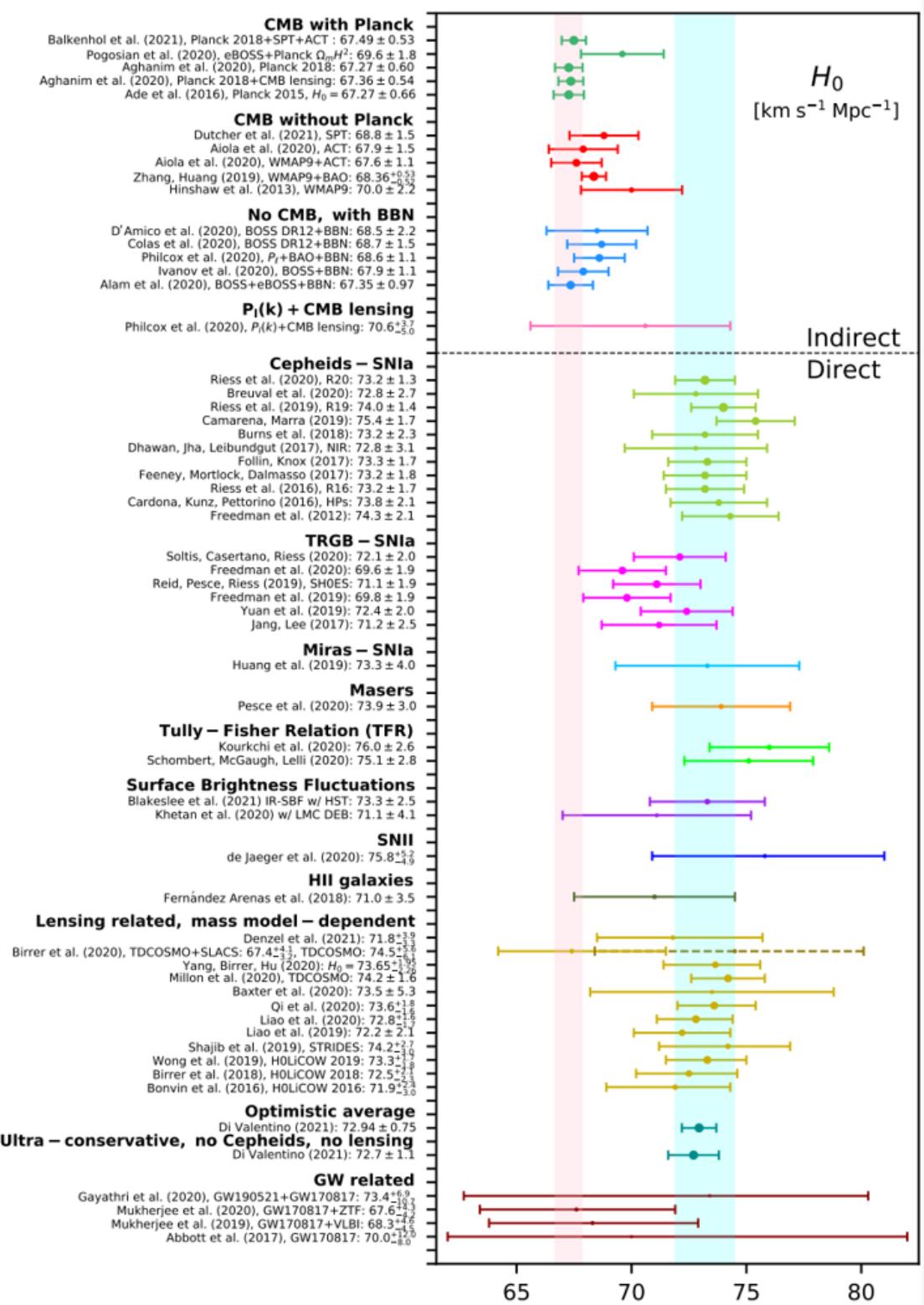


Fig. 1: Di Valentino et al. (2021) Figure 1. depicts H_0 estimates from an expansive list of independent probes. One notes the ‘indirect’ parameter estimates are smaller, which, naively, could signal a problem with the assumed Λ CDM cosmology. Both indirect and direct methods are rooted in the founding assumption of uniformity, employing the FLRW metric (see Appendix A). The former encodes nearly the entire expansion history of the Universe, whereas the latter is limited to evolution under $z \lesssim 1$. Colour bands depict early/late H_0 values, Eqs. (1, 2).

Commenting on the Hubble tension, [Efstathiou \(2021\)](#) broadly divides physical solutions to Λ CDM departures into (1) Modified gravity theories, (2) New early physics ($z \gg 10^3$), (3) New recombination physics (i.e. $z \sim 10^3$), and (4) New late physics ($z \lesssim 1$). However, alternative to new physics, this work seeks to investigate late-time systematic uncertainties, that may be contributing towards the tensions in Λ CDM. In this report, we focus on SNe Ia standardisation, and consider also the potential for magnitude offsets in Cepheid variables.

1.2 SNe Ia as Probes of Cosmological Parameters

SNe Ia occur when a white dwarf star accreting near the Chandrasekhar limit ($\approx 1.4M_\odot$) explodes in a runaway thermonuclear reaction. Now mature cosmological probes, SNe Ia are employed in a variety of ways to place constraints on various cosmological parameters. Therefore, controlling systematic uncertainties tied to SN magnitudes is an important enterprise, and an ongoing community-wide effort. This section briefly surmises how SNe are employed as cosmological probes.

For measuring H_0 , SNe are part of a 3-rung distance ladder ([Riess et al., 2009, 2016, 2021](#)). Milky-Way Cepheid variable star magnitudes are calibrated with a period-luminosity relation using parallax measurements. Extragalactic Cepheids, that also host SNe, are then used to calibrate SN magnitudes. These sources reside at $z \lesssim 0.01$. The third rung is the Hubble flow set of SNe Ia, at $0.01 \lesssim z \lesssim 0.1$. As well as Cepheids, standard candles such as Tip of the Red Giant Branch stars and Mira variables have been used in the distance ladder, in tandem with SNe Ia ([Soltis et al., 2021; Huang et al., 2020](#)).

By learning the absolute magnitude, accurate distances can be obtained, which ultimately encode cosmological parameters. The distance modulus, μ , is the difference between the measured apparent magnitude, m , and the inferred absolute magnitude, M :

$$\mu = m - M = 5 \log_{10} \left(\frac{D_L(z, \bar{z})}{\text{Mpc}} \right) + 25. \quad (3)$$

Here, the (z, \bar{z}) arguments of the luminosity distance, D_L , are the observed and cosmological redshift, the former encapsulating all effects that act to modify wavelength, including observer/source peculiar motion, gravitational redshift, and the expansion of space. The latter encapsulates only the expansion of space.

A cosmographic expansion of a cosmological luminosity distance, $\bar{D}_L(\bar{z})$, is related to luminosity distance by,

$$D_L(z, \bar{z}) = \frac{1+z}{1+\bar{z}} \cdot \bar{D}_L(\bar{z}), \quad (4)$$

which is most strongly dependent on H_0 , where the leading terms of the expansion are independent of the Universe curvature:

$$\bar{D}_L(\bar{z}) = \frac{c\bar{z}}{H_0} \left(1 + \frac{1}{2}(1-q_0)\bar{z} + \mathcal{O}(\bar{z}^2) \right) \equiv \frac{c\bar{z}}{H_0} (1 + \delta(\bar{z})), \quad (5)$$

(q_0 is a deceleration parameter evaluated today; [Visser \(2005\)](#)). In this way, the ‘Hubble diagram’ of distance modulus as a function of cosmological redshift can be plotted. The cosmographic expansion - rooted only in the assumption of uniformity - is fitted to the Hubble diagram, to constrain cosmological parameters. In practice, observed redshift z is corrected to the ‘heliocentric’ rest-frame, z_{Helio} , removing Doppler effects from the Earth rotation, and orbit around the Sun; this makes observations ‘time-of-year-independent’. Similarly, cosmological redshift is estimated as z_{CMB} , by correcting to the CMB rest-frame, removing the total Earth peculiar motion (e.g. peculiar motions of the Solar System, Milky-Way, Local Group, Local Supercluster; [Davis](#)

et al. (2019)). The procedure utilises the CMB $l = 1$ dipole, assuming the anisotropy manifests entirely due to our peculiar motion: $v_p^\odot = (369.82 \pm 0.11) \text{ km s}^{-1}$ (Planck Collaboration et al., 2020a), equivalent to $z_p^\odot = 1.2343(\pm 4) \times 10^{-3}$.

The dark energy equation of state parameter, w , is also probed, by contrasting a local Hubble flow SN set ($0.01 \lesssim z \lesssim 0.1$) with a high-redshift set ($0.5 \lesssim z \lesssim 1.5$) (Scolnic et al., 2018; Brout et al., 2019). Moreover, standardised SN magnitudes are used to probe peculiar velocities of local galaxies falling towards overdensities, and hence the underlying gravitational model. The parameterisation for this effect is $f\sigma_8$, where σ_8 is the RMS mass fluctuation within a radius $8 \text{ Mpc } h^{-1}$, and the growth rate is $f \approx \Omega_{m,0}^\gamma$; adopting ΛCDM , general relativity predicts a growth rate index $\gamma \approx 0.55$ (Howlett et al., 2017; Huterer et al., 2017; Graziani et al., 2020). Further still, high- z SNe can probe the matter distribution of the Universe through weak lensing magnifications (Holz & Wald, 1998; Smith et al., 2014; Scolnic et al., 2019), and strong lensing time-delays can be used to measure the Hubble constant (Refsdal, 1964; Bonvin et al., 2017). Searches for a direction dependent H_0 with SNe can also test the Cosmological Principle (Krishnan et al., 2021).

1.3 Late Time Systematics

1.3.1 Standardisable Candles

There are many challenges to overcome when utilising standardisable candles to yield accurate distance estimates. Following Eq. 3, we require accurate and precise measurements of apparent magnitude, an accurate prediction of the intrinsic absolute magnitude, and a good understanding of how photon flux is affected when transiting from source to observer (on a source-by-source basis).

Cross-calibration systematics - those that arise from observing sources with different instruments at different locations - are important, but being controlled for, following a community-wide effort to observe large SN samples on a single observatory, e.g. Pan-STARRS observatory for Foundation and the Young Supernova Experiment that provides a low- z SN anchor in the optical (Foley et al., 2018; Jones et al., 2019), SIRAH and RAISIN both employing the Hubble Space Telescope (HST) for a differential comparison of low and high redshift samples in the NIR (Jones et al. in prep), and the Legacy Survey of Space and Time (LSST) that will observe on the Vera C. Rubin Observatory, and discover $\sim 10^5$ SNe Ia over the projected 10-year life span¹. Similarly, the SH0ES collaboration observe Milky-Way Cepheids and extragalactic Cepheids/SNe with HST (Riess et al., 2016, 2019).

Instead, the effects of host-galaxy dust, the potential for a redshift evolution of absolute magnitude, and correlations between standardisable candle luminosity with host-galaxy properties are the dominant systematic uncertainties at play in late-time cosmology. Consider the 3-rung distance ladder. A heuristic equation for propagating apparent magnitudes up the rungs, **assuming absolute magnitude M does not evolve with redshift**, from the Milky-Way, MW , to Cepheid-SN hosts, $host$, to Hubble flow SNe, $flow$, is written:

$$\mu_{flow} \equiv \mu_{parallax} - m_{Ceph;MW} + m_{Ceph;host} - m_{SN;host} + m_{SN;flow}. \quad (6)$$

This toy-equation shows that if host-Cepheids/flow-SNe are too bright, or Milky-Way-Cepheids/host-SNe too dim, a correction $\Delta m \sim 0.1 - 0.2 \text{ mag}$ would resolve the Hubble tension. This is explored in Efstathiou (2020), where it is concluded $\Delta M_{E20} \sim 0.1 - 0.14 \text{ mag}$ in the SH0ES Wesenheit magnitudes of host-Cepheids is a resolving offset; this is investigated in § 4. Cepheid outlier rejection procedures may also introduce a bias (Efstathiou, 2014, 2020).

¹<https://www.lsst.org/about>

1.3.2 SN Ia Colour and Host Galaxy Dust Systematics

Current research in SN Ia cosmology is strongly focused on identifying, quantifying, and factoring out empirical correlations for magnitude standardisation (and using correlations to inform target selection). As in Rigault et al. (2020), theoretical astrophysical model precision is not “anywhere near the precision required for cosmological measurements”, hence, the empirical approach is the standard paradigm. The most widely adopted correction for magnitude standardisation is the Phillips ‘broader-brighter’ correlation (Phillips, 1993), that brighter SNe in the optical B -band take a longer time to decay in brightness. This is quantified with $\Delta m_{15}(B)$, the difference between apparent magnitudes at peak and 15 rest-frame days later, which is smaller for brighter SNe. Another dominant effect is an empirical ‘redder-dimmer’ relation, whereby SNe with redder (more positive) colours are optically dimmer. ‘Colour’ is a technical term referring to the difference in magnitudes at a single time across two passbands, most often $B - V$.

Simplistic models such as *SALT2* (Guy et al., 2007) apply these two corrections to the B -band apparent magnitude m_B with the ‘Tripp’ formula, using a ‘stretch’ term, x_1^s , which encodes light curve shape, and a single colour term, c^s , which is a proxy for $B - V$ peak apparent colour, for each supernova, s :

$$\mu^s = m_B^s - M_B + \alpha x_1^s - \beta c^s. \quad (7)$$

Here, the parameters (M_B, α, β) are population hyper-parameters; in particular, M_B is the mean standard candle absolute magnitude. However, this purely empirical relation - which ignores the physical origins of the ‘redder-dimmer’ trend - can lead to biased distance estimates.

In fact, there are two distinct origins of the ‘redder-dimmer’ trend. Dust along the line of sight in the SN Ia host galaxy acts to reduce transiting flux by scattering and/or absorbing & re-radiating incident photons in a random direction (Salim & Narayanan, 2020). Dust grains selectively scatter wavelengths that are comparable to the grain size, $\sim 0.1\mu\text{m}$, thus optical light can be significantly extinguished by host-galaxy dust, with an extinction magnitude $A_V \sim 0.1 - 1$ mag, whereas NIR extinction following the Fitzpatrick (1999) dust law, $\xi(\lambda, R_V)$, can be up to $\sim 5-11$ times smaller (Cardelli et al., 1989). The R_V parameter here controls the shape of the dust law, and is related to the dust grain size (Salim & Narayanan, 2020). Thus, in addition to extinction, there is a reddening effect, whereby *wavelength-dependent extinction* leads to a non-negligible difference between *intrinsic* and *apparent* colour, c . Specifying both (A_V, R_V) under the Fitzpatrick law thus provides a complete description of dust. In all, extinguished magnitudes, M_b^{ext} , and apparent colours, $c_{b,b'}^{app}$, resulting from dust action on intrinsic magnitudes, M_b^{int} , in the generic b -band, are described as:

$$\frac{A_\lambda}{A_V} = \xi(\lambda, R_V) ; \quad A_b = \frac{\int_{\lambda_{min}}^{\lambda_{max}} A_\lambda \mathcal{T}_b(\lambda) d\lambda}{\lambda_{max} - \lambda_{min}} \quad (8)$$

$$M_b^{ext} = M_b^{int} + A_b ; \quad c_{b,b'}^{app} = c_{b,b'}^{int} + A_b - A_{b'}. \quad (9)$$

where $\mathcal{T}_b(\lambda)$ is the passband transmission function.

The second effect that manifests as a ‘redder-dimmer’ trend is a physical variation of *intrinsic* colour with absolute magnitude within of a population of SNe Ia, i.e. *in the absence of dust*. Because this effect is distinct from dust, the slopes of these two effects in the plane of extinguished absolute magnitude against apparent colour are - in all likelihood - different (this is demonstrated in Mandel et al. (2017)). Thus, their combination leads to a *curve* in the $M^{ext} - c^{app}$ plane. **Fitting this curve with a linear relation as in the Tripp formula leads to biased distance estimates in the tail-ends of the fit**, with offsets of order $\sim 0.1\text{mag}$ (Mandel et al., 2017). Moreover, Brout & Scolnic (2021) demonstrate improper treatment of intrinsic colour and dust effects can lead to a systematic $\Delta w \sim 0.04$, which is now the dominant systematic

uncertainty in SN cosmology. To conclude, the Tripp formula groups together the two distinct effects of intrinsic colour variations and dust into one parameter, which can lead to biased distance estimates, and inaccurate constraints on cosmological parameters.

Clearly, the effects of dust are problematic for supernova cosmology. Despite this, SNe are conventionally observed at optical wavelengths, which are highly sensitive to dust; here, SNe are *standardisable* candles. However, at near-infrared wavelengths ($1 - 1.6\mu\text{m}$), where dust is transparent to photons, SNe are *nearly standard* candles, exhibiting a low variance in absolute magnitude *without* applying any stretch/colour corrections. [Avelino et al. \(2019\)](#) demonstrate that distance estimates using unstandardised SN NIR absolute magnitudes outperform standardised optical magnitudes. However, utilising the full optical-NIR spectrum in a hierarchical way is optimal and provides the most stringent constraints on dust ([Mandel et al. 2020](#), § 2.1).

1.3.3 SN-Host Galaxy Correlations

In addition to stretch/colour effects, quantifying less significant ($\lesssim 0.1\text{mag}$) SN correlations is an active area of research. In particular, star formation rate may be tied to SN absolute magnitudes, resulting in a ‘younger-fainter’ correlation ([Rigault et al., 2020](#)). A ‘redder-faster’ colour-velocity relation has also been reported, and is investigated in § 3 ([Foley & Kasen, 2011](#); [Siebert et al., 2020](#)). SN correlations may also exist between: host-galaxy type, age, distance to host centre, metallicity, redshift etc. ([Sullivan et al., 2010](#); [D’Andrea et al., 2011](#); [Foley & Kasen, 2011](#); [Rigault et al., 2013](#); [Gagliano et al., 2020](#); [Rigault et al., 2020](#)).

A controversial SN-host correlation is the ‘mass-step’, whereby SNe Ia in higher stellar-mass host galaxies ($\log_{10}(M_*/M_\odot) \gtrsim 10$) appear to be optically brighter post-standardisation ($\Delta m \sim 0.1\text{mag}$; [Kelly et al. \(2010\)](#); [Childress et al. \(2013\)](#); [Smith et al. \(2020\)](#); [Johansson et al. \(2021\)](#); [Brout & Scolnic \(2021\)](#); [Thorp et al. \(2021\)](#)). [Brout & Scolnic \(2021\)](#) test the hypothesis that there is no intrinsic mass-step in the absolute magnitudes of SNe Ia, and the observed mass-step is an apparent effect resulting from distinct R_V distributions, that differ between low and high stellar-mass hosts. By performing forward simulations, they replicate the observed mass-step in the Hubble residuals, by assuming that low/high-mass hosts have different Gaussian distributions in R_V centred on (2.75, 1.5), respectively. On the other hand, a full Bayesian analysis with *BayeSN* is performed in [Thorp et al. \(2021\)](#), and the population hyperparameters of the low/high mass R_V distributions are hierarchically constrained. The mean R_V hyperparameter values are consistent (~ 2.7) across a persistent mass-step, implying the mass-step is an intrinsic effect that is not explained by dust. Moreover, [Uddin et al. \(2020\)](#) and [Ponder et al. \(2020\)](#) both report a NIR mass-step; given the NIR-bands are dust-insensitive, a NIR mass-step - if it exists - cannot be explained away by a difference in dust distributions in low/high mass hosts.

1.4 This Work

This work is motivated to improve SN Ia distance estimates for constraints on cosmological parameters, by exercising a rigorous treatment of dust, allowing us to investigate and quantify SN-host correlations in an accurate way. I also investigate the potential for apparent magnitude offsets in the Wesenheit magnitudes of Cepheids.

In this work we utilise a next-generation hierarchical Bayesian model of Type Ia spectral energy distributions, *BayeSN* ([Mandel et al. \(2020\)](#); [Thorp et al. \(2021\)](#); § 2.1). Thus far, we have utilised two datasets, the ‘Avelino’ sample, a size 79 sub-sample ($0.01 \lesssim z \lesssim 0.04$) from [Avelino et al. \(2019\)](#), observed in $uBVRIYJHK$ passbands spanning UV-Optical-NIR, predominantly from CfA/CSP surveys and others, e.g. ([Krisciunas et al., 2017](#); [Friedman et al., 2015](#); [Krisciunas](#)

(et al., 2007), and the Foundation DR1 sample of 157 sources ($0.015 \lesssim z \lesssim 0.08$) in *griz* passbands observed with Pan-STARRS (Foley et al., 2018; Jones et al., 2019). We are also active members of the new DEHVILS (§ 5.2) and YSE SN surveys (§ 5.3).

In § 2, we perform an empirical validation of *BayeSN* model predictions using Gaussian processes. In § 3, we investigate the ‘redder-faster’ relation in SNe Ia with *BayeSN*, examining the significance of this correlation, and investigating whether velocity measurements can be used to improve distance estimates. In § 4, I investigate the crowding bias in Cepheids, studying the assumption of flux-symmetry about the mean, and apparent shortcomings of the method of artificial star injection. In § 5, I detail the road-map forward for future research, that we will conduct in the next ~ 3 years.

2 Empirical Validation of *BayeSN* Predictions using Gaussian Processes

BayeSN is a state-of-the-art statistical model of the spectral energy distributions (SEDs) in Type Ia supernovae (Mandel et al., 2020), which has several built in advantages over the widely adopted *SALT2* model (§ 1.3 ; Guy et al. (2007)). *BayeSN* exercises: hierarchical Bayesian modelling of independent intrinsic colour and dust distributions at the fundamental level of the time-dependent SED, continuous optical through near-infrared (NIR) coverage, and the ability to learn astrophysical correlations (with e.g. host galaxy stellar-mass, metallicity, SN ejecta velocity etc.). The goal in this section is to empirically validate *BayeSN* model predictions using a model independent approach, that employs Gaussian processes (GPs) to interpolate observer-frame light curve data. In § 2.1, I provide an overview of the *BayeSN* model. In § 2.2, I motivate the model independent approach, and detail the mathematical foundation for Gaussian processes. Finally, I describe the work undertaken to compare *BayeSN* and GP fits in § 2.3.

2.1 *BayeSN*

BayeSN is a continuous-SED hierarchical Bayesian model for Type Ia supernovae, that fuses Bayesian statistics with functional data analysis techniques, and works to yield distance estimates to SNe with unprecedented accuracy and precision (Mandel et al., 2020; Thorp et al., 2021). Utilising the Hsiao (2009) spectral template as a prior, *BayeSN* trains on photometric time series to learn the intrinsic functional form of luminosity with time and wavelength. The forward model, taken from Mandel et al. (2020), is depicted in Fig. 2. *BayeSN* is implemented in the *Stan* probabilistic programming language (Carpenter et al., 2017), that employs advanced Hamiltonian Monte-Carlo algorithms to efficiently sample the posterior distribution (Hoffmann & Gelman, 2014; Betancourt, 2016).

A distribution of SEDs is modelled by creating a grid of free parameters at ‘knots’, with values corresponding to the height or flux value at that point. The knots are located at pre-defined coordinates in time and wavelength, spaced by a rough estimate of the variability scale: $\sim (10d, 0.1\mu m)$. Cubic spline interpolation between the knots is used to model a continuous function of time and wavelength. Then functional principal components (FPCs) describing the primary modes of SED variation are learned. In addition to these FPCs, the population distribution hyperparameters governing dust extinction (Mandel et al., 2020), and reddening (Thorp et al., 2021), can be learned.

For each supernova, s , the coefficient on the first functional principal component, θ_1^s , is a parameter describing the primary intrinsic SED shape. ($K = 1$ modes are sufficient to explain the dominant modes of variation, with residuals - described by Gaussian processes drawn from a non-stationary kernel - capturing the remaining variation). Variations in θ_1 manifest as empirically derived correlations, between, for example, optical brightness and decline-rate such as in the Philips relation (Phillips, 1993); smaller (more negative) θ_1 corresponds to a slow-declining, bright B -band light curve. Host galaxy dust extinction, A_V^s , is estimated from the SED fit. Finally, the distance modulus, μ^s , and δM^s , an intrinsic wavelength-independent magnitude offset, are constrained for each supernova.

As a consequence of continuous modelling, *BayeSN* removes the need for K-corrections. K-corrections are conventionally applied to correct for the effects of cosmological redshift, the action of which results in **our observer-frame passbands probing a bluer region of the source intrinsic SED**. For illustration, an observed V -band peak magnitude ($\lambda_{V;Centre} \approx 5400\text{\AA}$) from a source at $z \approx 0.25$, would originate from a B -band region ($\lambda_{B;Centre} \approx 4300\text{\AA}$) of the intrinsic SED. To apply conventional K-corrections, an intrinsic SED model is required, as well as

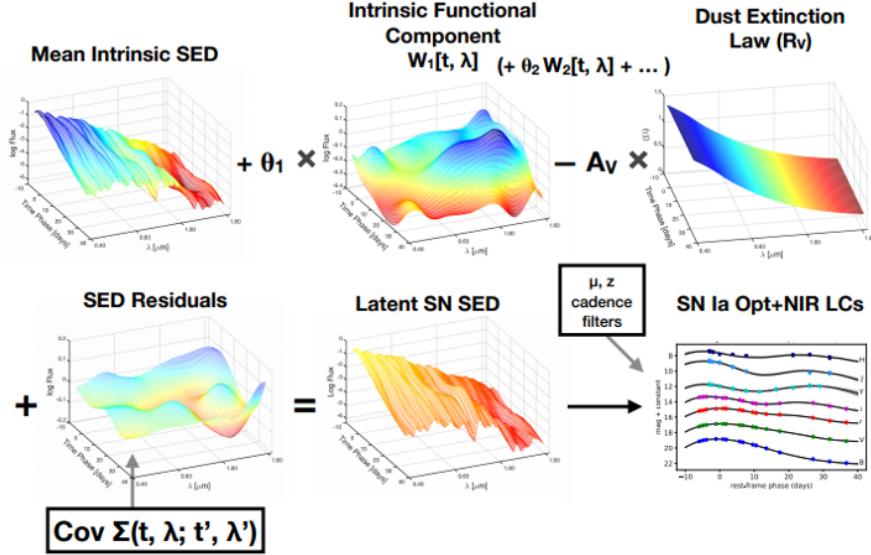


Fig. 2: Figure taken from [Mandel et al. \(2020\)](#) depicts the forward model for generating photometric time series. A population intrinsic sheet of flux as a function of time and wavelength, is warped on a SN-by-SN basis by functional principal components (FPCs) - scaled by the coefficients θ_i - dust extinction, and SN-specific SED residuals, to yield a latent SN model. This sheet is convolved with the transmission functions of the various photometric passbands, to yield a photometric time-series. Bayesian inference works to solve the inverse problem, learning the intrinsic sheets, FPC scores, dust parameters, and distance moduli, by training on real light curve data.

redshift measurements. *BayeSN* requires two redshifts as inputs, (z_{Helio}, z_{CMB}) , to encapsulate the effects of redshift ([§ 1.2](#)). *BayeSN* takes in transmission functions as inputs, and leverages the continuity of the SED model to fit and/or simulate light curves in any photometric system. In this way, cross-calibration uncertainties are minimised.

Cosmological time dilation is corrected for using z_{Helio} , yielding the source rest-frame passage of time, namely ‘phase’. For supernovae, the convention is to convert observer-frame Modified Julian Date, T_{MJD} , to phase, t , by zeroing at the *intrinsic* time of *B*-band maximum, $T_{B;max}$,

$$t = \frac{T_{MJD} - T_{B;max}}{1 + z_{Helio}}. \quad (10)$$

Re-iterating, $T_{B;max}$ is an MJD time in the *source* rest-frame - a ‘K-corrected peak time’ - and is fitted for in *BayeSN*. Unknown source peculiar redshift is modelled as a Gaussian distributed error, $\sigma_{pec} = 150 \text{ km s}^{-1}$ ([Carrick et al., 2015](#); [Mandel et al., 2020](#))

2.2 Gaussian Processes

2.2.1 Motivation

BayeSN is an ambitious and highly complex model, and a model-independent empirical validation of *BayeSN* predictions is warranted, to affirm the framework is functioning as intended. In this section we introduce Gaussian processes (GPs), an interpolation tool for assessing the

performance of *BayeSN* fits. The need for a model-independent approach is motivated by the use of the Hsiao (2009) template as a prior in *BayeSN*, which may introduce unknown systematic biases. Moreover, any model that makes predictions informed by a training set, is subject to internal biases that originate from training-set selection effects. Instead, single-task Gaussian processes provide for a completely data-driven means of interpolation, informed *only* by data in the single-band single-SN time series (and choice of ‘kernel’, see below). Here, ‘single-task’ means there is no sharing of information, neither between passbands, nor between SNe.

Contrary to population-driven models, the specification of a Gaussian process ‘kernel’ function is the *single* choice made to constrain the functional forms of the interpolation. This choice is informed simply by inspection of the population of time series, and a qualitative judgement of the variability characteristics. For our time series, one concludes the light curves (LCs) are smoothly varying on ~ 10 day timescales, and a stationary kernel function such as the ‘squared-exponential’ kernel is suitable...

2.2.2 Covariance Between Variates

So what is a Gaussian process? In the acclaimed text Rasmussen & Williams (2006), *Gaussian Processes for Machine Learning*, Gaussian processes are defined:

“A Gaussian process is a collection of random variables, any finite number of which have a joint Gaussian Distribution”.

For our purposes, this “collection of random variables” will be a set of flux values, $\mathbf{f}(\mathbf{t}_*)$, evaluated on a 1-D grid of evenly spaced points in time, \mathbf{t}_* . This ‘function’ that we use to interpolate between real, irregularly sampled, light curve data points, $(\mathbf{t}_o, \mathbf{y}(\mathbf{t}_o))$, is nothing more than a *vector* of values. Each entry in this vector is related to every other entry, according to the kernel function, which defines the **covariance** between any two variate values (Appendix. B). As an example, the aforementioned ‘squared exponential’ or ‘radial basis function’ (RBF) kernel, with generic input covariates (t, t') , is written;

$$Cov[f(t), f(t')] = k(t, t') = A^2 \exp\left(-\frac{(t - t')^2}{2\tau^2}\right). \quad (11)$$

Here, (A, τ) are the *hyper-parameters* governing the variate magnitude of variation, A , per unit variability timescale, τ . For larger τ , the time-elapsed for some fixed reduction in covariance *increases*, implying the ‘function’ is more *slowly* varying. Similarly, a larger A under fixed τ allows variate values to vary more per unit time-scale. A redacted case-study serves to demonstrate that **non-zero covariance between variates gives rise to ‘functions’, with a certain characteristic shape, governed by the kernel**.

In the most general case of a Gaussian process, vectors are drawn from a multivariate Gaussian distribution, characterised by a mean vector, \mathbf{m} , and a covariance-matrix, Σ , with *non-zero* off-diagonal elements, computed using the kernel function. Our GP is,

$$\mathbf{f}(\mathbf{t}) = \begin{pmatrix} f_1 \\ \vdots \\ f_N \end{pmatrix} \sim \mathcal{N}(\mathbf{m}(\mathbf{t}), \Sigma) \quad ; \quad \Sigma_{ij} = k(t_i, t_j), \quad (12)$$

which is succinctly denoted,

$$f(t) \sim GP(m(t), k(t, t')) \quad (13)$$

$$E[f(t)] = m(t) \quad (14)$$

$$Cov[f(t), f(t')] = E[(f(t) - m(t))(f(t') - m(t'))] = k(t, t'). \quad (15)$$

Appendix. B details how the GP is conditioned on the data, and how hyper-parameters are optimised to maximise the marginal log-likelihood.

2.3 Empirical Validation

The *T21 BayeSN* model from Thorp et al. (2021) is trained on 157 SN Ia *griz* light curves from Foundation DR1 (Foley et al., 2018; Jones et al., 2019), and is the first model for *z*-band SN Ia light curves. Unexpectedly, Thorp et al. (2021) found an inversion of the Phillips relation in the *z*-band, whereby optically brighter SNe are predicted to be *faster* declining in the *z*-band (see Fig. 3). As far as we know, there have not been any theoretical models that predict this inverse trend in *z*-band wavelengths. Moreover, the *iz*-bands are of particular interest, as they will be widely employed in current/future surveys, e.g. Foundation, YSE, LSST. Standardising SNe in *iz*-bands (the most dust insensitive of *griz*) will be important moving forward. Thus, validating the *BayeSN* model in this wavelength range is strongly motivated.

In this section, I extract features from Gaussian process fits to Foundation DR1 *griz* time series, and detail the rejection conditions (§ 2.3.1). In § 2.3.2, I perform a multiple linear regression on features, and affirm *BayeSN* trends. Following this, I directly compare GP and *BayeSN* features. The resulting discrepancies are then scrutinised by investigating GP fit dependence on hyper-parameters (§ 2.3.3). Finally, dust-law dependence on host-galaxy stellar mass is evaluated, and compared to results of Thorp et al. (2021) (§ 2.3.4).

Single-task Gaussian process fits are performed using the Python package `George`³ (Ambikasaran et al., 2015), which implements gradient-descent algorithms on the marginal log-likelihood, as in Eq. B.20, to optimise hyper-parameters. It is well documented that SN Ia light curves are smoothly varying, and stationary (kernel is time invariant). This motivates the squared-exponential kernel as the choice of GP covariance function, and is a popular choice in the literature (Kim et al., 2013; Lochner et al., 2016; Avelino et al., 2019). Example sets of GP light curve fits are shown in Fig. 4.

We correct light curves to rest frame phase, and slice so $t \in [-10, 40] d$, to match the *BayeSN* training phase range. Milky Way dust corrections are applied, with directional reddening $E(B - V)$, assuming the Fitzpatrick law (Fitzpatrick, 1999), and $R_{V,MW} = 3.1$. We then compute extinguished absolute peak magnitudes using apparent magnitudes and the distance moduli. Distance modulus is computed using cosmological redshift, and assuming a flat Λ CDM cosmology, using an astropy Python package⁴.

2.3.1 Rejection Conditions

Not all GP fits are robust, and the error envelope can sometimes far exceed the typical error bar scale ($\gg 5\sigma$). In such cases, it is appropriate to reject the GP fit. However, it is time-intensive to inspect and accept/reject all 4×157 light curves. Moreover, a poor GP fit is not without value, and its usefulness depends on the feature we wish to measure: if, for example, we are estimating Δm_{15} , and there are data points within $\Delta t \ll 1 d$ of each time of interest, $t = (0, 15) d$, then the quality of the GP fit is redundant, and the data useful. Here, the feature should not be rejected.

This motivates construction of an algorithm to loop through the light curves, given a feature of interest, and ‘decide’ whether the feature should be extracted for analysis. The parameters that characterise the rejection conditions are summarised as: (*Condition*, DT , G_{max} , N).

Condition: String is ‘Both’ or ‘Either’. This rejection parameter dictates whether a data point must reside on both sides of the time of interest (‘Both’), or whether a data point need only

³<https://george.readthedocs.io/en/latest/>

⁴<https://docs.astropy.org/en/stable/api/astropy.cosmology.FlatLambdaCDM.html>

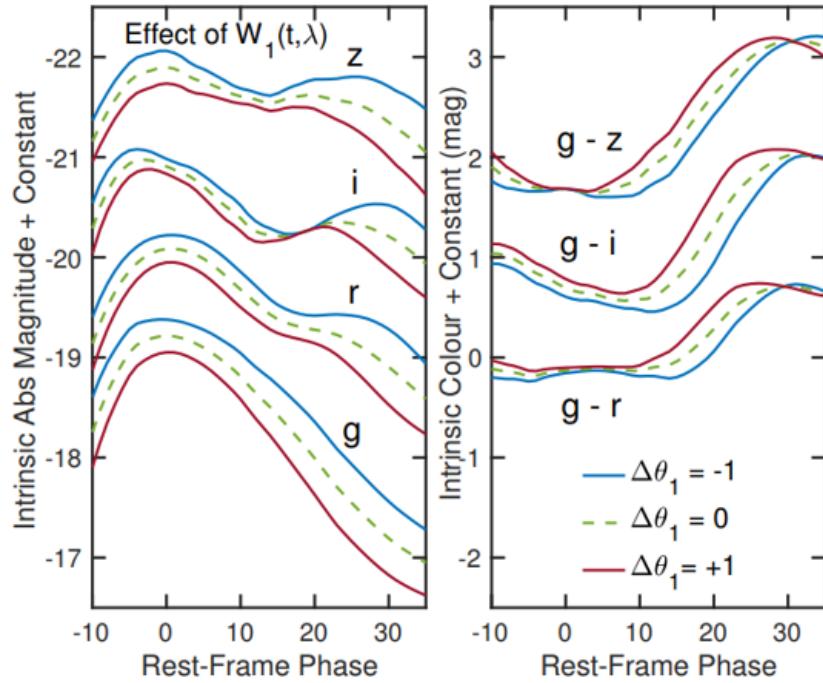


Fig. 3: Figure 2. from Thorp et al. (2021) uses the *BayeSN T21* model fitted to the Foundation DR1 *griz* dataset to synthesis light curves (LCs), and showcase the diversity in SN Ia light curve shapes. Plotted are the intrinsic absolute magnitudes (**Left**) and colours (**Right**) as a function of time, for three values of the coefficient on the first functional principal component, θ_1 . The familiar Philips relation at optical wavelengths is seen to manifest in the *BayeSN* *g*-band synthesised LCs, where the slow declining LCs are intrinsically brighter ('broader-brighter'). However, for longer-wavelengths, *BayeSN* has learned an inversion of this trend, where the slow-declining optically brighter SNe are *fast*-declining in the *z*-band.

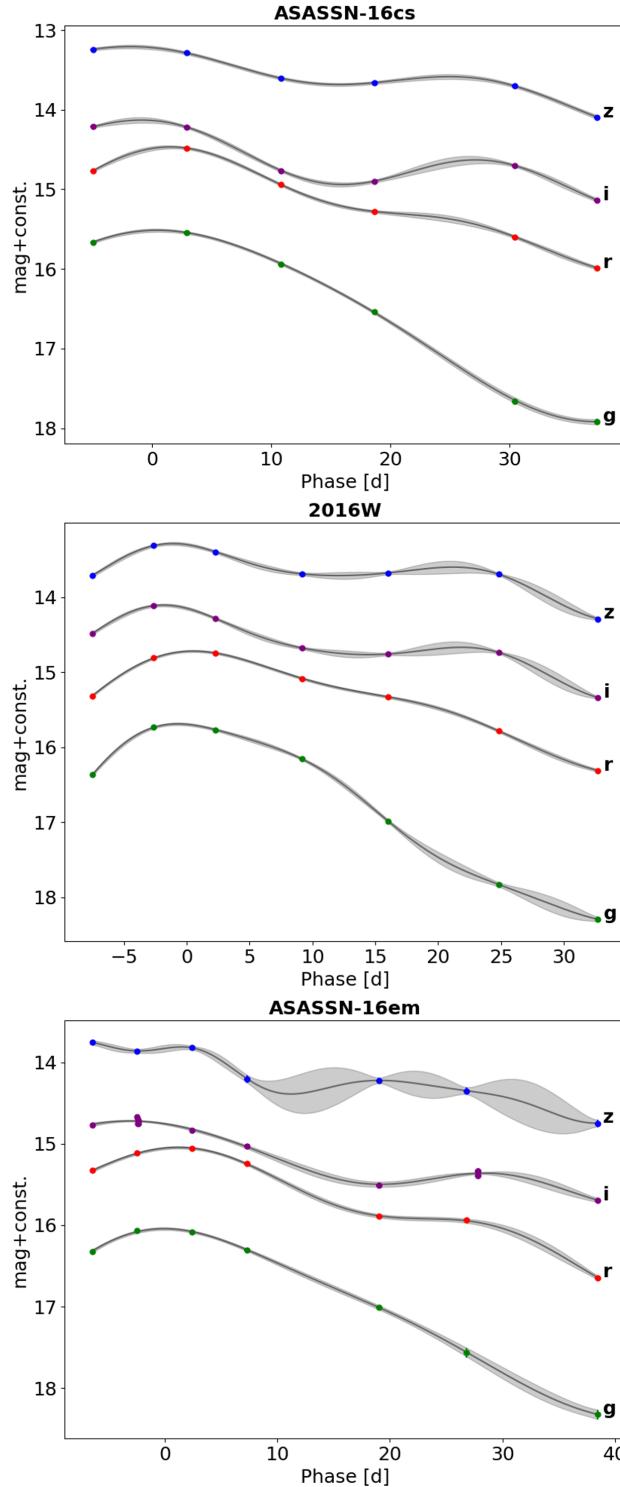


Fig. 4: Plotted are single-task Gaussian processes fits, computed using a constant mean and RBF kernel prior (Eqs. 11, B.12), and conditioning on the observer-frame *griz* photometric time series of Foundation DR1 SNe. Time is *rest*-frame phase, computed using heliocentric redshift (Eq. 10). GP fits are depicted by a dark grey mean function (Eq. B.17), and a light grey error envelope (Eq. B.18). **Upper:** SN ASASSN-16cs is an example of excellent GP fits. **Middle:** SN 2016W GP fits are good. Error envelope is seen to grow between data points, but only at a level $G < G_{max} = 3$. **Lower:** SN ASASSN-16em sees a rejection of $\Delta m_{15}(z)$, with $G \simeq 6.9$ from $t_o \simeq 19$ d to $t_* = 15$ d.

reside on one side ('Either').

DT: Float value in days. Tied to the *Condition* parameter, this setting governs how close a data point must reside to the time of interest.

G_{max} : Dimensionless float value. Here the maximum allowed growth parameter is set. The growth parameter is defined as the ratio of the GP envelope error at the time of interest to the error bar at the data: $G = \sigma_{GP}(t_*)/\sigma_{Data}(t_*)$.

N: Dimensionless integer. This is simply the minimum number of data points required in the single-band time series.

The settings we choose are: ('Either', 5, 3, 4). To accept a feature then, 'a data point must reside within 5 days of the time of interest, the growth parameter cannot exceed 3, and the light curve must have a minimum of 4 points'.

The 'Either' *Condition* setting is chosen to maximise acceptance: consider a scenario where the first data point occurs at $t = 1 d$, and is well sampled beyond this, meaning the GP extraction back to $t = 0$ is good. Under the 'Both' setting, peak magnitudes and decline-rates measured at $t = 0$ would be unnecessarily rejected. Following this, how far away should the first data point be from $t = 0$ before we judge the GP back-extraction to be inaccurate? The *DT* setting of $5 d$ is reasonable, given an estimated $\sim 10 d$ variability time scale; this setting is affirmed on inspection of GP hyper-parameters (see Fig. 10). For the growth parameter, a setting $G_{max} = 3$ strikes a balance between rejecting the globally poor GP fittings, but still accepting reasonable posterior variance growth when data is sparse. Finally, $N = 3$ is judged to be too sparse to yield reasonable feature estimates, particularly for *iz* curves that vary on short timescales. $N = 4$ is the minimum allowed number. An example rejection due to growth is shown in the lower panel of Fig. 4.

By applying rejections, typically $\sim 50 - 70\%$ of the 157 size sample is retained. Naturally, the number of retained multi-time features is exceeded by single-time features, e.g. peak time magnitudes. For interesting features located at times of peak/dip in *iz* curves, the number of accepted features drops to ~ 20 SNe. Here an inflection point in the light curve is accepted only if there is a *U*-turn in the GP error envelope, which requires ≥ 3 points nearby (i.e. within $\sim 10 d$).

2.3.2 Multiple Linear Regression

With accurate features extracted, a multiple linear regression can be performed. Kelly (2007) showed that a failure to account for covariate measurement errors and/or intrinsic dispersion in the linear model tends to bias linear slope parameters. Namely, a rudimentary ordinary least squares regression (OLS), or a chi-squared minimisation that includes *y*-measurement errors, returns underestimates of the slope. Conversely, a 'FITEXY' chi-squared statistic, that additionally incorporates the slope and covariate measurement errors, returns over-estimated slopes. Instead, an accurate estimate of regression parameters is obtained by performing a maximum likelihood estimation. Assuming our features are Gaussian random variables, the notation for a single covariate latent point, ξ_i , measured as x_i with measurement error $\sigma_{x,i}$, and a latent response variate, η_i , with a measurement and associated error $(y_i, \sigma_{y,i})$, is:

$$\eta_i = \alpha + \beta \xi_i + \epsilon_i \quad ; \quad \epsilon_i \sim \mathcal{N}(0, \sigma_{int}^2) \quad (16)$$

$$\xi_i \sim \mathcal{N}(\mu, \tau^2) \quad ; \quad x_i \sim \mathcal{N}(\xi_i, \sigma_{x,i}^2) \quad (17)$$

$$\eta_i \sim \mathcal{N}(\alpha + \beta \xi_i, \sigma_{int}^2) \quad ; \quad y_i \sim \mathcal{N}(\eta_i, \sigma_{y,i}^2) \quad (18)$$

Here, there is room for 'intrinsic scatter', σ_{int} , in the *latent* variables, i.e. even with zero measurement error, there is scatter about the linear relation. Drawing on the Gaussian identities

in Appendix. B, it can be shown the sampling distribution for a single data point is:

$$\begin{pmatrix} y_i \\ x_i \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} \alpha + \beta\mu \\ \mu \end{pmatrix}, \begin{pmatrix} \sigma_{int}^2 + \beta^2\tau^2 + \sigma_{y,i}^2 & \beta\tau^2 \\ \beta\tau^2 & \tau^2 + \sigma_{x,i}^2 \end{pmatrix} \right) \quad (19)$$

Collating all data points into vectors, the likelihood function - the probability of the data given the model (parameters) - is constructed: $p(\mathbf{y}, \mathbf{x} | \boldsymbol{\theta}, \boldsymbol{\psi})$, with model parameters $\boldsymbol{\theta} = (\alpha, \beta, \sigma_{int})$ and population hyperparameters $\boldsymbol{\psi} = (\mu, \tau)$. The full form of this likelihood function, extended to *multiple* covariate variables, is written in Equations (29-31) of Kelly (2007).

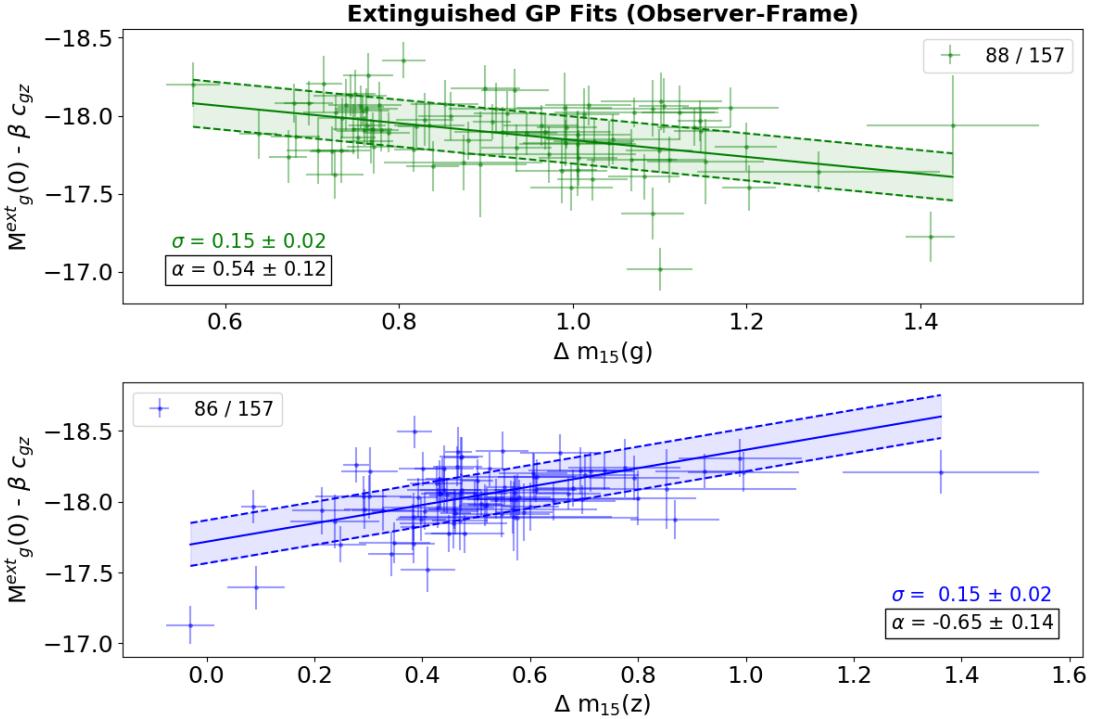


Fig. 5: Linear relations of Gaussian process fits, subject to rejection conditions, are consistent with expectations from *BayeSN* model predictions. Plotted are data points, associated errors, the *Stan* fitted model, and shaded envelope depicting the intrinsic scatter. This scatter, $\sigma_{int} \equiv \sigma$, and the slopes, α , are quoted with the sample average and standard deviation of MCMC samples. **Upper:** Extinguished g -band absolute magnitudes at peak time are standardised by an apparent colour at peak, c_{gz} , and a g -band decline-rate term, $\Delta m_{15}(g)$. **Lower:** Same as above, but the standardisation term is $\Delta m_{15}(z)$. More in Fig. 6.

Using gradient-descent algorithms, the likelihood function can be optimised to yield the MLE. However, by placing priors on the parameters, $\{\boldsymbol{\theta}, \boldsymbol{\psi}\}$, a full Bayesian analysis can be performed. I performed this analysis in *Stan*. At the time of analysis, no prior was specified in the *Stan* code, meaning the scheme defaulted to a uniform prior on $[-\infty, \infty]$. Although this is improper, it manifests only as a constant term in the log-posterior, which is ignorable (just like the Bayesian evidence $p(\mathcal{D})$). In this limit of a well behaved likelihood function, and data that are well described by the model, the improper prior is not deleterious to the analysis. For variances, (σ_{int}^2, τ^2) , a lower bound at zero is set trivially in *Stan*.

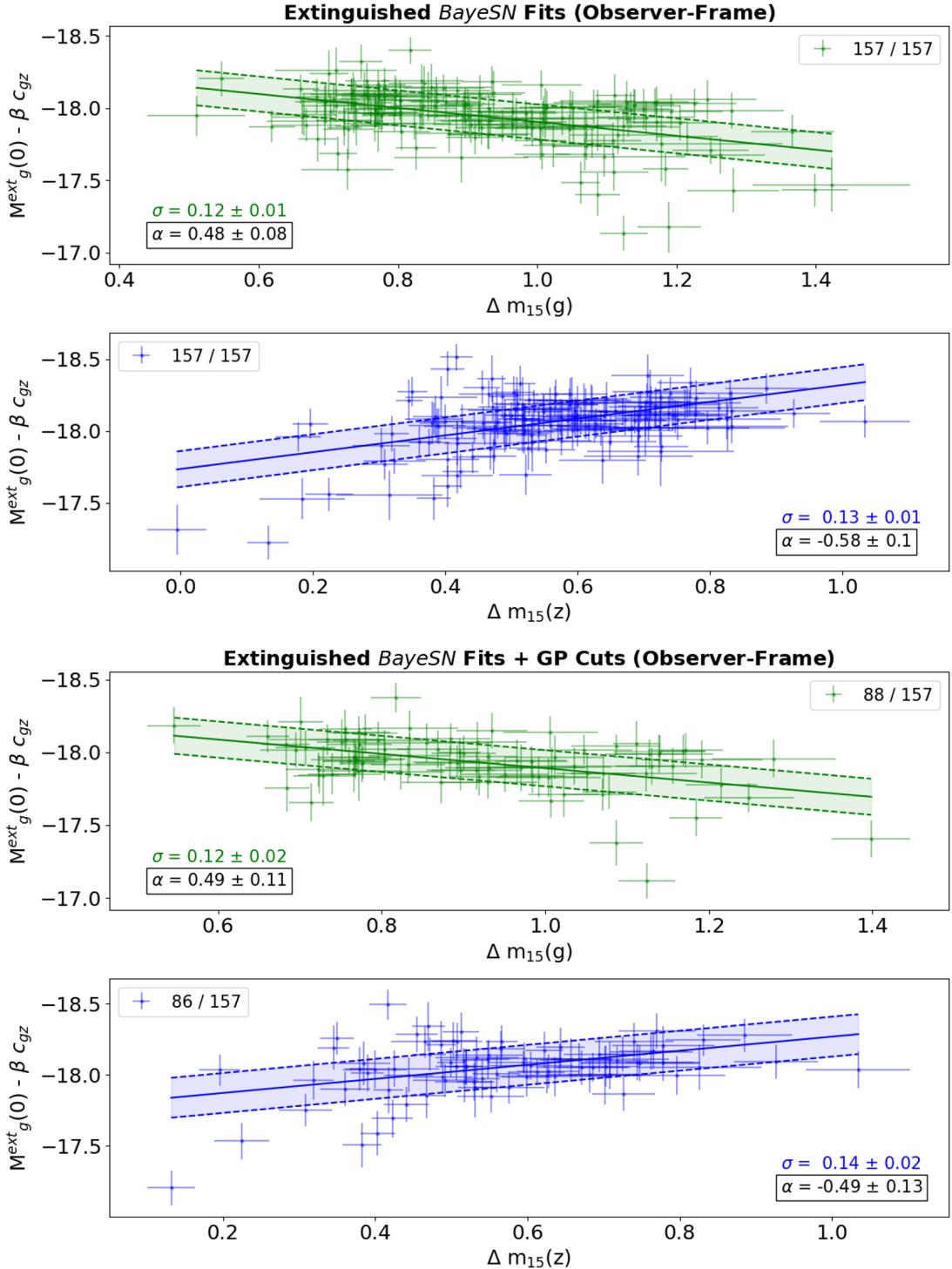


Fig. 6: Same as Fig. 5, but for *BayeSN* fits. Linear relations here, in tandem with Fig. 5, affirm bulk *BayeSN* trends are consistent with the data. **Upper:** *BayeSN* T21 model fits to all 157 Foundation DR1 SNe demonstrate learned trends, that optically brighter SNe are slow declining in the optical, but fast declining in the z -band. **Lower:** Same as upper panel, but SN sub-sample matches that following GP rejection conditions. Intrinsic scatter and α -slopes are all 1σ consistent across *BayeSN* and GP fits.

Fig. 3 shows a z -band inversion of the optical Philips relation learned by *BayeSN*. This is investigated by adapting the Tripp formulation (Eq. 7, Tripp (1998)), where we substitute magnitudes and colours for generic bands (X, Y, Z), and a decline-rate term in the generic Q -band $\Delta m_{15}(Q)$ measured from 0 to 15 days:

$$M_X^{ext}(t) = M_0 + \alpha \Delta m_{15}(Q) + \beta c_{YZ}^{app}(t) + \epsilon_{int}. \quad (20)$$

The sign change on the colour term - relative to Tripp - makes β positive. Results in Figs. (5, 6) show the z -band inverse-Philips-like trend (Fig. 3) found in Thorp et al. (2021) is affirmed independently by GP fits.

2.3.3 Feature Comparison & Hyper-parameter Dependence

With bulk *BayeSN* trends affirmed by the data, features are now directly compared, on a supernova-by-supernova basis. Features for comparison are peak *griz* magnitudes, and decline-rates, Δm_{15} . Quantitative comparisons of a feature, $\Delta x = x_{GP} - x_{BayeSN}$, are made by computing the precision weighted mean difference and associated error, (x_w, σ_w) , over all the accepted $\{x_i\}_{i=1}^N$ supernova features, where σ_i is the quadrature sum of the GP 1σ error envelope and the *BayeSN* 1σ credible interval.

$$w_i = \sigma_i^{-2} \quad ; \quad x_w = \frac{\sum_{i=1}^N w_i \Delta x_i}{\sum_{i=1}^N w_i} \quad ; \quad \sigma_w = \frac{1}{\sqrt{\sum_{i=1}^N w_i}}. \quad (21)$$

Results in Fig. 7 demonstrate g -band features are accurate, with discrepancies at a level $< 1\sigma$, whereas z -band features, in Fig. 8, are discrepant at $> 5\sigma$. This bias indicates light curve amplitudes at peak and trough are underestimated by the GP fits at $t = (0, 15)$ d.

Hyper-parameters are varied and resulting biases measured, to evaluate whether the GP fits - which have no knowledge of the SN latent process - are converging in such a way as to bias the fit. Bias dependence on the GP mean constant is negligible. Bias graphs for z peak magnitude and 15-day decline under varying natural-log amplitude, $\ln(A^2)$, and time-scale, $\ln(\tau^2)$, are depicted in Fig. 9. Results show that, while freezing τ^2 can reduce the bias and retain comparable sample size fractions, this fraction cannot be maintained while also reducing the bias to < 0.01 mag. These tests are also performed by freezing and offsetting the hyper-parameter distribution; results show a zero offset distribution is optimal.

Fig. 10 shows long cadence of z -band light curve sampling likely causes the GP to fail to capture extreme z -LC magnitudes, around peak and trough, resulting in the bias. This is shown using a kernel density estimate (KDE), i.e. by assigning a Gaussian at each data point, with standard deviation equal to Silverman's bandwidth, $bw = (4\hat{\sigma}^5/3N)^{1/5}$, with $(\hat{\sigma}, N)$ the sample standard deviation and size. With sampling characteristics (number of data points, typical phase-separation) the same across all 4-bands, it is to be expected that features of quickly varying z -band LCs are not as accurate as g -band features.

I visually inspect the 10 SNe from Fig. 8 with the most discrepant $\Delta m_{15}(z)$; an example SN is shown in Fig. 11. All of these supernovae exhibit the bias profile depicted in the lower panel, whereby peaks and troughs are under-estimated. It is the single-task nature of our Gaussian process fits, that lack population information, that causes the under-estimate: there is no incentive for the GP to peak higher at $t = 0$, or dip lower at $t = 15$. By contrast, as a hierarchical Bayesian model, *BayeSN* draws on the full SN Ia sample to build a model for z -band light curves, which is then fit to individual SNe Ia.

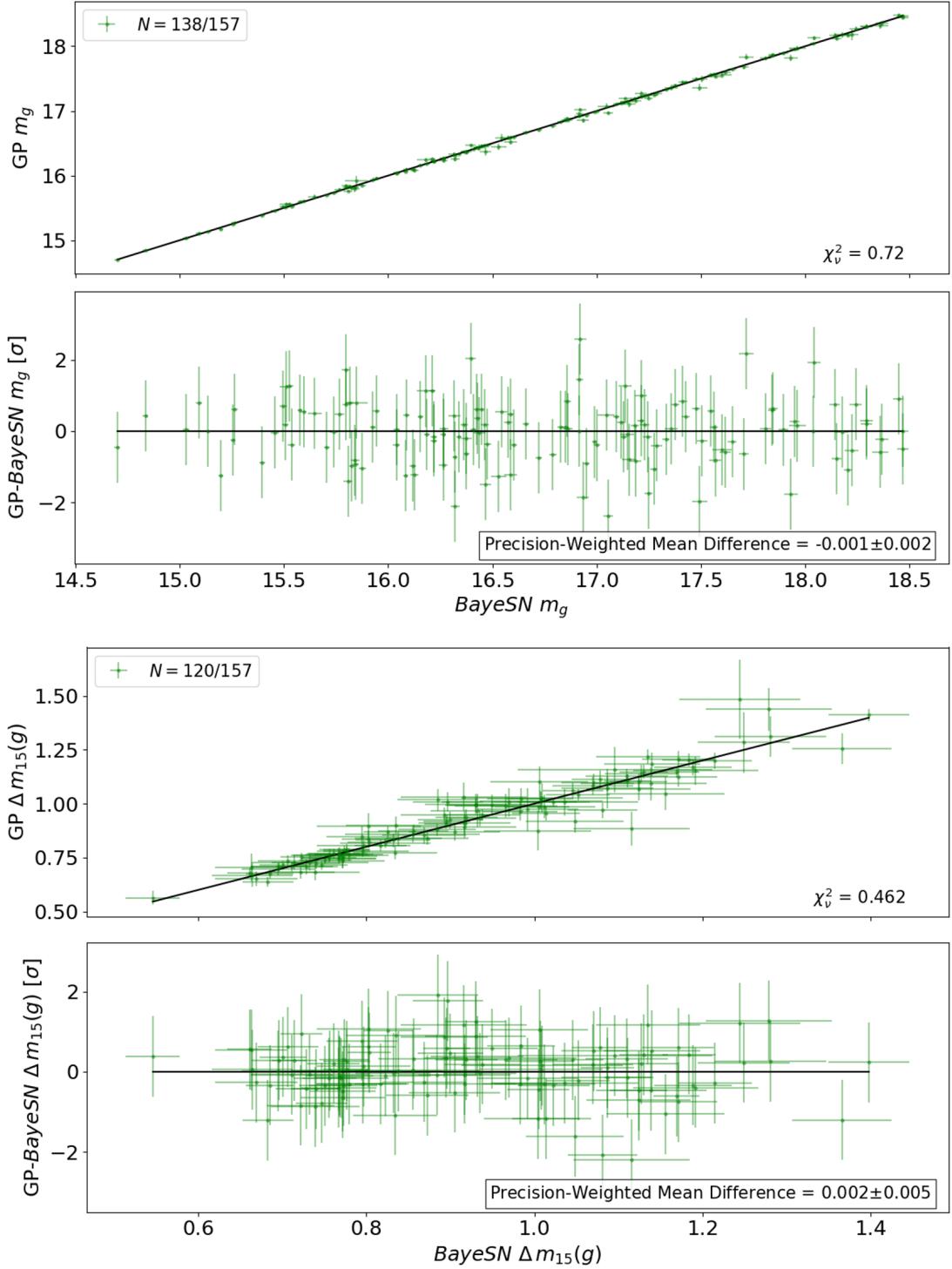


Fig. 7: Plots showcase comparisons of g -band extinguished observer-frame peak apparent magnitudes and decline-rates between *BayeSN* and Gaussian process fits. Features are accurate, with a precision weighted mean, as in Eq. 21, consistent with zero.

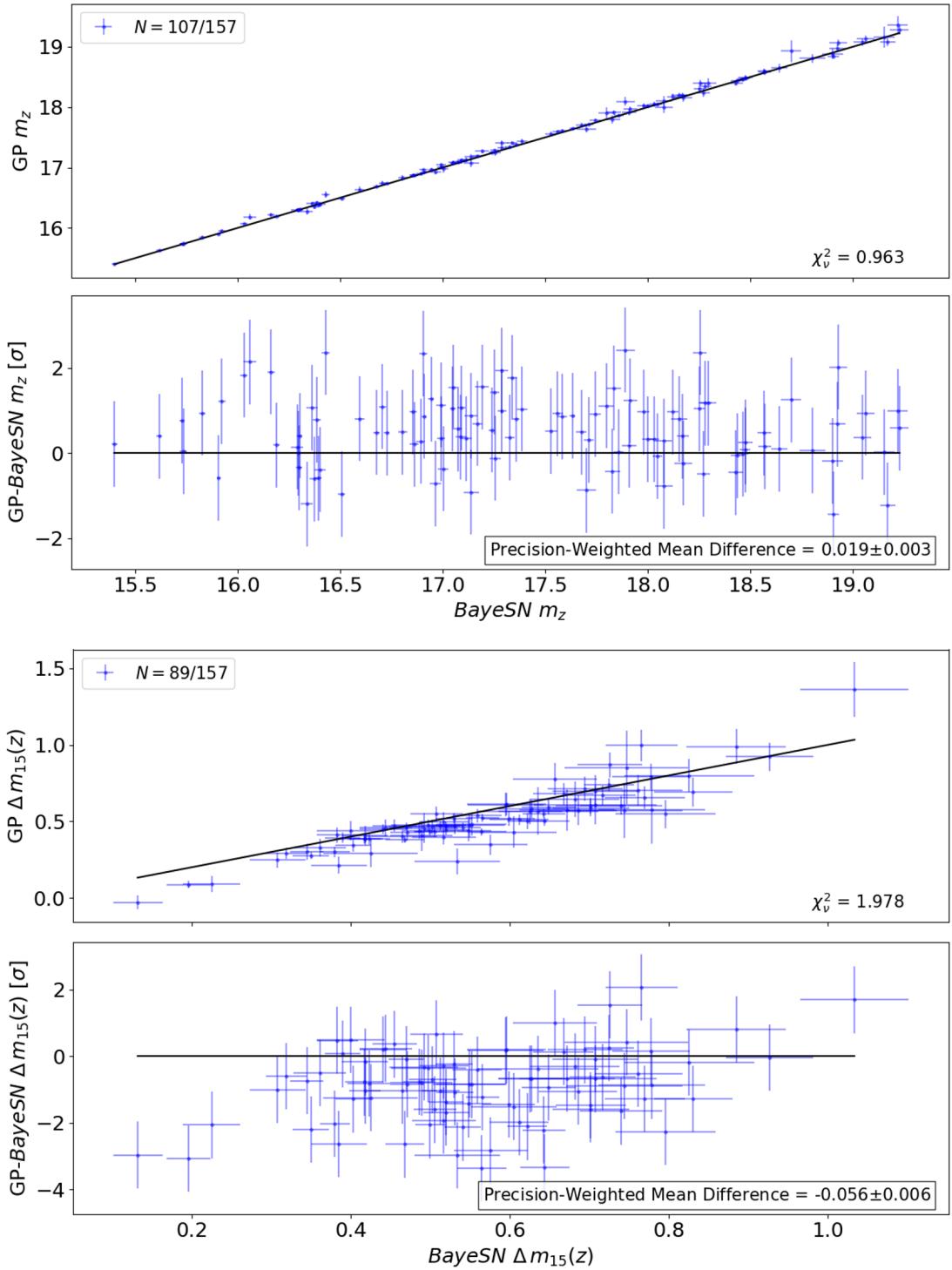


Fig. 8: Same as Fig. 7, for z -band fits. Features are biased, at a level $> 5\sigma$. Adopting $BayeSN$ fits as the null hypothesis, results indicate z -band GP fits aren't peaking high enough at $t = 0$, and aren't dipping low enough at $t = 15$.

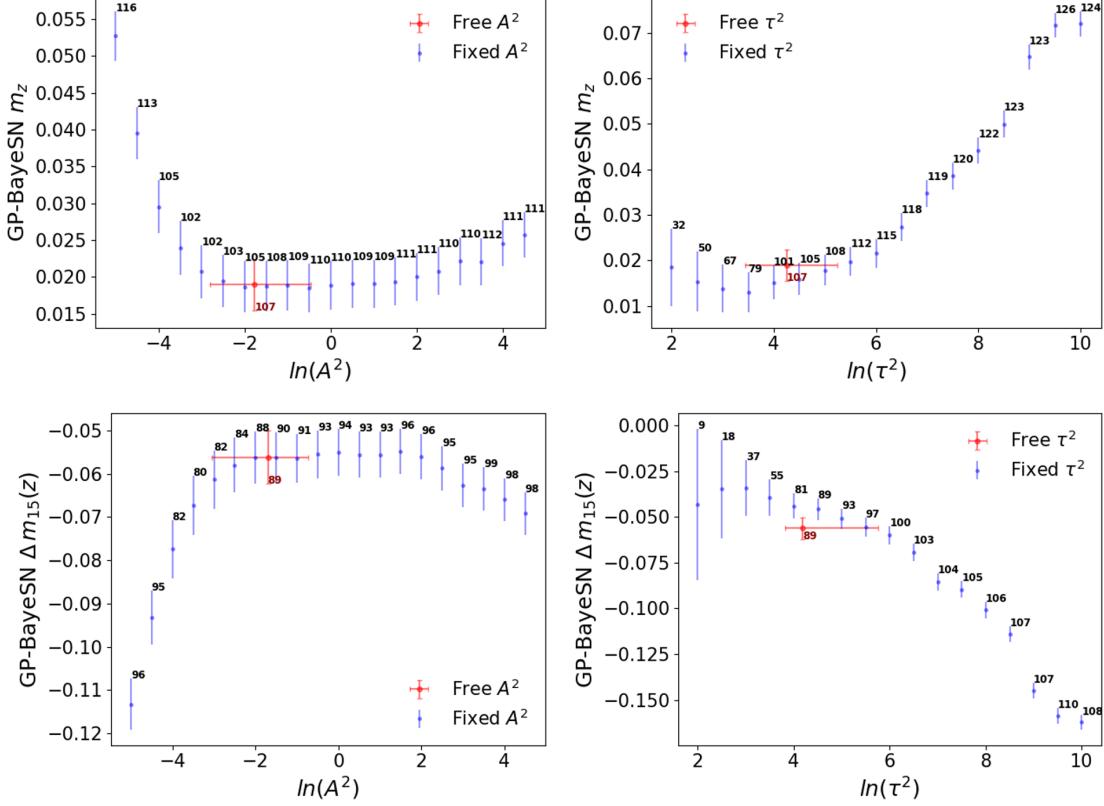


Fig. 9: **Left Column:** Bias on z -peak apparent magnitude and $\Delta m_{15}(z)$ is computed by freezing the RBF kernel amplitude, A^2 , before Gaussian process fits. Plotted at each data point is the number of SNe retained following rejection conditions. In red, is the sample mean and 68% confidence interval, estimated using a KDE, of the free amplitude GP fits. Plots demonstrate that bias does not improve by setting a universal amplitude. **Right Column:** Same as left column, but for varying characteristic timescale, τ^2 , in the RBF kernel. Here, we see GP fits benefit from a universal hyper-parameter, with a value close to the sample mean, striking a balance between minimising bias and retaining a large fraction of the 157 size sample. Nonetheless, bias is not reduced to $\lesssim 3\sigma$ without also reducing the accepted sample size substantially.

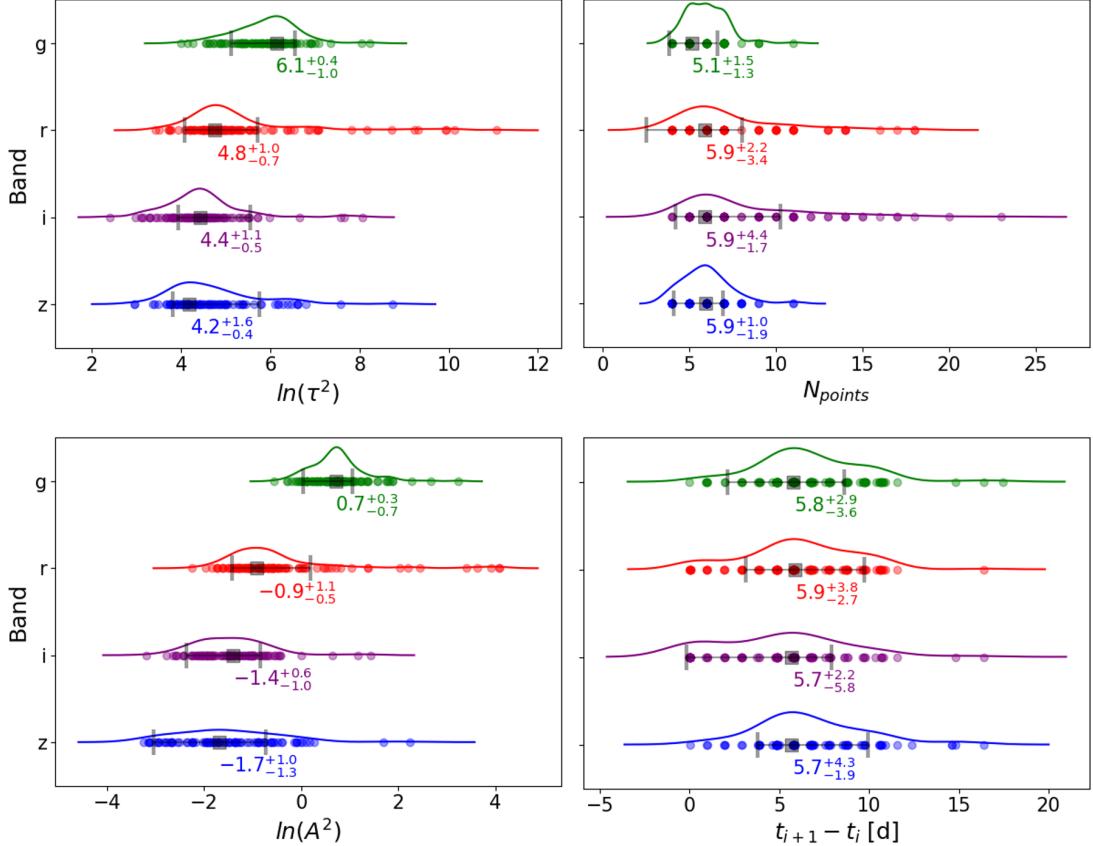


Fig. 10: **Left Column:** Kernel Density Estimates (KDEs) of the MLE distributions of GP RBF-kernel hyper-parameters (Eq. 11), subject to Δm_{15} rejections [$N_{kept;griz} = (120, 111, 96, 89)$]. Contrasting gz distributions, the z -band variations occur on shorter timescales, and are smaller. **Right Column:** Unlike the cross-band hyper-parameter distributions, the KDEs of the number of data points per light curve, N_{points} , and the average time separation between points, $t_{i+1} - t_i$, are strongly overlapping. Conclusively then, gz light curves are comparably sampled. Thus, low level of z -sampling *relative* to small variability on small time-scales can account for the bias in z -features in Fig. 8.

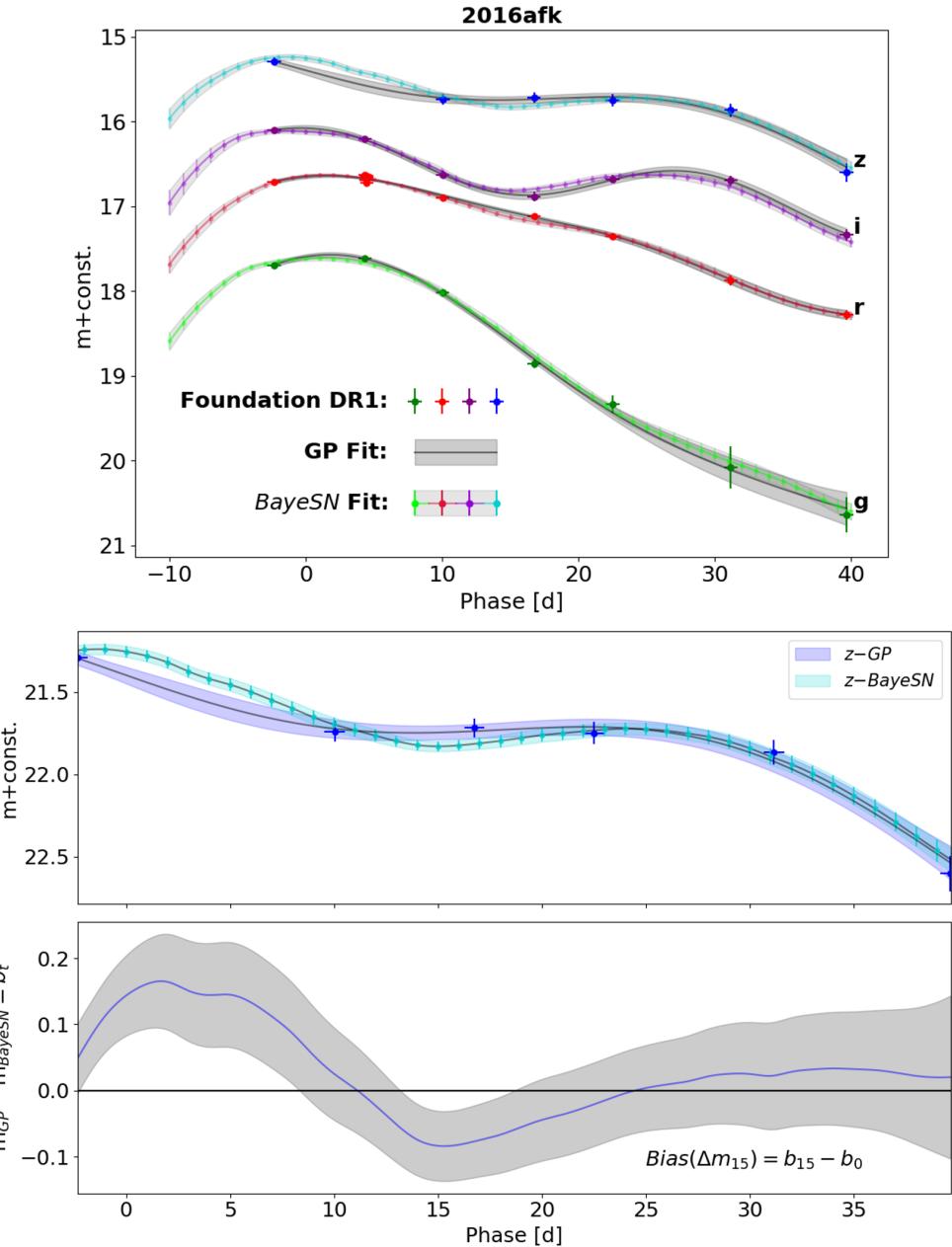


Fig. 11: SN 2016afk graphs of GP and *BayeSN* fits showcase an example of a highly discrepant $\Delta m_{15}(z)$, with a bias -0.228mag . The single-task Gaussian process does not utilise population information, and the available data does not pull the fit higher at peak, or lower at trough. This bias profile manifests in all 10 of the most discrepant features that are visually inspected.

2.3.4 Dust Laws across the Mass-Step

We briefly perform an empirical validation of the mass-step results reported in Thorp et al. (2021), that R_V distributions are likely consistent between low and high stellar-mass hosts (contrary to Brout & Scolnic (2021); for more details see § 1.3.3). We do this using Gaussian processes fitted to Foundation DR1, following the methods outlined in this chapter.

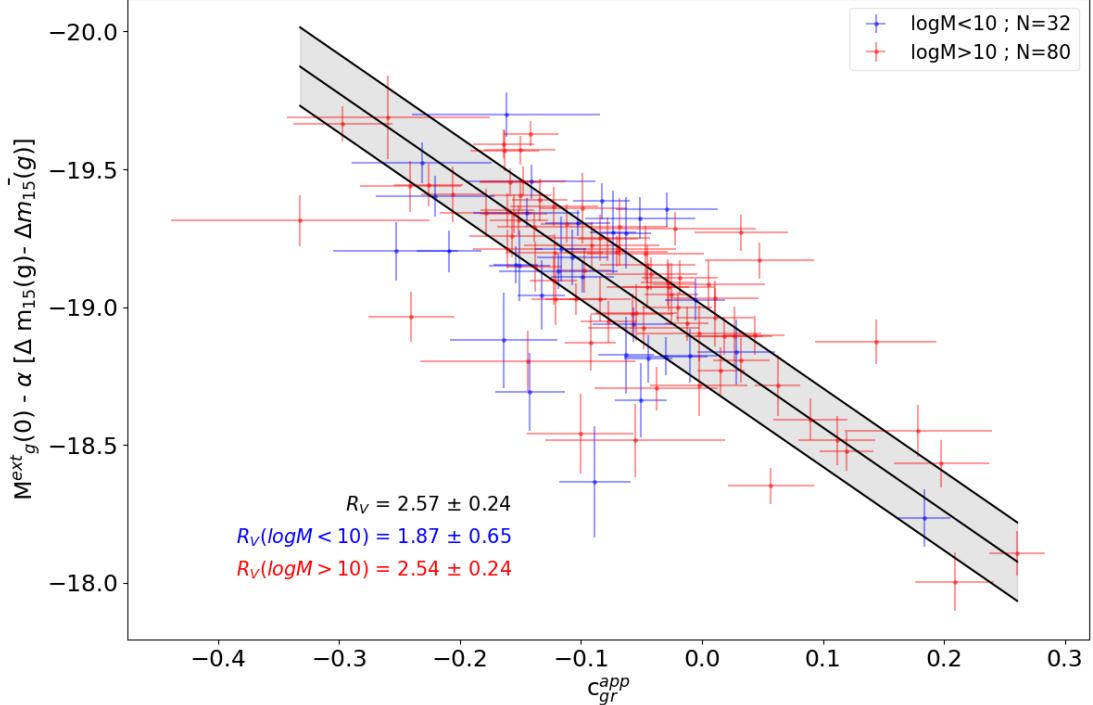


Fig. 12: Fitting to Philips-standardised observer-frame g -band extinguished magnitudes (using the sample mean decline-rate, $\Delta\bar{m}_{15}(g)$ Phillips (1993)), as a function of apparent colour $c_{gr}^{app} = m_g^{ext}(0) - m_r^{ext}(0)$, is plotted, where slope parameter roughly indicates the mean population R_V (adopting various assumptions [see text]). Magnitudes are extracted/rejected according to the procedures detailed in this Chapter. R_V is fitted to the 112 size sample, and to 32:80 size low/high mass sub-samples split at $\log_{10}(M_*/M_\odot) = 10$. Global fit [black/grey band] is dominated by the high mass sub-sample, and R_V 's are consistent with each other, and with $T21$ (both within $\sim 0.1\sigma$). Best fit low mass sample has a lower R_V ($\sim 1.3\sigma$ consistency with $T21$); owing to the small sample size, and a tight concentration of data along the covariate axis (i.e. very few SNe from low-mass hosts with $c_{gr}^{app} > 0$), the uncertainty in the low mass gradient, ± 0.65 , is large, and 1σ consistent with global/high-mass R_V . This simple experiment does not strongly evidence a mass-dependent R_V .

We fit a linear relation of extinguished absolute g -magnitude, M_g^{ext} , standardised by light curve shape, $\Delta m_{15}(g)$, and apparent colour, c_{gr}^{app} (following methods in § 2.3.1, 2.3.2). We fit for a gradient β_{gr} in Stan on the entire sample, and a individual sub-samples split at $\log_{10}(M_*/M_\odot) = 10$; Fig. 12.

Assuming all supernovae in a sub-sample have a common R_V , and no intrinsic colour dispersion, scatter about the best-fitting linear relation is interpreted as a residual distribution in

intrinsic magnitudes. Adopting these assumptions, we write $\beta_{gr} = A_g/(A_g - A_r)$, and compute R_V via 1:1 linear mapping using the [Fitzpatrick \(1999\)](#) law. To construct the mapping, rather than integrating the effect of the dust law on the SED under the passband, a first order approximation of (A_g, A_r) evaluates [Fitzpatrick \(1999\)](#) at the passband central wavelengths: $(\lambda_g, \lambda_r) = (4900, 6200)\text{\AA}$, which is valid given the dust law is linear over the passband width ($\Delta\lambda \sim 1000\text{\AA}$).

Fig. 12 shows there is no significant evidence for a mass-dependent R_V . If we focus only the best fit values, we tentatively find (at low significance) that R_V is lower in low mass galaxies. This is opposite to the trend claimed in [Brout & Scolnic \(2021\)](#), who propose that if R_V variations are responsible for the mass-step, a lower R_V is required in *higher* mass hosts. This analysis is preliminary; while we make no conclusive statements on the dependence of R_V on host stellar mass, this work is consistent with the findings of [Thorp et al. \(2021\)](#), and hard to reconcile with [Brout & Scolnic \(2021\)](#). In reality, there is a mixture of effects, namely, a locus of intrinsic magnitudes and colours, *and* a distribution of R_V , which should be modelled in future work.

3 Velocity-SED Correlations

In this section, we investigate whether velocity measurements of the Silicon-II line ($\lambda_{SiII} = 6355\text{\AA}$) - a trademark spectral feature of SNe Ia - contains valuable SED information, that can be used to improve distance estimates, for constraints on cosmological parameters. This work is ongoing (§ 5.4).

3.1 The Velocity-Colour Relation

A population variation in the velocity of SN explosion ejecta ($\sim 10^4 \text{ km s}^{-1}$) is found empirically. Heavy elements such as silicon are synthesized in the SN Ia thermonuclear explosion, and exhibit strong spectral features, which are used to measure blue-shift and estimate velocity e.g. the silicon-II absorption line (rest-frame wavelength $\lambda_{SiII} = 6355\text{\AA}$). [Foley & Kasen \(2011\)](#) used the Si-II line to estimate the ejecta velocity of 121 SNe Ia, permitting a sample split into normal and high velocity groups, (v_N, v_H) respectively, at a cut $-11,800 \text{ km s}^{-1}$ (following [Wang et al. \(2009\)](#)). By computing an intrinsic ($B - V$) colour with *SALT2*, a ‘redder-faster’ relationship is reported, that larger (more blue-shifted/negative) velocity SNe have redder (more positive) colours. Informed by this velocity-colour relation (VCR), the Hubble diagram scatter is reduced by allowing different intrinsic colours for different velocity groups, and when utilising only the normal velocity sample ($\sim 0.05\text{mag}$ reduction in ‘weighted residual scatter’). A more recent study also finds Hubble residual dependencies on velocity ([Siebert et al., 2020](#)), where two samples are analysed: 100 SNe Ia from the Carnegie Supernova Project (CSP, $BVgriYJH$, 66 overlapping with *M20*-fitted SNe Ia; [Hamuy et al. \(2006\)](#); [Contreras et al. \(2010\)](#); [Folatelli et al. \(2010\)](#); [Stritzinger et al. \(2011\)](#); [Kosciunas et al. \(2017\)](#)), as well as the second sample from [Wang et al. \(2009\)](#); [Foley & Kasen \(2011\)](#). They report a ‘velocity-step’ of $0.091 \pm 0.035\text{mag}$, with higher velocity SNe appearing over-luminous (negative Hubble residuals). These studies imply velocity information can be utilised to improve distance estimates.

[Foley & Kasen \(2011\)](#) provide a possible theoretical explanation for this VCR. For increased ejecta velocity, there is an increase in velocity *dispersion*, which leads to broader absorption lines, and thus a dimmer flux at the line location. Because the B -band has many more absorption lines from metals as compared to the V -band, it is expected higher velocity SNe should have *redder intrinsic colours* in optical $B - V$. Thus, combining measurements of apparent colour and ejecta velocity has the potential to help disentangle intrinsic colour from dust, thereby improving distance estimates for constraints on cosmological parameters. An open question then is how ejecta velocity correlates with intrinsic colour, and whether velocity information can be used to improve distance estimates.

Indeed, the existence of a VCR is still debated. For example, [Dettman et al. \(2021\)](#) search for VCRs in a variety of SN Ia samples from different surveys, observed in different passbands. In particular, they analyse 79 SNe Ia from the Foundation sample ($griz$, 72 overlapping with *T21*; [Foley et al. \(2018\)](#); [Jones et al. \(2019\)](#)), as well as the other two samples above from [Siebert et al. \(2020\)](#). For the Foundation sample, they find no significant *SALT2* colour offset between normal/high velocity sub-samples split at $-11,800 \text{ km s}^{-1}$, with $\Delta c = 0.005 \pm 0.014$, and a non-significant velocity step in the Hubble residuals: $0.015 \pm 0.049\text{mag}$. However, the other two surveys, and a combined set, all yield colour offsets $> 1\sigma$, the latter with $\Delta c = 0.017 \pm 0.007$.

The origin of these apparently survey-dependent VCRs is not clear. [Dettman et al. \(2021\)](#) point towards differences in colour offset results when using *SALT2*-fitted colour versus peak $B - V$ colour methodologies. The former tends to weaken colour offsets, implying velocity information encoded in photometry may have already been learned and built in to *SALT2*. They note also the Foundation survey is ‘untargeted’, therefore more likely to discover and observe SNe

in lower stellar-mass hosts, whereas the other two ‘targeted’ surveys select higher mass hosts, which typically host brighter SNe; thus the surveys may be probing different populations, which leads to the observed VCR differences between the surveys. To conclude then, whether velocity information can be used to improve distance estimates is still an open question.

3.2 *BayeSN* Methodology

BayeSN outputs can be used to investigate velocity colour relations in a robust way, with a rigorous treatment of dust, and employing the first continuous NIR SN Ia SED model. I take velocity measurements from Dettman et al. (2021), and analyse 66/100 SNe Ia from the CSP sample, and 72/79 SNe from Foundation DR1 (as listed above).

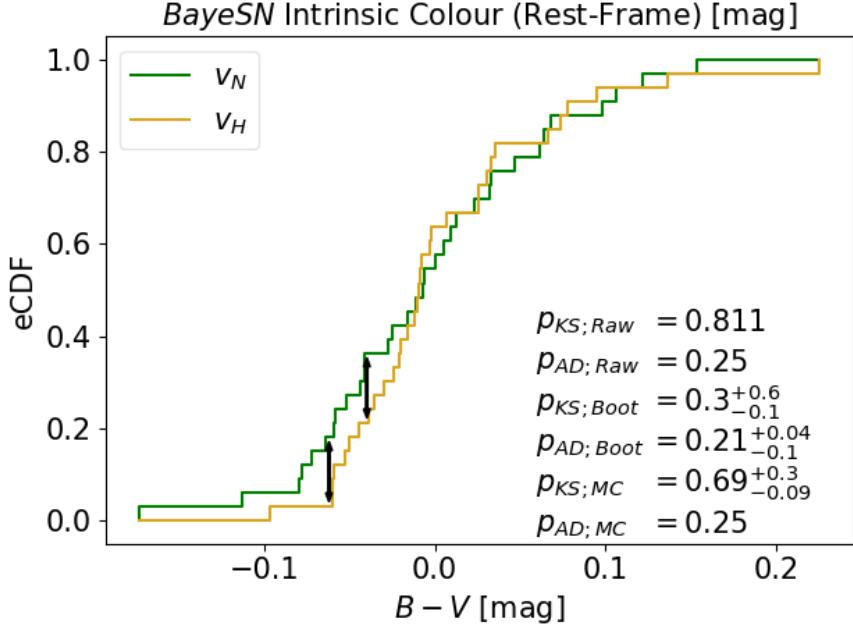


Fig. 13: The empirical cumulative distribution function (eCDF) is used to visually inspect population differences in intrinsic $B - V$ colours between velocity sub-samples from CSP, at a 33:33 split on the median velocity $-11,000 \text{ km s}^{-1}$. *Raw*, *Boot*, and *MC* p -values are respectively: the standard KS/AD values, KDEs of a bootstrap replacement distribution, and of a Monte Carlo distribution, assuming Gaussian errors. All p -values, even those computed using the *MC* method - that will over estimate confidence intervals - are $\gg 0.01$. Thus, there is no evidence for a velocity step in the intrinsic $B - V$ colours (at the median velocity). This result also holds for the apparent $B - V$ colours.

Using posterior estimates, and separating by the median velocity, p -values from Kolmogorov-Smirnoff (KS) and Anderson-Darling (AD) tests can be computed, to assess whether feature sub-samples are drawn from the same population. Computing KS/AD p -values with datasets *including error bars* is an open problem, and we trial 3 methods. As well as the raw p -values (*Raw*), we compute a distribution of p -values using bootstrap resampling to synthesis 1000 datasets, and use a KDE to estimate 68% confidence intervals (*Boot*); this method ignores error-bars. The final method assumes Gaussian errors, and draws 1000 random datasets from a normal

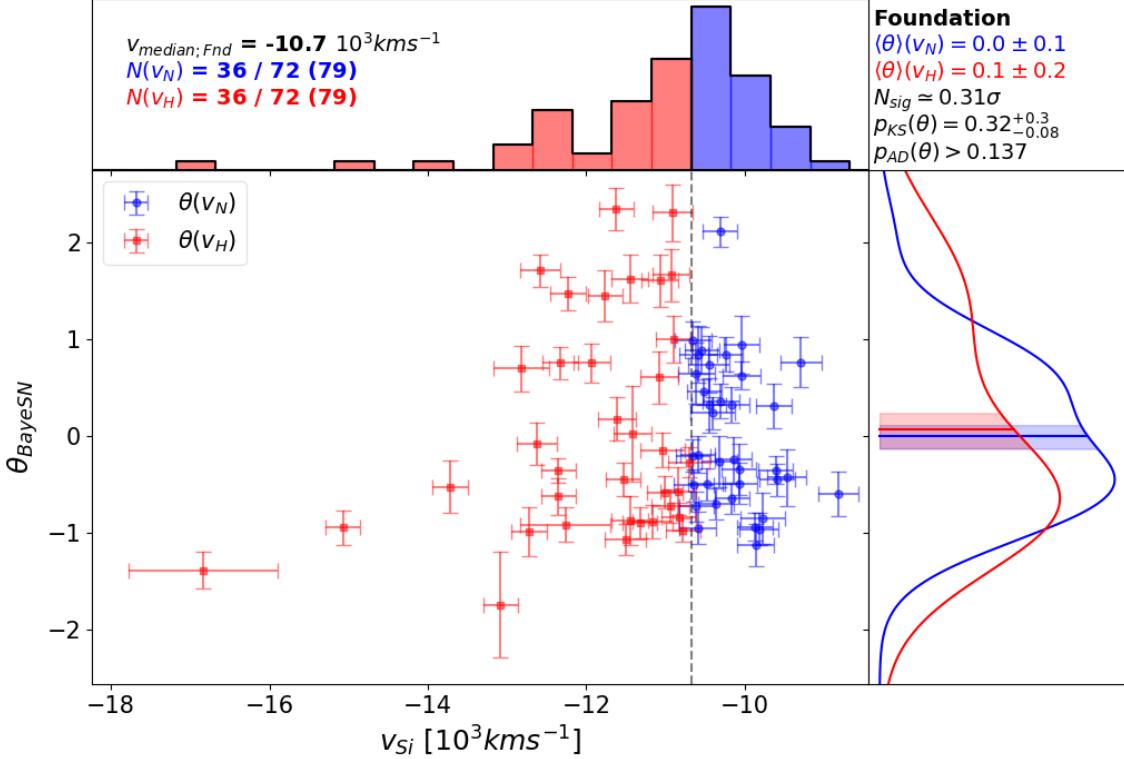


Fig. 14: Plot of the primary intrinsic SED shape parameter (the first FPC score), θ_{BayeSN} , against Silicon-II velocity, from the untargeted Foundation DR1 sub-sample fitted with the *T21 BayeSN* model. Right panel KDEs indicate the shape of the FPC distributions. Upper right p_{MC} values are $\gg 0.01$, indicating all FPC scores are drawn from the sample underlying population distribution. More in Fig. 15.

distribution, and again computes a KDE confidence interval (MC). However, this final method assumes samples are *un-correlated*, whereas *BayeSN* outputs are correlated: consequently, this method will overestimate the confidence interval. In future, we will use samples from the MCMC chain, and compute a p -distribution using these correlated ‘errors’ ($MCMC$).

3.3 Preliminary Results

Following preliminary work, we find no evidence for a $v_{SiII} - (B - V)$ relation (neither in intrinsic nor apparent colours, Fig. 13), when splitting at the sample median velocity, $-11,000 \text{ km s}^{-1}$. Given that *SALT2* confounds intrinsic colour with dust effects, this result is not necessarily surprising. Future work can consider a higher velocity split at $-11,800 \text{ km s}^{-1}$, however, care must be taken to rigorously account for the ‘look-elsewhere effect’, whereby a systematic search for an optimal value (in this case, a velocity-cut that minimises p) will invariably lead to a stationary point, that could be a statistical fluke. There is a weak indication of distinct velocity populations in the primary intrinsic SED shape parameter, θ_{BayeSN} , in the CSP sample, but not in the Foundation sample, which, interestingly, mimics Dettman et al. (2021) findings (§ 3.1; Figs (14, 15)). This will be investigated in future work, by considering p_{MCMC} and survey selection effects.

I perform a systematic search of velocity/single-feature correlations, testing colours, Hubble residuals, extinction, host-mass and (observer-frame) times of z -band first/second maximum and dip, and decline-rates. Work performed for a Hubble Space Telescope proposal motivates

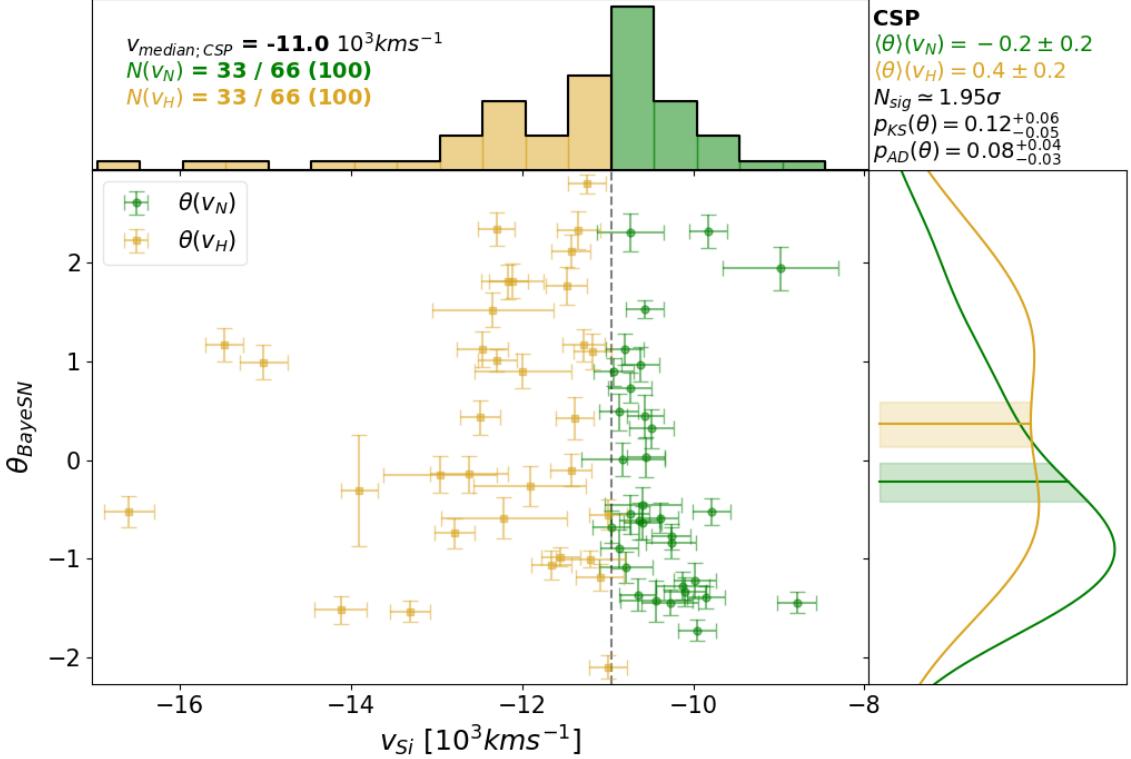


Fig. 15: Same as Fig. 14, but for the targeted CSP sample. Here, p -values are distinctly lower, and the KDE means are $\sim 2\sigma$ discrepant. Future work must analyse these distributions further, using p_{MCMC} , and evaluate the selection effects at play.

the use of near-infrared time-information to investigate velocity correlations. For this proposal, I found that NIR times of dip and second maximum are strongly correlated with the first functional principal component. Moreover, following systematic multiple linear regression fits of SN features (utilising tools/methods from § 2.3.2), I found that magnitude dispersion is minimised when NIR time information is exploited (see Appendix C).

Preliminary results indicate intrinsically distinct populations in $B - H$ colours, exhibiting the ‘redder-faster’ relation (Fig. 16). Other interesting correlations appear to emerge when using time-dependent declines, e.g. $\Delta m_{t_{dip}}(r)$ with $p_{KS;MC} = 0.02 \pm 0.01$. This Chapter indicates velocity information can be used to discriminate intrinsic colour populations, and in turn improve SN standardisation for enhanced distance estimates. This work is ongoing (see § 5.4).

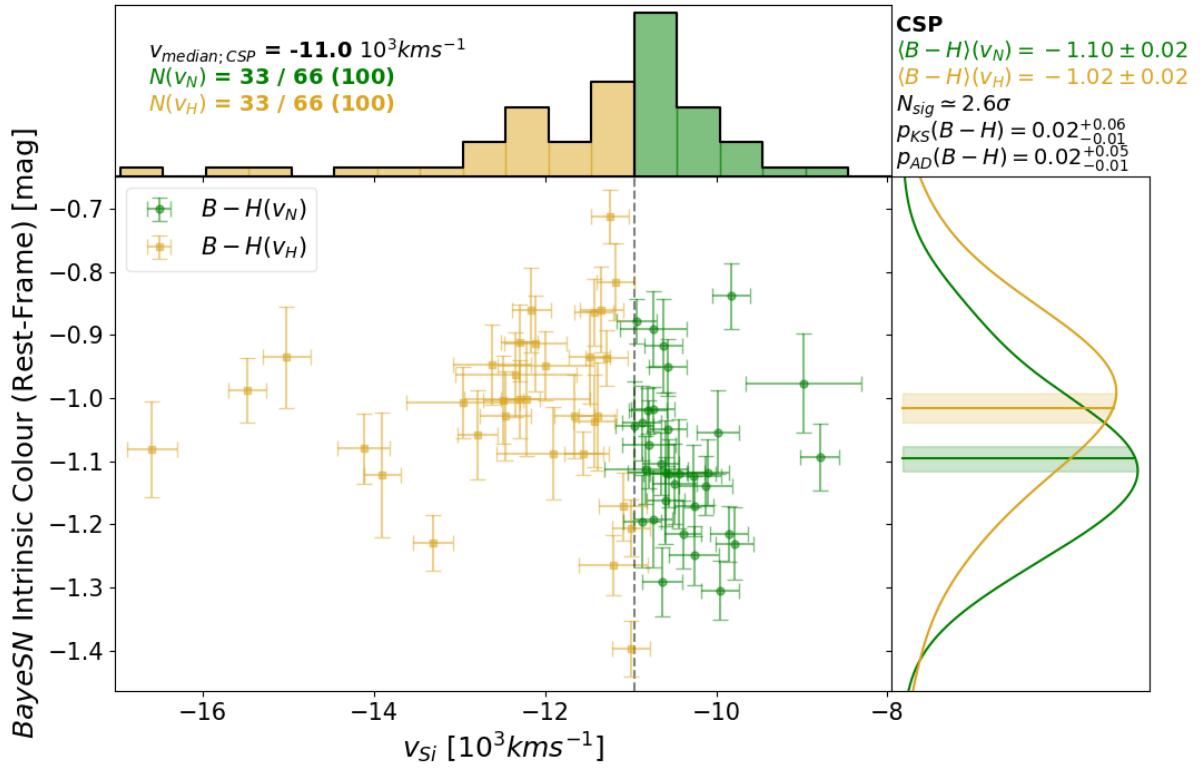


Fig. 16: Plot showcases a preliminary detection of a redder-faster relation in the intrinsic $B - H$ colours of the CSP sample. More in Fig. 17.

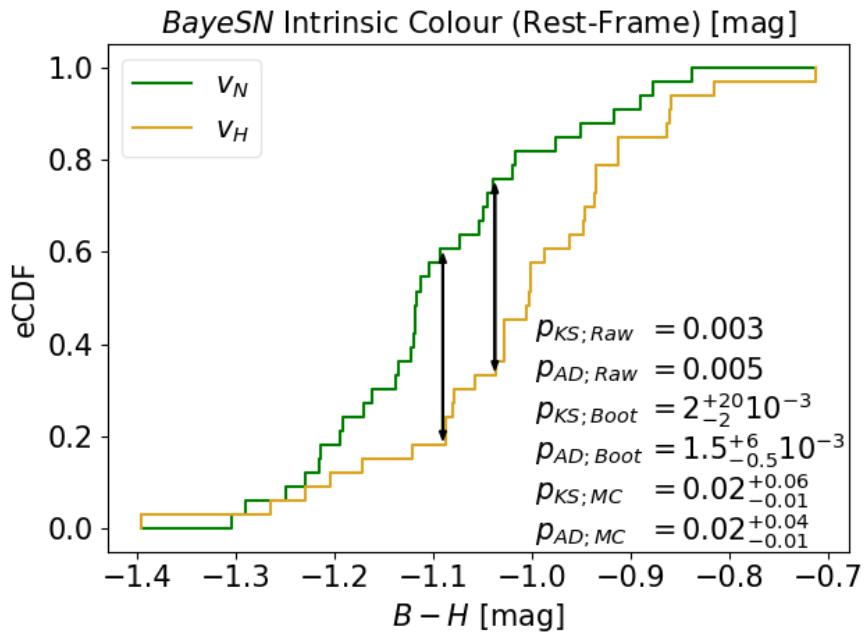


Fig. 17: Empirical CDF demonstrates the novel detection of a normal/high velocity population split in the intrinsic $B - H$ colours. Future work must use *BayeSN* MCMC chain samples.

4 Cepheid Amplitudes Encode Crowding Bias

Efstathiou (2020) demonstrates that a systematic offset of $\Delta M_{E20} \sim 0.1 - 0.14$ mag in Cepheid Wesenheit magnitudes would resolve the Hubble tension⁵ between results of CMB and Cepheid-Supernova analyses (Planck Collaboration et al., 2020b; Riess et al., 2019). The Wesenheit magnitudes are designed to be equal in the intrinsic and extinguished frames (Madore, 1982):

$$m_H^W = m_H - \tilde{R}(m_V - m_I), \quad (22)$$

where $\tilde{R} \equiv \frac{A_H}{A_V - A_I} \simeq 0.3 - 0.5$ is weakly dependent on R_V (Riess et al., 2016). It is argued that the crowding bias may be responsible, whereby multiple stellar sources residing in the smallest resolution element containing the Cepheid, may lead to a mis-estimate of the variable star brightness. Riess et al. (2020), hereafter R20, exploit the reduction in Cepheid light curve fractional amplitude in the presence of crowding to quantify the bias, and determine a crowding bias mis-estimate, $\Delta M_{R20} = -0.029 \pm 0.037$ mag, as compared to the bias inferred by employing the typical method of artificial star injection (Riess et al., 2016). R20 conclude these systematic errors “do not provide a plausible resolution to the Hubble tension”. However, what they have computed is a *multi-band* crowding bias mis-estimate *difference*, $\Delta M_H - \Delta M_V$, which should not be compared to the genuine crowding bias mis-estimate; $\Delta M_{E20} \equiv \Delta M_H - \tilde{R}(\Delta M_V - \Delta M_I)$. Moreover, an assumption of flux light curve symmetry about the mean is assumed throughout. Here, I show that removing this flux-symmetry simplifying assumption can lead to a non-negligible correction term. The size of this correction term depends on the complex interaction of $(F_R, \delta_{min}, \delta_{max})$ [see below]. Further study is warranted (§ 5.6).

4.1 Crowding Setup

In setting up the problem, I follow R20, and denote the total flux, F_{tot} , as

$$F_{tot}(t) = F_0(1 + \delta(t)) + F_1, \quad (23)$$

where the first term is the Cepheid flux, $F_{Ceph}(t)$, and F_1 is the constant stellar background flux within the resolution element. Averaged over time:

$$\langle \delta \rangle = 0 \quad ; \quad \langle F_{Ceph} \rangle = F_0 \quad ; \quad \langle F_{tot} \rangle = F_0 + F_1. \quad (24)$$

The time-independent parameter of interest is the crowding bias in magnitudes, Δm :

$$m_0 - m_{tot} = \Delta m = 2.5 \log_{10}(1 + F_R), \quad (25)$$

equivalent to,

$$10^{0.4\Delta m} = 1 + F_R, \quad (26)$$

where I denote F_R as the flux ratio, F_1/F_0 . From Eq. 25, we see correctly accounting for some significant crowding bias would yield a dimmer Cepheid flux, returning distance estimates that are further away, and thus a lower value of H_0 , in line with early-time results. The goal then is to measure F_R from crowded Cepheid light curves, by measuring the period, P , inferring the intrinsic amplitude, A , and comparing it to the observed amplitude, A' :

$$A = 2.5 \log_{10} \left(\frac{1 + \delta_{max}}{1 - \delta_{min}} \right) \quad ; \quad A' = 2.5 \log_{10} \left(\frac{1 + \delta_{max} + F_R}{1 - \delta_{min} + F_R} \right). \quad (27)$$

⁵Tension has since reduced from $\sim 4.3\sigma$ (Efstathiou, 2020), to $\sim 4.2\sigma$ (Riess et al., 2021).

4.2 A Compromising Approximation

The critical assumption in R20, that may significantly bias their results, is to assume $\delta_{min} = \delta_{max} = \alpha$. Adopting this assumption, and $\alpha + F_R < 1$, series expansions yield:

$$\frac{A'}{A} = \frac{\ln(\frac{1+(\alpha+F_R)}{1-(\alpha-F_R)})}{\ln(\frac{1+\alpha}{1-\alpha})} = \frac{\alpha}{(1+F_R)(\alpha+\beta)} + \frac{\beta}{\alpha+\beta} \approx \frac{1}{1+F_R} = 10^{-0.4\Delta m} \quad (28)$$

with $\beta = \sum_{n_{odd}=3}^{\infty} \frac{\alpha^n}{n} = \mathcal{O}(\alpha^3)$. However, on relaxing this assumption of symmetry, it is not obvious Eq. 28 holds. I introduce $\Delta m_{b;sym}$, the apparent crowding bias in a generic band, b , measured using the symmetry approximation of Eq. 28. This apparent bias is related to the true crowding bias, Δm , by a crowding bias correction, ΔM ;

$$\frac{A'}{A} = 10^{-0.4\Delta m_{b;sym}} ; \quad \Delta m = \Delta m_{b;sym} + \Delta M. \quad (29)$$

The performance is quantified by defining an error term, γ ;

$$\gamma(F_R, \delta_{min}, \delta_{max}) = \frac{A'}{A} - \frac{1}{1+F_R} = 10^{-0.4\Delta m_{b;sym}} (1 - 10^{-0.4\Delta M}). \quad (30)$$

If ΔM is $\sim 0.1 - 0.14$ mag, a robust crowding correction is significant enough to resolve the Hubble tension.

4.3 Parameter Space Evaluation

In exploring parameter space, I include reasonable estimates of where in parameter space the Cepheids are likely to reside. As noted in R20, Cepheids are optically bright, $L \sim 10^3 - 10^5 L_\odot$, and their flux dominates the resolution element: R20 cite Hoffmann et al. (2016) that $F_R(V) \sim 2\%$. Moreover, HST angular resolution is high in the optical (plate scale 0.04"/pixel) which minimises the number of crowding sources. However, Cepheids are dimmer in the NIR, and HST resolution poorer (0.13"/pixel), meaning crowding flux is $F_R(H) \sim 1$. On inspection of R20 Figure 5., crowding can be as high as $F_R \sim 3 - 5$ ($\Delta m_H - \Delta m_V \approx \Delta m_H \approx 1.5 - 2$, Eq. 26). A reasonable range in $F_R(V)$ is $0.01 - 0.05$, and for $F_R(H)$, I inspect the range $0.1 - 5$. Following R20, I take intrinsic amplitudes in V to be $A^V \sim \mathcal{N}(1, 0.2)$, and $A^H \sim \mathcal{N}(0.4, 0.1)$.

Given input parameters $(F_R, \delta_{min}, \delta_{max})$, the ratio of crowded to intrinsic amplitude is computed, and the crowding bias mis-estimate due to the assumption of asymmetry determined;

$$\Delta M = -2.5 \log_{10}(1 - \gamma 10^{0.4\Delta m_{b;sym}}) = 2.5 \log_{10} \left((1+F_R) \cdot \frac{A'}{A} \right). \quad (31)$$

Given typically observed amplitudes A , the profile in $(\delta_{min}, \delta_{max})$ space is:

$$\delta_{min} = 1 - (1 + \delta_{max}) 10^{-0.4A}. \quad (32)$$

Contours are plotted in Figs. (18, 19, 20).

V -Band symmetry correction is roughly surmised as,

$$-0.01 < \Delta M_{V;sym} \lesssim -0.001, \quad (33)$$

with hard bound negative values only achievable with high optical crowding ($F_R \sim 0.05$) and $\Delta\delta < 0$. H -Band correction is highly dependent on the true underlying flux-asymmetry, and could reasonably range,

$$|\Delta M_{H;sym}| \lesssim 0.03 \lesssim |\Delta M_{H;sym}| < 0.05, \quad (34)$$

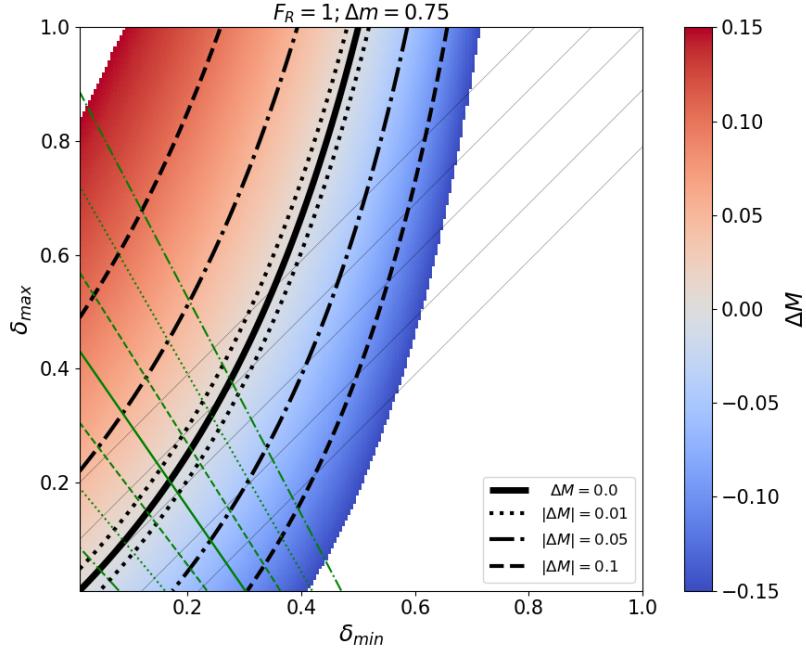


Fig. 18: H -band crowding bias mis-estimate, owing to the assumption of flux symmetry about the mean, is computed for input values $(\delta_{min}, \delta_{max}) \in [0.01, 1]$ at a crowding level $F_R = 1$. Colour contours are clipped at $\Delta M = 0.15$ mag. Black dot/dot-dash/dashed lines mark out ΔM contours, with the solid curve at $|\Delta M| = 0$. Green lines depict amplitudes, $A = A^H \sim \mathcal{N}(0.4, 0.1)$, with the solid line at $A^H = 0.4$, and dash/dot/dot-dash lines at $\pm[1, 2, 3]\sigma$. Fine black lines mark out asymmetry levels, $\Delta\delta = \pm[0, 0.1, 0.2]$. At this level of crowding, crowding bias of perfectly symmetric light curves is mis-estimated at $\Delta M \sim -0.01$ mag. For highly asymmetric H -band light curves, $\Delta\delta \approx 0.2$, this mis-estimate is roughly $\Delta M \approx 0.04$ mag.

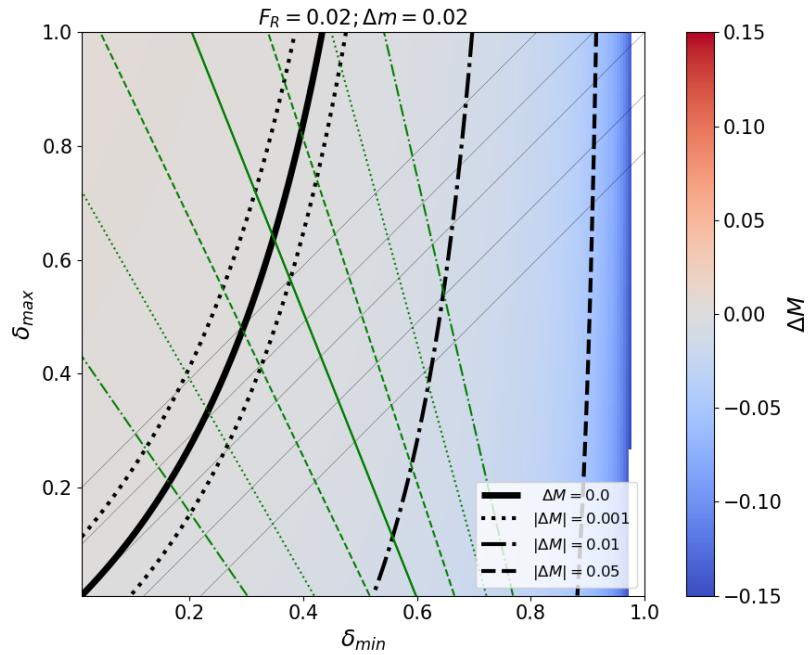


Fig. 19: Same as for Fig. 18, except for V -band light curves, with amplitudes $A^V \sim \mathcal{N}(1, 0.2)$, and $F_R = 0.02$. Here, for the mean amplitude under all $|\Delta\delta| \lesssim 0.2$, we have $-0.01 \lesssim \Delta M \lesssim -0.001$.

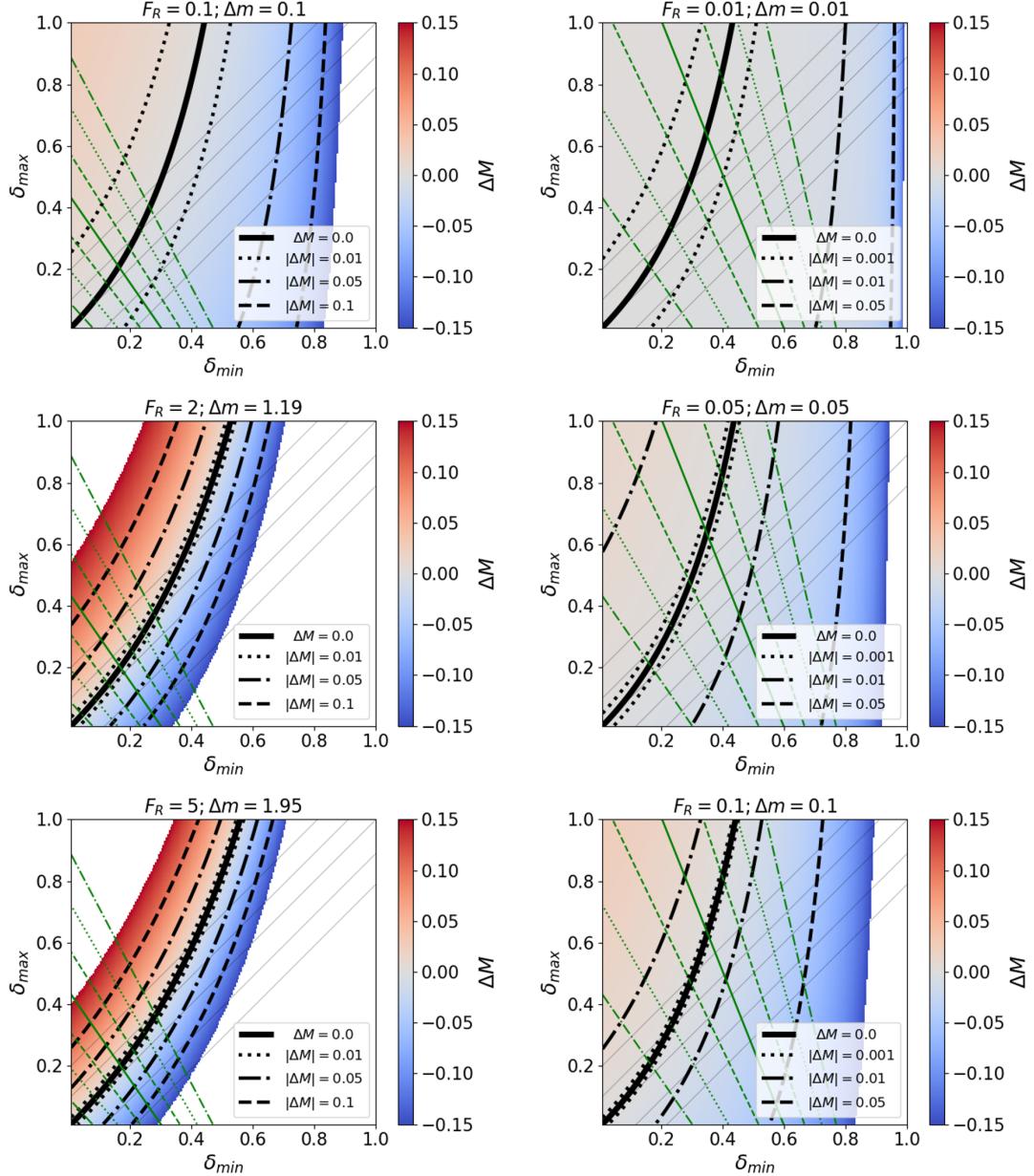


Fig. 20: Same as Figs. (18, 19), for wider ranges in F_R , with **Left** column denoting H -band amplitudes, and **Right** column denoting V -band. ΔM_H reasonably approaches ~ 0.05 for high crowding, provided only that typical H -band LCs are highly asymmetric: $\Delta\delta \simeq 0.2$. Interestingly, ΔM_H can tend to ~ -0.05 if $\Delta\delta \simeq -0.1$; in this direction, the Hubble tension would widen. ΔM_H does not reasonably approach 0.1mag, which would require $F_R > 1$ and $\Delta\delta > 0.2$.

with high corrections only possible with extremely asymmetric light curves, $\Delta\delta \sim 0.2$, and symmetric light curves returning corrections in the region $|\Delta M_{H;sym}| \lesssim 0.01$. Evaluating, I conclude the inclusion of asymmetry in the amplitude analysis cannot reasonably solve the Hubble tension, as this would require $F_R(H) \gtrsim 1$ and $\Delta\delta \gtrsim 0.2$ on average. However, this evaluation adopts a naive interpretation of R20 results (more in § 4.4).

4.4 Interpretation

Equation (15) in R20 is written in our notation as,

$$\chi^2(\Delta M_{R20}) = \sum_{i=1}^n \left(\frac{A'^H}{A'^V} - \frac{A^{H,MW}}{A^{V,MW}} 10^{-0.4(\Delta m_{H;AIS} - \Delta m_{V;AIS} + \Delta M_{R20})} \right)^2 \sigma_i^{-2} = \sum_{i=1}^n \frac{R_i^2(\Delta M_{R20})}{\sigma_i^2}, \quad (35)$$

where $\Delta m_{b;AIS}$ is the crowding bias estimate in the generic b -band using artificially injected stars (AIS), which I allow to differ from the true crowding bias by $\Delta M_{b;AIS}$, similar to Eq. 29. The Milky-Way amplitude ratio, denoted by MW , depends only on measured Cepheid period, and is summarised as a function $F(P)$. Thus, in complete notation, the residual function is,

$$R_i(\Delta M_{R20}) = F(P) 10^{-0.4(\Delta m_H - \Delta m_V)} \left(10^{0.4(\Delta M_{H;sym} - \Delta M_{V;sym})} - 10^{0.4(\Delta M_{H;AIS} - \Delta M_{V;AIS} - \Delta M_{R20})} \right). \quad (36)$$

Consider that $\Delta M_{R20} = -0.029 \pm 0.037$; under all R20 assumptions, $\Delta M_{H;sym} = \Delta M_{V;sym} = 0$, thus an R20 view of this result is: $\langle \Delta M_{H;AIS} - \Delta M_{V;AIS} \rangle = -0.029 \pm 0.037$. Even if all Cepheid flux light curves satisfy symmetry, R20 results allow for individually biased ΔM_{AIS} . Now consider that $\Delta M_{H;sym} - \Delta M_{V;sym} \neq 0$. In such a scenario, on average,

$$\langle \Delta M_{H;AIS} - \Delta M_{V;AIS} \rangle \approx \langle \Delta M_{H;sym} - \Delta M_{V;sym} \rangle - 0.029 \pm 0.037, \quad (37)$$

where I grossly assume $\langle f(x) \rangle \approx f(\langle x \rangle)$ for the χ^2 -minimisation (Eq. 35). If all $\Delta M = 0$, R20 conclusions are robust to scrutiny. If there are band-independent systematics for both AIS and symmetric-amplitudes methods, multi-band difference may cancel the common systematic. Considering Eq. 31, this is not possible, as it requires γ_H to *decrease* and counteract the increased $\Delta m_{H;sym}$ - relative to the V -band - to give comparable (V, H) crowding corrections on average. However, with increased crowding, γ necessarily *increases*; in all likelihood, $\langle |\Delta M_{H;sym}| \rangle > \langle |\Delta M_{V;sym}| \rangle$.

Ambiguities in interpreting R20 results also arise from: (1) The use of χ^2 -minimisation, rather than an MLE or Bayesian approach, which may bias slope parameters (e.g. Keller et al. (2007)), (2) Our assumption of $\langle f(x) \rangle = f(\langle x \rangle)$, which likely neglects an offset in Eq. 37, (3) The constrained R20 functional form for measuring H -band amplitudes: a sine wave, which may lead to A'^H underestimates (e.g. Figure 4., R20 Table. 3 amplitude measure equal -0.03mag), (4) The R20 inclusion of high-crowding sources, that would normally be excluded in a cosmological analysis (Riess et al., 2020), which may bias the χ^2 results away from an equivalent ‘cosmology sample’ analysis (Eq. 35). Note also the I -band is ignored entirely.

4.5 Artificial Star Injection

Where could a crowding bias mis-estimate be introduced in the artificially injected stars method (AIS), employed in e.g. Riess et al. (2009, 2011, 2016), and an independent study Javanmardi et al. (2021), hereafter J21? Both authors compute a crowding bias by *averaging a set of crowding bias estimates*. For Riess, they take the mean of means across the entire galaxy (which as J21

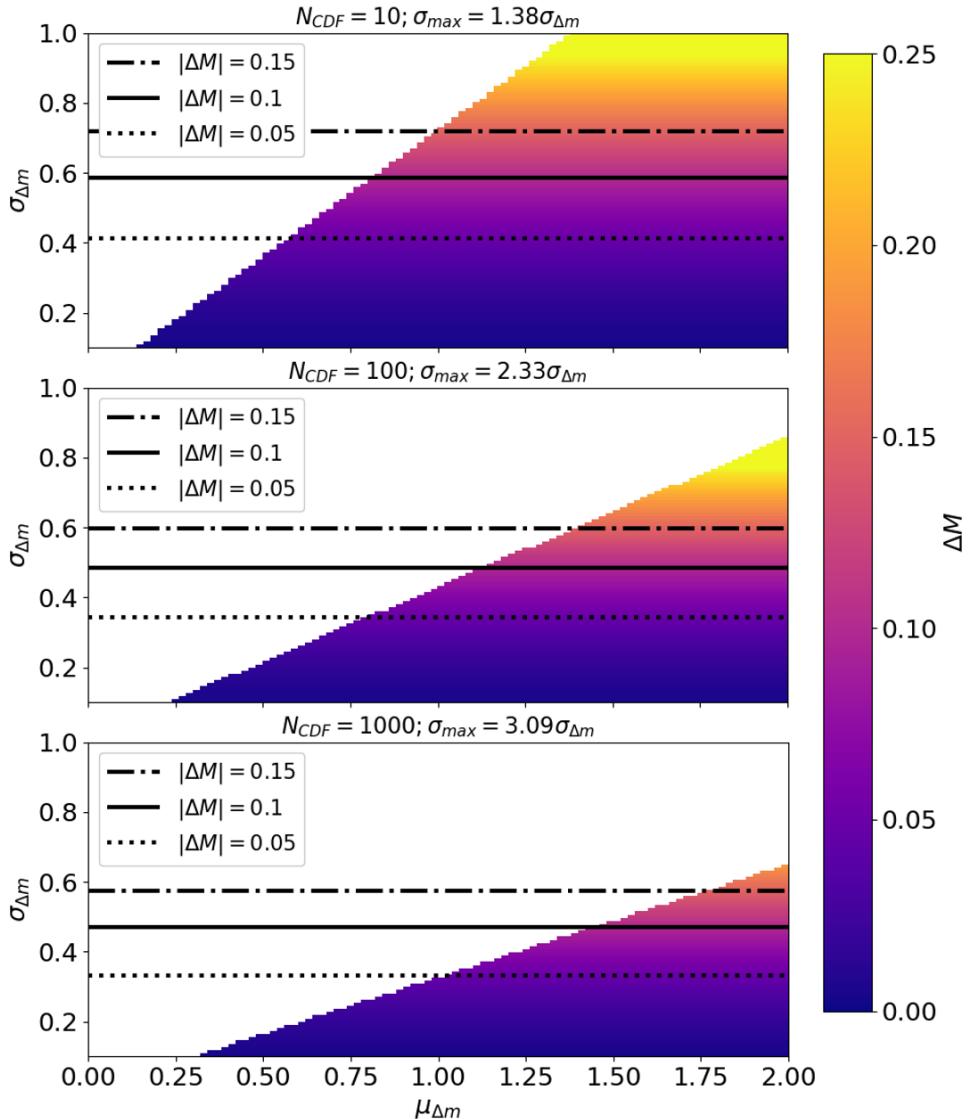


Fig. 21: Plots showcase the bias introduced by averaging over the crowding bias distribution, rather than computing the average crowding flux, and transforming to a magnitude (Eq. 38, replacing expectations for sample averages). These panels simulate the Δm_{AIS} distribution by drawing from a Gaussian CDF, with sample size $N_{CDF} = (10, 100, 100)$ in **Upper**, **Middle**, **Lower** panels, respectively. Adopting that $F_R > 0$, the simulation is cut if any $\Delta m_{AIS} < 0$. Panels demonstrate the number of injected stars, and their resulting distribution, strongly influence the level of bias. The $\Delta M = 0.1$ threshold is reached for $\sigma_{\Delta m} \simeq (0.59, 0.49, 0.47)$ at $\mu_{\Delta m} = (0.8, 1.1, 1.4)$

describe, will over(under)-estimate Cepheid crowding across a particular galaxy, but with a net offset $\rightarrow 0$); for J21, a crowding bias specific to the Cepheid is computed and applied. Riess et al. (2016) and J21 distance moduli to the Large Magellanic Cloud are consistent to $\sim 0.4\sigma$.

In this section, I briefly investigate whether averaging *magnitudes* rather than *fluxes* at the location of the Cepheid returns a significant crowding bias mis-estimate. As described in R20, crowding dispersion from *AIS* tests is log-normal in flux, and Gaussian in magnitudes, “out to a few standard deviations”. Quantitatively⁶:

$$\Delta M_{AIS} = 2.5 \log_{10}(1 + E[F_R]) - E[2.5 \log_{10}(1 + F_R)]. \quad (38)$$

Artificial stars, [numbering (1000, 100, 20 – 50) in Riess et al. (2009, 2011), Riess et al. (2016), J21], are injected at varying angular distances, [(0.05 – 0.75, $< 5, N/A$)” in Riess et al. (2009, 2011), Riess et al. (2016), J21], from the location of the Cepheid, with brightness comparable to the Cepheid (assigned using a period-luminosity relationship). For us, the known unknown is the standard deviation of the Δm distribution, $\sigma_{\Delta m}$ – that is apparently Gaussian.

I perform simulations to estimate ΔM_{AIS} , using Eq. 38, where Δm_{AIS} is drawn from a Gaussian, $\mathcal{N}(\mu_{\Delta m}, \sigma_{\Delta m})$ [Figs. (21, 22)]. Results show the *H*-band expectation correction, with a typical $\Delta m_H \equiv \mu_{\Delta m} \approx 0.5 – 1$, can ease and potentially resolve the Hubble tension, provided the Gaussian width, $\sigma_{\Delta m}$, is large enough ($\sigma_{\Delta m} \gtrsim 0.5$). These simulations are idealised, either drawing precisely from the Gaussian CDF (Fig. 21), or randomly from the distribution (Fig. 22). How the *real* distributions (and their resulting biases) differ, cannot be determined without direct access to *AIS* data. Nonetheless, it is clear this averaging method could significantly bias cosmological parameters.

Finally, it is not obvious how dust is accounted for when performing *AIS* tests. Indeed, to inject a star of known magnitude, both the distance, and the level of dust extinction acting to diminish the flux, must be accounted for. Dust can only be ignored when using Wesenheit magnitudes, but ***AIS* tests are performed in single-band frames**. Riess papers perform some iterative procedure, estimating Δm by *AIS*, correcting the Cepheid magnitude, repeating, all the while updating the period-luminosity relation. It is not obvious what the methodology is, nor whether it is robust; an ignorance of dust when converging to a single-band period-luminosity relation would map intrinsic properties to uncorrelated dust, and could introduce a bias, i.e. $m_b^{int} + A_b = Function(Intrinsic)$.

Additionally, consider that $m_{AIS} \equiv \hat{f}\hat{F}_0$, and it is measured to be $m_{app} \equiv \hat{f}\hat{F}_0 + fF_1$, where $(\hat{f}, f) < 1$ is (estimated, true) dust reduction in flux. The difference between injected magnitude and measured magnitude differs from the true crowding bias by the following:

$$\Delta M_{AIS} = 2.5 \log_{10} \left(\frac{1 + F_R}{1 + F_R \cdot \frac{\hat{f}}{f} \cdot \frac{L_0}{\hat{L}_0} \cdot \frac{\hat{D}^2}{D^2}} \right). \quad (39)$$

Only when dust, Cepheid luminosity, and distance are perfectly estimated, is the crowding bias mis-estimate zero. However, distance is the very object we are trying to infer.

To conclude then, the *AIS* method is lacking a mathematical description (Riess et al., 2009, 2011, 2016, 2020; Javanmardi et al., 2021), and is thus opaque. R20 is not a rigorous proof that crowding bias is well-estimated. In tandem with J21, there is the potential that averaging crowding bias estimates from artificially injected stars leads to an over-estimate of crowding. Moreover, there seems to a requirement for dust, Cepheid luminosity, and distance to be estimated when quantifying crowding, but References only report updating the luminosity term;

⁶Unbiased log-normal distribution estimator, $\mu_{\Delta m} \ln(10)/2.5 = \ln(1 + \bar{F}_R) - \ln(\sqrt{1 + Var(\Delta m)}/[1 + \bar{F}_R]^2)$, reduces ΔM by $\sim 0.02 – 0.03$ mag.

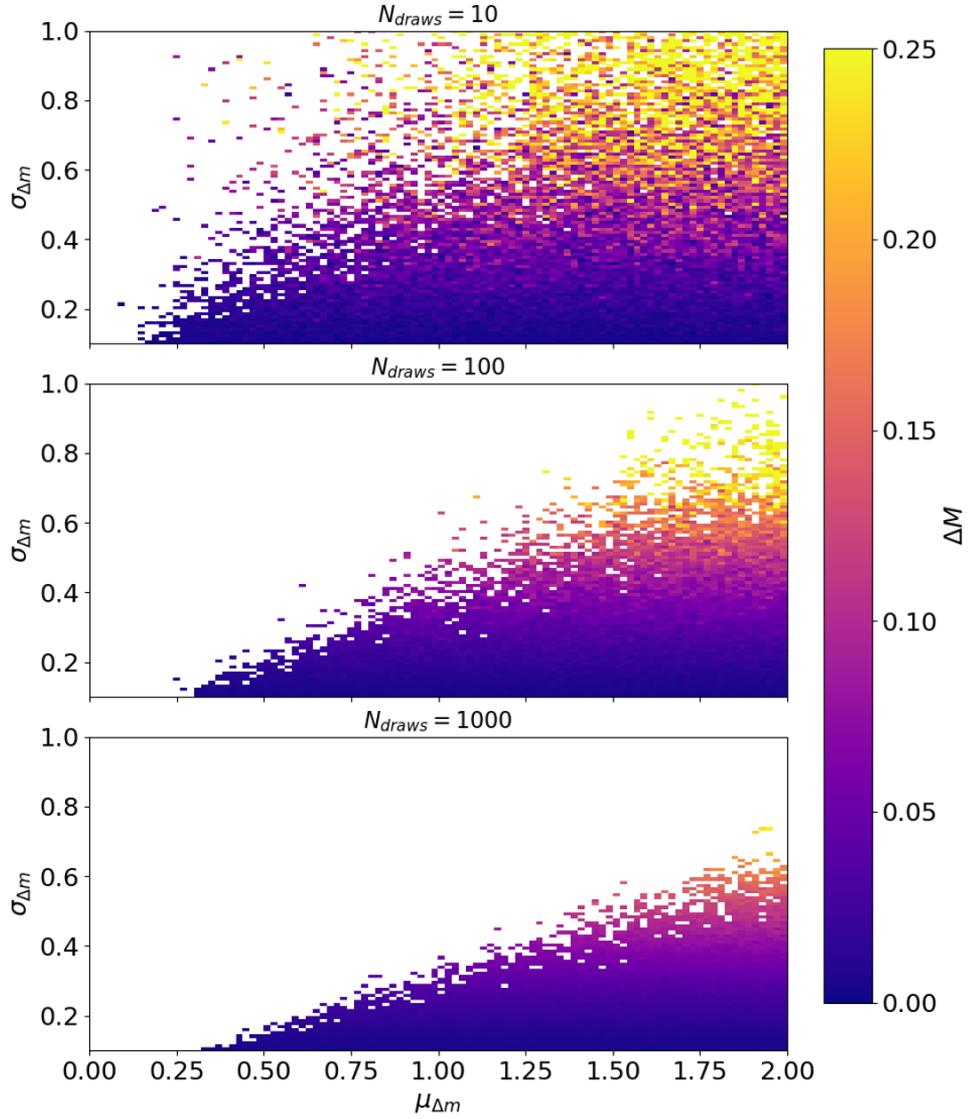


Fig. 22: Same as Fig. 21, except ‘real’ distributions are assumed, by drawing randomly from the Gaussian distribution, $\Delta m_{AIS} \sim \mathcal{N}(\mu_{\Delta m}, \sigma_{\Delta m})$, thus populating previously blank areas. These plots are a single set of random realisations, but bulk trends are nonetheless illustrated. Here, even lower $\mu_{\Delta m}$ - as compared to Fig. 21 - can reasonably ease and/or resolve the Hubble tension (as low as $(\mu_{\Delta m}, \sigma_{\Delta m}) \approx (0.3, 0.4)$ for $N_{draws} = 10$).

ignorance of dust - and an assumed distance - may be introducing a significant bias. Further study is warranted (§ 5.6).

5 Ongoing and Future Work

In this section, I describe ongoing work and future projects for the next ~ 3 years, for an Easter 2024 thesis submission.

- The Supernova Triplet in NGC 3147 (§ 5.1). We analyse a new SN Ia triplet as a model robustness test, and will evaluate the strength of host galaxy correlations. Pending a Cepheid distance modulus ([Riess et al. in prep.](#)), we will report the first *BayeSN* measurement of H_0 . This work will be completed by \sim Dec 2021 ([Ward et al. in prep.](#)), and will form a thesis chapter.
- We will use the DEHVILS optical-NIR *coYJH* dataset to investigate dust-law systematics, and SED correlations with host galaxy properties with *BayeSN*, in particular the ‘mass-step’ (§ 5.2). We will complete this thesis work by \sim Dec 2022.
- We will train a new *BayeSN* model on the Pan-STARRS+DECam *grizy* data from YSE. Complementary to DEHVILS, and analyses of host-galaxy dust systematics, this work will contribute towards an YSE cosmology analysis to place constraints on w (§ 5.3). Thesis work will be completed by \sim Dec 2022.
- Velocity-SED correlations (§ 5.4). Continuing § 3, we will: use MCMC chains to compute p -values, report our analysis of the optical-NIR $B-H$ velocity colour relation, and incorporate velocity measurements into *BayeSN*. Following this, we will train *BayeSN* on the public Kaepora spectral database to learn continuous spectral correlations and improve SN Ia distance estimates (§ 5.4.2). Work from this thesis chapter will be completed \sim Dec 2023.

Time-permitting, I propose two projects:

- New empirical methods that model R_V -insensitive *VJHK* extinguished magnitudes as an exponentially modified Gaussian random variable, can complement mass-step investigations (§ 5.5).
- The *BayeSN* functional principal component architecture can be utilised to train on Milky-Way Cepheids to robustly quantify the crowding bias in extragalactic Cepheids (§ 5.6).

The remainder of this section describes these projects in more detail (expected completion dates in headers).

5.1 The Supernova Triplet in NGC 3147 [~ Dec 2021]

Observations of supernova ‘siblings’ - those sharing a common host galaxy - provide stringent constraints on the intrinsic scatter of the underlying SN spectral energy distribution. Sharing a common distance, peculiar velocity, and host-galaxy type, various common systematic uncertainties, that act to widen empirical SED dispersion, are eliminated through the analysis of SN siblings. Siblings have been used to assess the accuracy of SN standardisation, e.g. (Burns et al., 2020; Biswas et al., 2021).

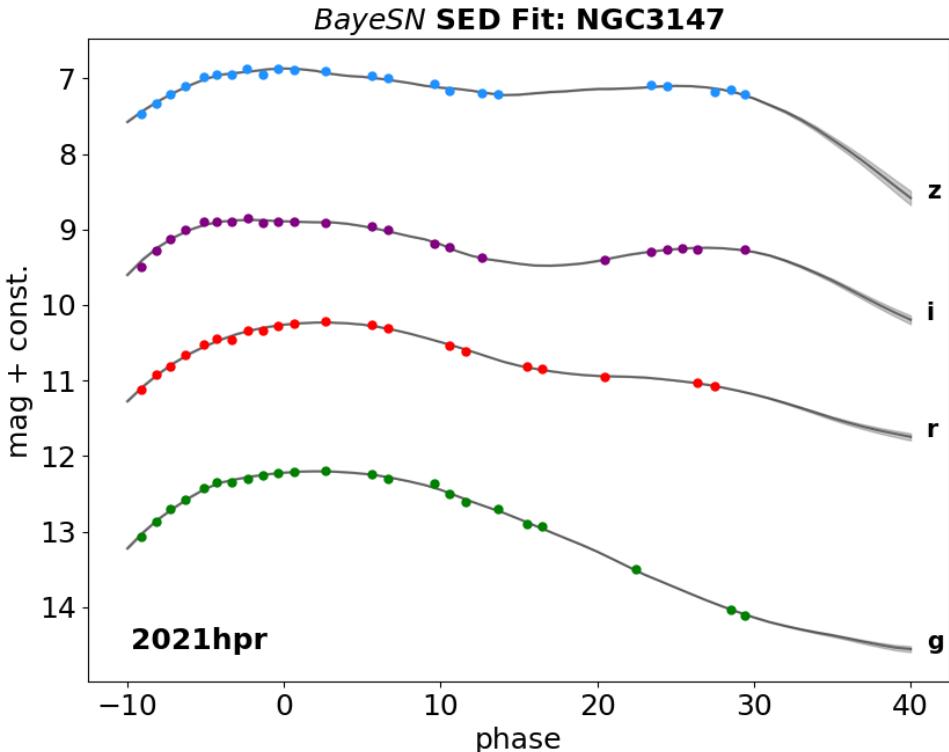


Fig. 23: *BayeSN M20* model fits to the *griz* light curves of 2021hpr, the third observed Type Ia SN in NGC 3147, out to $\simeq 30$ d. Grey shaded regions denote a 1σ credible envelope, computed by averaging 100 posterior functional draws. This YSE data is exclusively reserved for our group.

The recent discovery (08/04/21) of SN 2021hpr (21hpr) (Foley et al., 2018; Jones et al., 2019) in the Cepheid host NGC 3147 ($z_{Heli} = 0.00935^7$), along with SN 1997bq (97bq) (Jha et al., 2006) and SN 2008fv (08fv) (Biscardi et al., 2012; Friedman et al., 2015; Tsvetkov & Elenin, 2010) Type Ia siblings, will be the first example of a SN Ia triplet in a host with a Cepheid derived distance modulus. As such, a triplet analysis will: provide a 3-fold independent test of *BayeSN* distance modulus model predictions, enhance constraints on (A_V, θ) for a within-galaxy variation of SNe/dust study (by applying a shared distance constraint), and - subject to the acquisition of

⁷<https://ned.ipac.caltech.edu/>

a Cepheid distance modulus (Riess et al. in prep) - yield the first *BayeSN* measurement of H_0 , in tandem with a Hubble flow SN set. I am leading the analysis this investigation within YSE, and the new 21hpr Pan-STARRS data is exclusively reserved for our group.

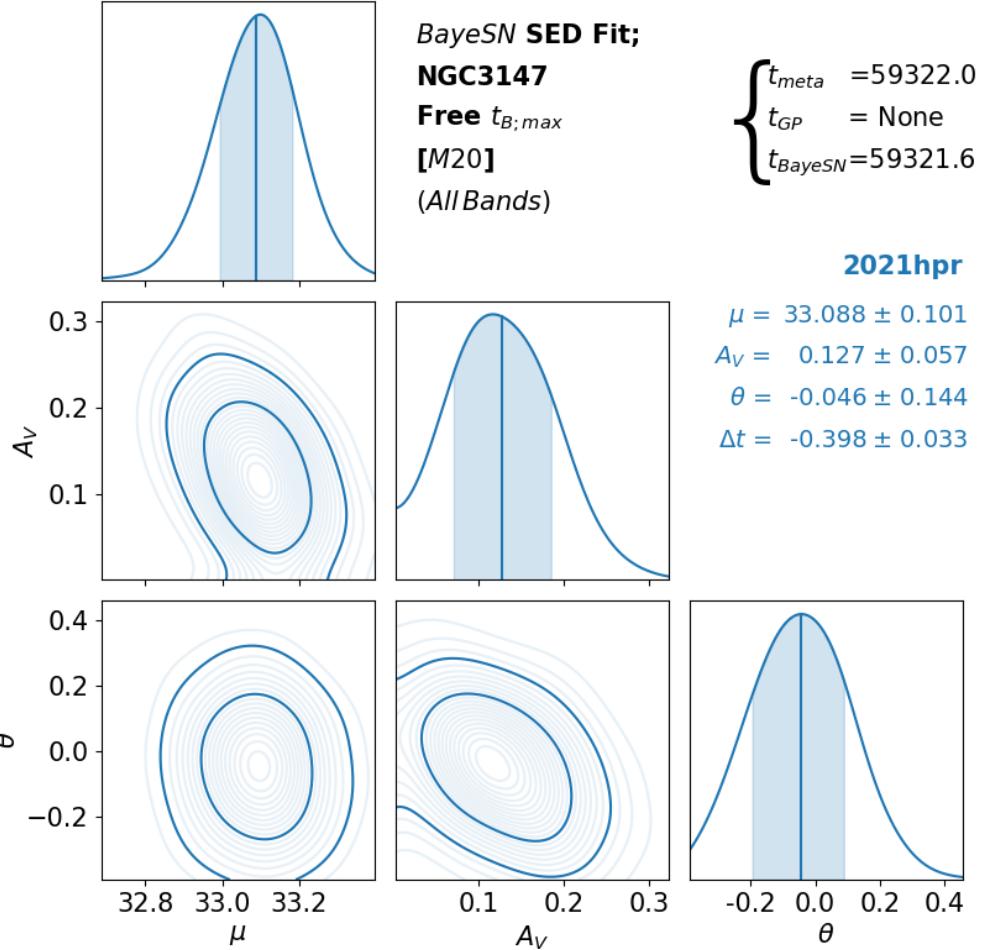


Fig. 24: Corner plot of *BayeSN* posterior estimates of key parameters: distance modulus, μ , extinction, A_V , and the light curve shape parameter (the first FPC score), θ , fitted to the novel data of 2021hpr. Shaded regions on marginal KDEs mark 68% credible intervals, with vertical lines marking the mean of posterior samples. 2-D contours denote (68, 95)% regions. Reflection KDE is employed for A_V , which is bounded to be non-zero. The time of B -maximum is fitted for, using a Gaussian prior width $\pm 5 d$ centred on the quoted ‘meta’ time. A GP estimate of $T_{B;\max}$ is not required. *BayeSN* fits are excellent.

Fig. 23 depicts the *BayeSN* M20 model fits to 21hpr out to $17/05 \simeq 30 d$, with Fig. 24 showing the posterior constraints on (μ, A_V, θ) . Figs. (25, 26) showcase independent *BayeSN* posterior

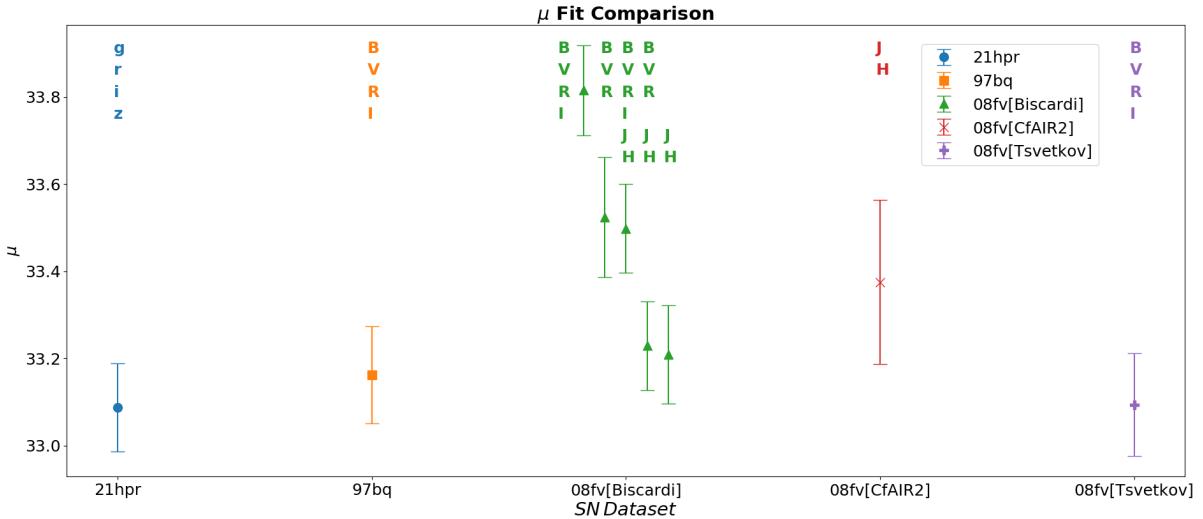


Fig. 25: Distance modulus constraints from *BayeSN* fits to five datasets of the SN triplet in NGC 3147; colour letters indicate passbands of data fitted. Biscardi dataset of 08fv in *BVRJH* passbands is most discrepant [green triangle]. Preliminary work indicates that there is a mis-match between standard *BayeSN* and real Biscardi transmission functions. In particular, there is a strong *I*-band atmospheric feature which could be absent in Biscardi data. Plot demonstrates removing certain bands brings Biscardi μ more in line with expectations; in particular, removing only the *I*-band reduces μ by ~ 0.2 mag. More in Fig. 25.

distance modulus estimates for 5 datasets across the triplet, the latent variables of which should all be equal. Biscardi et al. (2012) dataset of 08fv (hereafter Biscardi) is most discrepant ($> 2\sigma$), which is likely a product of a transmission function mis-match. *SNooPy* fits also yield consistently high distances (Fig. 26); this will be further investigated.

Next steps will be to construct a joint model in *BayeSN* that will fit for a shared distance constraint. It will be informative to examine how (μ, A_V, θ) constraints shift as a consequence. Improved extinction estimates will be contrasted with dust population hyper-parameters (Thorp et al., 2021), and qualitatively contrasted with local galaxy properties, e.g. density, distance to galactic centre etc. using *GHOST* outputs (Gagliano et al., 2020). For high extinction, R_V can be estimated. Comparisons of μ with *SALT2* will also be performed. As already noted (§ 1.4), there are various SN correlations with galaxy morphology. Characterising NGC 3147 and contrasting our triplet distribution in (A_V, θ) with a population distribution can provide insight into these correlations. For example, Scolnic et al. (2020) find weak evidence that the *SALT2* stretch parameter has a tighter distribution in SN siblings (8 pairs studied) as compared to the global population distribution. A tight grouping may signal a need to strongly consider galaxy selection effects.

With access to a Cepheid distance modulus, and a Hubble-flow sample, H_0 can be estimated. Future *BayeSN* cosmology investigations will employ multiple calibrators, given individual μ estimates are scattered about the best fitting cosmological model; this work is a stepping stone towards that larger goal.

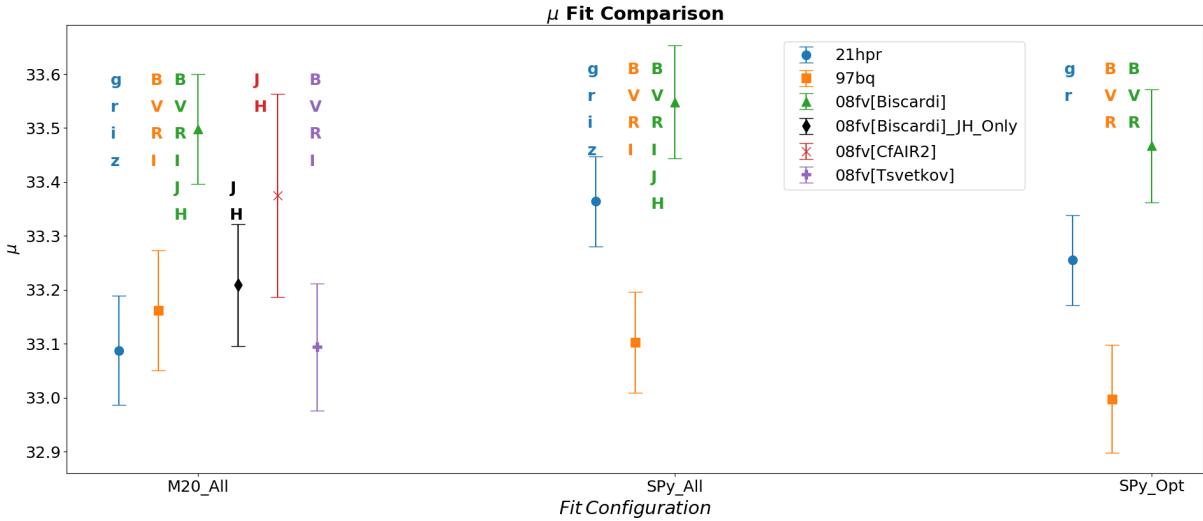


Fig. 26: Following from Fig. 25, the discrepant Biscardi dataset is inspected further by fitting with *SNoPy* (*SPy*), which also returns an anomalously high distance modulus, when fitting to all passbands, and optical only. Plotted on the horizontal axis are the fits using different model configurations, namely using the *M20 BayeSN* model fitted to all the data (*M20_ALL*), and *SNoPy* fitted to all the data and optical passband subsets (*SPy_All* and *SPy_Opt* respectively). These results does not rule out a model systematic common to both *BayeSN* and *SNoPy*. The next step will be to obtain real Biscardi transmission functions (Biscardi et al., 2012), that differ from the standard transmission functions used in *BayeSN*, and re-fit.

5.2 DEHVILS & Dust-Laws Across the Mass-Step [~Dec 2022]

The first year of DEHVILS data (~ 125 SNe Ia) provides an excellent opportunity for a *BayeSN* study of host-galaxy correlations, and dust laws across the mass-step. DEHVILS is a 3-year survey that began in 2020, observing in Optical-NIR *coYJH* with ATLAS and UKIRT, that will yield ~ 700 - 1000 SNe Ia ($0.01 \lesssim z \lesssim 0.08$). We have proprietary data access, and have fitted the test subject SN 2020sjo; Figs. (27, 28). *BayeSN* will be retrained on DEHVILS data. We will adopt the methodology of Thorp et al. (2021), and use *BayeSN* to model the dust in mass-split sub-samples. Given the *YJH*-NIR bands are less sensitive to dust (as compared to the optical, and to the *z*-band, the reddest passband analysed in Thorp et al. 2021), we will improve upon the *BayeSN* analysis in Thorp et al. (2021), and shed further light on the mass-step, using a new and independent sample.

5.3 Pan-STARRS + DECam & Constraints on w [~Dec 2022]

Complementary to DEHVILS, YSE survey observations using Pan-STARRS and the Dark Energy Camera (DECam) - two of the best-calibrated ground based instruments used in supernova cosmology - will observe and construct the premier low- z anchor of SN Ia sample in *grizy* passbands, for use in cosmological analyses. YSE, which began November 2019, is a 3.5 year survey that will observe $\sim 10^4$ SNe Ia (10^3 spectroscopically confirmed). This low- z survey will anchor high- z samples (e.g. LSST) for constraints on the dark energy equation of state parameter, w , to better than 3%. As key members of the YSE Type Ia cosmology team, we have access to

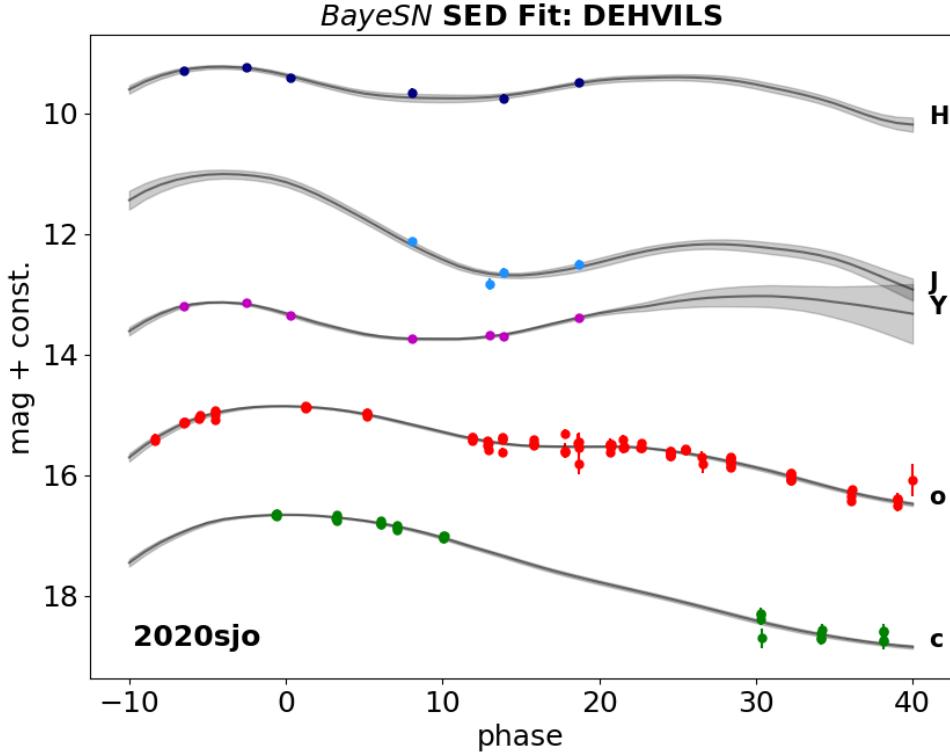


Fig. 27: *BayeSN M20* model fits to the ATLAS+UKIRT *coYJH* proprietary DEHVILS light curve data of a test subject SN 2020sjo. Leveraging the continuity of the NIR SEDs - a feature unique to *BayeSN* - *BayeSN* successfully fits these data. Adopting the *BayeSN* fiducial training cosmology: Λ CDM with $H_0 = 73.24 \text{ km s}^{-1}$, the Hubble residual is 0.06mag, a discrepancy of $\sim 0.5\sigma$.

Pan-STARRS/DECam data. We will train a new *BayeSN* model on these data, investigate systematics owing to host-galaxy dust, and contribute towards state-of-the-art cosmology analyses to constrain w .

5.4 Velocity-SED Correlations [~Dec 2023]

5.4.1 Silicon-II Velocity-SED Correlations

To investigate further whether velocity information can improve distance estimates, we will use correlated MCMC samples from *BayeSN* to compute a p -value distribution (§ 3). We will test velocity-dependencies on host properties, using outputs from the *GHOST* database (Gagliano et al., 2020). *BayeSN* also has functionality to learn astrophysical correlations (Mandel et al., 2020); velocity measurements will be incorporated into *BayeSN*, and the improvement in distance estimate accuracy will be quantified.

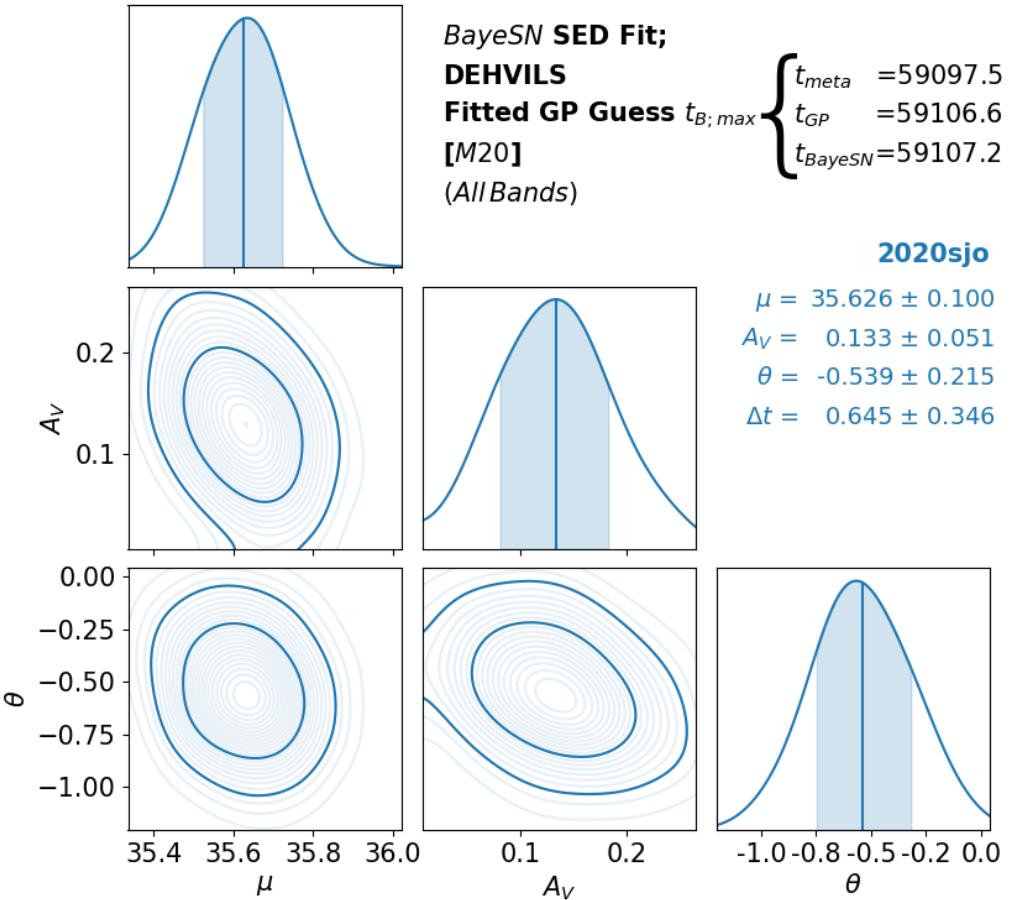


Fig. 28: Corner plot of (μ, A_V, θ) fitted to the proprietary DEHVILS data of SN 2020sjo. An algorithm is constructed to guess the time of intrinsic B -band maximum, by fitting a Gaussian process to the optical passbands. The $\pm 5 d$ prior in *BayeSN* on $T_{B; \max}$ is then centred on the guess. Here, *BayeSN* benefits from the GP guess, with a best fitting peak time +10 d after the quoted ‘meta’ time.

5.4.2 Kaepora Spectra & Spectroscopic Training

Advancing upon measurements of a single spectral line, we will use the public Kaepora database, containing 4975 spectra from 777 SNe Ia (Siebert et al., 2019), for spectroscopic training of *BayeSN*. Currently, *BayeSN* yields accurate distances by learning the broadband ($\Delta\lambda = 1000\text{\AA}$) population variations from SN Ia photometry. However, an improved *BayeSN* will be trained simultaneously on photometric and spectroscopic time series to learn fine-resolution ($\Delta\lambda = 50\text{\AA}$) spectroscopic correlations, which will further improve the usefulness of SNe Ia as cosmological probes. Other spectroscopic datasets available to us for training will include 620 NIR spectra of

149 SNe Ia from [Hsiao et al. \(2019\)](#), and HST optical/NIR photometry and spectroscopy from the SIRAH program, which we are members of. This improved *BayeSN* model will be important for fully exploiting future SN Ia data from LSST and the Roman Space Telescope for cosmological analyses.

5.5 Adjacent Empirical Analyses

New model-independent empirical techniques can be developed to separate intrinsic absolute magnitudes from dust, that supports *BayeSN* analyses. Assuming magnitudes such as $VJHK$ are weakly-dependent ‘enough’ on R_V (as motivated by the [Fitzpatrick \(1999\)](#) law), extinguished magnitudes can be modelled as the sum of a Gaussian random variable, the intrinsic absolute magnitude, M_V^{int} , and an exponentially-distributed random variable, the dust extinction, A_V . Adopting R_V -independence, the resulting extinguished magnitude, $M_V^{ext} = M_V^{int} + A_V$, is drawn from the exponentially-modified-Gaussian distribution. Unbiased frequentist sample estimators of this distribution can then be deployed to learn the population hyperparameters. With these estimates, mean intrinsic magnitudes and dust properties in low and high mass galaxies can be contrasted, for investigations into the mass-step. Such a methodology would be computationally inexpensive. Future work must quantify the R_V -dependence of $VJHK$ bands, and evaluate the viability of this novel methodology.

5.6 Bayesian Crowding

With Milky-Way Cepheid light curves as a training set, and extragalactic Cepheids as a test set, the crowding flux F_R can enter as a parameter, that is inferred and informed by the reduction in Cepheid light curve fractional amplitude, relative to the period. This analysis can robustly infer the crowding bias. The *BayeSN* architecture can be adapted to learn Cepheid functional principal components, and hierarchically infer $F_R \equiv \Delta m$. Moreover, Cepheids are better suited for within-galaxy variation of dust studies, numbering $\sim 10 - 100$ in a single host, as compared to SNe $\sim 1 - 3$.

Acknowledgements

As of 25/11/20, I am privileged to be guided and supported by the *BayeSN* team, comprising Kaisey Mandel, Stephen Thorp, Suhail Dhawan & Gautham Narayan. *BayeSN* was devised by KM, and has been implemented in *Stan* by ST. Figs. (24, 28) adapt ST’s script [Contour/Reflection KDEs]. ST compiled all NGC 3147 data, bar 21hpr, reserved for us by YSE (Foley et al., 2018; Jones et al., 2019). SD performed *SNooPy* fits (Fig. 25). Host galaxy properties (§ 3) were compiled by Alex Gagliano (Gagliano et al., 2020).

References

- Abbott T. M. C. e., 2021, FERMILAB-PUB-21-221-AE,
- Ambikasaran S., Foreman-Mackey D., Greengard L., Hogg D. W., O’Neil M., 2015, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **38**, 252
- Avelino A., Friedman A. S., Mandel K. S., Jones D. O., Challis P. J., Kirshner R. P., 2019, arXiv e-prints,
- Betancourt M., 2016, arXiv e-prints, p. arXiv:1601.00225
- Biscardi I., et al., 2012, *A&A*, **537**, A57
- Biswas R., et al., 2021, arXiv e-prints, p. arXiv:2103.16978
- Bonvin V., et al., 2017, *MNRAS*, **465**, 4914
- Brout D., Scolnic D., 2021, *ApJ*, **909**, 26
- Brout D., et al., 2019, *ApJ*, **874**, 150
- Burns C. R., et al., 2020, *ApJ*, **895**, 118
- Cardelli J. A., Clayton G. C., Mathis J. S., 1989, *ApJ*, **345**, 245
- Carpenter B., et al., 2017, *Journal of Statistical Software*, **76**, 1
- Carrick J., Turnbull S. J., Lavaux G., Hudson M. J., 2015, *MNRAS*, **450**, 317
- Childress M., et al., 2013, *ApJ*, **770**, 108
- Contreras C., et al., 2010, *AJ*, **139**, 519
- D’Andrea C. B., et al., 2011, *ApJ*, **743**, 172
- Davis T. M., Hinton S. R., Howlett C., Calcino J., 2019, *MNRAS*, **490**, 2948
- Dettman K. G., et al., 2021, arXiv e-prints, p. arXiv:2102.06524
- Di Valentino E., et al., 2021, arXiv e-prints, p. arXiv:2103.01183
- Efstathiou G., 2014, *MNRAS*, **440**, 1138
- Efstathiou G., 2020, arXiv e-prints, p. arXiv:2007.10716
- Efstathiou G., 2021, arXiv e-prints, p. arXiv:2103.08723

- Fitzpatrick E. L., 1999, *PASP*, **111**, 63
- Folatelli G., et al., 2010, *AJ*, **139**, 120
- Foley R. J., Kasen D., 2011, *ApJ*, **729**, 55
- Foley R. J., et al., 2018, *MNRAS*, **475**, 193
- Friedman A. S., et al., 2015, *ApJS*, **220**, 9
- Gagliano A., Narayan G., Engel A., Carrasco Kind M., 2020, arXiv e-prints, p. arXiv:2008.09630
- Graziani R., et al., 2020, arXiv e-prints, p. arXiv:2001.09095
- Guy J., et al., 2007, *A&A*, **466**, 11
- Hamuy M., et al., 2006, *PASP*, **118**, 2
- Hoffmann M. D., Gelman A., 2014, *J. Machine Learning Res.*, **15**, 1593
- Hoffmann S. L., et al., 2016, *ApJ*, **830**, 10
- Holz D. E., Wald R. M., 1998, *Phys. Rev. D*, **58**, 063501
- Howlett C., Robotham A. S. G., Lagos C. D. P., Kim A. G., 2017, *ApJ*, **847**, 128
- Hsiao E. Y., 2009, Ph.D. Thesis, Univ. Victoria
- Hsiao E. Y., et al., 2019, *PASP*, **131**, 014002
- Huang C. D., et al., 2020, *ApJ*, **889**, 5
- Huterer D., Shafer D. L., Scolnic D. M., Schmidt F., 2017, *JCAP*, **2017**, 015
- Javanmardi B., et al., 2021, *ApJ*, **911**, 12
- Jha S., et al., 2006, *AJ*, **131**, 527
- Johansson J., et al., 2021, arXiv e-prints, p. arXiv:2105.06236
- Jones D. O., et al., 2019, *ApJ*, **881**, 19
- Kasen D., 2006, *ApJ*, **649**, 939
- Keller S. C., et al., 2007, , **24**, 1
- 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
- Kim A. G., et al., 2013, *ApJ*, **766**, 84
- Krisciunas K., et al., 2007, *AJ*, **133**, 58
- Krisciunas K., et al., 2017, *AJ*, **154**, 211
- Krishnan C., Mohayaee R., Colgáin E. Ó., Sheikh-Jabbari M. M., Yin L., 2021, arXiv e-prints, p. arXiv:2106.02532

- Lochner M., McEwen J. D., Peiris H. V., Lahav O., Winter M. K., 2016, *ApJS*, **225**, 31
- Madore B. F., 1982, *ApJ*, **253**, 575
- Mandel K. S., Scolnic D. M., Shariff H., Foley R. J., Kirshner R. P., 2017, *ApJ*, **842**, 93
- Mandel K. S., Thorp S., Narayan G., Friedman A. S., Avelino A., 2020, arXiv e-prints, [p. arXiv:2008.07538](#)
- Nadolny T., Durrer R., Kunz M., Padmanabhan H., 2021, arXiv e-prints, [p. arXiv:2106.05284](#)
- Perivolaropoulos L., 2014, *Galaxies*, **2**, 22
- Perivolaropoulos L., Skara F., 2021, arXiv e-prints, [p. arXiv:2105.05208](#)
- Perlmutter S., et al., 1999, *ApJ*, **517**, 565
- Phillips M. M., 1993, *ApJ*, **413**, L105
- Planck Collaboration et al., 2020a, *A&A*, **641**, A1
- Planck Collaboration et al., 2020b, *A&A*, **641**, A6
- Ponder K. A., Wood-Vasey W. M., Weyant A., Barton N. T., Galbany L., Garnavich P., Matheson T., 2020, arXiv e-prints, [p. arXiv:2006.13803](#)
- Rasmussen C. E., Williams C. K. I., 2006, MIT Press
- Refsdal S., 1964, *MNRAS*, **128**, 307
- Riess A. G., et al., 1998, *AJ*, **116**, 1009
- Riess A. G., et al., 2009, *ApJ*, **699**, 539
- Riess A. G., et al., 2011, *ApJ*, **730**, 119
- Riess A. G., et al., 2016, *ApJ*, **826**, 56
- Riess A. G., Casertano S., Yuan W., Macri L. M., Scolnic D., 2019, arXiv e-prints,
- Riess A. G., Yuan W., Casertano S., Macri L. M., Scolnic D., 2020, *ApJ*, **896**, L43
- Riess A. G., Casertano S., Yuan W., Bowers J. B., Macri L., Zinn J. C., Scolnic D., 2021, *ApJ*, **908**, L6
- Rigault M., et al., 2013, *A&A*, **560**, A66
- Rigault M., et al., 2020, *A&A*, **644**, A176
- Salim S., Narayanan D., 2020, *ARA&A*, **58**, 529
- Scolnic D. M., et al., 2018, *ApJ*, **859**, 101
- Scolnic D., et al., 2019, Astro2020: Decadal Survey on Astronomy and Astrophysics, **2020**, 270
- Scolnic D., et al., 2020, arXiv e-prints, [p. arXiv:2002.00974](#)
- Siebert M. R., et al., 2019, *MNRAS*, **486**, 5785

- Siebert M. R., Foley R. J., Jones D. O., Davis K. W., 2020, *MNRAS*, **493**, 5713
- Smith M., et al., 2014, *ApJ*, **780**, 24
- Smith M., et al., 2020, arXiv e-prints, p. arXiv:2001.11294
- Soltis J., Casertano S., Riess A. G., 2021, *ApJ*, **908**, L5
- Stritzinger M. D., et al., 2011, *AJ*, **142**, 156
- Sullivan M., et al., 2010, *MNRAS*, **406**, 782
- Thorp S., Mandel K. S., Jones D. O., Ward S. M., Narayan G., 2021, arXiv e-prints, p. arXiv:2102.05678
- Tripp R., 1998, *A&A*, **331**, 815
- Tsvetkov D. Y., Elenin L., 2010, *Peremennye Zvezdy*, **30**, 2
- Uddin S. A., et al., 2020, *ApJ*, **901**, 143
- Visser M., 2005, *General Relativity and Gravitation*, **37**, 1541
- Wang X., et al., 2009, *ApJ*, **699**, L139

A Intro Cosmology

Objects have their usual meaning; (χ, x) is comoving coordinate/distance. The FLRW metric, Eq. A.4, has a *constant* spatial curvature, k (as in Friedmann equation; Eq. A.2).

$$x = \int_{t_e}^{t_r} \frac{c dt}{a(t)}, \quad (\text{A.1})$$

$$\frac{H^2}{H_0^2} = \frac{\Omega_{m,0}}{a^3} + \frac{\Omega_{\gamma,0}}{a^4} + \Omega_{\Lambda,0} - \frac{kc^2}{H_0^2 a^2} = E^2(a, H_0, k, \{\Omega_{i,0}, w_i\}) = E^2(a, \Phi) \quad (\text{A.2})$$

$$1+z = \frac{\lambda_o}{\lambda_e} ; \quad 1+\bar{z} = \frac{1}{a}, \quad (\text{A.3})$$

$$ds^2 = -c^2 dt^2 + a^2(t) \left(\frac{d\chi^2}{1-k\chi^2} + \chi^2(d\theta^2 + \sin^2\theta d\phi^2) \right) = -c^2 dt^2 + a^2(t) dx^2. \quad (\text{A.4})$$

$$x = \int_{t_e}^{t_0} \frac{c dt}{a} = \int_0^{S_k(x)} \frac{d\chi}{\sqrt{1-k\chi^2}} = \begin{cases} \frac{\arcsin(\sqrt{k}S_k(x))}{\sqrt{k}} & k > 0 \\ S_k(x) & k = 0 \\ \frac{\operatorname{arcsinh}(\sqrt{|k|}S_k(x))}{\sqrt{|k|}} & k < 0 \end{cases}. \quad (\text{A.5})$$

$$S_k(x) = \begin{cases} \frac{\sin(\sqrt{k}x)}{\sqrt{k}} & k > 0 \\ x & k = 0 \\ \frac{\sinh(\sqrt{|k|}x)}{\sqrt{|k|}} & k < 0 \end{cases}. \quad (\text{A.6})$$

$$dA = \sqrt{\det(g)} d\theta d\phi = a^2(t) \chi^2 \sin\theta d\theta d\phi, \quad (\text{A.7})$$

$$F = \frac{L}{4\pi S_k^2(x)(1+z)^2} = \frac{L}{4\pi D_L^2(z)}, \quad (\text{A.8})$$

$$D_L(z) = S_k(x)(1+z). \quad (\text{A.9})$$

$$\mu = m - M = 5 \log_{10} \left(\frac{D_L(z)}{Mpc} \right) + 25 \quad (\text{A.10})$$

B Gaussian Cheat Sheet

$$p(x|\mu, \Sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} ; \quad X \sim \mathcal{N}(\mu, \sigma^2). \quad (\text{B.1})$$

$$P(\mathbf{f}) = p(\mathbf{f}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{\sqrt{\det[2\pi\boldsymbol{\Sigma}]}} e^{-\frac{1}{2}(\mathbf{f}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{f}-\boldsymbol{\mu})} ; \quad \mathbf{f} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}). \quad (\text{B.2})$$

$$\mathbf{f} \equiv \begin{pmatrix} \mathbf{U} \\ \mathbf{V} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \boldsymbol{\mu}_U \\ \boldsymbol{\mu}_V \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_U & \boldsymbol{\Sigma}_{UV} \\ \boldsymbol{\Sigma}_{VU} & \boldsymbol{\Sigma}_V \end{pmatrix}\right) \equiv \begin{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} \\ \begin{bmatrix} u \times u & u \times v \\ v \times u & v \times v \end{bmatrix} \end{bmatrix} \quad (\text{B.3})$$

$$P(\mathbf{U}) = \mathcal{N}(\mathbf{U}|\boldsymbol{\mu}_u, \boldsymbol{\Sigma}_U), \quad (\text{B.4})$$

$$P(\mathbf{U}) = \int P(\mathbf{U}, \mathbf{V}) d\mathbf{V} = \int P(\mathbf{U}|\mathbf{V}) P(\mathbf{V}) d\mathbf{V}, \quad (\text{B.5})$$

$$\mathbf{U}|\mathbf{V} \sim \mathcal{N}(E[\mathbf{U}|\mathbf{V}], Var[\mathbf{U}|\mathbf{V}]) \quad (\text{B.6})$$

$$E[\mathbf{U}|\mathbf{V}] = \boldsymbol{\mu}_U + \boldsymbol{\Sigma}_{UV} \boldsymbol{\Sigma}_V^{-1} (\mathbf{V} - \boldsymbol{\mu}_V) \quad (\text{B.7})$$

$$Var[\mathbf{U}|\mathbf{V}] = \boldsymbol{\Sigma}_U - \boldsymbol{\Sigma}_{UV} \boldsymbol{\Sigma}_V^{-1} \boldsymbol{\Sigma}_{VU}. \quad (\text{B.8})$$

$$\mathbf{V} \sim \mathcal{N}(\boldsymbol{\mu}_V, \boldsymbol{\Sigma}_V) \quad (\text{B.9})$$

$$\mathbf{U}|\mathbf{V} \sim \mathcal{N}(\mathbf{U}_0 + \mathbf{X}\mathbf{V}, \boldsymbol{\Sigma}_{U|V}) \quad (\text{B.10})$$

$$\begin{pmatrix} \mathbf{U} \\ \mathbf{V} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mathbf{U}_0 + \mathbf{X}\boldsymbol{\mu}_V \\ \boldsymbol{\mu}_V \end{pmatrix}, \begin{pmatrix} \mathbf{X}\boldsymbol{\Sigma}_V \mathbf{X}^T + \boldsymbol{\Sigma}_{U|V} & \mathbf{X}\boldsymbol{\Sigma}_V \\ \boldsymbol{\Sigma}_V \mathbf{X}^T & \boldsymbol{\Sigma}_V \end{pmatrix}\right). \quad (\text{B.11})$$

[Properties of Gaussian Distributions, Dr K. Mandel, Part III Astrostatistics]. Accompanying text for GP-conditioning/hyperparameter optimisation is redacted ([Rasmussen & Williams, 2006](#)).

$$\begin{pmatrix} \mathbf{y} \\ \mathbf{f}_* \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mathbf{1}_c \\ \mathbf{1}_c \end{pmatrix}, \begin{pmatrix} \mathbf{K}_{o,o} + \mathbf{W} & \mathbf{K}_{o,*} \\ \mathbf{K}_{*,o} & \mathbf{K}_{*,*} \end{pmatrix}\right). \quad (\text{B.12})$$

$$y(t_i) = f(t_i) + \epsilon_i ; \quad \epsilon_i \sim \mathcal{N}(0, \sigma_i^2) \quad (\text{B.13})$$

$$Var[y_i] = Var[f_i] + Var[\epsilon_i] + 2\underbrace{Cov[f_i, \epsilon_i]}_0 = K_{i,i} + \sigma_i^2 \quad (\text{B.14})$$

$$Cov[y_i, y_j] = Cov[f_i, f_j] + Cov[\epsilon_i, \epsilon_j] + 0 + 0 = K_{i,j} + \sigma_{ij}^2 \quad (\text{B.15})$$

$$\boldsymbol{\Sigma}_o = \mathbf{K}_{o,o} + \mathbf{W} ; \quad W_{ii} = \sigma_i^2 ; \quad W_{ij} = \sigma_{ij}^2. \quad (\text{B.16})$$

$$\mathbf{f}_*|\mathbf{y} \sim \mathcal{N}(\mathbf{1}_c + \mathbf{K}_{*,o}(\mathbf{K}_{o,o} + \mathbf{W})^{-1}(\mathbf{y} - \mathbf{1}_c), \quad (\text{B.17})$$

$$\mathbf{K}_{*,*} - \mathbf{K}_{*,o}(\mathbf{K}_{o,o} + \mathbf{W})^{-1}\mathbf{K}_{o,*}) \quad (\text{B.18})$$

$$p(\mathbf{y}) = \mathcal{N}(\mathbf{1}_c, \mathbf{K}_{o,o} + \mathbf{W}) \equiv p(\mathbf{y}|\boldsymbol{\theta}), \quad (\text{B.19})$$

$$\ln(p(\mathbf{y}|\boldsymbol{\theta})) = -\frac{N}{2}\ln(2\pi) - \frac{1}{2}\ln(\det[\mathbf{K}_{o,o} + \mathbf{W}]) - \frac{1}{2}(\mathbf{y} - \mathbf{1}_c)^T(\mathbf{K}_{o,o} + \mathbf{W})^{-1}(\mathbf{y} - \mathbf{1}_c), \quad (\text{B.20})$$

C NIR-Times Correlate with the First Functional Principal Component

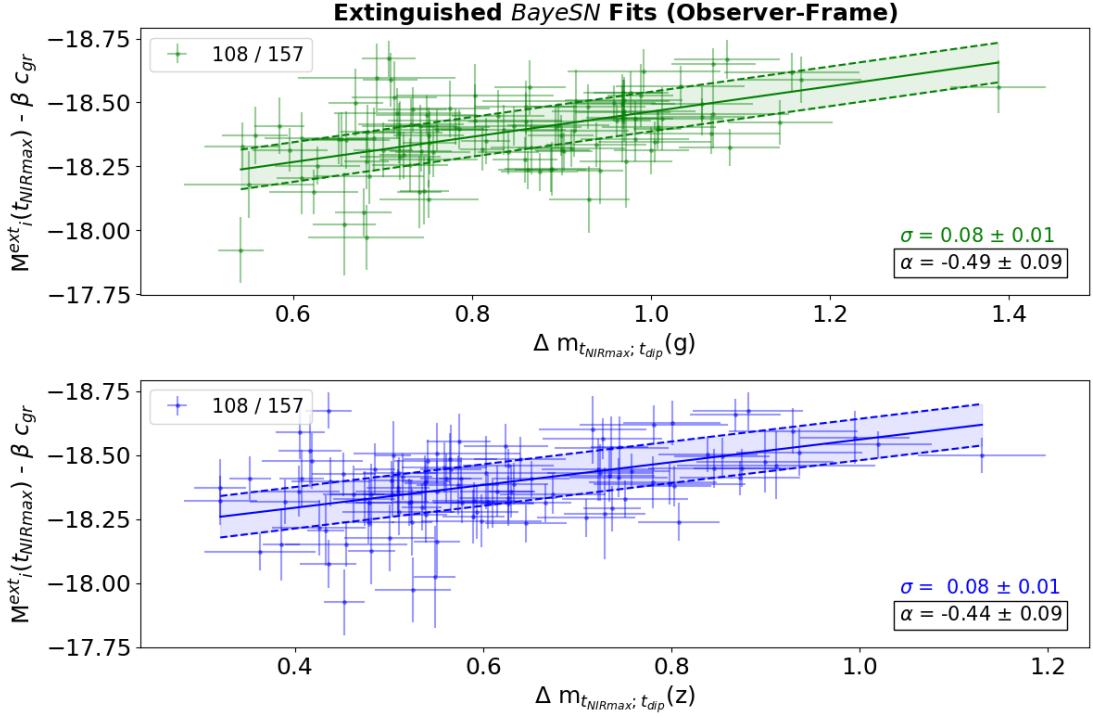


Fig. 29: The lowest scatters in multiple linear regressions are achieved when time-information is exploited. Plotted, are peak extinguished magnitudes and colours, extracted at the time of z -peak, t_{NIRmax} , and declines to the time of z -trough, t_{dip} , using *BayeSN* fits. 49 SNe are rejected that do not exhibit a 1σ significant peak/dip. A dispersion, $\sigma \sim 0.08$ mag, is also achieved for declines measured to time of 2nd z -peak, and for $M_r^{ext} - \beta c_{gi}|_{t=t_{NIRmax}}$.

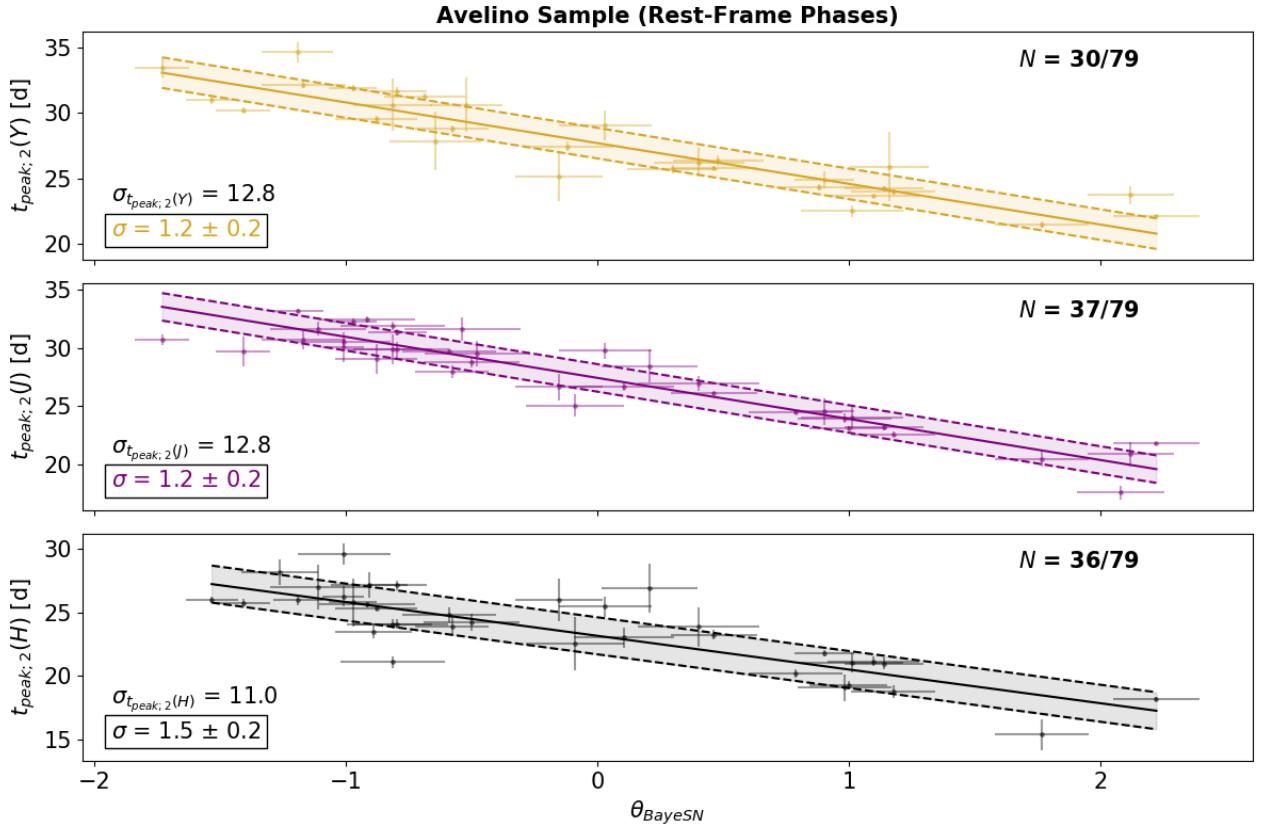


Fig. 30: Plots demonstrate the utility of time information when standardising SN magnitudes. Using the Avelino sample ([Avelino et al., 2019](#); [Mandel et al., 2020](#)), 2nd peak times in the near infrared bands of extinguished light curves are extracted, and plotted against the *BayeSN* coefficient on the first functional principal component. By contrasting the raw scatter of peak times, ~ 10 d, with the scatter in the time-FPC relation, ~ 1 d, conclusively, peak times encode FPC information, and can be used for standardisation. This finding accounts for the low scatter in Fig. 29, and matches theoretical expectations ([Kasen, 2006](#)).