Astronomy & Astrophysics manuscript no. Nicolas
February 17, 2020

Redshift Evolution of the Type Ia Supernova Population Stretch Distribution

N. Nicolas ⋆[1], M. Rigault ⋆⋆[2], R. Graziani[2], M. Briday[1], Y. Copin[1], and Y.-L. Kim[1]

[1] Université de Lyon, F-69622, Lyon, France; Université de Lyon 1, Villeurbanne; CNRS/IN2P3, Institut de Physique des Deux Infinis, Lyon
[2] Université Clermont Auvergne, CNRS/IN2P3, Laboratoire de Physique de Clermont, F-63000 Clermont-Ferrand, France.

Received 2 November 1992 / Accepted 7 January 1993

ABSTRACT

Context. ABSTRACT NOT DONE YET. Type Ia supernovae (SNe Ia) allow for the construction of the Hubble diagram, giving us information about the Universe's expansion and its fundamental components, one of which is dark energy. But systematic uncertainties are now starting to be limiting in our ability to measure those parameters. In particular, the physics of SNe Ia is still mostly unknown, and is thought not to change in time/with the redshift.
Aims. In an attempt to reduce those uncertainties, we try to find an empirical law describing SNe Ia's length of explosion (stretch) evolution with the redshift.
Methods. We started by getting a complete sample representing all of the stretch distribution that Nature can give us, before using LsSFR measurements, an age tracer which evolution with redshift is known, that has been shown to have a strong correlation with the stretch. We compare their AIC, an estimator of the relative quality of statistical models that includes the number of free parameters, to determine which ones describe best the data.
Results. Models with an evolution of the stretch with the redshift have a better AIC than the ones without.
Conclusions. We find that implementing these models allows us to fit the data better than models without stretch evolution.

Key words. Cosmology – Type Ia Supernova – Systematic uncertainties

1. Introduction

Type Ia supernovae (SNe Ia) are powerful cosmological distance indicators that have enabled the discovery of the acceleration of the Universe's expansion (Riess et al. 1998; Perlmutter et al. 1999). They remain today a key cosmological probe to understand the properties of dark energy (DE) as it is the only tool able to precisely map the recent expansion rate $z < 0.5$, when DE is driving it (e.g. Scolnic et al. 2019). They also are key to directly measure the Hubble Constant (H_0), provided one can calibrate their absolute magnitude (Riess et al. 2016; Freedman et al. 2019). Interestingly, the value of H_0 derived when the SNe Ia are anchored on Cepheids (the SH0ES project Riess et al. 2009, 2016) is $\sim 5\sigma$ higher than what is predicted by cosmic microwave background (CMB) data measured by Planck assuming the standard ΛCDM, or when the SN luminosity is anchored at intermediate redshift by the baryon acoustic oscillation (BAO) scale (Riess et al. 2019; Reid et al. 2019; Planck Collaboration et al. 2016; Feeney et al. 2019). While using the tip of the red giant branch technique in place of the Cepheids seem to favor an intermediate value of H_0 (Freedman et al. 2019), time delay measurements from strong lensing seem to also favor high H_0 values (Wong et al. 2019).

The H_0 tension has received a lot of attention as it could be a sign of new fundamental physics. Yet, no simple solution is able to accommodate this H_0 tension when accounting for all other probes (Knox & Millea 2019) and the current most promising scenario appears to be a burst of expansion at the matter-radiation decoupling moment caused by a (fine tuned) early dark energy (Poulin et al. 2019).

Alternatively, systematic effects affecting one or several of the aforementioned analysis could also explain the tension. In Rigault et al. (2015) we suggested that SNe Ia from the Cepheid calibrator sample differs by construction from the Hubble flow as it strongly favors young stellar population environments where one could find Cepheids. This selection effect would impact the derivation of H_0 if SNe Ia from young and older environments differ in average magnitudes.

For the last decades, numerous analysis have studied the relation between SNe Ia and host properties, finding first that the standardized SNe Ia magnitude significantly depends on the host stellar-mass, SNe Ia from high-mass host being brighter on average (e.g. Kelly et al. 2010; Sullivan et al. 2010; Childress et al. 2013; Betoule et al. 2014; Rigault et al. 2018; Kim et al. 2019). This mass-step correction is currently used in cosmological analyses (e.g. Betoule et al. 2014; Scolnic et al. 2018), including for deriving H_0 (Riess et al. 2016, 2019). Yet, the underlying connection between the SN and the host remains unclear when using global properties such as the host stellar mass, which raises the question of the accuracy of the correction. More recently, studies have used the local SN environment to probe more direct connections between the SN and its environments (Rigault et al. 2013), showing that local age-tracers such as the local specific star formation rate or the local color are more strongly correlated with the standardized SN magnitude (Rigault et al. 2018; Roman et al. 2018; Kim et al. 2018), suggesting the age to be the driving parameter underlying the mass-step. If true, this would have a significant impact for cosmology because the redshift drift magni-

⋆ n.nicolas@ipnl.in2p3.fr, equal contribution
⋆⋆ m.rigault@ipnl.in2p3.fr, equal contribution

tude correction to apply might strongly vary (Rigault et al. 2013; Childress et al. 2014; Scolnic et al. 2018). Yet, the importance of local SN environmental study remains highly debated (e.g. Jones et al. 2015, 2019) and especially the impact it has on the derivation of H_0 (Jones et al. 2015; Riess et al. 2016, 2018; Rose et al. 2019).

In this paper, we take a step aside to probe the validity of our modeling of the SN population that we claim to be constituted of two age-populations (Rigault et al. 2013; Rigault et al. 2015, 2018): one old and one young, the former having on average lower lightcurve stretches and being brighter after standardization. We indeed use the correlation between the SN age probed with local specific star formation rate in Rigault et al. (2018) and the SN stretch to model the expected evolution of the underlying SN stretch distribution as a function of redshift. This modeling relies on three assumptions: (1) there are indeed 2 populations of SNe Ia; (2) the relative fraction of each of these populations as a function of redshift follows the age-modeling made in (Rigault et al. 2018) and (3) the underlying distribution of stretch for each age-sample can be modeled and do not significantly drift. This paper aims at testing this with datasets from the literature.

The concept of the SNe Ia age dichotomy aroused when the SN Ia rate has first been studied. Mannucci et al. (2005); Scannapieco & Bildsten (2005); Sullivan et al. (2006); Aubourg et al. (2008) have shown that the relative SNe Ia rate in galaxies could only be explained if two populations existed, one young, following the host star formation activity, and one old following the host stellar mass (the so called "prompt and delayed" or "A+B" model). In Rigault et al. (2018) we used the specific Star Formation Rate at the SN location (Local sSFR or LsSFR) to classify which are the prompts (those with a high LsSFR) and which are the delayed (those with low LsSFR). Furthermore, since the first SNe Ia host analysis, the SN stretch has been known to be strongly correlated with the SN host properties (Hamuy et al. 1996, 2000), which has been extensively confirmed since (e.g. Neill et al. 2009; Sullivan et al. 2010; Lampeitl et al. 2010; Kelly et al. 2010; Gupta et al. 2011; D'Andrea et al. 2011; Childress et al. 2013; Rigault et al. 2013; Pan et al. 2014; Kim et al. 2019). Following the "A+B" model and the connection between SN stretch and host properties, Howell et al. (2007) first discussed the potential redshift drift of the SN stretch distribution. In this paper we revisit this analysis with the most recent SNe Ia dataset.

We present in section 2 the sample we are using for this analysis, which is based on the Pantheon dataset (Scolnic et al. 2018). We discuss the importance of obtaining a "complete" sample, i.e. representative of the true underlying SNe Ia distribution, and how we build one from the Pantheon sample. We then present in section 3 our modeling of the distribution of stretch as a function of redshift based stretch distribution of young and old SNe Ia. Our results are presented in section 4 where we test if the SN stretch does evolve as a function of redshift and if the age model is in good agreement with this evolution. Consequence for cosmology of our results are briefly discussed in section 5 and we conclude in section 6.

2. Complete Sample Construction

We base our analysis on the latest SNe Ia compilation: the Pantheon catalog from Scolnic et al. (2018). A naive approach to test the SN stretch redshift drift would be to simply compare the observed SN stretch distributions in a few bins of redshift. In practice, selection effects are affecting the observed SN stretch distributions. Indeed, because the observed SN Ia magnitude correlates with the lightcurve stretch (and color), the first SN Ia that a magnitude-limited survey will miss are the lowest-stretch (and reddest) ones. Consequently, if selection effects are not accounted for, one might confuse true population drift with selection effects, and conversely (e.g., Kim et al. 2019).

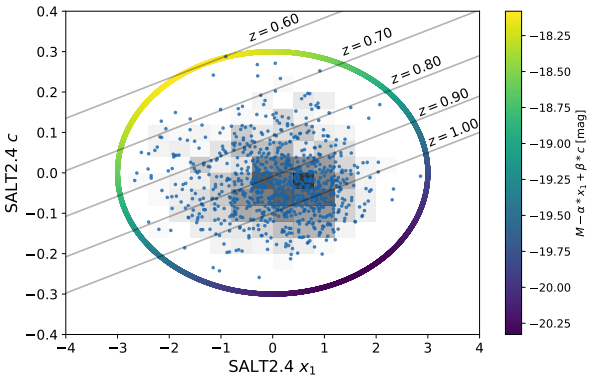

Fig. 1. SALT2.4 stretch (x_1) and color (c) lightcurve parameters of SNe Ia from the SDSS, PS1 and SNLS samples from the Pantheon catalog. The individual SNe are shown as blue dots and a 2D histogram is shown in gray to highlight point density. The ellipse ($x_1 = \pm 3, c = \pm 0.3$) is displayed, colored by the corresponding standardized absolute magnitude using the α and β coefficients from Scolnic et al. (2018). The lines represent the (x_1, c) evolution for $m = m_{lim}$ in Eq. 2, for z between 0.60 and $z = 1.00$ using SNLS's m_{lim} of 24.8 mag. The SNe above these lines are those for which $m > m_{lim}$, meaning they cannot be observed.

Assuming sufficient (and random unbiased?) spectroscopic follow-up for acquiring typing and host redshift, the selection effects of magnitude-limited surveys should be negligible bellow a given redshift where even the faintest SN Ia can be observed. In contrast, targeted surveys have highly complex selection functions and will be discarded from our analysis. Fortunately, modern SN cosmology samples such as the Pantheon dataset are now dominated by magnitude limited surveys.

We present in Fig. 1 the lightcurve stretch and color distributions of SNe Ia from the following surveys: PanStarrs (PS1 Rest et al. 2014), the Sloan Digital Sky Survey (SDSS Frieman et al. 2008) and the SuperNovae Legacy Survey (SNLS Astier et al. 2006). An ellipse in the SALT2.4 (x_1, c) plan with $x_1 = \pm 3$ and $c = \pm 0.3$ encapsulates the full distribution (Guy et al. 2007; Betoule et al. 2014); see also Bazin et al. (2011) and Campbell et al. (2013) for similar contours, the second using a more conservative $|c| \leq 0.2$ cut. Assuming the SN absolute magnitude with $x_1 = 0$ and $c = 0$ is $M_0 = -19.36$ (Kessler et al. 2009; Scolnic et al. 2014), we can derive the absolute standardized magnitude at maximum of light $M = M_0 - \alpha x_1 + \beta c$ along the aforementioned ellipse given the standardization coefficient $\alpha = 0.156$ and $\beta = 3.14$ from Scolnic et al. (2018): the faintest SN Ia is that with $(x_1 = -1.66, c = 0.25)$ and an absolute standardized magnitude at peak in Bessel-B band of $M_{min}^{t_0} = -18.31$ mag. Since one need to detect this object typically a week before and 10 days after peak to build a suitable lightcurve, the limiting standardized absolute magnitude effectively is approximately $M_{min} = -18.00$ mag. Hence, given the ~~5σ point source detection~~ magnitude limit m_{lim} of a magnitude limited survey, one can derive the maximum redshift z_{lim} above which the faintest SNe Ia will be missed using the relation between apparent magnitude, redshift and absolute magnitude given by:

$$m = \mu(z) + \underbrace{M_{min}^{t_0 - 10}(x_1, c)}_{=-18.00 \text{ mag}} < m_{lim} \quad (1)$$

$\mu(z_{lim}) = m_{lim} - M_{min}$ min vs. lim??? (2)

SNLS typically acquires SNe Ia in the redshift range $0.4 < z < 0.8$. At these redshifts the rest-frame Bessel-B band roughly corresponds to the SNLS-i filter that has a 5σ depth of 24.8 mag[1]. This converts to a $z_{lim} = 0.60$, in perfect agreement with Neill et al. (2006), Perrett et al. (2010) and Bazin et al. (2011). Fig. 14 of Perrett et al. 2010 suggests a lower limit of $z_{lim} = 0.55$; both will be considered as discussed below.

Similarly, PS1 observes SNe Ia in the range $0.2 < z < 0.4$, their g-band 5σ depth is 23.1 mag (Rest et al. 2014), which yields to $z_{lim} = 0.30$ in agreement with, e.g., Fig. 6 of Scolnic et al. (2018). This figure could also suggest a more conservative z_{lim} of 0.27.

In similar redshift range, SDSS has a limiting magnitude of 22.5 (Dilday et al. 2008; Sako et al. 2008), which would lead to a $z_{lim} = 0.24$. However, the SDSS surveys were more sensitive to limited spectroscopic resources. Kessler et al. (2009) pointed out that during year-1 of SDSS, SNe Ia with $r-mag < 20.5$ were favored for spectroscopic follow up, corresponding to a redshift cut at 0.15. For the rest of the SDSS survey, additional spectroscopic resources were used, and Kessler et al. (2009) and Dilday et al. (2008) show a relative completeness up to $z_{lim} = 0.2$. Following these analyses we will use $z_{lim} = 0.2$ as the baseline SDSS redshift limit ~~for the rest of this study~~.

For the rest of the analysis, we will consider ~~both cases,~~ the aforementioned redshift limits as well as more conservative cuts for each of these surveys, namely $z_{lim} = 0.15$ for SDSS, $z_{lim} = 0.27$ for PS1 and $z_{lim} = 0.55$ for SNLS.

The complete sample selection is summarized in Table 1, and the redshift distribution of these three surveys are shown in Fig. 2. We see there that the redshift limits we have selected roughly correspond to the peak of these histograms. As the peaks indicate that we are starting to miss the faintest SNe Ia, it seems logical to have the redshift cuts fall near said peaks.

In addition, we use the SNe Ia from the Nearby Supernova Factory (SNfactory Aldering 2004) published in Rigault et al. (2018) and that have been discovered from non-targeted searches (114 SNe Ia, see their section 3 and 4.2.2). Their spectroscopic follow-up were done on a redshift range of $0.02 < z < 0.09$ (as in Rigault et al. 2018), while the searches were much deeper. As such, these SNe Ia should also be a random sampling of the underlying SN population. The SNf sample is particularly useful for studying SN property drift as it enables us to have a large SN Ia sample at $z < 0.1$.

Finally, we included the HST sample from Pantheon, that follows the same logic of having a search deeper than the follow-up and we therefore kept it entirely (Strolger et al. 2004), see Table 1. We present the stretch distribution and redshift histogram of these five surveys in Fig. 3.

3. Modeling the redshift drift

In Rigault et al. (2018) we presented a modeling of the evolution of the fraction of prompt and delayed SNe Ia as a function of redshift following former work on rates and delay time distributions (e.g., Mannucci et al. 2005; Scannapieco & Bildsten 2005; Sullivan et al. 2006; Aubourg et al. 2008; Childress et al. 2014; Maoz et al. 2014). In short, we assumed that the number of prompt SNe Ia follows the star formation activity (SFR) in the

[1] CFHT final release website.

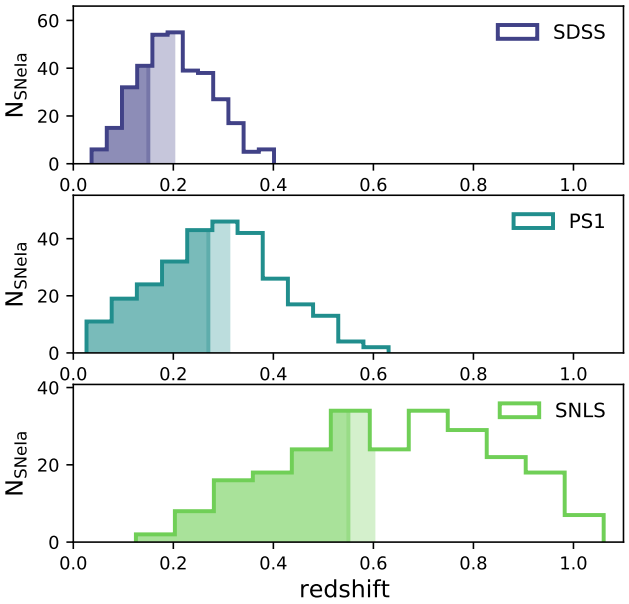

Fig. 2. From top to bottom: histogram of SN Ia redshifts for SDSS, PS1 and SNLS from the Pantheon dataset. The colored part of the distribution shows which SNe Ia has been kept in our analysis for they are supposedly free from selection bias (see section 2). The dark (resp. light) color responds to the conservative (resp. fiducial) selection cut.

Table 1. Source surveys and number of SNe Ia used in our "complete" sample. Conservative cuts are indicated in brackets.

Survey	z_{lim}	N_{SN}
SNf	–	114
SDSS	0.20(0.15)	167(82)
PS1	0.30(0.27)	160(122)
SNLS	0.60(0.55)	102(78)
HST	–	26
Total	–	569(422)

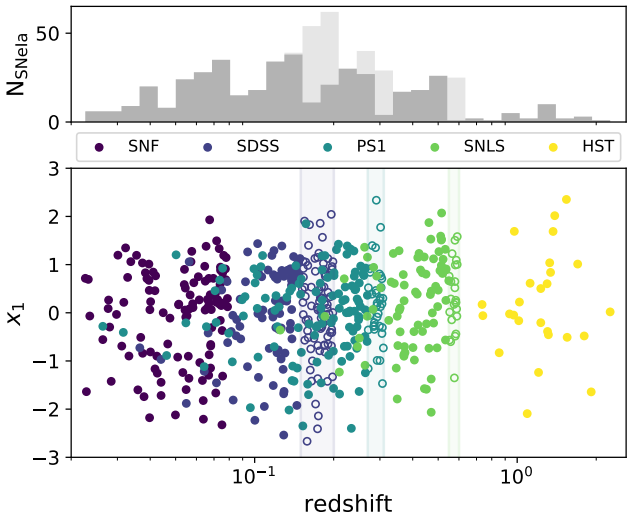

Fig. 3. *Main*: SN stretches in each survey as a function of redshift, using solid (resp. empty) markers for the conservative (resp. fiducial) selection. *Top*: combined redshift histogram for all surveys, in dark (resp. light) gray for the conservative (resp. fiducial) selection cuts.

Universe while the number of delayed SNe Ia follows the number of Gyr-old stars in the Universe, i.e. the stellar mass (M). Hence, if we denote $\delta(z)$ (resp. $\psi(z) = 1 - \delta(z)$) the fraction of young (resp. old) SNe Ia in the Universe as a function of redshift, then their ratio δ/ψ is expected to follow the evolution of the specific star formation rate (SFR/M) which goes as $(1+z)^{2.8}$ until $z \sim 2$ (e.g., Tasca et al. 2015). Since $\delta(0.05) \sim \psi(0.05)$ (Rigault et al. 2013; Rigault et al. 2018; Wiseman et al. 2020), in agreement with rate expectations (Mannucci et al. 2006; Rodney et al. 2014), we conclude in Rigault et al. (2018) that

$$\delta(z) = \left(K^{-1} \times (1+z)^{-2.8} + 1\right)^{-1} \quad (3)$$

with $K = 0.87$. This model is comparable to the evolution predicted by Childress et al. (2014) based on SN rates in galaxies depending on their quenching time as a function of their stellar mass.

3.1. Base underlying distribution

To model the evolution of the SN stretch as a function of redshift, given our aforementioned model of the evolution of the fraction of young and old SNe Ia with redshift, we need to model the SN stretch distribution for each age subsample.

In Rigault et al. (2018), we presented the relation between SNe stretches and LsSFR measurements, which trace progenitor age, from the SNf sample, and is shown in Fig. 4. Given the structure of the stretch-LsSFR scatter plot, our model is defined as follows:

- for the young population the underlying stretch distribution is modeled as a normal distribution $\mathcal{N}(\mu_1, \sigma_1^2)$;
- the old population stretch distribution is modeled as $a \times \mathcal{N}(\mu_1, \sigma_1^2) + (1-a) \times \mathcal{N}(\mu_2, \sigma_2^2)$, i.e. a Gaussian mixture where one mode is the same as the young population one. explain meaning of a

When combining our prediction of the evolution of young SNe Ia as a function of redshift (Eq. 3), and the stretch distribution of both young and old SNe Ia, our model for the underlying distribution of stretch $X_1(z)$ as a function of redshift is:

$$X_1(z) = \delta(z) \times \mathcal{N}(\mu_1, \sigma_1^2) + (1 - \delta(z)) \times \left[a \times \mathcal{N}(\mu_1, \sigma_1^2) + (1-a) \times \mathcal{N}(\mu_2, \sigma_2^2)\right]. \quad (4)$$

The maximum-likelihood estimate of the 5 free parameters $\theta \equiv (\mu_1, \mu_2, \sigma_1, \sigma_2, a)$ of the model is obtained by minimizing a pseudo-χ^2 $\mathcal{L}$ defined as:

$$\mathcal{L} = -2 \sum_i \ln \mathcal{P}\left(x_1^i \mid \theta; \mathrm{d}x_1^i, y^i\right), \quad (5)$$

with

$$\mathcal{P}\left(x_1^i \mid \theta; \mathrm{d}x_1^i, y^i\right) = y^i \times \mathcal{N}\left(x_1^i \mid \mu_1, \sigma_1^2 + \mathrm{d}{x_1^i}^2\right) + (1 - y^i) \times \left[a \times \mathcal{N}\left(x_1^i \mid \mu_1, \sigma_1^2 + \mathrm{d}{x_1^i}^2\right) + (1-a) \times \mathcal{N}\left(x_1^i \mid \mu_2, {\sigma_2}^2 + \mathrm{d}{x_1^i}^2\right)\right] \quad (6)$$

where i is the index of the SN Ia, x_1^i, $\mathrm{d}x_1^i$ and y^i are the `SALT2.4` stretch, its associated error and the probability that the SN is young, respectively. In practice, to ensure fraction a is constrained between 0 and 1, we fit for α such that $a = \arctan(\alpha)/\pi + 0.5$, which results into asymmetric error on a.

Depending on whether y^i can be estimated directly from LsSFR measurements or not, there are two ways to proceed.

3.1.1. With LsSFR measurements

For the SNfactory sample, we can readily set $y^i = {p_y}^i$, the probability to have $\log \mathrm{LsSFR} \geq -10.82$ (see Fig. 4), to minimize Eq. 5 with respect to θ. Results on fitting the SNf SNe with this model are shown Table 2 and illustrated in Fig. 4.

3.1.2. Without LsSFR measurements

In lack of direct LsSFR measurements (i.e. p_y^i), we can extend the analysis to non-SNfactory samples by using the redshift-evolution of the fraction $\delta(z)$ of young SNe Ia (Eq. 3) as a proxy for the probability of a SN to be young. This still corresponds to minimizing Eq. 5 with respect to the parameters $\theta \equiv (\mu_1, \mu_2, \sigma_1, \sigma_2, a)$ of the stretch distribution X_1 (Eq. 4), but this time assuming $y^i = \delta(z^i)$ for any given SN i.

For the rest of the analysis, we will therefore minimize Eq. 5 using ${p_y}^i$ – the probability for the SN i to be young – when available (i.e. for SNf dataset) and $\delta(z^i)$ – the expected fraction of young SNe Ia at the SN redshift z^i – otherwise.

Results on fitting this model on all the 569 (resp. 422) SNe from the fiducial (resp. conservative) sample are given Table 2 and illustrated in Fig. 5. We see in this figure that the measured mean SN Ia stretch per bin of redshift closely follows our redshift drift modeling; that is, when considering selection-bias free SN Ia samples, SNe Ia at higher redshift have on average larger stretch (0.34 ± 0.10 at $z \sim 0.65$) than those at lower redshift (-0.17 ± 0.10 at $z \sim 0.05$). This is indeed what is expected if old environments favor low SN stretches (e.g. Howell et al. 2007) *and* if the fraction of old SNe Ia reduces as a function of redshift. See section 4 for a more quantitative discussion.

3.2. Alternative models

In section 3.1, we have modeled the underlying stretch distribution following Rigault et al. (2018), i.e. as a single Gaussian for the young SNe Ia and a mixture of two Gaussians for the old SNe Ia population, the same as the young-population one plus another for the fast declining SNe Ia that seem to only exist in old populations. This is our so-called "Base" model.

However, to test different modeling choices, we have implemented a suite of alternative parametrizations that we also adjust to the data following the procedure described section 3.1.2.

Howell et al. (2007) used a simpler model with a normal distribution for each young and old populations. We thus consider a "Howell+drift" model, with one single Gaussian per age group. one has to explain the "drift" part!!!

Alternatively, since we aim at probing the existence of a redshift drift, we are also testing non-drifting models by restricting the "Base" and "Howell" models to use a supposedly redshift-independent fraction $\delta(z) = f$ of young SNe; these models are labeled "Base constant" and "Howell constant". fix labels!!!

We are also considering other non-drifting models, and notably the one developed for the Beam with Bias Correction (BBC, Scolnic & Kessler 2016; Kessler & Scolnic 2017), currently used in recent SN cosmological analyses (e.g. Scolnic et al. 2018; Abbott et al. 2019; Riess et al. 2016, 2019) to account for Malmquist biases. The BBC formalism assumes sample-based (hence intrinsically non-drifting) asymmetric Gaussian stretch distributions: $\mathcal{N}\left(\mu, \sigma_-^2 \text{ if } x_1 < \mu \text{ else } \sigma_+^2\right)$. The idea behind this sample-based approach is twofold: (1) Malmquist biases are driven by survey properties and (2) because current surveys cover limited redshift ranges, doing so covers some po-

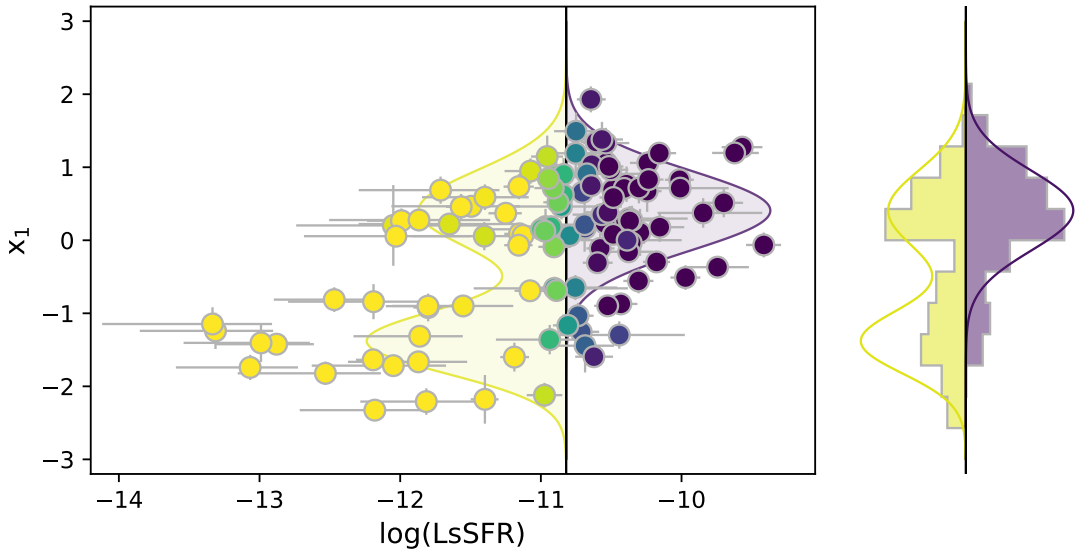

Fig. 4. *Main*: SALT2.4 lightcurve stretch (x_1) as a function of the local specific star formation rate (LsSFR) for SNf SNe used in this analysis. The color corresponds to the probability p_y for the SNe Ia to be young, i.e. to have $\log \mathrm{LsSFR} \geq -10.82$; see Rigault et al. (2018). *Right*: histogram of p_y-weighted distribution of the SN stretchs. On both panel, we illustrate the Base model fitted on the SNf SNe. The young and old population contributions are shown in purple and yellow, respectively.

Table 2. Best fit values of the parameters for the Base stretch distribution model when applied to the SNf dataset only (114 SNe Ia), the reference 569 SN Ia sample or the conservative one (422).

Sample	μ_1	σ_1	μ_2	σ_2	a
SNf	0.41 ± 0.08	0.55 ± 0.06	-1.38 ± 0.10	0.44 ± 0.08	$0.48^{+0.08}_{-0.08}$
Reference	0.37 ± 0.05	0.61 ± 0.04	-1.22 ± 0.16	0.56 ± 0.10	$0.51^{+0.09}_{-0.10}$
Conservative	0.38 ± 0.05	0.60 ± 0.04	-1.26 ± 0.13	0.53 ± 0.08	$0.47^{+0.09}_{-0.08}$

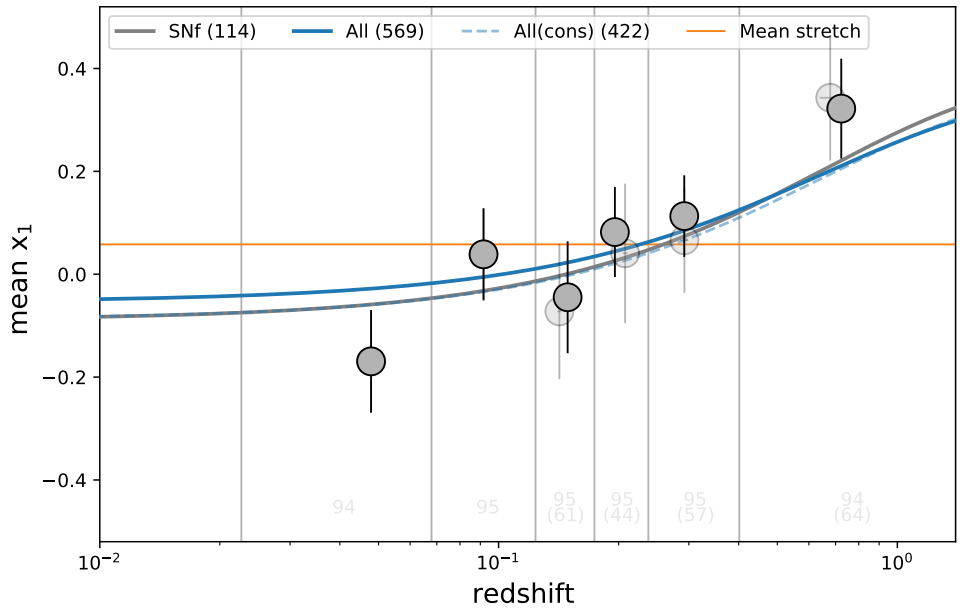

Fig. 5. Evolution of the mean SN SALT2.4 stretch (x_1) as a function of redshift. Markers show the mean stretch measured in redshift bins of equal sample size indicated in light gray at the bottom of each redshift bins. Full and light markers are used when considering the fiducial or the conservative samples, respectively. The orange horizontal line represents the mean redshift of the reference sample illustrating the expectation if the SN stretch distribution is not evolving with redshift. Best fits of our Base drifting model are shown as blue, dashed-blue and gray, when fitted on the fiducial sample, the conservative one or the SNf dataset only, respectively; all are compatible. The light-blue illustrates the amplitude of the error of the best fit model when considering the reference dataset.

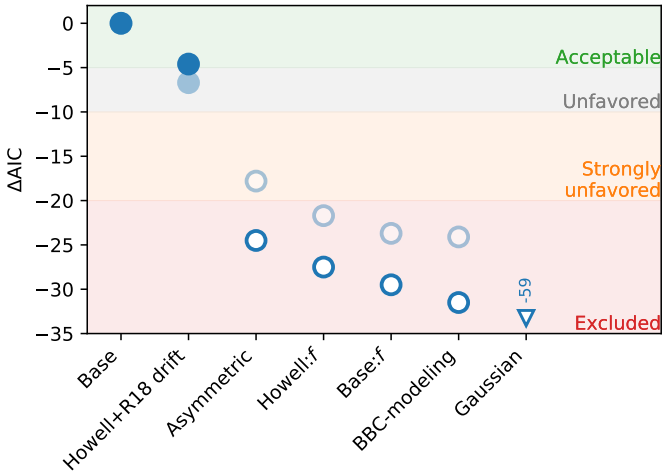

Fig. 6. ΔAIC between the reference "Base" model and the other models (see also Table 3). The full and open blue markers correspond to model with and without redshift drift, respectively. Light markers show the results when this analysis is performed on the conservative sample rather than the fiducial one. Color-bands illustrate the validity of the models, from Acceptable (ΔAIC > −5) to Excluded (ΔAIC < −20), see text This is not explained in the text!. According to the AIC, all non-drifting models (open symbols) are excluded for the reference sample.

tential redshift evolution information (Scolnic & Kessler 2016; Scolnic et al. 2018). See further discussion concerning the BBC in section 5.

Finally, for the sake of completeness, we also consider simple "Gaussian" and "asymmetric" Gaussian non-drifting models.

4. Results

We applied each model on both the reference and conservative samples (cf. section 2). Because the models presented in section 3 have different degrees of freedom, we use the Akaike Information Criterion (AIC, e.g. Burnham & Anderson 2004) to compare them. This estimator penalizes extra degrees of freedom to avoid over-fitting the data. It is defined as follow:

$$\mathrm{AIC} = \mathcal{L}_{min} + 2k \tag{7}$$

where $\mathcal{L}_{min}$ is the minimum value of the pseudo-χ^2 as defined Eq. (5) and k is the number of free parameters. The reference model is the one with the smallest AIC, and the probability for another model to be at least as representative of the data as this one is given by:

$$p(\mathrm{other} \geq \mathrm{ref}) = \exp(\Delta\mathrm{AIC}/2) \tag{8}$$

with ΔAIC < 0. Results are presented in Table 3 and are illustrated Fig. 6.

We first note that the best model (with smallest AIC) is the so-called Base model, implementing Rigault et al. (2018) description of the stretch distribution and its redshift evolution. This is therefore the reference model, both on the fiducial and conservative samples.

Furthermore, we find that underlying stretch distributions not including a redshift drift are all excluded to describe the data as well as the Base model. In fact, the best non drifting model (Asymmetric) has a very marginal chance ($p = 2 \times 10^{-6}$) not same number as in table!!! to describe the data as well as the Base model. This result is just a quantitative assessment of qualitative facts clearly visible in Fig. 5: the mean SN stretch per bin of redshift strongly suggests a redshift evolution rather than having a constant value, and this evolution is compatible with the redshift drift expected from Eq. 3.

Are the next 2 paragraphs (now commented out) still pertinent? The model does not appear in table nor fig. Mickael, please rephrase and remove what is not needed anymore.

This paragraph is not clear: what is the impact, and its importance, on which case? Isn't it in contradiction with the conclusions of BBC section, where it is said that proper handling of stretch correction is of paramount importance for accurate cosmology??? To the best of our knowledge, the SN Ia stretch redshift drift has never been explicitly accounted for in cosmological analyses. Not doing so is a second order issue for SN cosmology as it only affects the way one does Malmquist bias correction. Indeed, as long as the Phillip's relation standardization parameter α is not redshift dependent (a study behind the scope of this paper, but see e.g. Scolnic et al. 2018), if a sample is free from selection effects, the stretch-corrected SNe Ia magnitudes used for cosmology are blind to the underlying stretch distribution. However, since the first missed SNe Ia are the faintest, which are also those with the lowest stretch, once we include SNe Ia affected by selection effects we need to account for the fact that only the brightest SNe Ia exist in the low-stretch end of the sample after standardization. The modeling of the SN stretch then enters the analysis to understand which SNe Ia are missing given the selection functions to mimic a selection-bias free sample.

5. BBC discussion

The state-of-the-art tool for doing so what? is the BBC described in Scolnic & Kessler (2016) and Kessler & Scolnic (2017), which is used in recent SN cosmological analyses (Jones et al. 2018; Scolnic et al. 2018; Brout et al. 2019; Abbott et al. 2019) including the direct measurement of H_0 (Riess et al. 2016, 2019).

Surprisingly, the BBC technique, which assumes an individual asymmetric distribution per sample (Scolnic & Kessler 2016; Kessler & Scolnic 2017), is one of the worst modelisation in our analysis. While its pseudo-χ^2 $\mathcal{L}$ is the smallest of all non-drift models (but still $\Delta\mathcal{L} = -11.5$ worse than the reference Base model), it is penalized by requiring 15 free parameters (3 for each of the 5 samples of the analysis). Scolnic & Kessler (2016, section 2) and Scolnic et al. (2018, section 5.4) highlighted that, because traditional surveys span limited redshift ranges, the per sample approach can somewhat account for implicit redshift drifts. We stress here that, as measurements of modern surveys cover increasingly larger redshift ranges to reduce calibration systematic uncertainties, this approach is becoming less valid, notably for PS1, DES and, soon, LSST redshift ranges? Why not SNLS too?. Already with the current surveys, the BBC sample-based technique is ruled out to be an as good representation of the data as our best model ($p = 4 \times 10^{-8}$). A more detailed discussion of the consequence of this results for cosmology follows in section 5.

We report in Table 4 the samples' $\sigma_\pm$ and μ_0 that we find in close agreement with Scolnic & Kessler (2016) for SNLS and SDSS and with Scolnic et al. (2018) for PS1. This validates our construction of a selection-bias free sample, see section 2. WHY??? The developed BBC for *incomplete* samples! If we were to use Scolnic & Kessler (2016) and Scolnic et al. (2018) best fit values for SNLS, SDSS and PS1, respectively, the ΔAIC between our Base drifting model and the BBC modeling would go even deeper from −32 to −47. It would seem that our minimization of these parameters is closer to our Base

Table 3. Comparison of the relative ability for a model to describe the data using AIC criterion. See the model definition section 3. The drift column indicates if the model contains the age-drifting model ($\delta(z)$) or not (constant f, i.e. non-drifting). The BBC is sample based which somewhat includes redshift dependency information. $\mathcal{L}$ is the pseudo-χ^2 defined Eq. 5 and Proba is the probability that a model is at least as good as the reference model to describe the data. See also Fig. 6.

			Fiducial sample (569 SNe)				Conservative sample (422 SNe)			
Name	drift	k	$\mathcal{L}$	AIC	ΔAIC	Proba	$\mathcal{L}$	AIC	ΔAIC	Proba
Base	$\delta(z)$	5	1456.7	1466.7	–	–	1079.5	1089.5	–	–
Howell	$\delta(z)$	4	1463.3	1471.3	−4.6	1.0×10^{-1}	1088.2	1096.2	−6.7	3.4×10^{-2}
Asymmetric	–	3	1485.2	1491.2	−24.5	4.7×10^{-6}	1101.3	1107.3	−17.8	1.4×10^{-4}
Howell	f	5	1484.2	1494.2	−27.5	1.0×10^{-6}	1101.2	1111.2	−21.7	1.9×10^{-5}
Base	f	6	1484.2	1496.2	−29.5	3.9×10^{-7}	1101.2	1113.2	−23.7	7.1×10^{-6}
BBC	per sample	3x5	1468.2	1498.2	−31.5	1.5×10^{-7}	1083.6	1113.6	−24.1	5.7×10^{-6}
Gaussian	–	2	1521.8	1525.8	−59.1	1.5×10^{-13}	1142.6	1146.6	−57.1	4.0×10^{-13}

Table 4. Best fit parameters for our sample-based asymmetric modeling of the underlying stretch distribution.

Asymmetric	σ_-	σ_+	μ_0
SNf	1.34 ± 0.13	0.41 ± 0.10	0.68 ± 0.15
SDSS	1.31 ± 0.11	0.42 ± 0.09	0.72 ± 0.13
PS1	1.01 ± 0.11	0.52 ± 0.12	0.38 ± 0.16
SNLS	1.41 ± 0.13	0.15 ± 0.13	1.22 ± 0.15
HST	0.76 ± 0.36	0.79 ± 0.35	0.11 ± 0.44

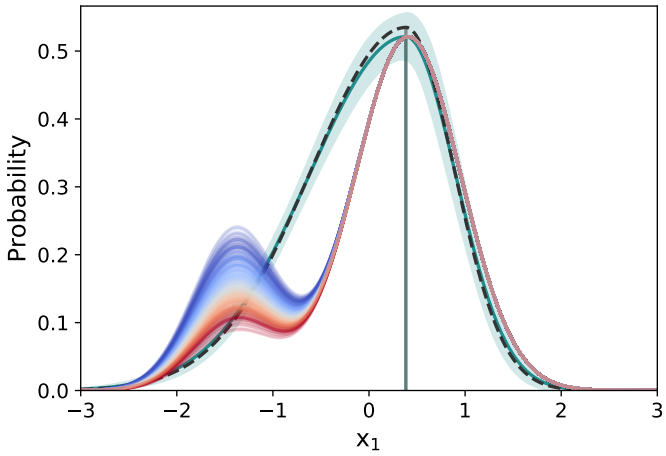

Fig. 7. Underlying SN SALT2.4 stretch (x_1) distributions for the PS1 sample. The Green line (band) shows our best fitted asymmetric Gaussian distribution (its error) mimicking the BBC approach. The dashed line shows the noe ??? estimated by Scolnic et al. (2018). The blue to red curves display our Base drifting stretch model estimated at each PS SN redshift; bluer for lower redshifts.

model than theirs. ??? Last, fixing the Base model parameters to those derived using SNf data only (top row Table 2) and using the BBC parameters from the aforementioned literature for PS1, SDSS and SNLS, we find ΔAIC = −31.7, but removing the SNf sample reduces this difference to ΔAIC = −5 not clear: the SNf sample is removed or not???. Such a dramatic apparent improvement is expected for two reasons. First, it removes the nearby SNe Ia, and thereby significantly reduces the redshift lever arm to probe the impact of redshift drifts. Second, the SNf data are better at discriminating the individual stretch distributions by age since the probability y^i for a SN to be young is known ($y^i = p_y{}^i$) while other samples rely on a probabilistic proxy for y^i, namely the fraction $\delta(z^i)$ of young SNe Ia at SN redshift. While in any case the Base model is favored, the relative weight of the SNf sample highlights the importance of adding non-targeted nearby surveys, such as the Zwicky Transient Facility (ZTF, Bellm et al. 2019; Graham et al. 2019) to unambiguously confirm our model.

We illustrate in Fig. 7 the prediction difference in the underlying stretch distribution between the BBC modeling and our Base drifting model for the PS1 sample. Our model is bimodal and the relative amplitude of each mode depends on the redshift-dependent fraction of old SNe Ia in the sample: the higher the fraction of old SNe Ia, the higher the amplitude of the low-stretch mode. As expected the two approaches strongly differ in modeling the negative part of the SN stretch distribution. The BBC asymmetric distribution goes in the middle of the bimodal distribution, largely over-estimating the number of SNe Ia at $x_1 \sim -0.7$ and largely under-estimating it at $x_1 \sim -1.7$ for typical PS1 SN redshifts. This means that the bias-correction of the individual SN stretch $\bar{\delta}_{x_1}$ (Kessler & Scolnic 2017) and consequently the magnitude bias correction μ_B is most likely inaccurate. However, given the complexity of the BBC analysis, a full study using our Base model given Eq. 4 in place of the sample-based asymmetric modeling would be necessary to have an exact understanding of the effect this inaccurate modeling has on cosmology.

Nonetheless, to have a rough estimation of the expected amplitude of the effect, we show Fig. 8 the difference of mean stretch correction $\alpha \times (\langle x_1 \rangle_{\rm BBC} - \langle x_1 \rangle_{\rm Base})$ as a function of redshift, using $\alpha = 0.156$ from Scolnic et al. (2018). We see in this figure that the amplitude of the magnitude bias caused by an inaccurate modeling of the underlying stretch distribution is of the order of few tens of millimag, which is the typical scale that exotic forms of dark energy could have on $\mu(z)$. In the era of modern cosmology where we aim at probing w_0 at a sub-percent and w_a at a few percent (e.g., **?**), it is therefore of paramount importance to further understand the exact modeling of the lightcurve parameters and especially their potential redshift drift; see also discussions about the impact of stretch and color distribution modeling for cosmology in Rubin et al. 2015 and Rubin & Hayden 2016.

6. Conclusion

We have presented a study of the drift of the underlying I still prefer "intrinsic" SNe Ia stretch distribution as a function of red-

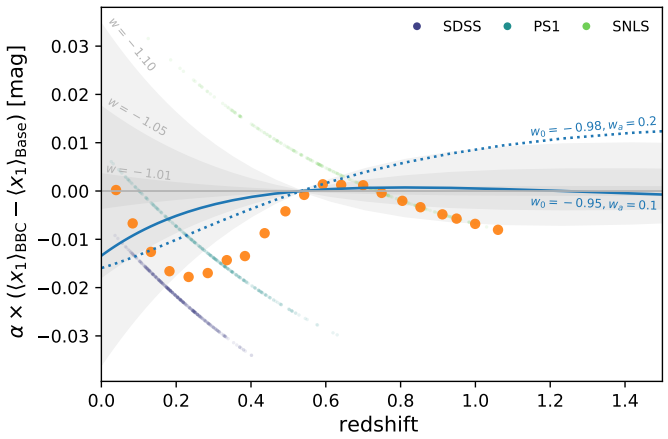

Fig. 8. Difference of mean stretch standardization (in mag) between the BBC algorithm used in Scolnic et al. (2018) and our Base drifting model. The small colored points show the SDSS, PS1 and SNLS evolution, while the orange markers represent their mean evolution in 30 regular redshift bins. For comparison, we show in gray bands the expected magnitude difference between ΛCDM and wCDM cosmologies ($w = -1 \pm 10$, 5 and 1%) or, in blue, more exotic w_0, w_aCDM examples. This figure illustrates that the amplitude of the potential magnitude bias caused by inaccurate modeling of the underlying stretch distribution is comparable to the observational effect of exotic form of dark energy. Is this fig. still relevant?

shift. We used SNe Ia from magnitude-limited surveys from the Pantheon dataset (Scolnic et al. 2018, SDSS, PS1 and SNLS) as well as HST, to which we added SNfactory data from Rigault et al. (2018) for the low-redshift bin. We only considered the supernovae that have been discovered in redshift ranges of each surveys where selection effect are negligible. This way the observed SNe Ia stretch are random sampling of the true underlying distribution. This corresponds to 569 SNe Ia (422 when considering more conservative cuts).

Following observations made in Rigault et al. (2018) we introduced a redshift drift modeling which depends on the expected fraction of young and old SNe Ia as a function of redshift, each age-population having its own underlying stretch distribution. We have studied various suites of distribution, including non-drifting modeling. We also studied the stretch distribution modeling of the current version of the BBC algorithm used for Malmquist bias correction by recent Dark energy or Hubble constant cosmological analyses.

Our conclusions are the following.

1. Non-drifting models are excluded to be as good descriptions of the data as our Base model. This model assumes that: (1) the young population has an unimodal Gaussian stretch distribution while the old population stretch one is bimodal, one mode being the young one; (2) the evolution of the relative fraction of young and old SNe Ia follows the prediction made in Rigault et al. (2018).
2. Given 1., we conclude that the SNe Ia stretch is indeed drifting, as already suggested by e.g. Howell et al. (2007).
3. The BBC approach, which assumes each survey to have its own asymmetric Gaussian distribution, is excluded to be an as good description of the data as our drifting model ($p = 4\times10^{-8}$). Hence, the BBC sample-based approach does not accurately account for redshift drift contrary to the suggestion of Scolnic & Kessler (2016) and Scolnic et al. (2018).
4. An inaccurate underlying stretch distribution modeling is estimated to bias the mean standardized SNe Ia to the order of a few percents, which is comparable in scale to the observational signature of exotic forms of dark energy.
5. We suggest to use the following stretch population modeling:

$$X_1(z) = \delta(z) \times \mathcal{N}(\mu_1, \sigma_1^2) + (1-\delta(z)) \times \left[a \times \mathcal{N}(\mu_1, \sigma_1^2) + (1-a) \times \mathcal{N}(\mu_2, \sigma_2^2)\right] \quad (9)$$

with: $a = 0.50, \mu_1 = 0.37, \mu_2 = -1.22, \sigma_1 = 0.61, \sigma_2 = 0.56$ (see Table 2) and using the age-population drift modelization $\delta(z)$ defined in Rigault et al. (2018) with $K = 0.87$:

$$\delta(z) = \left(K^{-1} \times (1+z)^{-2.8} + 1\right)^{-1}. \quad (10)$$

Acknowledgements. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement n 759194 - USNAC).

References

Abbott, T. M. C., Abdalla, F. B., Alarcon, A., et al. 2018, Phys. Rev. D, 98, 043526
Abbott, T. M. C., Allam, S., Andersen, P., et al. 2019, ApJ, 872, L30
Aldering, G. 2004, APS April Meeting Abstracts 2004, V4.003
Astier, P., Guy, J., Regnault, N., et al. 2006, A&A, 447, 31
Ata, M., Kitaura, F.-S., Chuang, C.-H., et al. 2017, MNRAS, 467, 3993
Aubourg, É., Tojeiro, R., Jimenez, R., et al. 2008, A&A, 492, 631
Bazin, G., Ruhlmann-Kleider, V., Palanque-Delabrouille, N., et al. 2011, A&A, 534, A43
Bellm, E. C., Kulkarni, S. R., Graham, M. J., et al. 2019, PASP, 131, 018002
Betoule, M., Kessler, R., Guy, J., et al. 2014, A&A, 568, A22
Brout, D., Scolnic, D., Kessler, R., et al. 2019, ApJ, 874, 150
Burnham, K., Anderson, D., 2004, Sociological Methods & Research, 33, 2
Campbell, H., D'Andrea, C. B., Nichol, R. C., et al. 2013, ApJ, 763, 88
Chabanier, S., Millea, M., & Palanque-Delabrouille, N. 2019, MNRAS, 489, 2247
Childress, M., Aldering, G., Antilogus, P., et al. 2013, ApJ, 770, 108
Childress, M. J., Wolf, C., & Zahid, H. J. 2014, MNRAS, 445, 1898
Coles, P., & Jones, B. 1991, MNRAS, 248, 1
D'Andrea, C. B., Gupta, R. R., Sako, M., et al. 2011, ApJ, 743, 172
Dilday, B., Kessler, R., Frieman, J. A., et al. 2008, ApJ, 682, 262
Feeney, S. M., Peiris, H. V., Williamson, A. R., et al. 2019, Phys. Rev. Lett., 122, 061105
Freedman, W. L., Madore, B. F., Hatt, D., et al. 2019, ApJ, 882, 34
Frieman, J. A., Bassett, B., Becker, A., et al. 2008, AJ, 135, 338
Graham, M. J., Kulkarni, S. R., Bellm, E. C., et al. 2019, PASP, 131, 078001
Gupta, R. R., D'Andrea, C. B., Sako, M., et al. 2011, ApJ, 740, 92
Guy, J., Astier, P., Baumont, S., et al. 2007, A&A, 466, 11
Hamuy, M., Phillips, M. M., Suntzeff, N. B., et al. 1996, AJ, 112, 2391
Hamuy, M., Trager, S. C., Pinto, P. A., et al. 2000, AJ, 120, 1479
Howell, D. A., Sullivan, M., Conley, A., et al. 2007, ApJ, 667, L37
Jones, D. O., Riess, A. G., & Scolnic, D. M. 2015, ApJ, 812, 3 1
Jones, D. O., Riess, A. G., Scolnic, D. M., et al. 2018, ApJ, 867, 108
Jones, D. O., Scolnic, D. M., Riess, A. G., et al. 2018, ApJ, 857, 51
Jones, D. O., Scolnic, D. M., Foley, R. J., et al. 2019, ApJ, 881, 19
Kelly, P. L., Hicken, M., Burke, D. L., et al. 2010, ApJ, 715, 743
Kessler, R., Becker, A. C., Cinabro, D., et al. 2009, ApJS, 185, 32
Kessler, R., & Scolnic, D. 2017, ApJ, 836, 56
Kim, Y.-L., Smith, M., Sullivan, M., et al. 2018, ApJ, 854, 24
Kim, Y.-L., Kang, Y., & Lee, Y.-W. 2019, Journal of Korean Astronomical Society, 52, 181
Knox, L., & Millea, M. 2019, arXiv e-prints, arXiv:1908.03663
Lampeitl, H., Smith, M., Nichol, R. C., et al. 2010, ApJ, 722, 566
Mannucci, F., Della Valle, M., Panagia, N., et al. 2005, A&A, 433, 807
Mannucci, F., Della Valle, M., & Panagia, N. 2006, MNRAS, 370, 773
Maoz, D., Mannucci, F., & Nelemans, G. 2014, ARA&A, 52, 107
Neill, J. D., Sullivan, M., Balam, D., et al. 2006, AJ, 132, 1126
Neill, J. D., Sullivan, M., Howell, D. A., et al. 2009, ApJ, 707, 1449
Pan, Y.-C., Sullivan, M., Maguire, K., et al. 2014, MNRAS, 438, 1391
Perlmutter, S., Aldering, G., Goldhaber, G., et al. 1999, ApJ, 517, 565
Perrett, K., Balam, D., Sullivan, M., et al. 2010, AJ, 140, 518
Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2016, A&A, 594, A13
Poulin, V., Smith, T. L., Karwal, T., et al. 2019, Phys. Rev. Lett., 122, 221301
Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2018, arXiv e-prints, arXiv:1807.06209
Reid, M. J., Pesce, D. W., & Riess, A. G. 2019, arXiv e-prints, arXiv:1908.05625

Rest, A., Scolnic, D., Foley, R. J., et al. 2014, ApJ, 795, 44
Riess, A. G., Filippenko, A. V., Challis, P., et al. 1998, AJ, 116, 1009
Riess, A. G., Macri, L., Casertano, S., et al. 2009, ApJ, 699, 539
Riess, A. G., Macri, L. M., Hoffmann, S. L., et al. 2016, ApJ, 826, 56
Riess, A. G., Casertano, S., Yuan, W., et al. 2018, ApJ, 861, 126
Riess, A. G., Casertano, S., Yuan, W., et al. 2019, ApJ, 876, 85
Rigault, M., Copin, Y., Aldering, G., et al. 2013, A&A, 560, A66
Rigault, M., Aldering, G., Kowalski, M., et al. 2015, ApJ, 802, 20
Rigault, M., Brinnel, V., Aldering, G., et al. 2018, arXiv:1806.03849
Rodney, S. A., Riess, A. G., Strolger, L.-G., et al. 2014, AJ, 148, 13
Roman, M., Hardin, D., Betoule, M., et al. 2018, A&A, 615, A68
Rose, B. M., Garnavich, P. M., & Berg, M. A. 2019, ApJ, 874, 32
Rubin, D., Aldering, G., Barbary, K., et al. 2015, ApJ, 813, 137
Rubin, D., & Hayden, B. 2016, ApJ, 833, L30
Sako, M., Bassett, B., Becker, A., et al. 2008, AJ, 135, 348
Scannapieco, E., & Bildsten, L. 2005, ApJ, 629, L85
Scolnic, D., Rest, A., Riess, A., et al. 2014, ApJ, 795, 45
Scolnic, D., & Kessler, R. 2016, ApJ, 822, L35
Scolnic, D. M., Jones, D. O., Rest, A., et al. 2018a, ApJ, 859, 101
Scolnic, D. M., Lochner, M., Gris, P., et al. 2018, arXiv e-prints, arXiv:1812.00516
Scolnic, D., Perlmutter, S., Aldering, G., et al. 2019, Astro2020: Decadal Survey on Astronomy and Astrophysics, 2020, 270
Strolger, L.-G., Riess, A. G., Dahlen, T., et al. 2004, ApJ, 613, 200
Sullivan, M., Le Borgne, D., Pritchet, C. J., et al. 2006, ApJ, 648, 868
Sullivan, M., Conley, A., Howell, D. A., et al. 2010, MNRAS, 406, 782
Tasca, L. A. M., Le Fèvre, O., Hathi, N. P., et al. 2015, A&A, 581, A54
Wiseman, P., Smith, M., Childress, M., et al. 2020, arXiv e-prints, arXiv:2001.02640
Wong, K. C., Suyu, S. H., Chen, G. C.-F., et al. 2019, arXiv e-prints, arXiv:1907.04869