Astronomy & Astrophysics manuscript no. Nicolas
February 8, 2021
©ESO 2021

Redshift Evolution of the Underlying Type Ia Supernova Stretch Distribution

N. Nicolas ⋆[1], M. Rigault ⋆⋆[1],
Y. Copin[1], R. Graziani[2], G. Aldering[3], M. Briday[1], Y.-L. Kim[1], J. Nordin[4], Saul Perlmutter[3], and M. Smith[1,5]

[1] Univ Lyon, Univ Claude Bernard Lyon 1, CNRS, IP2I Lyon / IN2P3, IMR 5822, F-69622, Villeurbanne, France
[2] Université Clermont Auvergne, CNRS/IN2P3, Laboratoire de Physique de Clermont, F-63000 Clermont-Ferrand, France.
[3] Physics Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA, 94720
[4] Institut fur Physik, Humboldt-Universität zu Berlin, Newtonstr. 15, 12489 Berlin
[5] University of Southampton: Southampton, GB

Submitted to A&A the 19th of May 2020

ABSTRACT

The detailed nature of type Ia supernovae (SNe Ia) remains uncertain, and as survey statistics increase, the question of astrophysical systematic uncertainties rises, notably that of SN Ia population evolution. In this paper, we study the dependence with redshift of the SN Ia `SALT2.4` lightcurve stretch, a purely intrinsic SN property, to probe its potential redshift drift. The SN stretch has been shown to strongly correlate with the SN environment, notably with stellar age tracers. We model the underlying stretch distribution as a function of redshift, using the evolution of the fraction of young and old SNe Ia as predicted by Rigault et al. (2020), and assuming constant underlying stretch distribution for each age population consisting of Gaussian mixtures. We test our prediction against published samples cut to have marginal magnitude selection effects, so that any observed change is indeed of astrophysical and not observational in origin. In this first study, there are indications that the underlying SN Ia stretch distribution is evolving as a function of redshift, and that the young/old drifting model is a better description of the data than any time-constant model, including the sample-based asymmetric distributions often used to correct Malmquist bias **at a significance of over 5** σ. The favored underlying stretch model is the bimodal one derived from Rigault et al. (2020), composed of a high-stretch mode shared

⋆ n.nicolas@ip2i.in2p3.fr, equal contribution
⋆⋆ m.rigault@ip2i.in2p3.fr, equal contribution

by both young and old environments, and a low-stretch mode exclusive to old environments. The precise impact of the redshift evolution of the SN Ia population intrinsic properties on cosmology remains to be studied. Yet, the astrophysical drift of the SN stretch distribution does affect current Malmquist bias corrections and thereby distances derived from SN affected by observational selection effects. We highlight that such a bias will increase with surveys covering increasingly larger redshift ranges, which is particularly important for LSST.

Key words. Cosmology – Type Ia Supernova – Systematic uncertainties

Use \titlerunning to supply a shorter title and/or \authorrunning to supply a shorter list of authors.

1. Introduction

Type Ia supernovae (SNe Ia) are powerful cosmological distance indicators that 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 to Cepheids (the SH0ES project, Riess et al. 2009, 2016) is $\sim 5\sigma$ higher than what is predicted from cosmic microwave background (CMB) data measured by Planck assuming the standard ΛCDM (Planck Collaboration et al. 2020; Riess et al. 2019; Reid et al. 2019), or when the SN luminosity is anchored at intermediate redshift by the baryon acoustic oscillation (BAO) scale (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, 2020), time delay measurements from strong lensing seem to also favor high H_0 values (Wong et al. 2020).

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 2020). Alternatively, systematic effects affecting one or several of the aforementioned analyses might explain at least some of this tension. Rigault et al. (2015) suggested that SNe Ia from the Cepheid calibrator sample differ by construction from those in the Hubble flow sample, as the former strongly favors young stellar populations, while the latter does not. This selection effect would impact the derivation of H_0 if SNe Ia from young and older environments differ in average standardized magnitudes.

The relation between SNe Ia and their host galaxy properties has been studied extensively. The first key finding was that the standardized SNe Ia magnitudes significantly depend on the host galaxy stellar mass, with SNe Ia from high-mass host galaxies 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.

2020; 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 SNe and their host galaxies remains , particularlyunclear when using global properties such as the host stellar . This in turnmass raises the question of the accuracy of such corrections since host galaxy global properties evolve with redshift. More recently, studies have used the local SN host galaxy environment to probe more direct connections between the SNe and their host galaxy environments (Rigault et al. 2013), showing that local age tracers such as the Local specific Star Formation Rate (LsSFR) or the local color are more strongly correlated with the standardized SN magnitude (Roman et al. 2018; Kim et al. 2018; Rigault et al. 2020). Further reinforcing this connection is the identification of SN Ia spectral features that are correlated with LsSFR (Nordin et al. 2018). These results suggest age as the driving parameter underlying the mass-step. If true, this would have a significant impact for cosmology, since the environmental correction to apply to SN standardization could strongly vary with redshift. (Rigault et al. 2013; Childress et al. 2014; Scolnic et al. 2018). Yet, the importance of local SN environmental studies remains highly debated (e.g. Jones et al. 2015, 2019) and especially the impact of such an astrophysical bias has on the derivation of H_0 (Jones et al. 2015; Riess et al. 2016, 2018; Rose et al. 2019).

The concept of the SN Ia age dichotomy arose with the study of the SN Ia rate. 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 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). Rigault et al. (2020) used the LsSFR to classify which are the younger (those with a high LsSFR) and which are the older (those with low LsSFR). Furthermore, since the first SNe Ia host analyses, the SN stretch has been known to be strongly correlated with the SN host galaxy properties (Hamuy et al. 1996, 2000). This correlation that 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 galaxy properties, Howell et al. (2007) first discussed the potential redshift drift of the SN stretch distribution. In this paper we revisit this question using the most recent SNe Ia datasets.

In this paper, we take a step aside from the cosmological analyses to probe the validity of our modeling of the SN population, which we claim to be constituted of two age-populations (Rigault et al. 2013, 2015, 2020): one old and one younger, the former having on average lower lightcurve stretches and being brighter after standardization. We use the correlation between the SN age, as probed by the LsSFR, 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 two distinct populations of SNe Ia; (2) the relative fraction of each of these populations as a function of redshift follows the model presented in Rigault et al. (2020) and (3) the underlying

distribution of stretch for each age sample is constant. This paper aims at testing this specific model with datasets from the literature.

We present in Section 2 the sample we are using for this analysis, derived from the Pantheon catalog (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 and our results are presented in Section 4. In this section, we test whether the SN stretch distribution evolves as a function of redshift and if the aforementioned age model is in good agreement with this evolution. We briefly discuss these results in the context of SN cosmology in Section 5 and we conclude in Section 6.

2. Complete Sample Construction

The ideal SN Ia sample for studying this question would be a very deep, large-area, volume-limited sample. This would capture the true underlying stretch distribution function, and we would then study how it evolves with redshift. No such sample exists, so first we must construct subsamples that are as near to volume-limited as possible from existing high-redshift SN Ia samples.

2.1. Applying redshift cuts

We base our analysis on the most recent comprehensive 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, however, differential observational selection effects will affect the observed SN stretch distributions. Indeed, because the observed SN Ia magnitude correlates with the lightcurve stretch (and color), the first SNe Ia that a magnitude-limited survey will miss are the lowest-stretch (and reddest) ones. Consequently, if magnitude-related observational selection effects are not accounted for, one might confuse true population drift with survey properties, and conversely.

Assuming sufficient (and unbiased) spectroscopic follow-up for acquiring SN types and host galaxy redshifts, the observation selection effects of magnitude-limited surveys should be negligible below a given redshift at which even the faintest normal SNe Ia can be observed. Aiding in the construction of nearly volume-limited subsamples is the fact that the SN Ia population trails off towards fainter SNe Ia. A complication is that complete spectroscopic follow-up has not always been the norm, as discussed below. In contrast, targeted surveys have highly complex observational selection functions and so are discarded from our analysis. High-redshift SN cosmology samples, such as Pantheon, are predominately from magnitude-limited surveys from which volume-limited SN Ia subsamples can be constructed.

We present in Fig. 1 the lightcurve stretch and color 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) plane

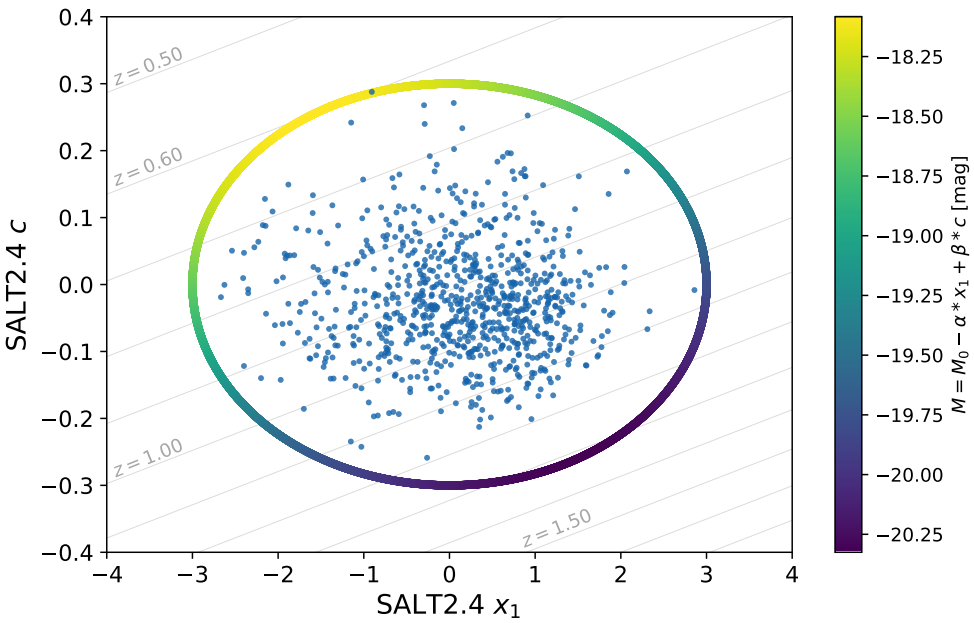

Fig. 1. SALT2.4 stretch (x_1) and color (c) lightcurve parameters of SNe Ia from the SDSS, PS1 and SNLS samples of the Pantheon catalog. The individual SNe are shown as blue dots. The ellipse ($x_1 = \pm 3, c = \pm 0.3$) is displayed, colored by the corresponding standardized absolute magnitude using the α and β coefficients from Scolnic et al. (2018). The grey diagonal lines represent the (x_1, c) evolution for $m = m_{\rm lim}$, for z between 0.50 and $z = 1.70$ using SNLS's $m_{\rm lim}$ of 24.8 mag.

with $x_1 = \pm 3$ and $c = \pm 0.3$ encapsulates the full parent 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 $(x_1 = -1.65, c = 0.25)$ and an absolute standardized magnitude at peak in Bessel B band of $M_{\rm min}^{t_0} = -18.31$ mag. Since one ought to detect this object typically 5 days before and a week after peak to build a suitable lightcurve, the effective limiting standardized absolute magnitude is approximately $M_{\rm lim} = -18.00$ mag. Hence, given the magnitude limit $m_{\rm lim}$ of a magnitude limited survey, one can derive the maximum redshift $z_{\rm lim}$ above which the faintest SNe Ia will be missed using the relation between apparent magnitude, redshift and absolute magnitude $\mu(z_{\rm lim}) = m_{\rm lim} - M_{\rm lim}$.

We will thus consider a set of cuts that will define a first fiducial sample, taking the limits as initially suggested by the previous completeness analysis. However, as this solution might be an overly simplified way to create a complete sample, e.g., because it ignores spectroscopic follow-up in efficiency for redshifts below z_{lim}, we also consider another set of cuts to define a so-called "conservative" sample. The latter is smaller and therefore less statistically constraining, but also even less prone to observational selection effects. If the redshift drift is still significant in the conservative sample, it would be even more meaningful in a carefully-tailored selection-free sample. These samples are adequate for the goal of this study, which is to develop a first implementation of

Table 1. Composition of the SNe Ia dataset used in this analysis. Conservative cuts are indicated in parentheses. The SNf limit is set by Rigault et al. (2020), see text.

Survey	$z_{\rm lim}$	$N_{\rm SN}$
SNf	0.08	114
SDSS	0.20 (0.15)	167 (82)
PS1	0.31 (0.27)	160 (122)
SNLS	0.60 (0.55)	102 (78)
HST	–	26
Total	–	569 (422)

a model for drift in SN Ia properties. If fruitful, the sample selection can later be refined, if needed, with a more detailed model of the observational selection, e.g., using the SNANA package (Kessler et al. 2009).

SNLS typically acquired 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, which has a 5σ depth of 24.8 mag[1]. This converts to a $z_{\rm lim} = 0.60$, in agreement with Neill et al. (2006), Perrett et al. (2010) and Bazin et al. (2011). Fig. 14 of Perrett et al. (2010, see their Section 5) suggests however a lower limit of $z_{\rm lim} = 0.55$. We will therefore use $z = 0.60$ and $z = 0.55$ as redshift limits for the fiducial and conservative samples, respectively, for SNLS.

Similarly, PS1 observed 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 $z_{\rm lim} = 0.31$, in agreement with, e.g., Fig. 6 of Scolnic et al. (2018). If we were to be conservative, this figure would also suggest of a more stringent $z_{\rm lim} = 0.27$ cut; hence, we will use 0.31 and 0.27 for our fiducial and conservative samples, respectively, for PS1.

In a similar redshift range, SDSS has a limiting magnitude of 22.5 (Dilday et al. 2008; Sako et al. 2008), which would lead to $z_{\rm lim} = 0.24$. However, the SDSS surveys had to contend with limited spectroscopic resources. As discussed in Kessler et al. (2009, Section 2), during the first year of SDSS, SNe Ia with $r < 20.5$ mag 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 available, and Kessler et al. (2009) and Dilday et al. (2008) show a reasonable completeness up to $z_{\rm lim} = 0.2$. Based on this, we will use $z_{\rm lim} = 0.20$ and $z_{\rm lim} = 0.15$ for our fiducial and conservative samples, respectively, for SDSS.

The sample selection is summarized in Table 1, and the redshift distribution of these three surveys is shown in Fig. 2. As expected, the selected redshift limits are roughly located slightly before the peak of these histograms. In Section 2.2 we validate that these redshift limits are effective for constructing nearly volume-limited subsamples from samples that were initially more closely magnitude-limited in their search or spectroscopic follow-up.

In addition, we use the SNe Ia from the Nearby Supernova Factory (SNfactory, Aldering et al. 2002) published in Rigault et al. (2020) and that have been discovered from non-targeted searches (114 SNe Ia, see their sections 3 and 4.2.2; SNe Ia time series are published in Saunders et al. 2020, see also Aldering et al. 2020). For this dataset, spectroscopic screening was done for candidates

[1] CFHT final release website.

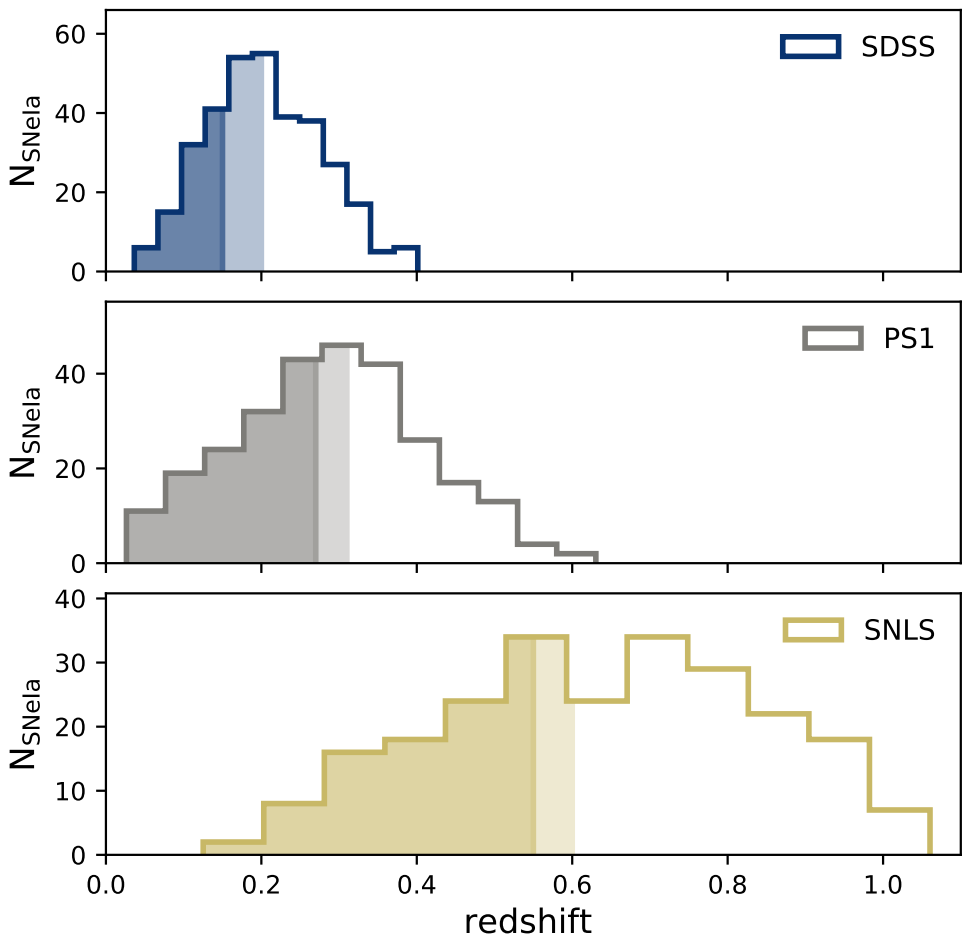

Fig. 2. *From top to bottom*: Redshift histograms of SNe Ia from the SDSS, PS1 and SNLS dataset respectively (data from Pantheon, Scolnic et al. 2018). The colored parts represent the distribution of SNe Ia kept in our analysis for they are supposedly free from observational selection bias (see Section 2). The darker (resp. lighter) color responds to the conservative (resp. fiducial) selection cut.

with $r \lesssim 19.5$; redshift cuts were then applied when selecting which SN Ia to follow, resulting in a redshift range of $0.02 < z < 0.09$, further reduced to < 0.08 in Rigault et al. (2020) for extracting local host properties. These 114 SNfactory SNe Ia are thus in the volume-limited part of the survey (Aldering et al., in prep.), and are therefore assumed to be a random sampling of the underlying SN population. The SNfactory sample is particularly useful for studying SN property drift, as it enables us to have a large complete SN Ia sample at $z < 0.1$.

Finally, we include the HST sample from Pantheon (Strolger et al. 2004). These high-redshift SNe are of great interest as they provide the greatest leverage for testing evolution. While at these redshifts the supernovae typing is challenging, the target classification was robust enough to include them within the cosmological analysis (Scolnic et al. 2018) and we do not impose further cuts. Section 4 highlights that, while compatible with it, our results are not dependent on the inclusion of this dataset.

We present the stretch distribution and redshift histogram of these five surveys up to their respective $z_{\rm lim}$ in Fig. 3. We observe here that the fraction of low-stretch SNe (typically $x_1 < -1$) appears to decrease as a function of redshift; this is confirmed in Fig. 6, in which the evolution of

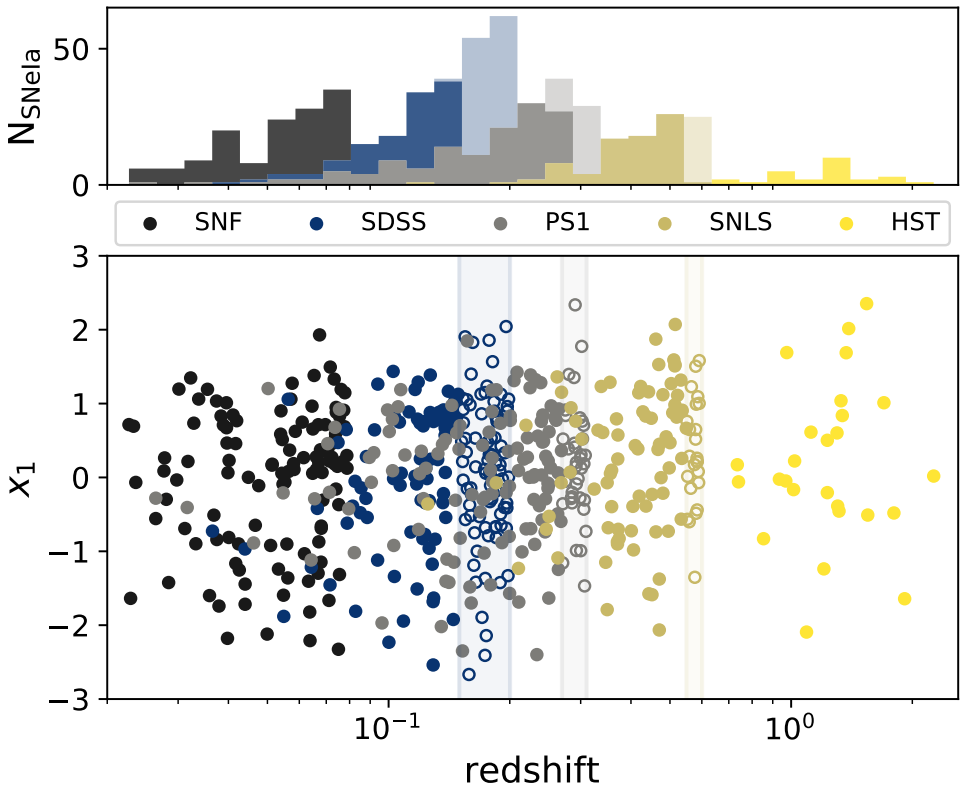

Fig. 3. *Bottom:* SALT2.4 lightcurve stretch as a function of redshift for each survey considered in this analysis (see legend). Solid (resp. open) markers correspond to the conservative (resp. fiducial) redshift cuts. *Top:* stacked redshift histograms in dark (resp. light) colors for the conservative (resp. fiducial) redshift cuts.

the mean stretch is shown, with the data split in redshift bins of regular sample size. We see that 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 $z \sim 0.05$), suggesting that the underlying stretch distribution is evolving with redshift.

2.2. Testing the construction of a volume-limited sample

In section 2.1, we have built volume-limited samples from a set of magnitude-limited ones, using simple redshift cuts. This simplified approach is statistically sub-optimal, but should be enough to test our key question: whether redshift evolution of stretch is compatible with the Rigault et al. (2020) model. However, the possibility remains that a complex observational selection function related to spectroscopic follow-up efficiencies below our fiducial (or even conservative) redshift cuts could still affect our sample, making it not fully volume-limited; this would, in turn, bias our conclusion on the astrophysical drift of the SNe Ia population. We now look at this possibility.

To test for the existence of potential leftover observational selection biases in our sample, we compare the stretch and color distributions of the SNe Ia originating from different datasets having overlapping redshift ranges: these distributions should be similar if they reflect the underlying parent population. We note that the redshift range has to be narrow enough so that any drift would be negligible.

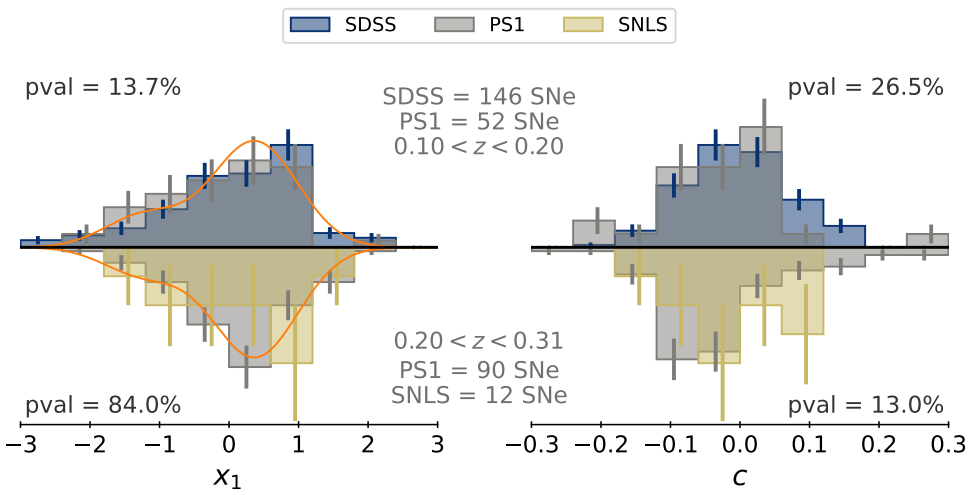

Fig. 4. x_1 (left) and c (right) distribution histograms of different surveys overlapping in redshift. *Facing up*: SDSS and PS1 within the redshift range $0.10 < z < 0.20$; *facing down*: PS1 and SNLS within the redshift range $0.20 < z < 0.31$. Error bars show the Poisson noise. Our stretch "Base" model is illustrated in orange at the mean redshift of the redshift ranges, 0.15 and 0.25, respectively. Kolmogorov-Smirnov test p-values are indicated on the top (resp. bottom) of each panel showing no sign that the SDSS and PS1 (resp. PS1 and SNLS) x_1 and c distributions are not drawn from the same underlying distributions

The two samples that overlap the most in redshift are PS1 and SDSS in the redshift range $0.10 < z < 0.20$ (see Fig 3). This overlapping subsample consists of the 146 SNe Ia at the high-redshift end of SDSS and thus the most likely to be affected by residual observational selection effects (see the corresponding discussion in section 2.1). Over that same redshift range, PS1 has 52 SNe Ia, which are in the lowest redshift bins and thus unlikely to have any observational selection issue. To identify potential inconsistency between the PS1 and SDSS subsamples, Fig. 4 (upper panels) compares the stretch and color distribution of both these surveys. The resulting Kolmogorov-Smirnov (KS) similarity test p-values ($p > 10\%$) do not support any inconsistency, in agreement with the visual impression from Fig. 4.

We perform a similar analysis for PS1 and SNLS over the redshift range $0.20 < z < 0.31$ (Fig. 4, lower panels), where the same conclusion can be drawn: there is no substantial sign of discrepancy in the stretch and color distributions between the low- and high-end of our fiducial SNLS and PS1 samples, respectively. Nonetheless, the small size of the SNLS dataset at $z < 0.31$ (12 SNe Ia vs. 90 for PS1) limits the sensitivity of this test, and only a strong deviation would be noticeable. Extending the redshift range to $0.20 < z < 0.40$ (though we have no PS1 data above 0.3) allows to increase the SNLS subsample to 31, yet the stretch p-values remains high (34%) showing no sign of inconsistency.

We finally highlight that the SNe Ia color is more prone to observational selection effects than stretch as illustrated in Fig. 1; see also e.g., Fig. 3 of Kessler & Scolnic (2017). Hence, since the comparison of color distributions shows no significant hint of leftover observational selection effect, it further supports our claim that our simple redshift-based selection criteria are sufficient to build the complete SNe Ia samples required to test the redshift evolution of the stretch distribution.

3. Modeling the redshift drift

Rigault et al. (2020) presented a model for the evolution of the fraction of younger and older 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, it was assumed that the number of "young" SNe Ia follows the star formation rate (SFR) in the Universe, while the number of "old" 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 the 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, 2020; Wiseman et al. 2020), in agreement with rate expectations (Mannucci et al. 2006; Rodney et al. 2014), Rigault et al. (2020) concluded that

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

with $K = 0.87$. This model is comparable to the evolution subsequently 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 stretch distribution

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

Rigault et al. (2020) presented the relation between SN stretch and LsSFR measurement, a progenitor age tracer, using the SNfactory sample. This relation is shown in Fig. 5 for the SNfactory SNe used in the current analysis. Given the structure of the stretch-LsSFR scatter plot, our model of the underlying SN Ia stretch distribution is defined as follows:

- for the younger population (i.e., log(LsSFR) ≥ -10.82), the stretch distribution is modeled as a single normal distribution $\mathcal{N}(\mu_1, {\sigma_1}^2)$;
- the older population (i.e., log(LsSFR) < -10.82) stretch distribution is modeled as a bimodal Gaussian mixture $a \times \mathcal{N}(\mu_1, {\sigma_1}^2) + (1-a) \times \mathcal{N}(\mu_2, {\sigma_2}^2)$, where one mode is the same as for the young population, a representing the relative influence of the two modes.

The stretch probability distribution function (pdf) of a given SN will be the linear combination of the stretch distributions of these two population weighted by its probability y^i to be young (see Section 3.2). But generally, the fraction of young SNe Ia as a function of redshift is given by $\delta(z)$ (see Eq. 1) and therefore, our redshift drift model of the underlying stretch distribution of SNe Ia

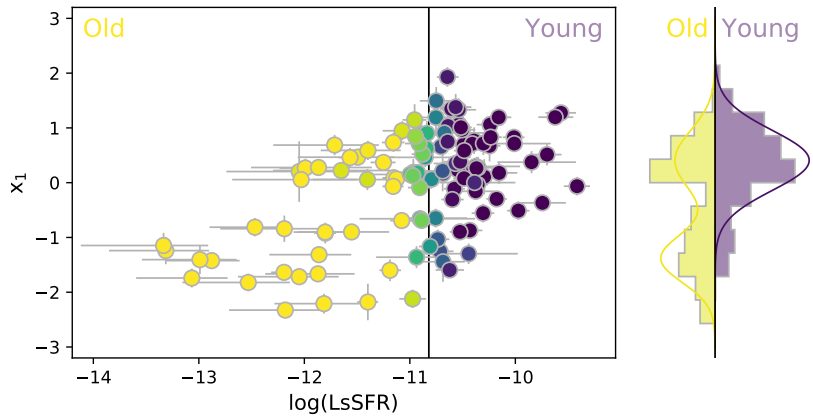

Fig. 5. *Main*: SALT2.4 lightcurve stretch (x_1) as a function of the local specific star formation rate (LsSFR) for SNfactory SNe used in this analysis. The color corresponds to the probability, p_y, for the SNe Ia to be young, i.e. to have $\log \mathrm{LsSFR} \geq -10.82$ (see Rigault et al. 2020). *Right*: p_y-weighted histogram of the SN stretches, as well as the adjusted Base model; the younger and older population contributions are shown in purple and yellow, respectively.

as a function of redshift $X_1(z)$ is given by:

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

This is our Base drifting model.

3.2. Comparison to data

Given the probability y^i that a given SN is young, and assuming our Base model (see Section 3.1), the probability to measure a SALT2.4 stretch x_1^i with an error $\mathrm{d}x_1^i$ is given by:

$$\mathcal{P}\left(x_1^i \mid \boldsymbol{\theta}; \mathrm{d}x_1^i, y^i\right) = y^i \times \mathcal{N}\left(x_1^i \mid \mu_1, {\sigma_1}^2 + {\mathrm{d}x_1^i}^2\right) + (1 - y^i) \times \left[a \times \mathcal{N}\left(x_1^i \mid \mu_1, {\sigma_1}^2 + {\mathrm{d}x_1^i}^2\right) + (1 - a) \times \mathcal{N}\left(x_1^i \mid \mu_2, {\sigma_2}^2 + {\mathrm{d}x_1^i}^2\right)\right] \quad (3)$$

The maximum-likelihood estimate of the 5 free parameters $\boldsymbol{\theta} \equiv (\mu_1, \mu_2, \sigma_1, \sigma_2, a)$ of the model is obtained by minimizing the following:

$$-2\ln(L) = -2 \sum_i \ln \mathcal{P}\left(x_1^i \mid \boldsymbol{\theta}; \mathrm{d}x_1^i, y^i\right). \quad (4)$$

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

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

Sample	μ_1	σ_1	μ_2	σ_2	a
SNfactory	0.41 ± 0.08	0.55 ± 0.06	-1.38 ± 0.10	0.44 ± 0.08	0.48 ± 0.08
Fiducial	0.37 ± 0.05	0.61 ± 0.04	-1.22 ± 0.16	0.56 ± 0.10	0.51 ± 0.09
Conservative	0.38 ± 0.05	0.60 ± 0.04	-1.26 ± 0.13	0.53 ± 0.08	0.47 ± 0.09

3.2.1. With LsSFR measurements

For the SNfactory sample, we can readily set $y^i = p^i_y$, the probability to have log(LsSFR) ≥ -10.82 (see Fig. 5), to minimize Eq. 4 with respect to $\boldsymbol{\theta}$. Results on fitting the SNf SNe with this model are shown Table 2 and illustrated in Fig. 6.

3.2.2. Without LsSFR measurements

When lacking direct LsSFR measurements (i.e. p^i_y), we can extend the analysis to non-SNfactory samples by using the redshift-evolution of the fraction $\delta(z)$ of young SNe Ia (Eq. 1) as a proxy for the probability of a SN to be young. This still corresponds to minimizing Eq. 4 with respect to the parameters $\boldsymbol{\theta} \equiv (\mu_1, \mu_2, \sigma_1, \sigma_2, a)$ of the stretch distribution X_1 (Eq. 2), 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. 4 using p^i_y — the probability for the SN i to be young — when available (i.e. for SNfactory dataset), and $\delta(z^i)$ — the expected fraction of young SNe Ia at the SN redshift z^i — otherwise.

Results of fitting this model to all the 569 (resp. 422) SNe from the fiducial (resp. conservative) sample are given Table 2, and the predicted redshift evolution of mean stretch (expected x_1 given the distribution of Eq. 2) illustrated as a blue band in Fig. 6 accounting for parameters errors and their covariances. We see in this figure that the measured mean SN Ia stretch per redshift bins of equal sample size closely follows our redshift drift modeling. 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 declines as a function of redshift. See Section 4 for a more quantitative discussion.

3.3. Alternative models

In Section 3.1, we have modeled the underlying stretch distribution following Rigault et al. (2020), i.e. as a single Gaussian for the "young" SNe Ia and a mixture of two Gaussians for the "old" SNe Ia population, one being the same as for the young population, plus another one for the fast-declining SNe Ia that seem to only exist in old local environments. 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 in Section 3.2.2.

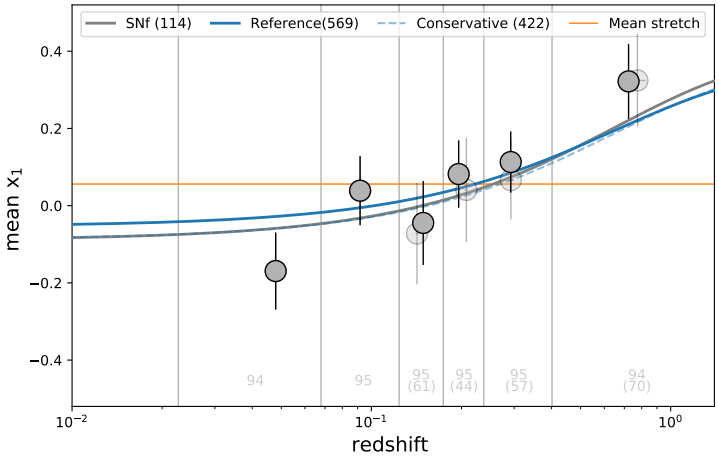

Fig. 6. Evolution of the mean SN `SALT2.4` stretch (x_1) as a function of redshift. Markers show the stretch plain mean (with error estimated from scatter) measured in redshift bins of equal sample size, indicated in light gray at the bottom of each redshift bin. Full and light markers are used when considering the fiducial or the conservative samples, respectively. The orange horizontal line represents the mean stretch for the non-evolving Gaussian model (last line of Table 3) on the fiducial sample. 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 SNfactory dataset only, respectively; all are compatible. The light-blue band illustrates the amplitude of the error (incl. covariance) of the best fit model when considering the fiducial dataset.

Howell et al. (2007) used a simpler unimodal model per age category, assuming a single normal distribution for each of the young and old populations. We thus consider a "Howell+drift" model, with one single Gaussian per age group and the $\delta(z)$ drift from Eq. 1.

Alternatively, since we aim at probing the existence of an evolution with redshift, we also test constant models by restricting the "Base" and "Howell" models to use a supposedly redshift-independent fraction $\delta(z) \equiv f$ of young SNe; these models are hereafter labeled "Base+constant" and "Howell+constant".

We also consider another intrinsically non-drifting model, the functional form developed for Beams with Bias Correction (BBC, Scolnic & Kessler 2016; Kessler & Scolnic 2017), 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) since current surveys cover limited redshift ranges, having a sample-based approach covers some potential redshift evolution information (Scolnic & Kessler 2016; Scolnic et al. 2018). See further discussion concerning BBC in Section 5.

Finally, for the sake of completeness, we also consider redshift-independent pure and asymmetric Gaussian models.

4. Results

We adjusted each of the models described above on both the fiducial and conservative samples (cf. Section 2); results are gathered in Table 3, and illustrated in Fig. 7.

Table 3. Comparison of the relative ability of each model to describe the data. For each considered model, we report if the model is drifting or not, its number of free parameters and, for both the fiducial and the conservative cuts, $-2\ln(L)$ (see Eq. 4), the AIC and the AIC difference (ΔAIC) between this model and the Base model used as reference for it has the lowest AIC.

			Fiducial sample (569 SNe)			Conservative sample (422 SNe)		
Name	drift	k	$-2\ln(L)$	AIC	ΔAIC	$-2\ln(L)$	AIC	ΔAIC
Base	$\delta(z)$	5	1456.7	1466.7	–	1079.5	1089.5	–
Howell+drift	$\delta(z)$	4	1463.3	1471.3	−4.6	1088.2	1096.2	−6.7
Asymmetric	–	3	1485.2	1491.2	−24.5	1101.3	1107.3	−17.8
Howell+const	f	5	1484.2	1494.2	−27.5	1101.2	1111.2	−21.7
Base+const	f	6	1484.2	1496.2	−29.5	1101.2	1113.2	−23.7
Per sample Asym.	per sample	3×5	1468.2	1498.2	−31.5	1083.6	1113.6	−24.1
Gaussian	–	2	1521.8	1525.8	−59.1	1142.6	1146.6	−57.1

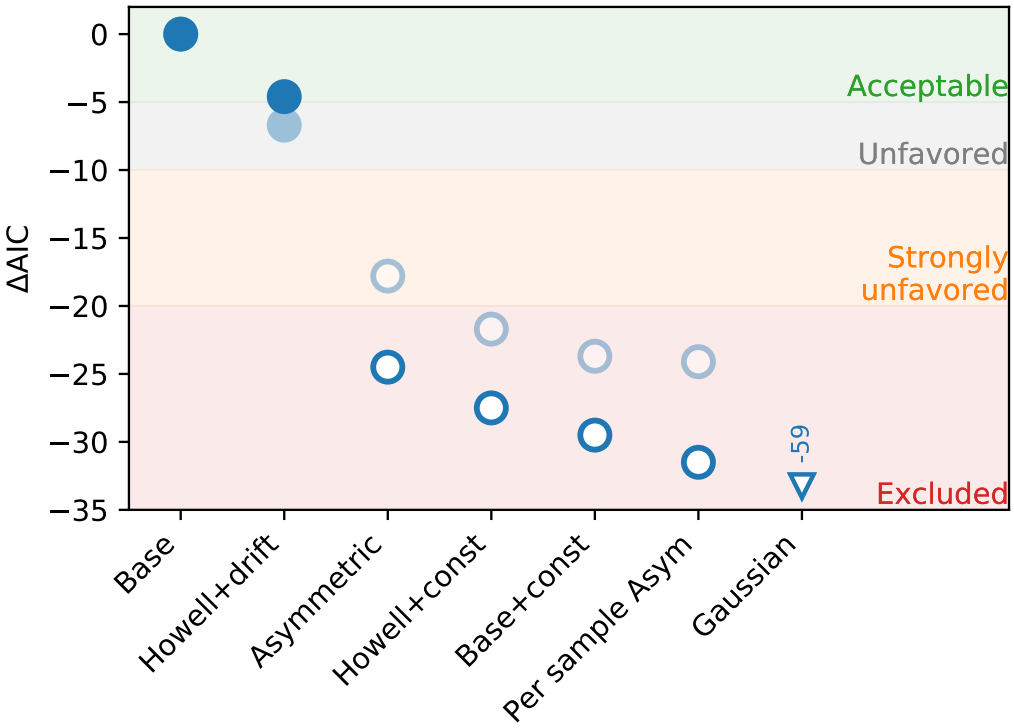

Fig. 7. ΔAIC between "Base" model (reference) and other models (see Table 3). Full and open blue markers correspond to models with and without redshift drift, respectively. Light markers show the results when the analysis is performed on the conservative sample rather than the fiducial one. Color-bands illustrate the validity of the models, from Acceptable (ΔAIC > -5) to Excluded (ΔAIC < -20), see text. According to the AIC, all non-drifting models (open symbols) are excluded to be as good a representation of the data as the Base (drifting) model.

Since the various models have different degrees of freedom, we use the Akaike Information Criterion (AIC, e.g. Burnham & Anderson 2004) to compare their ability to properly describe the observations. This estimator penalizes extra degrees of freedom to avoid over-fitting the data, and is defined as follow:

$$\mathrm{AIC} = -2\ln(L) + 2k \tag{5}$$

where $-2\ln(L)$ is derived by minimizing Eq. (4), and k is the number of free parameters to be adjusted. The reference model is the one with the smallest AIC; in comparison to this model, the models with ΔAIC < 5 are coined acceptable, the ones with $5 < \Delta$AIC < 20 are unfavored, and

those with ΔAIC > 20 are deemed excluded. This roughly corresponds to 2, 3 and 5 σ limits for a Gaussian probability distribution.

The best model (with smallest AIC) is the so-called Base model and thus is our reference model; this is true both on the fiducial and conservative samples. The Base model also has the smallest $-2\ln(L)$, making it the most likely model even ignoring the over-fitting issue accounted for by the AIC formalism.

Furthermore, we find that redshift-independent stretch distributions are all excluded as suitable descriptions of the data relative to the Base model. In fact, the best non-drifting model (the Asymmetric one) has a very marginal chance ($p \equiv \exp(\Delta\mathrm{AIC}/2) = 5 \times 10^{-6}$) to describe the data as well as the Base model. This result is just a quantitative assessment of qualitative facts clearly visible in Fig. 6: the mean SN stretch per bin of redshift strongly suggests a significant redshift evolution rather than a constant value, and this evolution is well described by Eq. 1.

Surprisingly, the sample-based Gaussian asymmetric modeling used by current implementations of the BBC technique (Scolnic & Kessler 2016; Kessler & Scolnic 2017) has one of the highest AIC value in our analysis (see Section 4). While its $-2\ln(L)$ is the smallest of all redshift-independent models (but still -11.5 worse than the reference Base model), it is strongly penalized for requiring 15 free parameters ($\mu_0, \sigma_\pm$ for each of the 5 samples of the analysis). Hence, its ΔAIC < -20, which could be interpreted as a probability $p = 2 \times 10^{-7}$ of being as good a representation of the data as the Base model.

We note that, when comparing models adjusted on individual subsamples rather than globally, the Bayesian Information Criterion (BIC $= -2\ln(L) + k\ln(n)$, with n the number of data points) might be better suited than AIC, since it explicitly accounts for the fact that each subsample is fitted separately: the sample-based model BIC is rightfully the sum of the BIC for each sample. We find ΔBIC $= -48$, again refuting the sample-based asymmetric Gaussian model as being as pertinent as the Base model.

In order to ensure that our results are not driven by the incompletely-modeled HST subsample, we recompute ΔAIC for each model excluding this dataset; we find that it does not change ΔAIC by more than few tenths. The consistency of these values with those in Table 3 show that the HST subsample does not drive our conclusions.

We report in Table 4 our determination of μ_0 and $\sigma_\pm$ for each sample when implementing an asymmetric Gaussian model, adjusted on the nominally selection-free samples using our fiducial cuts (see Section 2). We find our results in close agreement with Scolnic & Kessler (2016) for SNLS and SDSS and with Scolnic et al. (2018) for PS1, who derived these model parameters using the full BBC formalism, using numerous simulations to model the observational selection effects (see details e.g., Section 3 of Kessler & Scolnic 2017). The agreement between our fit of the asymmetric Gaussians on the supposedly selection-free part of the samples and the results derived using the BBC formalism supports our approach to constructing a sample with negligible observational selection effects. If we were to use Scolnic & Kessler (2016) and Scolnic et al. (2018)

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

Asymmetric	σ_-	σ_+	μ_0
SNfactory	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

best fit values of the $\mu_0, \sigma_\pm$ asymmetric parameters 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. We further discuss the consequence of this result for cosmology in Section 5.

We also performed tests allowing the high-stretch mode of the old population to differ from the young population mode, hence adding two degrees of freedom. The corresponding fit is not significantly better, with a ΔAIC of -0.4. This reinforces our assumption that the young and old populations indeed appear to share the same underlying high-stretch mode. Furthermore, one might wonder whether a low-stretch mode might also exist in the young-population (see Fig. 5). We test for this by allowing this population to also be bimodal, finding the amplitude of such a low-stretch mode to be compatible with 0 ($< 2\%$) in this young population. More generally, this raises the question of how well a given environmental tracer (here LsSFR) traces the age. An analysis dedicated to this question will be presented in Briday et al. in prep.

Finally, ignoring the LsSFR measurements — available only for the SNfactory dataset (see Section 3) — reduces the significance of the results presented in this section, as expected. Even so, non-drifting models remain strongly disfavored. For instance, the best fitting sample-based Gaussian asymmetric model is still $\Delta\text{AIC} < -10$ less representative of the data than our Base drifting model.

5. Discussion

To the best of our knowledge, a SN Ia stretch redshift drift modeling has never been explicitly used in cosmological analyses, though Bayesian hierarchy formalism such as UNITY (Rubin et al. 2015), BAHAMAS (Shariff et al. 2016) or Steve (Hinton et al. 2019) can easily allow it (see e.g., sections 1.3 and 2.5 of Rubin et al. (2015)). Not doing so is a second order issue for SN cosmology, as it only affects the way one accounts for Malmquist bias. Indeed, as long as the Phillips relation (Phillips 1993) standardization parameter α is not redshift dependent (a study beyond the scope of this paper, but see e.g. Scolnic et al. 2018), the stretch-corrected SNe Ia magnitudes used for cosmology are blind to the underlying stretch distribution for complete samples. However, surveys usually do have significant Malmquist bias for the upper half of their SN redshift distribution. As a consequence, mismodeling of the underlying stretch distribution will bias the SN magnitudes derived from such surveys.

Commonly used Malmquist bias correction techniques, such as the BBC-formalism, assume per sample asymmetric Gaussian functions to model the underlying stretch and color distributions.

Yet, as shown in Section 4, such a sample-based distribution is excluded in comparison to our drifting model. Then, contrary to what Scolnic & Kessler (2016, Section 2) and Scolnic et al. (2018, Section 5.4) have suggested — i.e., that traditional surveys span sufficiently limited redshift ranges such that the per-sample approach accounts for implicit redshift drifts — a direct modeling of the redshift drift is more appropriate than a sample-based approach. We add here that, as measurements of modern surveys try to cover increasingly larger redshift ranges in order to reduce calibration systematic uncertainties, this sample-based approach becomes less valid, notably for PS1, DES and, soon, LSST.

We illustrate in Fig. 8 the prediction difference in the underlying stretch distribution between the per-sample asymmetric 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 and young SNe Ia in the sample: the higher the fraction of old SNe Ia (at lower redshift), the higher the amplitude of the old-specific low-stretch mode. This redshift dependency of the underlying stretch distributions is shown running from blue to red in Fig. 8 for the redshift range covered by PS1. The observed x_1 histogram follows the model we defined using the sum of individual underlying SN-redshift distributions. As expected, the two modeling approaches differ mostly in the negative part of the SN stretch distribution. The asymmetric Gaussian distribution goes through the middle of the bimodal distribution, over-estimating the number of SNe Ia at $x_1 \sim -0.7$ and under-estimating it at $x_1 \sim -1.7$ in comparison to our Base drifting model for typical PS1 SN redshifts. This means that the SN bias-corrected standardized magnitude estimated at a redshift suffering from by observational selection effects would be biased by a mismodeling of the true underlying stretch distribution.

Assessing the amplitude of this magnitude bias for cosmology is beyond the scope of this paper given the complexity of the BBC analysis. It would require a full study using our Base model (Eq. 2) in place of the sample-based asymmetric modeling as part of the BBC simulations. However, we already highlight that even if a non-drifting sample-based model could provide comparable result in the volume-limited part of the various samples, these models would differ when extrapolating to higher redshifts, precisely where the underlying distribution will matter for correcting Malmquist biases.

In the era of modern cosmology, where we aim to measure w_0 at a sub-percent level and w_a with ten-percent precision (e.g., Ivezić et al. 2019), we stress that correct modeling of potential SN redshift drift should be further studied and care should be taken when using samples affected by observational selection effects.

6. Conclusion

We have presented an initial study of the drift of the underlying SNe Ia stretch distribution as a function of redshift. We built effectively volume-limited SN Ia subsamples from the Pantheon dataset (Scolnic et al. 2018, SDSS, PS1 and SNLS), to which we added HST and SNfactory data

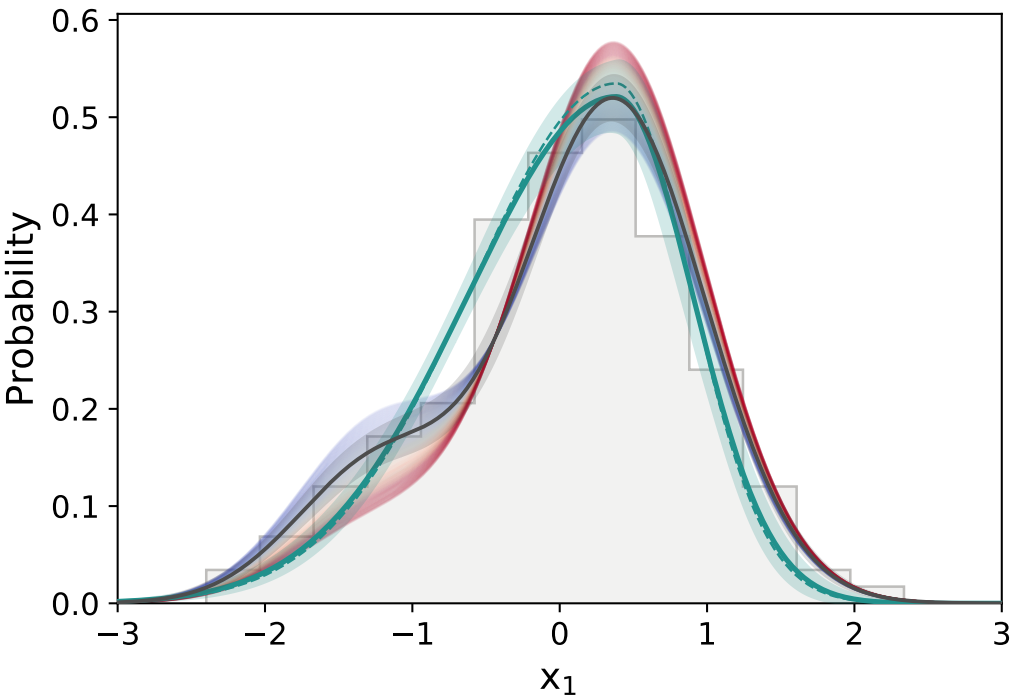

Fig. 8. Distribution of the PS1 SN Ia `SALT2.4` stretch (x_1) after the fiducial redshift limit cut (grey histogram). This distribution is supposed to be a random draw from the underlying stretch distribution. The green lines show the BBC model of this underlying distribution (asymmetric Gaussian). The full line (band) is our best fit (its error); the dashed line shows the Scolnic et al. (2018) result. The black line (band) shows our best fitted Base-modeling (its error, see Table 2) that includes redshift drift. For illustration, we show as colored (from blue to red with increasing redshifts) the evolution of the underlying stretch distribution as a function of redshift for the redshift range covered by PS1 data.

from Rigault et al. (2020) for the high- and low-redshift bins. We only considered the SNe that have been discovered in the redshift range of each survey where observational selection effects are negligible, so that the observed SNe Ia stretches are a random sampling of the true underlying distribution. This resulted in a fiducial sample of 569 SNe Ia (422 SNe when more conservative cuts were applied).

Following predictions made in Rigault et al. (2020), we introduced a redshift drift model which depends on the expected fraction of "young" and "old" SNe Ia as a function of redshift, with each age population having its own underlying stretch distribution.

In addition to this "base" model, we have studied various distributions, including redshift independent models; we also studied the prediction from a per-sample asymmetric Gaussian stretch distribution used, for instance, by the Beams with Bias Correction Malmquist bias correction algorithm (Scolnic & Kessler 2016; Kessler & Scolnic 2017).

Our conclusions are the following:

1. The underlying SN Ia stretch distribution is significantly redshift dependent, as previously suggested by e.g. Howell et al. (2007), in a way that observational selection effects alone cannot explain. This result is largely independent of details on each age-population model.
2. Redshift-independent models are quantitatively excluded as suitable descriptions of the data relative to our Base model. This model assumes that: (1) the younger population has a unimodal

Gaussian stretch distribution, while the older population stretch distribution is bimodal, one mode being the same as the young one; (2) the evolution of the relative fraction of younger and older SNe Ia follows the prediction made in Rigault et al. (2020). This second result further supports the existence of both young and old SN Ia populations, in agreement with rate studies Mannucci et al. (2005); Scannapieco & Bildsten (2005); Sullivan et al. (2006); Aubourg et al. (2008).

3. Models using survey-based asymmetric Gaussian distributions, e.g., as employed in the current implementation of BBC, are excluded as a good description of the data relative to our drifting model. Hence, the sample-based approach does not accurately account for redshift drift, a problem that will be exacerbated as surveys span increasingly larger redshift ranges. As a result, even if the necessary extra degrees of freedom might be acceptable given the large number of SNe Ia in cosmological studies, extrapolating the SN property distributions from the volume-limited part of a survey to its Malmquist-biased magnitude-limited part would still be inaccurate because of the redshift evolution.

4. Given the current dataset, we suggest the use of the following stretch population model as a function of redshift:

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

with $a = 0.51$, $\mu_1 = 0.37$, $\mu_2 = -1.22$, $\sigma_1 = 0.61$, $\sigma_2 = 0.56$ (see Table 2), and using the age-population drift model

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

with $K = 0.87$.

In this paper, we considered a simple two-population Gaussian mixture modeling. Additional data free from significant Malmquist bias would enable us to refine it, as necessary. We note that samples at the low- and high-redshift ends of the Hubble diagram would be particularly helpful for this drifting analysis; fortunately this will soon be provided by the Zwicky Transient Facility (low-z, Bellm et al. 2019; Graham et al. 2019), and Subaru and SeeChange SNe Ia programs (high-z), respectively.

The next step in this line of analysis will incorporate our model into the SNANA framework (Kessler et al. 2009), both to more accurately account for observational selection functions and to test the impact of our model on the derivation of cosmological parameters; this study is underway.

Acknowledgements. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation program (grant agreement no 759194 - USNAC). This work was supported in part by the Director, Office of Science, Office of High Energy Physics of the U.S. Department of Energy under Contract No. DE-AC025CH11231. This project is partly financially supported by Région Rhône-Alpes-Auvergne.

References

Abbott, T. M. C., Allam, S., Andersen, P., et al. 2019, ApJ, 872, L30

Aldering, G., Adam, G., Antilogus, P., et al. 2002, Proc. SPIE, 61

Aldering, G., Antilogus, P., Aragon, C., et al. 2020, Research Notes of the American Astronomical Society, 4, 63

Astier, P., Guy, J., Regnault, N., et al. 2006, A&A, 447, 31

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

Childress, M., Aldering, G., Antilogus, P., et al. 2013, ApJ, 770, 108

Childress, M. J., Wolf, C., & Zahid, H. J. 2014, MNRAS, 445, 1898

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

Freedman, W. L., Madore, B. F., Hoyt, T., et al. 2020, ApJ, 891, 57. doi:10.3847/1538-4357/ab7339

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

Hinton, S. R., Davis, T. M., Kim, A. G., et al. 2019, ApJ, 876, 15

Howell, D. A., Sullivan, M., Conley, A., et al. 2007, ApJ, 667, L37

Ivezić, Ž., Kahn, S. M., Tyson, J. A., et al. 2019, ApJ, 873, 111

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., Bernstein, J. P., Cinabro, D., et al. 2009, PASP, 121, 1028

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. 2020, Phys. Rev. D, 101, 043533. doi:10.1103/PhysRevD.101.043533

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

Nordin, J., Aldering, G., Antilogus, P., et al. 2018, A&A, 614, A71

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

Phillips, M. M. 1993, ApJ, 413, L105

Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2020, A&A, 641, A6. doi:10.1051/0004-6361/201833910

Reid, M. J., Pesce, D. W., & Riess, A. G. 2019, ApJ, 886, L27. doi:10.3847/2041-8213/ab552d

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. 2029, A&A, 644, A176

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

Saunders, C., Aldering, G., Antilogus, P., et al. 2020, VizieR Online Data Catalog, J/ApJ/869/167

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., Perlmutter, S., Aldering, G., et al. 2019, Astro2020: Decadal Survey on Astronomy and Astrophysics, 2020, 270

Shariff, H., Jiao, X., Trotta, R., et al. 2016, ApJ, 827, 1

Strolger, L.-G., Riess, A. G.,

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, MNRAS, 495, 4040. doi:10.1093/mnras/staa1302

Wong, K. C., Suyu, S. H., Chen, G. C.-F., et al. 2020, MNRAS, 498, 1420. doi:10.1093/mnras/stz3094