Astronomy & Astrophysics manuscript no. referee_answer
September 17, 2020

Redshift Evolution of the Underlying Type Ia Supernova Stretch Distribution

N. Nicolas ⋆[1], M. Rigault ⋆⋆[2],
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[4]

[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.
[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 true nature of type Ia supernovae (SNe Ia) remains largely unknown, and as survey statistics increase, the question of astrophysical systematic uncertainties rises, notably that of the 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 **?**, and assuming constant underlying stretch distribution for each age population made of Gaussian mixtures. We test our prediction against published samples chosen to have negligible magnitude selection effects, so that any observed change is indeed of astrophysical and not observational origin. We clearly demonstrate that the underlying SN Ia stretch distribution is evolving as a function of redshift, and that the young/old drifting model is a much better description of the data than any time-constant model, including the sample-based asymmetric distributions usually used to correct Malmquist bias. The favored underlying stretch model is the bimodal one derived from **?**: a high-stretch mode shared 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 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

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 (**??**). 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. **?**). They also are key to directly measure the Hubble Constant (H_0), provided one can calibrate their absolute magnitude (**??**). Interestingly, the value of H_0 derived when the SNe Ia are anchored on Cepheids (the SH0ES project, **??**) is $\sim 5\sigma$ higher than what is predicted from 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 (**?**). While using the tip of the red giant branch technique in place of the Cepheids seem to favor an intermediate value of H_0 (**??**), time delay measurements from strong lensing seem to also favor high H_0 values (**?**).

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 (**?**), but see e.g. **?**. Alternatively, systematic effects affecting one or several of the aforementioned analyses could also explain the tension. **?** suggested that SNe Ia from the Cepheid calibrator sample differ by construction from the Hubble flow sample ones as the former strongly favors young stellar populations, while the latter not. This selection effect would impact the derivation of H_0 if SNe Ia from young and older environments differ in average standardized magnitudes.

For the last decades, numerous analyses 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. **??????**). This mass-step correction is currently used in cosmological analyses (e.g. **??**), including for deriving H_0 (**??**). Yet, the underlying connection between the SNe and their 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 (**?**), 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 magni-

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

tude (???), suggesting 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. (???). Yet, the importance of local SN environmental studies remains highly debated (e.g. ??) and especially the impact of such an astrophysical bias has on the derivation of H_0 (????).

The concept of the SN Ia age dichotomy arose with the study of the SN Ia rate. ???? 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). In ? we 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 properties (??), correlation that has been extensively confirmed since (e.g. ??????????). Following the "A+B" model and the connection between SN stretch and host properties, ? 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.

In this paper, we take a step aside to probe the validity of our modeling of the SN population, which we claim to be constituted of two age-populations (???): 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 ? 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. Note that the progenitor age as traced by the LsSFR seems to capture physical feature intrinsic to the progenitor and/or explosion mechanism that the stretch alone is not capturing (?).

We present in Section ?? the sample we are using for this analysis, derived from the Pantheon catalog (?). 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 ?? our modeling of the distribution of stretch and our results are presented in Section ??. 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 ?? and we conclude in Section ??.

2. Complete Sample Construction

2.1. Applying redshift cuts

We base our analysis on the most recent comprehensive SNe Ia compilation, the Pantheon catalog from ?. 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 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 SNe Ia that a magnitude-limited survey will miss are the lowest-stretch (and reddest) ones. Consequently, if magnitude-related selection effects are not accounted for, one might confuse true population drift with survey properties, and conversely.

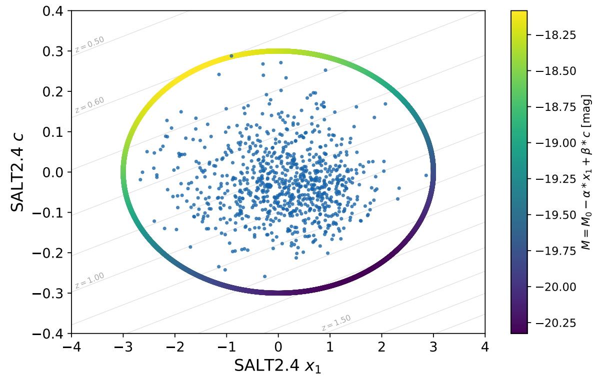

Fig. 1. `SALT2.4` stretch (x_1) and color (c) lightcurve parameters of SNe Ia from the SDSS, PS1 and SNLS samples of the Pantheon catalog. The individual SNe are shown as blue dots. The ellipse ($x_1 = \pm 3, c = \pm 0.3$) is displayed, colored by the corresponding standardized absolute magnitude using the α and β coefficients from ?. 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.

Assuming sufficient (and unbiased) spectroscopic follow-up for acquiring typing and host redshift, the selection effects of magnitude-limited surveys should be negligible below a given redshift at which even the faintest SNe 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 one are now dominated by magnitude-limited surveys.

We present in Fig. ?? the lightcurve stretch and color of SNe Ia from the following surveys: PanStarrs (PS1 ?), the Sloan Digital Sky Survey (SDSS ?) and the SuperNovae Legacy Survey (SNLS ?). An ellipse in the `SALT2.4` (x_1, c) plane with $x_1 = \pm 3$ and $c = \pm 0.3$ encapsulates the full distribution (??); see also ? and ? 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 $M_0 = -19.36$ (??), 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 ?: 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 $M^{t_0}_{\rm min} = -18.31$ 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 $\mu(z_{\rm lim}) = m_{\rm lim} - M_{\rm lim}$.

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, that has a 5σ depth of 24.8 mag[1]. This converts to a $z_{\rm lim} = 0.60$, in agreement with ?, ? and ?. Fig. 14 of (?, see their Section 5) suggests a lower limit of $z_{\rm lim} = 0.55$; both limits will be considered, as discussed below.

Similarly, PS1 observed SNe Ia in the range $0.2 < z < 0.4$, their g-band 5σ depth is 23.1 mag (?), which yields to $z_{\rm lim} = 0.30$

[1] CFHT final release website.

Table 1. Composition of the SNe Ia dataset used in this analysis. Conservative cuts are indicated in parentheses.

Survey	$z_{\rm lim}$	$N_{\rm SN}$
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)

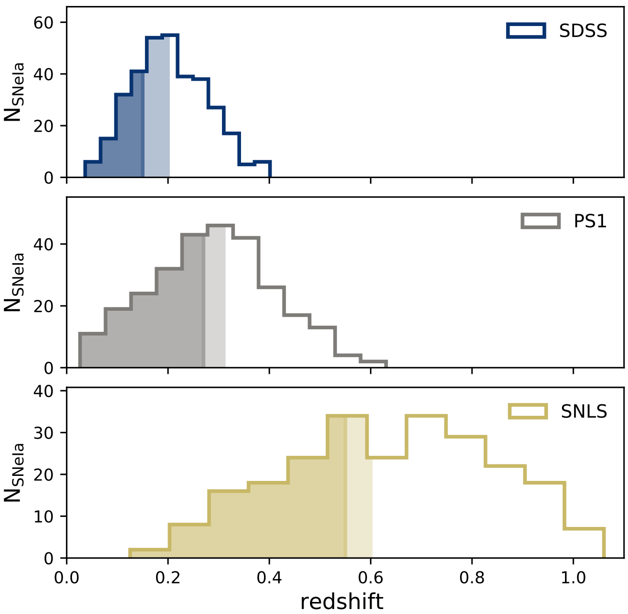

Fig. 2. *From top to bottom*: Redshift histograms of SNe Ia from the SDSS, PS1 and SNLS dataset respectively (data from Pantheon, **?**). The colored parts represent the distribution of SNe Ia kept in our analysis for they are supposedly free from selection bias (see Section **??**). The darker (resp.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

in agreement with, e.g., Fig. 6 of **?**. This figure is also suggestive of a more conservative $z_{\rm lim} = 0.27$.

In a similar redshift range, SDSS has a limiting magnitude of 22.5 (**??**), which would lead to a $z_{\rm lim} = 0.24$. However, the SDSS surveys were more sensitive to limited spectroscopic resources; (**?**, see their Section 2) pointed out that 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 **?** and **?** show a reasonable completeness up to $z_{\rm lim} = 0.2$. Following these analyses, we will use $z_{\rm lim} = 0.2$ as the baseline SDSS redshift limit.

The sample selection is summarized in Table **??**, and the redshift distribution of these three surveys is shown in Fig. **??**. As expected, the selected redshift limits roughly correspond to the peak of these histograms. We show in Section **??** that this sample selection indeed provides a subset of SNe Ia with insignificant selection effects when compared to state-of-the-art Malmquist correction techniques.

In addition, we use the SNe Ia from the Nearby Supernova Factory (SNfactory, **?**) published in **?** 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 **?**, see also

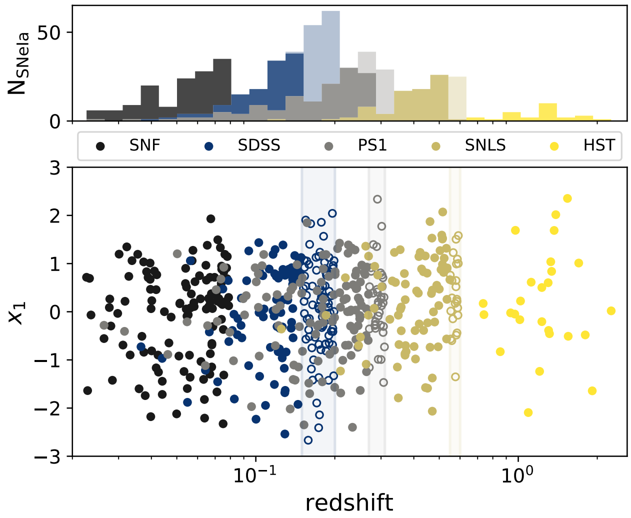

Fig. 3. *Bottom:* `SALT2.4` lightcurve stretch as a function of redshift for each survey considered in this analysis (see legend). Solid (resp. open) markers correspond to the conservative (resp. fiducial) redshift cuts. *Top:* stacked redshift histograms in dark (resp. light) colors for the conservative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

?). The spectroscopic follow-up was done over a redshift range of $0.02 < z < 0.09$ (as in **?**), while the search was much deeper. 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 SN Ia sample at $z < 0.1$.

Finally, we include the HST sample from Pantheon, that similarly have a search deeper than the follow-up and that we therefore kept entirely (**?**).

We present the stretch distribution and redshift histogram of these five surveys up to their respective $z_{\rm lim}$ in Fig. **??**. The evolution of the mean stretch is also shown in Fig. **??**, where the data are split in redshift bins of regular sample size. We see that SNe Ia at higher redshift have on average larger stretch ($0.34{\pm}0.10$ at $z \sim 0.65$) than those at lower redshift ($-0.17{\pm}0.10$ at $z \sim 0.05$), suggesting that the underlying stretch distribution is drifting.

2.2. Testing the construction of a volume limited sample

In this analysis, we have built a volume limited sample from a set of magnitude limited ones, using a simple redshift limit. An ongoing follow up analysis will make use of the SNANA package (**?**) to account for more subtle selection functions. However, the strength of the current approach is that it is independent of potential inaccuracy in the modeling of the complex selection functions of the aforementioned dataset. Yet, it is necessary to test if the sample we have created could show sign of leftover sample selections.

To do so, we compare the stretch and color distributions of the SNe Ia in our sample between datasets of overlapping redshifts. If indeed we have built a volume limited sample, the distributions of any SNe Ia parameter at a given redshift should not depend of their origin surveys. When considering a redshift range, this has to be narrow enough such that any drift would be negligible. The two samples that overlap the most in redshift

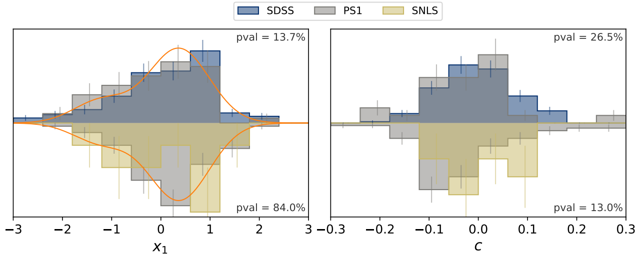

Fig. 4. x_1 and c distribution histograms of different surveys overlapping in redshift. *Facing up* SDSS and PS1 within the redshift range $0.10 < z < 0.20$; *facing down* PS1 and SNLS within the redshift range $0.20 < z < 0.3$. Error bars show the Poisson noise. Our stretch "Base" modeling is illustrated in orange at the mean redshift of the redshift ranges, 0.15 and 0.25, respectively. P-values of the Kolmogorov-Smirnov tests are indicated on the top (bottom) of each panel showing no sign that the SDSS and PS1 (PS1 and SNLS) x_1 and c distributions are not drawn from the same underlying distributions.

are PS1 and SDSS in the redshift range $0.10 < z < 0.20$ (see Fig **??**). This corresponds to the 146 SNe Ia from SDSS at the highest redshift end and thus potentially the most affected by potential selection effect leftover (see the corresponding discussion in section **??**). In that same redshift range, PS1 has 52 SNe Ia, which are at in the lowest redshift bin and thus unlikely to have any selection issue. The Fig.

We highlight that the SNe Ia color is more prone to selection effects than stretch as illustrated in Fig. **??**; see also e.g., Fig. 3 of **?**. Hence, since the color shows no sign of leftover selection effects, it further supports our claim that our simple redshift-based selection criteria is sufficient to test the redshift drift of the SN Ia stretch distribution.

3. Modeling the redshift drift

? 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., **??????**). 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., **?**). Since $\delta(0.05) \sim \psi(0.05)$ (**???**), in agreement with rate expectations (**??**), **?** 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 predicted by **?** 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.

? presented the relation between SN stretch and LsSFR measurement, a progenitor age tracer, using the SNfactory sample. This relation is shown in Fig. **??** 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., $\log(\mathrm{LsSFR}) \geq -10.82$), the stretch distribution is modeled as a single normal distribution $\mathcal{N}(\mu_1, \sigma_1{}^2)$;
- the older population (i.e., $\log(\mathrm{LsSFR}) < -10.82$) 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 **??**). But generally, the fraction of young SNe Ia as a function of redshift is given by $\delta(z)$ (see Eq. **??**) and therefore, our redshift drift model of the underlying distribution of SNe Ia as a function of redshift $X_1(z)$ is given by:

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

3.2. Base model applied to data

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

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

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:

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

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

3.2.1. 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. **??**), to minimize Eq. **??** with respect to $\boldsymbol{\theta}$. Results on fitting the SNf SNe with this model are shown Table **??** and illustrated in Fig. **??**.

3.2.2. Without LsSFR measurements

When lacking 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. **??**) as a proxy for the probability of a SN to be young. This still corresponds to minimizing Eq. **??** with respect to the parameters $\boldsymbol{\theta} \equiv (\mu_1, \mu_2, \sigma_1, \sigma_2, a)$ of the stretch distribution X_1 (Eq. **??**), but this time assuming $y^i = \delta(z^i)$ for any given SN i.

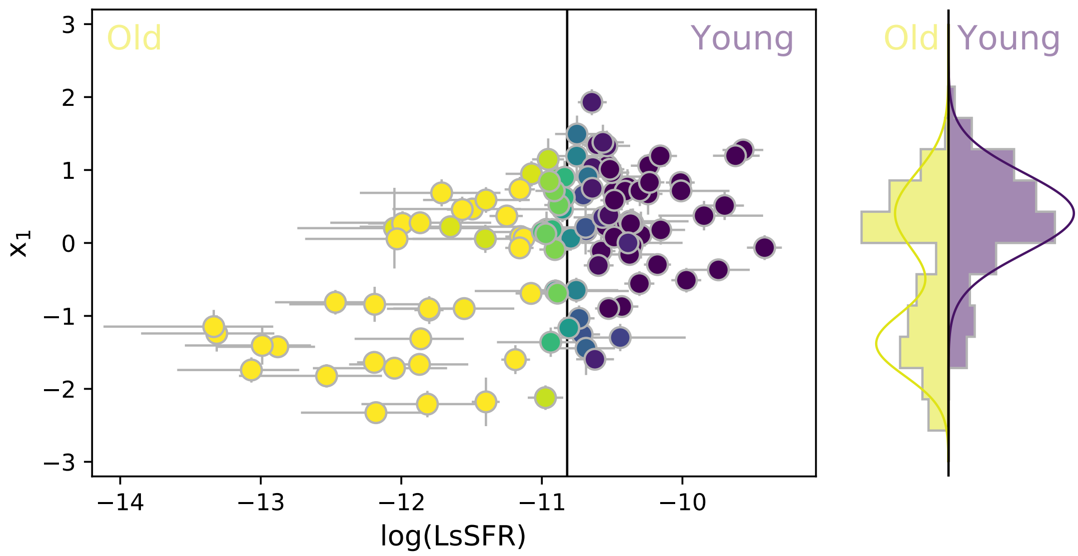

Fig. 5. *Main*: `SALT2.4` lightcurve stretch (x_1) as a function of the local specific star formation rate (LsSFR) for SNfactory SNe used in this analysis. The color corresponds to the probability, p_y, for the SNe Ia to be young, i.e. to have $\log \mathrm{LsSFR} \geq -10.82$ (see **?**). *Right*: p_y-weighted 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.

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

For the rest of the analysis, we will therefore minimize Eq. **??** using p_y^i – 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 **??**, and the predicted redshift evolution of mean stretch (expected x_1 given the distribution of Eq. **??**) illustrated as a blue band in Fig. **??** 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. **?**) *and* if the fraction of old SNe Ia declines as a function of redshift. See Section **??** for a more quantitative discussion.

3.3. Alternative models

In Section **??**, we have modeled the underlying stretch distribution following **?**, 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 **??**.

? 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. **??**.

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, **??**), used in recent SN cosmological analyses (e.g. **????**) 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 (**??**). See further discussion concerning BBC in Section **??**.

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 **??**); results are gathered in Table **??**, and illustrated in Fig. **??**.

Since the various models have different degrees of freedom, we use the Akaike Information Criterion (AIC, e.g. **?**) to com-

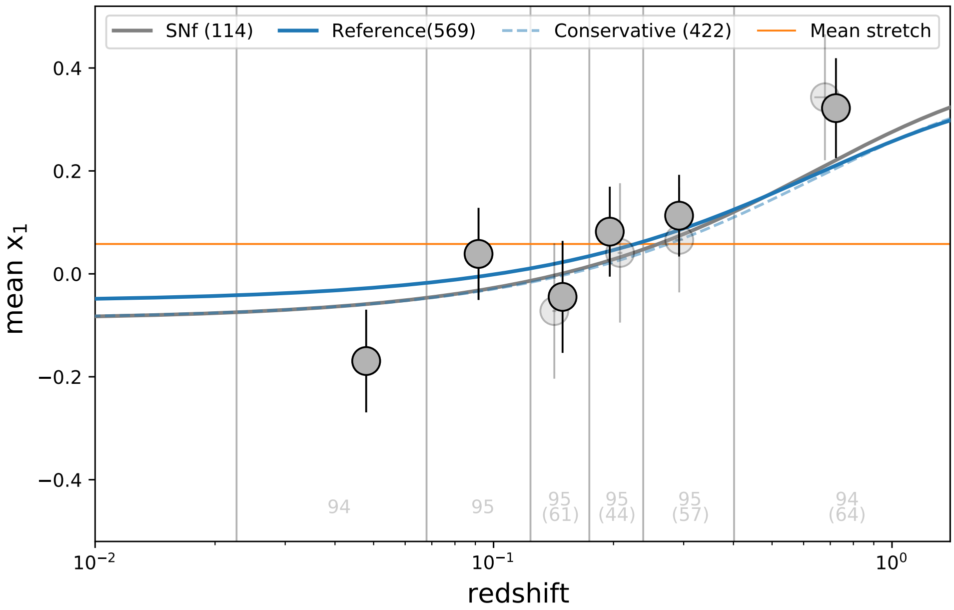

Fig. 6. Evolution of the mean SN `SALT2.4` stretch (x_1) as a function of redshift. Markers show the mean stretch measured in redshift bins of equal sample size, indicated in light gray at the bottom of each redshift bin. Full and light markers are used when considering the fiducial or the conservative samples, respectively. The orange horizontal line represents the mean stretch of the fiducial 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 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.

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. **??**), 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

pare 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:

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

where $-2\ln(L)$ is derived by minimizing Eq. (**??**), 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 $\Delta\mathrm{AIC} < 5$ are coined acceptable, the ones with $5 < \Delta\mathrm{AIC} < 20$ are unfavored, and those with $\Delta\mathrm{AIC} > 20$ 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 ($p \equiv \exp(\Delta\mathrm{AIC}/2) = 5\times10^{-6}$) to describe the data as well as the Base model. This result is just a quantitative assessment of qualitative facts clearly visible in Fig. **??**: 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. **??**.

Surprisingly, the sample-based Gaussian asymmetric modeling used by current implementations of the BBC technique (**??**) has one of the highest AIC value in our analysis (see Section **??**). 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 $\Delta\mathrm{AIC} < -20$, which could be interpreted as a probability $p = 2\times10^{-7}$ of being an as good representation of the data as the Base model.

Remark that, when comparing models adjusted on individual subsamples rather than globally, the Bayesian Information Criterion ($\mathrm{BIC} = -2\ln(L) + k\ln(n)$, 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

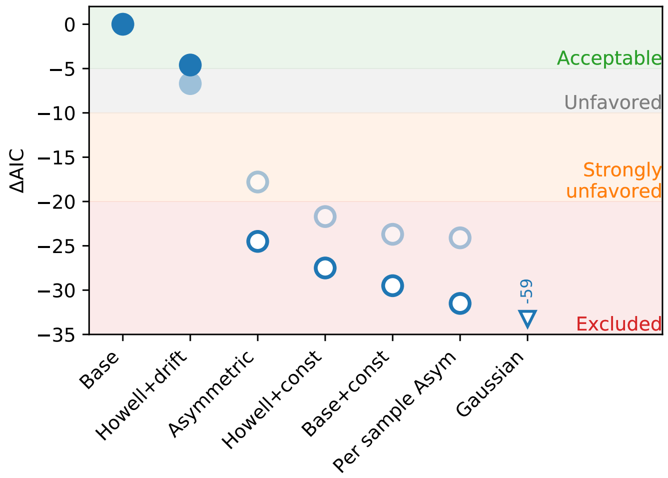

Fig. 7. ΔAIC between "Base" model (reference) and other models (see Table ??). Full and open blue markers correspond to models with and without redshift drift, respectively. Light markers show the results when the analysis is performed on the conservative sample rather than the fiducial one. Color-bands illustrate the validity of the models, from Acceptable ($\Delta\text{AIC} > -5$) to Excluded ($\Delta\text{AIC} < -20$), 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.

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

each sample. Doing so, we find $\Delta\text{BIC} = -48$, again refuting the sample-based asymmetric Gaussian model as being as pertinent as the Base model.

We report in Table ?? the samples' μ_0 and $\sigma_\pm$ adjusted on the nominally selection-free samples using our fiducial cuts (see Section ??). We find our results in close agreement with ? for SNLS and SDSS and with ? for PS1, who derived these model parameters using the full BBC formalism, using numerous simulations to model the selection effects (see details e.g., Section 3 of ?). 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 get a sample with negligible selection effects. If we were to use ? and ? 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 ??.

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 strengthens the fact 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. ??. We tested that by allowing this population to also be bi-modal, finding the amplitude of this low-stretch mode to be compatible with 0 in this young population (< 2%). More generally, this raises the question of inaccurate tracing of age by a given environmental tracer (here the LsSFR). A dedicated analysis will be presented in Briday et al. in prep.

Finally, ignoring the LsSFR measurements – available only for the SNfactory dataset, see Section ?? – reduces the significance of the results presented in this section, as expected. Yet, non-drifting models remain strongly disfavored, and for instance, the best fitted sample-based Gaussian asymmetric modeling still is $\Delta\text{AIC} < -10$ less representative of the data than our Base drifting modeling.

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 (?), BAHAMAS (?) or Steve (?) can easily allow it; see e.g., section 1.3 and 2.5 of ?. 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 Phillips' relation (?) standardization parameter α is not redshift dependent (a study behind the scope of this paper, but see e.g. ?), 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 at least for the upper half of their SN redshift distribution. As a consequence, an ill-modeling of the underlying stretch distribution will bias the SN magnitude derived from such surveys.

Commonly used Malmquist bias correction techniques, such as the BBC-formalism, assume per sample asymmetric Gaussian functions for modeling the underlying stretch and color distributions. Yet, as shown in Section ??, such a sample-based distribution is excluded as being as good as our drifting model. Then, unlike what ?, Section 2 and ?, Section 5.4 suggested, i.e. that traditional surveys span limited redshift ranges and that therefore 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 stress 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 is becoming less valid, notably for PS1, DES and, soon, LSST.

We illustrate in Fig. ?? 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 is shown as blue to red underlying distributions in Fig. ?? for redshift ranges covered by PS1. The observed x_1 histogram follows our modeling defined as 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 affected by selection effect would be biased by an ill-modeling of the true underlying stretch distribution.

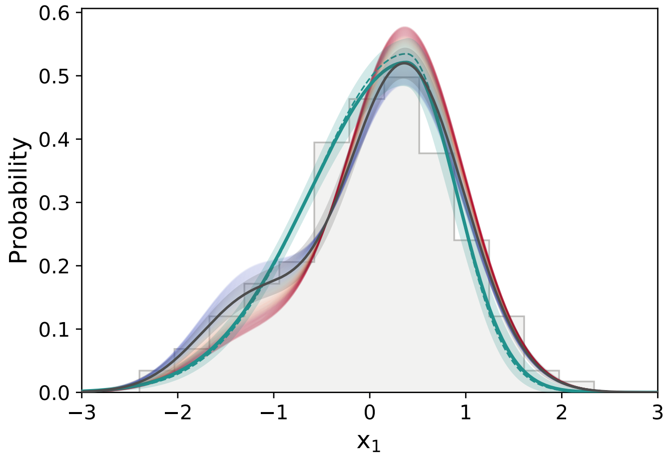

Fig. 8. Distribution of the PS1 SN Ia SALT2.4 stretch (x_1) after the fiducial redshift limit cut (grey histogram). This distribution is supposed to be a random draw from the underlying stretch distribution. The green lines show the BBC model of this underlying distribution (asymmetric Gaussian). The full line (band) is our best fit (its error); the dashed line shows the **?** result. The black line (band) shows our best fitted Base-modeling (its error, see Table **??**) 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.

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. **??**) 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 at higher redshifts, precisely where the underlying distribution will matter for correcting Malmquist biases.

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

6. Conclusion

We have presented a study of the drift of the underlying SNe Ia stretch distribution as a function of redshift. We built a magnitude-limited SN Ia sample from the Pantheon dataset (**?**, SDSS, PS1 and SNLS), to which we added HST and SNfactory data from **?** for the high- and low-redshift bins. We only considered the SNe that have been discovered in the redshift range of each survey where selection effects are negligible, so that the observed SNe Ia stretches are random sampling of the true underlying distribution. This resulted in a 569 SN Ia fiducial sample (422 SNe when more conservative cuts were considered).

Following predictions made in **?**, we introduced a redshift drift model 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.

In addition to this "base" modeling, 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 (**??**).

Our conclusions are the following:

1. The underlying SN Ia stretch distribution is significantly redshift dependent, as previously suggested by e.g. **?**. This result is largely independent of details on each age-population modeling.
2. Redshift-independent models are indeed 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 **?**. This second result strongly supports the existence of both young and old SN Ia populations, in agreement with rate studies **????**.
3. Models using survey-based asymmetric Gaussian distributions, as done, e.g., in the current implementation of the BBC, are excluded to be good descriptions of the data relative to our drifting model. Hence, the sample-based approach does not accurately account for redshift drift and even less so as survey span increasingly larger redshift ranges. We stipulate that, even if 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 one 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:

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

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

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

with $K = 0.87$.

In this paper, we considered a simple Gaussian mixture modeling, but 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, **??**), and Subaru and SeeChange SNe Ia programs (high-z), respectively.

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). 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 Region Rhone-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, arXiv e-prints, arXiv:2002.01550
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. 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
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. 2018, arXiv e-prints, arXiv:1807.06209
Poulin, V., Smith, T. L., Karwal, T., et al. 2019, Phys. Rev. Lett., 122, 221301
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
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., 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