

Questions about redshift evolution of SNe Ia

Generic comments

[JN] Suggestion to add a reference to the U-band paper, since the trends between spectroscopic (i.e. explosion) properties and LsSFR we found there shows that there is a real physical difference between SNe in the two LsSFR modes (young vs old), which also better reflects the SN luminosity (Fig 9 and 10). We are thus not talking about pure statistical effects where different environments have different probabilities of generating SNe with different x_1 values, but rather that the SN age directly influences the explosion in a way that x_1 cannot fully capture.

Answer: New sentence near the end of the introduction added.

Section 2: Complete sample

1. [GA] Saying that the spectroscopic follow-up for acquiring typing and host redshift is sufficient and unbiased is *definitely not true for SDSS*

Answer: Indeed and we say so in the text right after.

2. [GA] Targeted surveys wait *volume-limited*, not magnitude-limited

Answer: Indeed, and to get that, we take magnitude-limited surveys below their magnitude limit. Targeted surveys could never be converted into volume-limited surveys.

3. [JN] The discussion on spectroscopic completeness is fairly shallow. It would obviously be very challenging to study this in detail, but is it really through that both SNLS and PS were really spectroscopic complete up to the target redshift? Especially for the results here, any bias in terms of e.g. getting spectra of transients in complex or dusty (young) could offset the model. Are there some direct paper sections you can reference about this?

Answer: It is indeed an important discussion and this is why we provide so many references. For SNLS, Perrett 2010 is all about that; you can look at the whole paper, and specifically Section 5. For SDSS, Dilday 2008 is the reference, and we also pointed to the discussion from Kessler 2009 about it (more specifically in Section 2). For PS, we directly refer to Rest and a figure of Scolnic. But because it is indeed an important step, we also added in this section that this is further tested in section 4, where we confirm the accuracy of our technique.

Changes in text: Precision on which section of Perrett 2010 deals with selection bias, and which section of Kessler 2009 speaks about SDSS follow-up. Added sentence at the end of “The sample selection is summarized in Table 1” paragraph: “We show in section 4 that this sample selection indeed provide a subset of SNe Ia with insignificant selection effects when compared to state-of-the-art Malmquist correction techniques.”




Section 3: Modeling

1. [JN] Looking at the Young population in Figure 4, it seems obvious that this also contains a fraction of the low-stretch mode. Could this bias the model?

Answer: We tested it indeed, but when leaving the low-stretch mode having a non-zero amplitude for the young population, the best fit found it non-significantly different than 0. We added in the text (result section) that we did so.

2. [GA] (About the bump in the young population in Fig. 4) May need to discuss, e.g., maybe it's actually in the old populations but that the LsSFR is not 100% pure near boundary?
Answer: See response to Jakob just before. We did test that, and we added a reference to Martin's work dedicated to the contamination issue (in prep).
3. [JN] Fig. 5. (*note: the person writing this is playing the devil's advocate. The idea is: who could we answer to such a comment?*) "Rather than a trend in redshift, this figure could just as well be explained by outlying behavior from SNfactory and HST. HST is a small sample and was hard to consistently type, so it could very well be off. Maybe SNfactory did something weird and got a biased sample? This would both explain the previous fit to the data, and anything that is presented here (this conclusion agrees with previous claims that the LsSFR effect does not bias H0 studies). The authors consider various analysis alternatives, but none where the SNfactory data is not included. This would be especially important since the model that is fitted was derived from SNfactory data."
Answer: I agree, but removing data because you don't like some points is dangerous. Especially, when testing for redshift drift, removing the outer bins seems very unfair. People might say so and ZTF + Subaru/SeeChange will prove us right or wrong. We now explicitly say so at the very end of the conclusion.
4. [GA] Error bars in Fig. 5 look good; too good? Meaning too large, giving χ^2_μ (*not sure about this part*) $\ll 1$?
Answer: Well, it is what it is. Mean ; STD/sqrt(N-1). The fact that the distributions are non-gaussian plays a role in the visual aspect of things.

Section 4: Results

1. [GA] Trouble with pseudo- χ^2 ; it seemed that we agreed that it should be $\sim N_{\text{SNe}}$, yet it's more like $1.6 \cdot N_{\text{SNe}}$, and $2.5 \cdot N_{\text{SNe}}$ in Table 3. Referees are going to ask about that, should be clearer.
 **Answer:** So $(x-\mu)^2/\text{variance}$ should follow the $\chi^2/\text{DoF} \sim 1$. $(x-\mu)^2/\text{variance} = -2\ln L + \ln(2\pi \text{var})$. When measuring it we have $\chi^2/\text{DoF} \sim 1.6$, which is a bit too large.
For clarity in the text, we removed the "pseudo- X^2 " to only have $-2\ln(L)$ that is not expected to follow χ^2/DoF
2. [GA] There seems to be a big drop in the (*previously χ^2*) between fiducial and conservative samples; we expected a $569-422=147$ difference in (*previously χ^2*), but got \sim twice (*from ~ 1460 to ~ 1080*). Also, note (*I think that's the previous google doc*) says $\chi^2_{\text{base}} \sim 900$; here it is much larger.
 **Answer:** Same answer as just before " $\chi^2 = -2\ln(L) - \sum(\ln(2\pi(dx_1^2 + \sigma^2))) = 907$ for fiducial and 710 for conservative, so -197, close to what is expected. (here $\sigma=0.5$)
3. [GA] It seems that adding more parameters would suit many people, as long as it results in less bias. The penalty doesn't seem as big as the 2k factor in the AIC suggests; it's only related to the efficiency of a model, which is fine for showing it is better, but we need to think about the fact that people might accept a $30/569 \sim 5\%$ loss in efficiency if it means that they don't have to worry about the details of the population model.
 **Answer:** Indeed. What this paper says here is that there is a drift. You can model it using our model, or do something else that will do part of the job or all of it at a price of many more parameters, but at the benefit of less modeling. The paper says nothing about that, the cosmological consequences being left to be studied. We simply highlight that once you are affected by Malmquist bias, you have

to assume a model to extrapolate from the volume-limited to the magnitude-limited part of your sample. We now explicitly say so in the conclusion.

Section 5: Discussion



1. [GA]: David has claimed that as long as the 1st and 2nd moments of a population are correct, there is very little bias. In Figure 8 and in the text you make the point that if there is drift, there will be bias versus redshift - good. But a lot of the effect in Figure 8 is in the mean offset per sample, and if David is right, the means per sample should not be biased. I worry that something is being held fixed that should not be, leading to mean offsets.

Answer: We clarified the text highlighting several times (including in the fig caption) that this is not the actual magnitude bias but rather the mean-stretch bias that has to be multiplied by the amplitude of the bias-correction which depends on both the surveys and redshift.

2. [GA]: In the Unity paper David says: *"The model must also allow for population drift (whether due to selection effects, or changes in the SN population with redshift), or risk a significant bias on cosmological parameters. Population drift can also be handled well with either framework, as Equation (2) handles each object independently (if the measurements are uncorrelated object to object) and our new framework solves for population drift simultaneously with other parameters (Section 2.5)." and elsewhere in the paper: "The prior must be able to vary in redshift more rapidly than the cosmological fit in order to not introduce bias."* It would be good for the text to reflect the fact that this issue is known to be important in *principle* (but then of course emphasize that in this paper you show there is a real drift and so it is important in *practice*).

Answer: Added in the text.

3. [JN]: The BBC model could capture any drift, if existing, right? And with enough data each sample could be split into multiple slices, we are just waiting for LSST. It is not correct that BBC models are excluded - their χ^2 fit is just as good as the Base model. It is correct that the BBC fits more parameters, but they are necessary and transparent. The Base model rather hides extra free parameters in terms of a model which is also determined by data (but in a different paper). **** Note: The key phrase, that the per sample approach does not work as each covers large redshift ranges, comes way too late and is hidden in 5.1. I would find a way to mention this already in the introduction so that the reader has this information all the way. ****

Answer: The conclusion has been clarified. See response to Greg's question above. The issue comes from extrapolation of the volume-limited part of the sample (where you are complete and you set your parameters) to the magnitude-limited part (where you are using these parameters while the distribution has drifted). This is added in the conclusion.

4. [SP]: *About the fact that if α does not depend on z , the stretch-corrected magnitudes are blind to the underlying stretch distribution:* Doesn't the recent Manifold Twins analysis of Boone et al. (SNf collab) shows that there are regions of the x_1 parameter space where SALT2 gets the magnitude wrong for a subpopulation of SNe Ia? Couldn't the relative populations of these subpopulations (only visible with the spectroscopy-derived Manifold Twins components) change with redshift, and wouldn't this change this statement?

Answer: We say in the text that $\alpha(z)$ is outside the scope of this paper. It is indeed very interesting and we might want to look at that in the near future.

Main changes from the previous version

Overall

1. We kept “underlying [distribution]” instead of “intrinsic [distribution]”
2. Changed “In Rigault et al. (2018), **we** presented [...]” to “Rigault et al. (2018) **presented** [...]”.

Section 1: Introduction

1. Paragraph on the H_0 tension simplified.
2. Moved paragraph “The concept of the SNe Ia age dichotomy [...]” before “In this paper, we [...]”

Section 2: Complete Sample

1. No more 2D histogram on Fig. 1
2. Changed “Since one ought to detect this object typically a week before and 10 days after peak [...]” to “Since one ought to detect this object typically **5 days** before and **a week** after peak [...]”
3. Deleted small and unnecessary paragraph: “For the rest of the analysis [...]” mentioning the conservative cuts a second time. Direct reference to the table instead.
4. Measure of mean stretch at low- and high- z moved from the end of 3.1.2 to the end of Section 2., along with referencing to Fig. 5 directly. The aim is to show that with the complete sample only, there is a mean stretch evolution that we can see and that we attempt to model afterward.

Section 3: Modeling

1. Made the distinction between the Base model underlying distribution, $X_1(z)$, and the PDF of a given SN, $P(x_1^i)$. The former is what we plot in Fig. 5 and what is the heart of the paper (the only equation mentioned again in the conclusion); the latter explains *how* we compute the parameters. Hence there is a new subsubsection, and a new paragraph just before it.
2. No more pseudo- χ^2 , only $-2\ln(L)$, to avoid confusion with the value of $-2\ln(L)/\text{DoF} \neq 1$.

Section 4: Results

1. More information on alternative modelings that we didn’t discuss or show in the paper (for the sake of being concise); see the last 2 paragraphs of this section.
2. Information about Per Sample Asym modeling (Delta AIC) moved from section 5 to the middle of section 4.

Section 5: Discussion

1. Rewording and shortening of the introductory paragraph. Overall, we tried to make it less about the BBC than about the sample-based approach.
2. Removed subsubsections
3. Ex 5.1 First paragraph removed
4. Removed comparison with $\overline{\delta}_{x_1}$
5. Stressed the lack of robust cosmological impact of this study

Section 6: Conclusion

1. Moved sentences about ZTF and Subaru after conclusions listing
2. Swapped point 1. and 2. of the conclusions listing
3. Clarification of the issue with BBC modeling.