These different models can be combined to test the quality of the underlying assumptions and possible bias due to incorrectly corrected SNe Ia. These different modelling runs can be combined with each other to test the quality of the underlying assumptions and the possible bias due to incorrectly correcting the SNe Ia. We now turn to the goodness of fit of the simulated data to the real data.

I.3 Comparison of simulated data to real data

Before comparing the cosmological results, we were interested in the correspondence between the simulated and real data in order to appreciate the implications for the distributions of the different models. In this section we present the different diagnostics that allow us to compare both graphically and numerically the agreement between simulated and real data.

First, we show in Figure I.6 the fraction of young stars as a function of mass for the reference mass model, SNfsupp, and for the NR model.

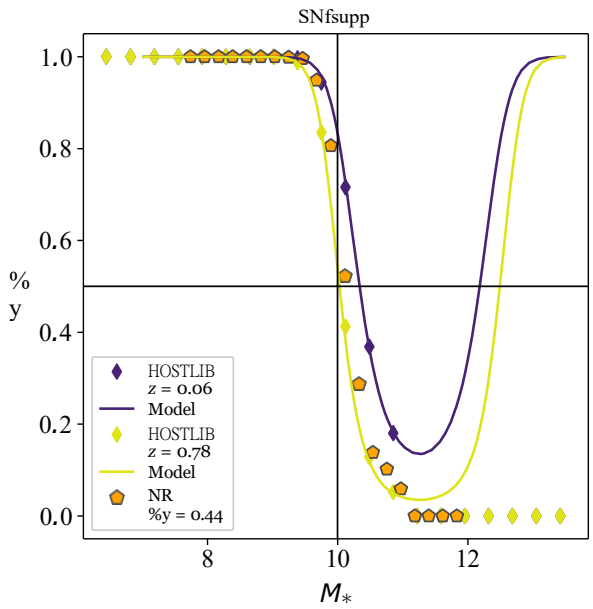


Figure I.6 - *In violet (yellow)*: fraction of young stars as a function of mass for the SNfsupp mass model at medium redshift of the HOSTLIB used at high (low) redshifts.

In orange: same fraction but for the simulated NR sample. We can see here that the model is suppressed for $M_* > 10^{11}$ M_0 .

Then, to have an effective comparison, we randomly select the data from each survey to reproduce the expected ratios. The amounts of data in these measures, expressed as percentages of the Pantheon sample, are equal to the ratios of the smallest simulated survey of unsampled data: those in the
The "SDSS" in Table I.5.

We recall that the BP and SK models use parameter distributions specifically fitted to the data; BP uses asymmetric Gaussian distributions, with 3 free parameters, in intervals of 0.2 ^{1010} M_0 (10 for LOWZ, 20 for the others) to reproduce the stretching of the SNe Ia; SK also includes asymmetric Gaussian distributions, one for each of the SDSS, PS1 and SNLS soundings, and the distribution given in Equation I.2 with 6 free parameters, for a total of $k = 15$. In contrast, the NN and NR models rely on forward modelling, based on a stretch model with 5 free parameters and a

mass with 4 free parameters. The evolution of the fraction of young stars relies on 2 parameters (K, Φ) that are fixed.

We first look at the one-dimensional adjustment of the parameters (Section I.3.1) before dealing with the two-dimensional aspect (Section I.3.2).

I.3.1 Agreement between the data: uni-dimensional analysis

Here we present the results of the simulations of the redshift, stretch and mass parameters of the different models, the graphical representations of which are given in Figure I.7.

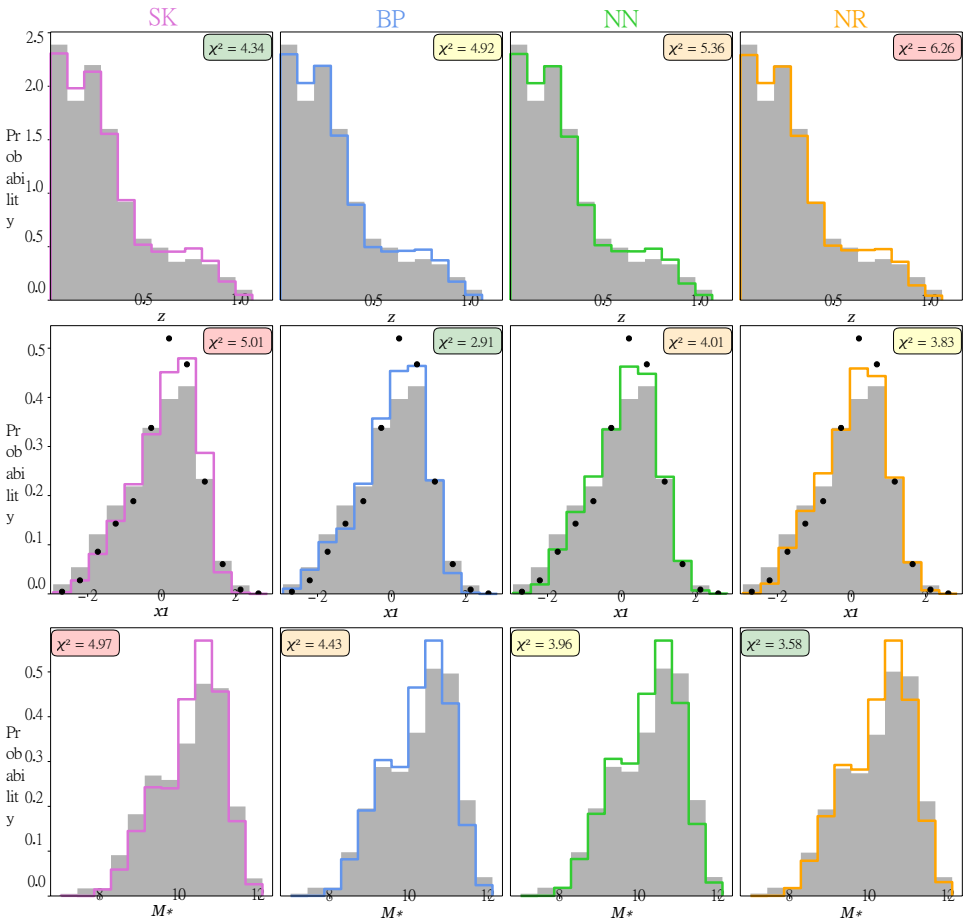


Figure I.7 - Normalized histograms of simulated data (coloured solid lines) and real data (grey) by model and parameter. *From left to right*: results for the SK, BP, NN and NR models, respectively. *From top to bottom*: number of simulated data as a function of redshift, stretch and mass, respectively. The χ^2 values between simulated and real data are shown in the upper right corner of each figure, and from green to red from smallest to largest. We show in black dots the NN stretch model at the mean redshift of the Pantheon sample.

For each of the comparisons, we calculate a value of χ^2. To do this, we normalise the histograms of the simulated data to the number of Pantheon data,

then calculate :

$$\chi 2 = \frac{1}{N} \times \sum_{i=o}^{N-1} \frac{(d_i - s_i)^2}{d_i + s_i} \quad \text{(I.5)}$$

where N is the number of histogram intervals and d_i (s_i) is the number of real (simulated) data in interval i. The best agreement is described by the smallest χ^2. The values are given in Table I.6.

Table I.6 - Values of χ^2 giving a comparison of the ability of each simulation to represent redshift, stretch and mass data.

	$\chi 2$			
Parameter	SK	BP	NN	NR
Redshift	4.34	4.92	5.36	6.26
Stretch	5.01	2.91	4.01	3.83
Mass	4.97	4.43	3.96	3.58
Sum	14.32	12.26	13.33	13.67
Probability	0.36	1.00	0.59	0.49

Notes. For each simulation, a selection of data is made to match the ratios of the Pantheon data, and the calculation of the χ^2 is the average over 500 of these draws each time. The probability is given over the best model (BP), such that $\text{Pmodel} = \exp(\chi^2_{BP} - \chi^2_{model})/2$.

Overall, the BP model appears to be the best description of the data, NN and NR give similar results and SK has the worst agreement. For redshift, however, the SK model is the best fit to the data; this is in line with our expectations since their stretch distributions divide by redshift. For mass, as all simulations use the same WEIGHTMAP, the differences are less noticeable. NR does however give a better representation of the data, but not significantly so. For stretch, BP describes the data best; this also corresponds to our expectations since their stretch distributions are large. We note however that for this parameter, the NN and NR models are well representative of the data. Looking by sample, we observe that in fact the NN and NR models perform on average much better than the other two but have some difficulty in reproducing the LOWZ distribution. Indeed, due to the targeted nature of the survey, the prediction of the NN model cannot apply, which leads to their χ^2 values which we detail in Table I.7. We present Figure I.8 the agreement between simulated and real data of the LOWZ sample for the different models.

In practice, however, the results of the different models are compatible with each other for all surveys, and we can consider them all as good representations of the data. For LOWZ specifically, we could improve the simulation of the survey by modifying the NN model for stretching, notably by including the ZTF data (Chapter ?? Section ??) or by using the expected fraction of young stars. Indeed, the model for the evolution of the fraction of young stars $\delta(z)$ gives a value of 50% young SNe Ia at $z = 0.05$; however, in our case, the survey is located at an average redshift of $z = 0.03$ but the tested data (see Chapter ??, Figure ??) of the

Table I.7 - Values of χ^2 giving a comparison of the ability of each simulation to represent the stretch data for each simulated borehole.

	χ^2			
Survey	SK	BP	NN	NR
LOWZ	14.63	10.29	37.05	36.84
SDSS	7.03	8.52	7.66	7.05
PS1	10.35	3.58	4.06	4.29
SNLS	15.14	23.13	15.03	15.00
Sum	47.15	45.52	63.80	63.18
Without LOWZ	35.52	35.23	26.7 5	26.34

Notes. No prints are made here.

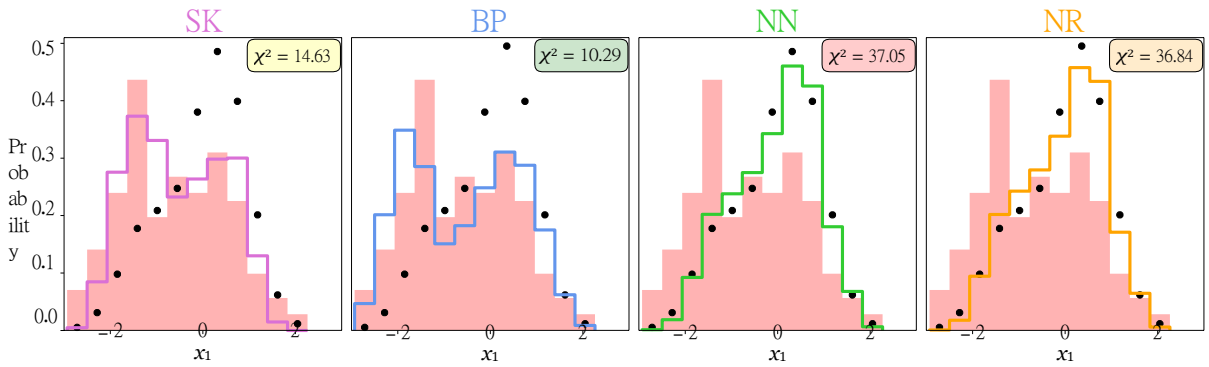


Figure I.8 - Normalized histograms of the stretches of the simulated data (coloured solid lines) and the real data (red) for the LOWZ survey according to the model. *From left to right*: results for the SK, BP, NN and NR models, respectively. The χ values^2 between simulated and real data are shown in the upper right corner of each figure, and from green to red from smallest to largest. We show in black dots the stretching pattern of NN at the average redshift of the LOWZ sample.

The NR model has only 20%, reduced to 15% in the retained data. As we have created our HOSTLIB using the redshift of each input, the agreement with the data is in fact wrong, and it is likely that the true fraction is even lower when looking at these results Figure I.9.

I.3.2 Agreement between the data: bi-dimensional analysis

We now present the stretch distributions as a function of redshift on the one hand and the stretch distributions as a function of host galaxy mass on the other. To determine the agreement between the real and simulated samples in a quantitative way, we used a kernel estimation to convert the simulated data into a two-dimensional probability density, allowing us to calculate a total probability reflecting the agreement between the real data and the kernel. Two examples are given in Figure I.10, where we represent the distributions of the simulated data *via* its kernel estimation in

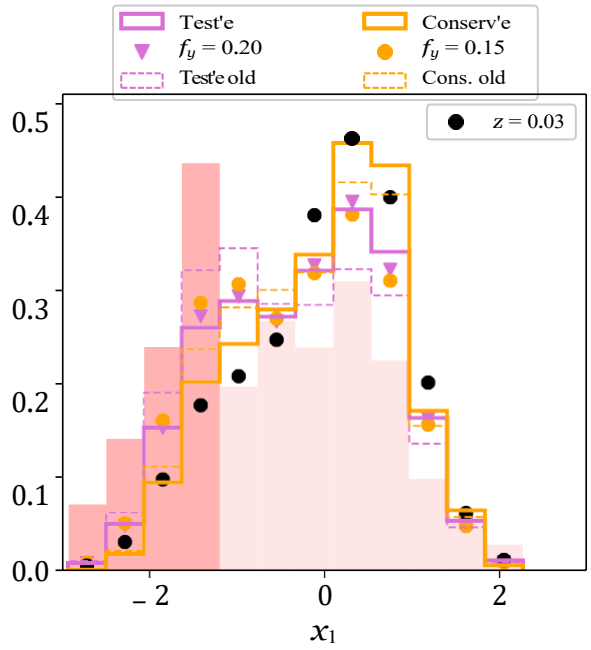


Figure I.9 - Histograms of the stretches of the LOWZ data: *in red* the Pantheon data; *in purple* the tested data and *in orange* the data retained for the NR model. The NN model evaluated at the fractions of the young SNe Ia for these two samples are plotted in markers of the corresponding colour; the model evaluated at the mean redshift of the distribution is plotted in black markers. The old parts of the tested and stored data are dotted.

The results are shown in Table I.8. The results are shown in Table I.8: higher probability represents better agreement.

Table I.8 - Comparison of the ability of each simulation to represent stretch and mass data on the one hand, and stretch and redshift data on the other.

	Probability		
Models	x_1 vs M_*	x_1 vs z	Sum
SK	103,03	252,57	355,60
BP	103,37	246,49	349,85
NN	102,35	236,25	338,60
NR	102,32	235,61	338,93

Notes. For each simulation, we calculate a two-dimensional kernel estimate on the simulated data and use it to determine each probability.

We observe that the SK model is the best of the four on the combination of these distributions; the BP model is second, and the NN and NR models have a similar score placing them as the worst-fitting models to the data. The results differ only slightly on the joint redshift and mass distributions, and most of the difference between the models comes from the joint modelling of stretch and redshift for which the probability values vary more rapidly due to the small size of the data (see coloured bars). However, we observe that the SK and BP models have two roughly equivalent probability clouds at $z < 0.1$, corresponding to the LOWZ survey, whereas for NN and NR the high stretch point cloud is more pronounced than the low stretch one. This naturally follows from the difference in the modelling of the stretch of this sounding,

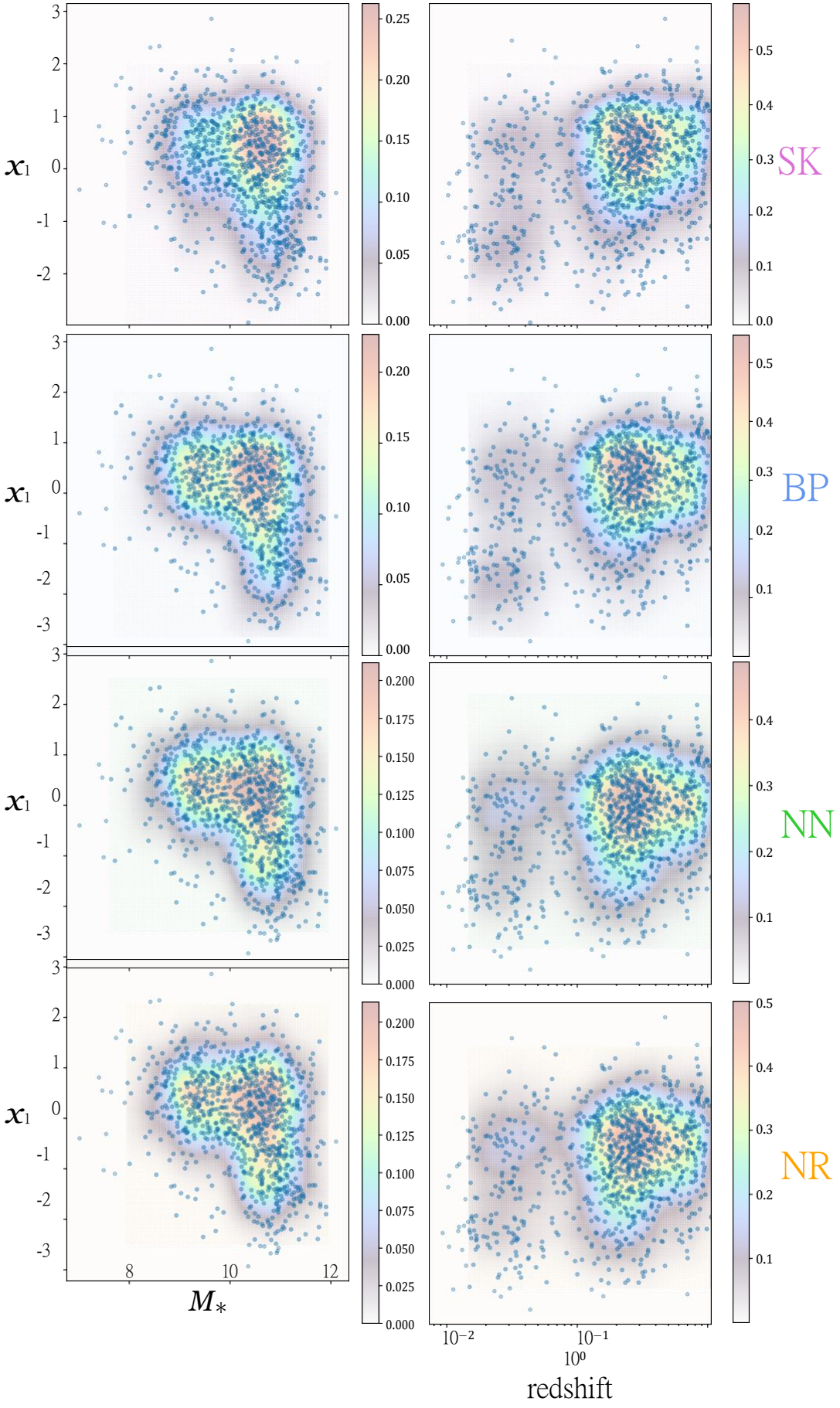


Figure I.10 - Agreement between real (blue transparent dots) and simulated (coloured) 2D data for all models. *On the left*: stretching on the ordinate and mass on the abscissa. *Right:* Stretch on the ordinate and redshift on the abscissa. *From top to bottom*: SK, BP, NN and NR models, respectively, see legend; colours correspond to those in Figure I.7.

as discussed earlier. We note, however, that all four models give similar agreement to the real data.

I.4 Impact on cosmology

Now that we have observed the agreement of each of the models to the real data, we can study the cosmological bias of dealing with data with their own physics with a potentially different correction. To do this, as introduced in Section I.2.5, we apply the BBC7D method (Popovic et al. 2021, see Chapter ?? Section ??) on the data of the different models with each of the BiasCor samples. This gives us 16 corrected samples (whose number of data is given in Table I.5), each with a value of w, γ_{masse} , α and β; as a reminder, we set Ω_M to 0.315 ± 0.005, and we are therefore not interested in the variation of this parameter.

We present Section I.4.1 the results for the standardisation parameters α, β and γ, which are needed to compare the results on the dark energy state parameter w, whose results are presented in Section I.4.2. We discuss the impact of the choice of mass model Section I.4.3.

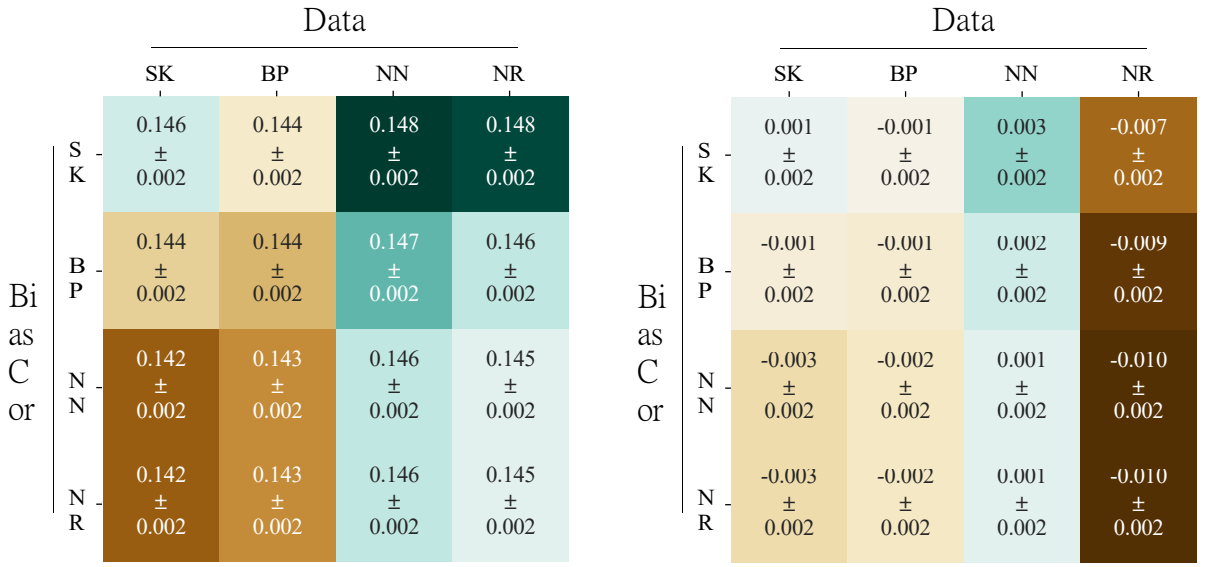
I.4.1 Standardisation results

In order to have meaningful cosmological results, we were interested in the fitted α and β parameter values of our simulations. In order to reproduce the results of the Pantheon analysis (Scolnic et al. 2018), we want the output values to be consistent with the input values, $\alpha_{\text{ref}} = 0.145$ and $\beta_{\text{ref}} = 3.1$. In our first simulations, NR gave a reduced α value of 0.135. We had to raise the input value for this model to $\alpha_{\text{ref,NR}} \approx 0.155$ in order to obtain the results in Figure I.11 where the colours represent the deviation from the reference values. The values of β and γ are shown in Figure I.12, with $\gamma_{\text{ref}} = 0.05$.

We note that the fact that the last two lines are the same is due to the
the fact that the BBC7D process does not take into account the variation in the magnitude of the NR model when the sample is used as a BiasCor, since it is evaluated as θ for the data.

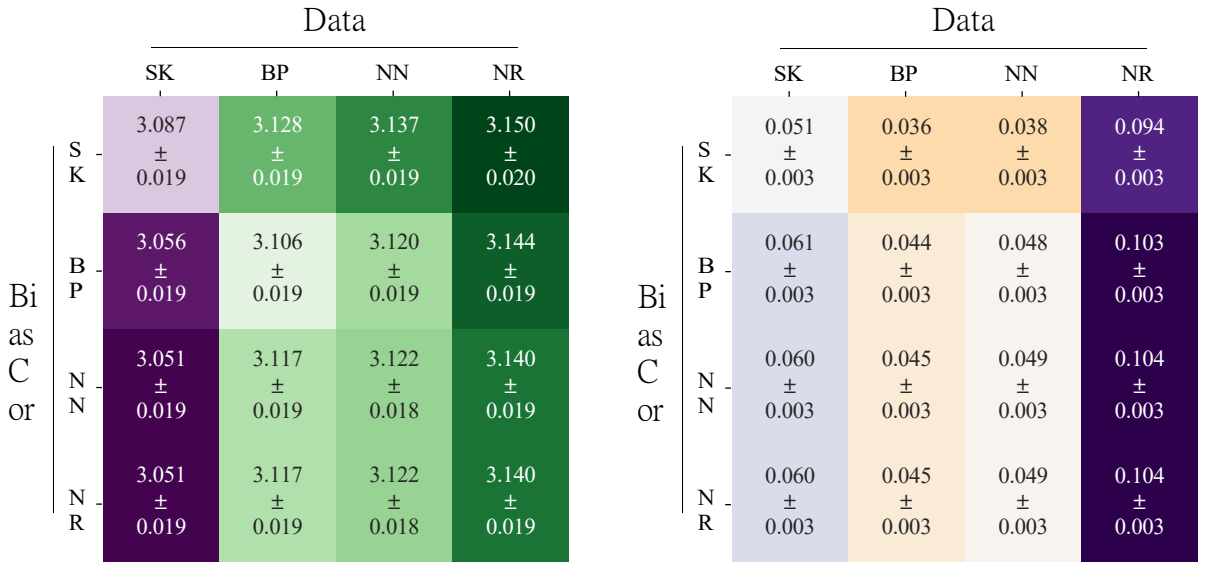
In these figures we see that the SK, BP and NN models are consistent with themselves since the output values are compatible with the input values. For NR, we assume that the fact that one has to change the reference value comes from a misinterpretation of the correlation of magnitude with age by SNANA (here the age-based magnitude walk). Part of this correlation is then associated with α, part with γ and the rest in β. However, we did not adjust the reference β value for NR, although they are 2σ away from the reference value, because the other mass models (SNf, SEDSNf) give a better agreement with this parameter (which is not the case for α); but this does not change the w values significantly as we will see in the following.

With the adjusted mass-based magnitude walk values of Figure I.12b, we observe that the BBC7D method gives a value of $\gamma_{\text{masse}} \approx 0.100$, twice as high as the values usually found for NR. This deviation is inherent to the use of mass as a tracer of the magnitude walk, whereas in the



(a) Values of α. **(b)** Values of $\Delta\alpha$.

Figure I.11 - Cosmological results: values of α and $\Delta\alpha$ determined by fitting with the BBC7D method (see Chapter **??**). The figure on the right highlights that the NR model perceives a reference value of $\alpha = 0.155$.



(a) Values of β. **(b)** Values of γ.

Figure I.12 - Cosmological results: values of β and γ determined by fitting with the BBC7D method (see Chapter **??**).

NR modelling this step is based on the age of the SN. This result is however in full agreement with the study of Briday et al. (2022) who report that a less efficient tracer to differentiate SNe Ia populations will assign a lower magnitude step than a more discriminating tracer; in particular, in their figure 4 (recopied Figure I.13), a value of $\gamma = 0.13$ mag with the LsSFR would result in a value of $\gamma \approx 0.11$ mag with

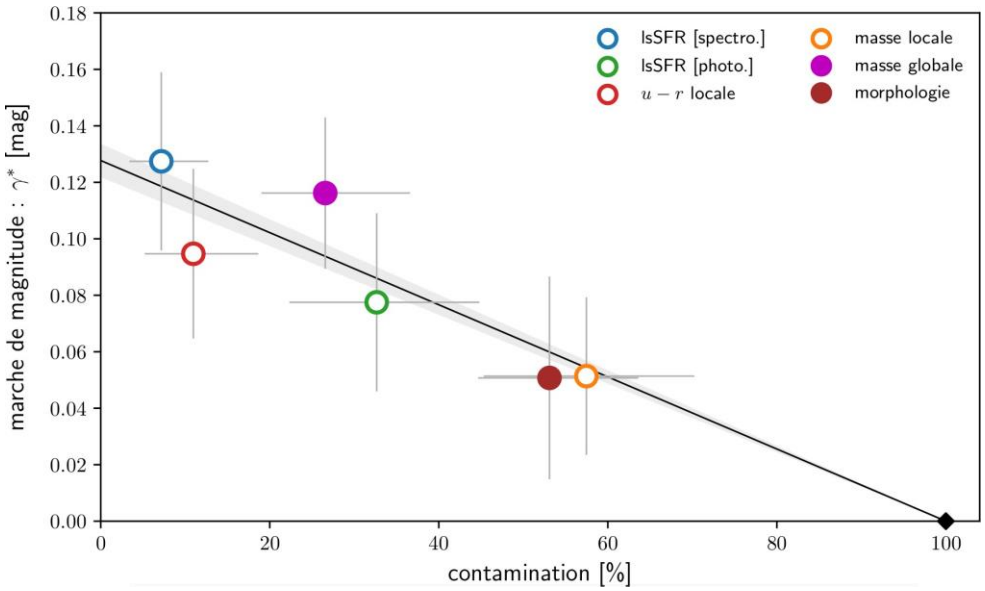


Figure I.13 - SNe Ia magnitude walk as a function of tracer. Figure by Briday (2021). Depending on the ability of the tracers to distinguish SNe Ia populations (reflected by their contamination), the magnitude step found varies. Taking LsSFR as a reference tracer with γ = 0.13 mag, the global mass would give a magnitude step around 0.10 mag.

the overall mass.

I.4.2 Cosmology results

We present in Figure I.14 the results of the fitted values of w, where the colours represent the deviation from the reference value: $w = -1.00$. To understand these results, we observed the evolution of the Hubble residuals and the $\delta\mu$ $\text{value}_{\text{biais}}$ (see Chapter ??) as a function of the redshift for each sample, shown in Figure I.15.

On the first, we observe that the first three models corrected in a manner consistent with their generation (on the diagonal, therefore) are all compatible with $w = -1.00$. On the other hand, the SK model renders the other models incompatible, whether using it for the data (left column) or as a BiasCor (top row), introducing a bias of up to 4% for the NN_SK sample. This results from a systematic under- or over-correction depending on the use of SK, as we can see in Figure I.15a: models giving $w = -1.00$ are balanced around 0, while an under-correction increases this value. The BP_NN and NN_BP samples, on the other hand, give values compatible with -1.00.

The NR model used as data (right-hand column) does not provide any results of w close to -1.00; we argue that this result stems partly from misinterpretations of the age-magnitude relationship as we have seen previously, but also from the consequence of the evolution of these parameters. Indeed, as we observe in Figure I.15a, it is from $z \sim 0.50$ onwards that the Hubble residuals of the NR column exceed the orange horizontal line, whereas those of NN are below it; at this redshift we expect that the average magnitude march starts to be felt between the young and old populations,

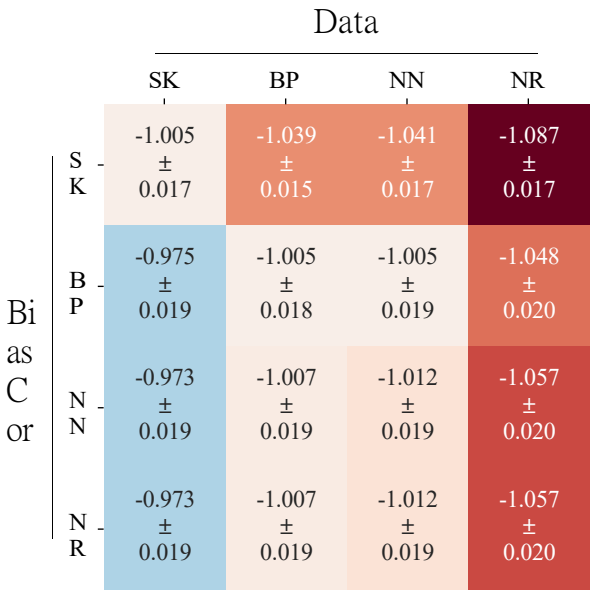


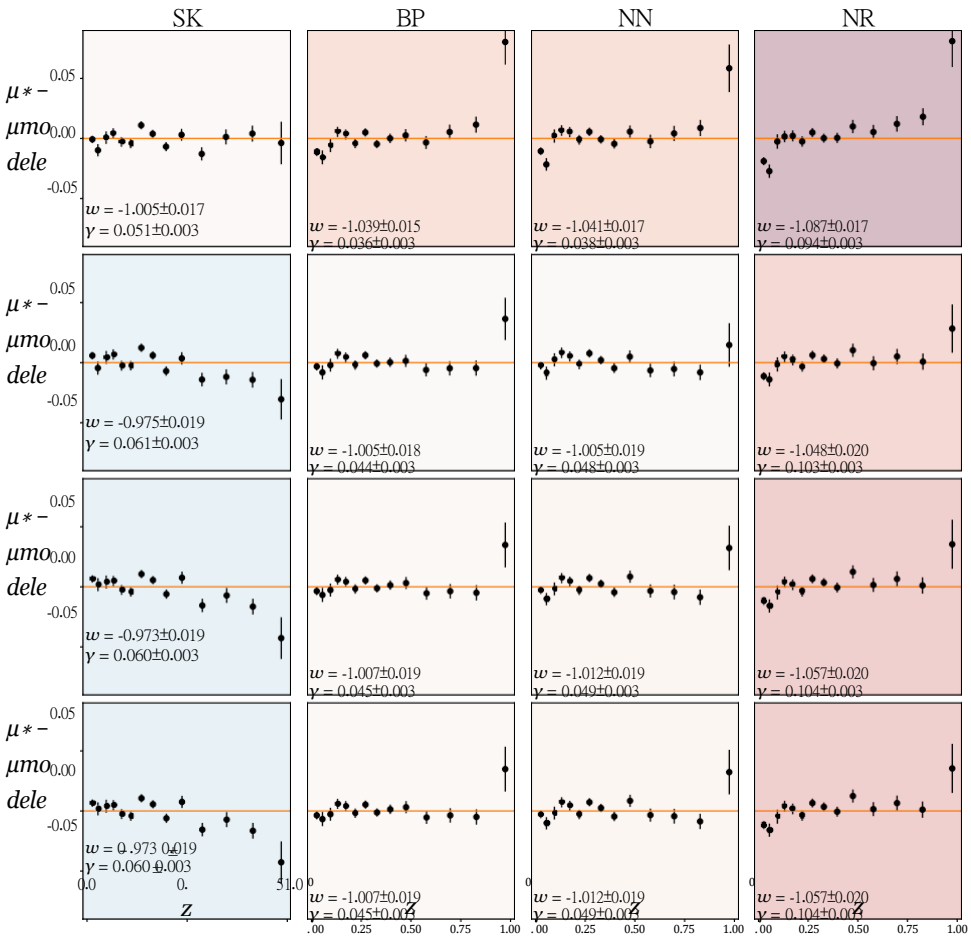
Figure I.14 - Values of w determined by fitting with the BBC7D method (see Chapter **??**).

from $(\gamma_{env})_{z=0.05} = 0$ to $(\gamma_{env})_{z=0.5} = 0.03$ and reaching $(\gamma_{env})_{z=1.0} = 0.05$.

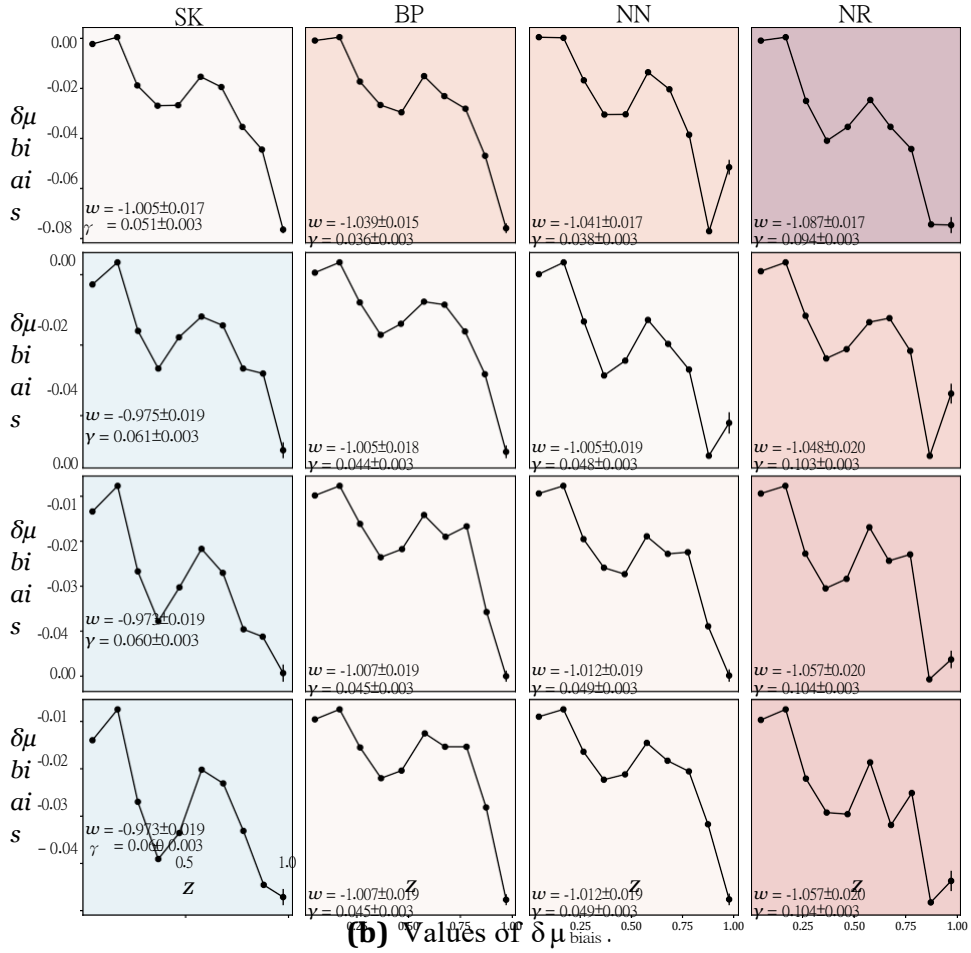
I.4.3 Systematic due to the choice of the mass model

The choice of mass model differs mainly in the fraction of young stars found as a function of mass, whose graphical representations we give in Figure I.16. We show here the cosmological differences due to the choice of mass model, presented on the values of β Figure I.17 and on the values of w Figure I.18. Between the different mass models, SNfsupp (Figure I.6) has the lowest fraction of young stars with 1% less than the other models. The impact on the standardisation only affects the values of β. Thus, all else being equal, it would appear that a higher fraction of young stars in a sample causes a some decrease in the β parameter.

This seems to be reflected in the values of w as shown in Figure I.18, where we observe an average increase in values in both right-hand columns.

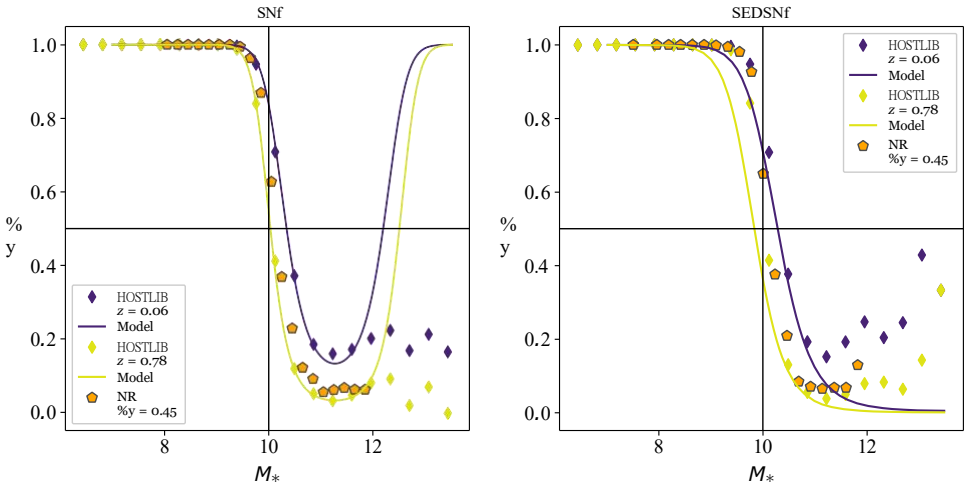


(a) Hubble residual values.



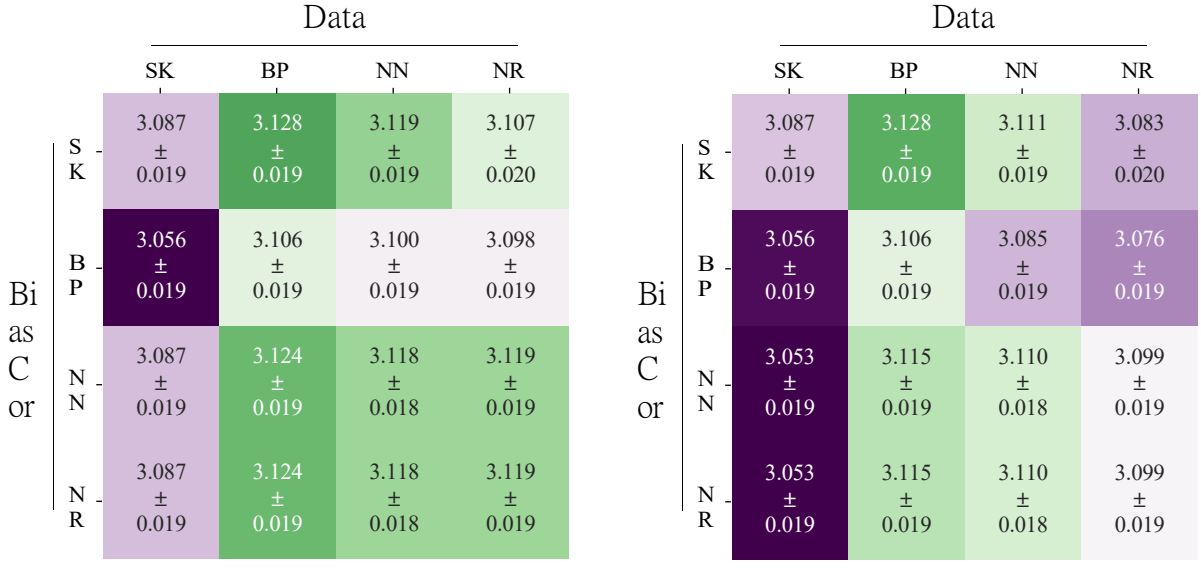
(b) Values of $\delta\mu_{biais}$.

Figure I.15 - Cosmological results: Hubble residuals and $\delta\mu_{biais}$ determined by fitting with the BBC7D method (see Chapter **??**).



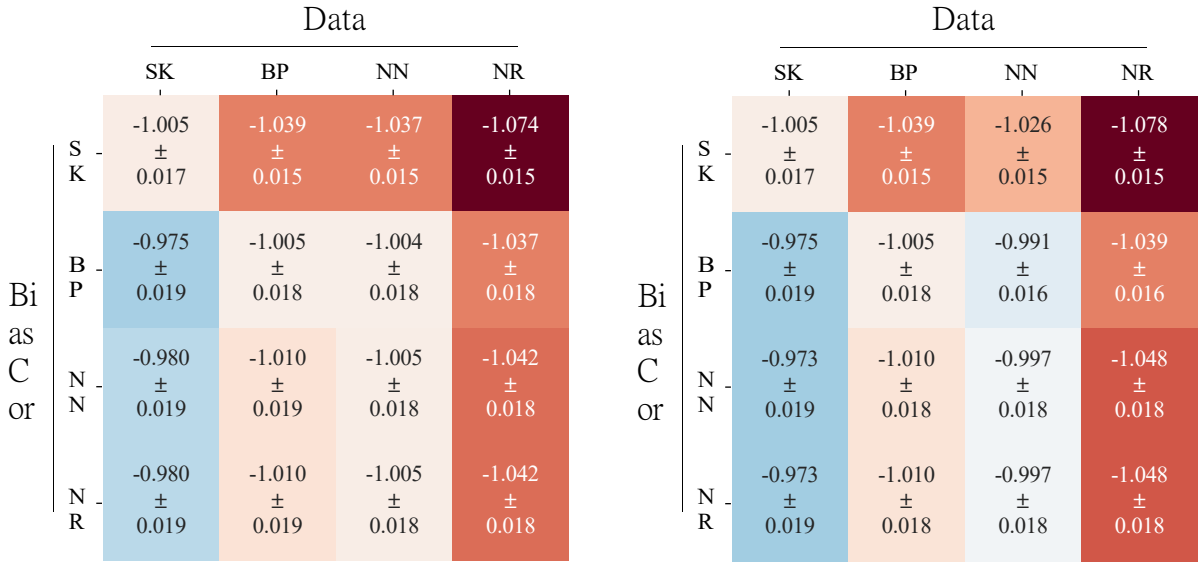
(a) Fraction for SNf. **(b)** Fraction for SEDSNf.

Figure I.16 - Evolution of the fraction of young stars as a function of mass for the SNf model *on the left* and the SEDSNf model *on the right*. *In purple (yellow)*: fraction of young stars as a function of mass for the SNfsupp mass model at the mean redshift of the HOSTLIB used at high (low) redshifts. *In orange*: same fraction but for the simulated NR sample.



(a) Values of β for SNf. **(b)** Values of β for SEDSNf.

Figure I.17 - Cosmological results: β values for SNF mass models *on the left* and SEDSNf *on the right*.



(a) Values of w for SNf.

(b) Values of w for SEDSNf.

Figure I.18 - Cosmological results: w values for SNF mass models *on the left* and SEDSNf *on the right*.

I.5 Conclusion

We have presented a continuation of our study of the drift of the underlying stretch distribution of SNe Ia as a function of redshift. We used sample mass data to establish a mass drift model that depends on the expected fraction of young and old SNe Ia as a function of redshift. This allowed us to augment the BP HOSTLIBs by associating an age to each of the entries, and thereby change the correlation with stretch and magnitude.

This implementation allowed us to test the cosmological results found by SNANA when the data correction can be made inconsistent with its generation, reproducing the correlations of SK and BP. We then obtain 16 samples for which we have values of α, β, γ_{masse} and w.

Through this analysis, we have been able to show the following elements:

1) Each of the models (SK, BP, NN, NR) are good representations of the data. The targeted nature of the LOWZ survey, however, makes it a sample that we were unable to replicate with our forward-looking stretch model approach;

2) We have highlighted the current limitation of the SNANA software package to take into account the age o f a SN as a tracer of its properties;

3) We have highlighted the existence of a bias of around 4% on the calculation of w according to the correction hypotheses.

This could be corrected with the implementation of the age as a parameter describing the intrinsic physics of SNe Ia, and would open the way to the study of the cosmological bias arising from the use of different tracers.

Figures

Tables

Bibliography

Briday M. 2021, "Étude de l'impact de l'environnement galactique sur la standardisation des Supernovae de Type Ia", Theses, Université Claude Bernard - Lyon I, HAL theses
↑ Section I.13

Briday M. , Rigault M. , Graziani R. et al. 2022, "Accuracy of environmental tracers and consequences for determining the Type Ia supernova magnitude step", A&A, 657, A22
↑ Section I.2, ↑ Section I.2.3, ↑ Section I.4.1

Burnham, Kenneth & Anderson D. R. 2004, "Multimodel Inference: Understanding AIC and BIC in Model Selection", Sociological Methods & Research, 33, 261 Sectipn I.2.2

Howell D. A. Sullivan M. , Conley A. and Carlberg R. 2007, 'Predicted and Observed Evolution in the Mean Properties of Type Ia Supernovae with Redshift', ApJ, 667, L37 ↑ Section I.2.1

Nicolas N. , Rigault M. , Copin Y. et al. 2021, "Redshift evolution of the underlying type Ia supernova stretch distribution", A&A, 649, A74 ↑ Section I.1.3, ↑ Section I.2.2,
↑ Section I.2.5, ↑ Section I.7, ↑ Section I.3.1, ↑ Section I.8, ↑ Section I.9, ↑ Section I.3.2

Popovic B. Brout D. , Kessler R. , Scolnic D. and Lu L. 2021, "Improved Treatment of Host-galaxy Correlations in Cosmological Analyses with Type Ia Supernovae", ApJ, 913, 49 ↑ Section I.1, ↑ Section I.1.2, ↑ Section I.2.4, ↑ Section I.2.5, ↑ Section I.3,
↑ Section I.3.2, ↑ Section I.4, ↑ Section I.5

Rigault M. , Brinnel V. Aldering G. et al. 2020, 'Strong dependence of Type Ia supernova standardization on the local specific star formation rate', A&A, 644, A176
↑ Section I.1.4, ↑ Section I.2.1, ↑ Section I.1

Scolnic D. and Kessler R. 2016, 'Measuring Type Ia Supernova Populations of Stretch and Color and Predicting Distance Biases', ApJ, 822, L35 ↑ Section I.1, ↑ Section I.1.1,
↑ Section ?? ↑ Section I.2, ↑ Section I.2.5, ↑ Section I.2.5, ↑ Section I.3, ↑ Section I.3.2,
↑ Section I.5

Scolnic D. M. , Jones D. O. , Rest A. et al. 2018, 'The Complete Light-curve Sample of Spectroscopically Confirmed SNe Ia from Pan-STARRS1 and Cosmological Constraints from the Combined Pantheon Sample', ApJ, 859, 101 ↑ Section ?? , ↑ Section I.1.1,
↑ Section ?? ↑ Section I.5, ↑ Section I.5, ↑ Section I.4.1

Taylor E. N. , Hopkins A. M. , Baldry I. K. et al. 2011, 'Galaxy And Mass Assembly (GAMA): stellar mass estimates', MNRAS, 418, 1587 ↑ Section I.2.1