



HAL
open science

Accuracy of environmental tracers and consequences for determining the Type Ia supernova magnitude step

M. Briday, M. Rigault, R. Graziani, Y. Copin, G. Aldering, M. Amenouche, V. Brinnel, A. G. Kim, Y.-L. Kim, J. Lezmy, et al.

► To cite this version:

M. Briday, M. Rigault, R. Graziani, Y. Copin, G. Aldering, et al.. Accuracy of environmental tracers and consequences for determining the Type Ia supernova magnitude step. *Astronomy and Astrophysics* - A&A, 2022, 657, pp.A22. 10.1051/0004-6361/202141160 . hal-03501082

HAL Id: hal-03501082

<https://hal.science/hal-03501082>

Submitted on 22 Dec 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Accuracy of environmental tracers and consequences for determining the Type Ia supernova magnitude step

M. Briday¹, M. Rigault¹, R. Graziani¹, Y. Copin¹, G. Aldering³, M. Amenouche², V. Brinnet⁵, A. G. Kim³, Y.-L. Kim¹, J. Lezmy¹, N. Nicolas¹, J. Nordin⁵, S. Perlmutter⁴, P. Rosnet², and M. Smith¹

¹ Univ Lyon, Univ Claude Bernard Lyon 1, CNRS, IP2I Lyon/IN2P3, IMR 5822, 69622 Villeurbanne, France
e-mail: briday@ipnl.in2p3.fr

² Université Clermont Auvergne, CNRS/IN2P3, Laboratoire de Physique de Clermont, 63000 Clermont-Ferrand, France

³ Lawrence Berkeley National Laboratory, 1 Cyclotron Rd., Berkeley, CA 94720, USA

⁴ Department of Physics, University of California Berkeley, 366 LeConte Hall MC 7300, Berkeley, CA 94720-7300, USA

⁵ Institute of Physics, Humboldt-Universität zu Berlin, Newtonstr. 15, 12489 Berlin, Germany

Received 23 April 2021 / Accepted 17 August 2021

ABSTRACT

Type Ia Supernovae (SNe Ia) are standardizable candles that allow us to measure the recent expansion rate of the Universe. Due to uncertainties in progenitor physics, potential astrophysical dependencies may bias cosmological measurements if not properly accounted for. The dependency of the intrinsic luminosity of SNe Ia with their host-galaxy environment is often used to standardize SNe Ia luminosity and is commonly parameterized as a step function. This functional form implicitly assumes two-populations of SNe Ia. In the literature, multiple environmental indicators have been considered, finding different, sometimes incompatible, step function amplitudes. We compare these indicators in the context of a two-populations model, based on their ability to distinguish the two populations. We show that local $H\alpha$ -based specific star formation rate (lsSFR) and global stellar mass are better tracers than, for instance, host galaxy morphology. We show that tracer accuracy can explain the discrepancy between the observed SNe Ia step amplitudes found in the literature. Using lsSFR or global mass to identify the two populations can explain all other observations, though lsSFR is favoured. As lsSFR is strongly connected to age, our results favour a prompt and delayed population model. In any case, there exists two populations that differ in standardized magnitude by at least 0.121 ± 0.010 mag.

Key words. distance scale – surveys – supernovae: general – cosmology: observations

1. Introduction

Type Ia supernovae (SNe Ia) are powerful empirically standardized distance indicators. They enabled the discovery of the acceleration of the expansion of the Universe (Riess et al. 1998; Perlmutter et al. 1999), and today are still key cosmological probes in the context of the new generation of surveys (Scolnic et al. 2019). SNe Ia play an important role in probing the nearby Universe ($z < 0.3$) and are the last step of the direct distance ladder to derive the Hubble-Lemaître constant H_0 (e.g., Freedman et al. 2001; Riess et al. 2009). Interestingly, when calibrating the SNe Ia absolute luminosity using the Cepheid period–luminosity relation, this direct H_0 measurement is 4.4σ higher than expectation based on the Lambda cold dark matter (Λ CDM) model anchored by Planck Collaboration VI (2020) data (Riess et al. 2019; Reid et al. 2019). This “tension” has received a great deal of attention as it could be a sign of new fundamental physics (Knox & Millea 2020). This finding is supported by analyses of strongly lensed quasars that are also reporting high H_0 measurements (e.g., Wong et al. 2020). However Freedman et al. (2019) find a lower H_0 value when using tip of the red giant branch (TRGB) distances in place of the Cepheids.

This raises the question of systematic uncertainties affecting direct H_0 measurements in particular, and the distances derived from the observation of the SNe Ia in general. Rigault et al. (2015) suggests that an unaccounted for astrophysical bias,

affecting the derivation of the absolute SNe Ia luminosity, could explain at least part of the tension. The SNe Ia from the calibrating sample significantly differ from the Hubble flow ones; those in the calibrating sample are selected such that their host galaxy also contains Cepheid stars and are thus star forming. Rigault et al. (2020) claim that SNe Ia from younger environments are 0.16 mag fainter than those from older environments, leading to a bias on H_0 because of this selection effect. Even so, Riess et al. (2019) have mimicked the Cepheid selection function onto the Hubble flow sample and find no variation in H_0 , suggesting that they are not affected by this astrophysical effect (see also Jones et al. 2015).

After more than a decade of analyses, the amplitude and the root causes of the astrophysical biases affecting the distance measurements from SNe Ia remain unclear. Early Ia rate studies have shown evidences that two populations of SNe Ia may exist: one arising from young (<100 Myr) progenitor systems and one related to older (Gyr), most evolved progenitors (i.e., the A+B or prompt and delayed models; see e.g., Mannucci et al. 2005, 2006; Scannapieco & Bildsten 2005; Sullivan et al. 2006; Aubourg et al. 2008; Smith et al. 2012; Maoz et al. 2014; Rodney et al. 2014). The first significant evidence of an astrophysical bias in the SN distance derivations was the observed dependency of the standardized SNe Ia magnitude (using the classical two-parameter light curve standardization method) with host galaxy stellar mass (e.g., Kelly et al. 2010; Sullivan et al. 2010; Lampeitl et al. 2010;

Gupta et al. 2011; Childress et al. 2013b; Betoule et al. 2014; Uddin et al. 2017; Ponder et al. 2020): SNe Ia from massive galaxies ($M_* > 10^{10} M_\odot$) are brighter, after standardisation, by ~ 0.1 mag.

We use the term magnitude-step to describe the difference in average standardized magnitudes between two SN Ia subsamples defined from an environmental tracer cutoff. This simple functional form has been shown by Childress et al. (2013b) to be the best fit to data in comparison to a linear trend or other theoretically inspired forms. We further highlight that, in practice, this data-driven step implies that there are two populations of ‘SN Ia+environment’ that are simultaneously present; this observation is the central point of this paper.

The term mass-step has been extensively used in the literature for the global host-stellar mass tracer. The mass-step is used as a third standardization parameter in many recent SNe Ia cosmological analyses (Sullivan et al. 2010; Betoule et al. 2014; Scolnic et al. 2018), including the direct H_0 measurements from Riess et al. (2016, 2019). The amplitude of this effect is ~ 0.08 mag. However, the underlying physics causing this magnitude dependency, and the proper way to account for such astrophysical biases, remain unclear.

In the last decade many host environmental studies seem to have converged towards either the age of the progenitor, or dust around the progenitor or in the host interstellar medium as the origin of the mass-step. Rigault et al. (2013, 2020), Roman et al. (2018), Kim et al. (2018) and Kelsey et al. (2021) would suggest age, while others, like Brout & Scolnic (2021), suggest that variable dust extinction curves affecting the observed color of the supernova can explain correlations with host galaxy properties. Rigault et al. (2020) show the most significant correlation between the luminosity of the SN and the properties of the environment. They split their SNe Ia as a function of the specific star formation rate (sSFR) derived from $H\alpha$ flux measured within a 1 kpc radius projected onto the local environment around the SN (local sSFR; lsSFR). The SNe Ia having a large lsSFR, hence a large fraction of young stars in their vicinity, are fainter than those from passive local environments by 0.163 ± 0.029 mag. Since high-mass galaxies favor older stellar populations, massive hosts favor SNe from old environments, and so are, on average, brighter after light curve standardization, resulting in the mass-step.

Surprisingly, while most SN samples now observe significant correlation between host properties and standardized SN magnitudes, the magnitude step amplitudes differ and seem incompatible. For instance, the Supernova Legacy Survey (SNLS) five-year data and the Sloan Digital Sky Survey (SDSS) data from Roman et al. (2018), updating the Joint Light-curve Analysis (JLA) catalog from Betoule et al. (2014), find a local U-V step of 0.091 ± 0.013 mag, seemingly incompatible with Rigault et al. (2020). Jones et al. (2018), using the low-redshift bin of the Pantheon dataset (Scolnic et al. 2018), reported that locally massive environments (i.e., having a large surface density of stars) are fainter by 0.067 ± 0.017 mag. While in agreement with the local mass-step reported in Rigault et al. (2020, 0.059 ± 0.024 mag), this effect is significantly weaker than that observed using the lsSFR indicator. Finally, using the Pantheon dataset Pruzhinskaya et al. (2020) found that SNe Ia from elliptical and lenticular galaxies are brighter (0.058 ± 0.019 mag), in agreement with Kim et al. (2019, using JLA, 0.018 ± 0.052 mag) and Henne et al. (2017, 0.04 ± 0.05 mag).

As already pointed out by Jones et al. (2018), this variety of results, made using different host tracers, local or global, brings confusion about how to best account for astrophysical biases in

SN cosmology, and notably on the derivation of H_0 . In this paper we try to clarify the situation by studying how well each environmental indicator is able (or not) to trace a given environmental property.

We start in Sect. 2 by presenting that, mathematically, if two SN Ia populations were to exist, the observed amplitude of their true standardized magnitude difference linearly depends on the ability of a tracer to accurately measure which population a SN belongs to. We present in Sect. 3 the data sample we use for this work, and we describe the methodology used to extract the measurements of the environmental tracers. Then we apply in Sect. 4 our two-populations model on these data and we present our findings in Sect. 5, comparing them with results from the literature. For this we used the $H\alpha$ -based lsSFR tracer as reference, and we test this hypothesis in Sect. 6. We discuss our findings and we conclude in Sect. 7.

2. The two-populations model

As discussed in the introduction, many SN cosmological analyses use step functions to account for environmental dependencies in the derivation of distances. These steps are the difference of average properties, for example q , between two sides of a boundary t_{cut} in a considered environmental property t . When applied to the previously mentioned mass-step, it corresponds to the difference of absolute magnitude (q is the magnitude) of SNe Ia from low- and high-mass hosts (t is the host stellar mass), conventionally split at the host stellar mass of $t_{\text{cut}} = 10^{10} M_\odot$.

The underlying assumptions when using a step function are the following: (1) there are two categories of SNe, for example a and b , that differ on average in q and (2) the tracer t is able to probe these categories. Consequently, the amplitude of the observed step depends on the intrinsic SN properties q and the quality of the tracer t .

In the following subsections, we describe our statistical model starting from an illustrative mock example, first explaining the mathematical concept without measurement errors, and then including them in the model, to finally introduce the probability function. A detailed mathematical derivation is given Appendix A.

2.1. Concept of contamination

Let us assume that the two SN populations a and b have a normally distributed quantity q , for example $N_a = \mathcal{N}(\mu_a, \sigma_a)$ and $N_b = \mathcal{N}(\mu_b, \sigma_b)$, for which they differ on average by $\gamma_0 = \mu_a - \mu_b$; this difference corresponds to the true step amplitude.

We now assume that we have access to a tracer t that is able to discriminate between the a and b populations, but not with perfect accuracy (using the statistical binary classification terminology, that is with neither perfect specificity nor sensitivity). The tracer classification is based on the cutoff value t_{cut} such that the SNe are classified as a or b if they are either above or below the cut, respectively.

Even assuming this tracer provides error-free measurements, we expect misclassifications from the tracer inaccuracy: some SNe from the a category will be measured below the t_{cut} and will thus be incorrectly classified as b , and vice versa.

This is illustrated in Fig. 1 for three cases with varying accuracy: perfect, medium, and null. This figure also illustrates how the estimation of the underlying q distribution for each a and b category is affected by the tracer’s inaccuracy, and consequently how the derived steps are potentially underestimated. As the tracer accuracy degrades and misclassification cases increase,

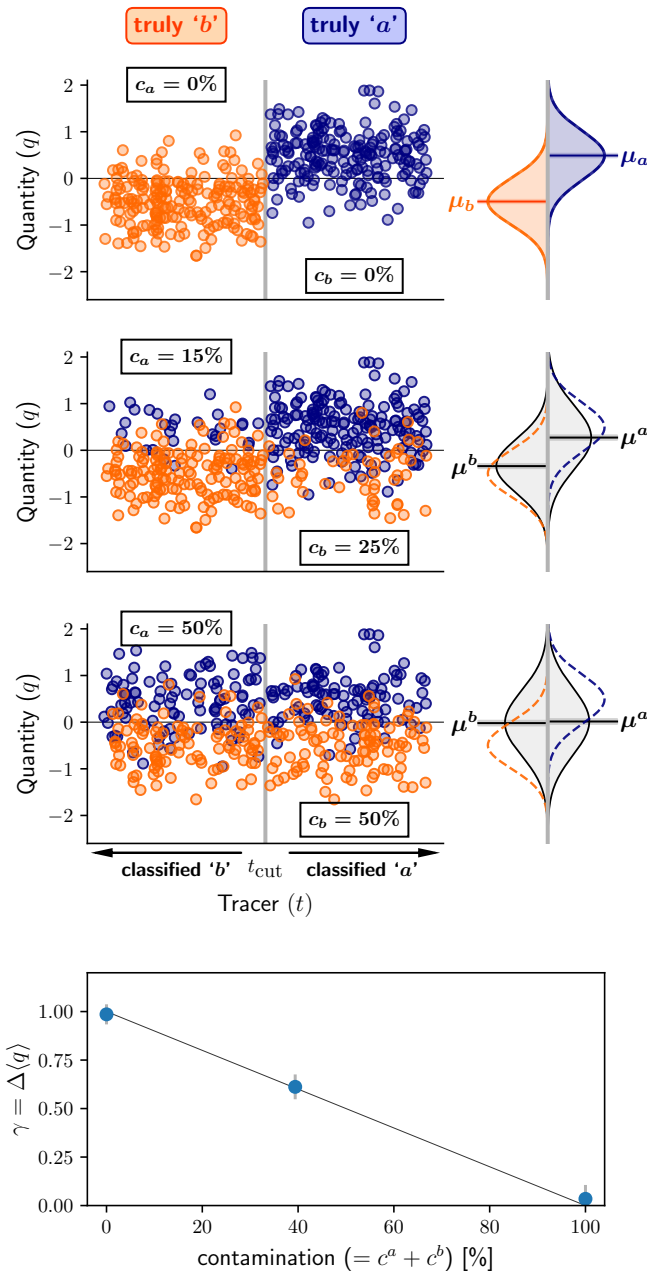


Fig. 1. Concept of tracer purity and its impact on the step measurement of an arbitrary quantity q . *Top three panels:* 400 mock data, 200 of category a (in blue), normally distributed over the quantity q by $N(\mu_a = 0.5, \sigma_a = 0.5)$, and 200 of category b (in orange) with a similar distribution $N(\mu_b = -0.5, \sigma_b = 0.5)$. The figure shows the mock q values as a function of the tracer values t for three different tracers. The data are classified as a or b given the tracer values if they are measured above or below t_{cut} , respectively. The *top, middle, and bottom panels* show tracers with perfect, medium, and null ability to track the two categories. The *right panels* show the tracer estimation of the q distribution for each tracer (in gray), to be compared to the true q distribution for each category (colored dashed lines). *Bottom panel:* evolution of the observed step $\gamma = \mu^a - \mu^b$, the difference of mean q values for targets classified as a or b by the tracers, as a function of the tracer contamination $c = c^a + c^b$; the circles show each of the three simulations illustrated in the *top panels*, while the line shows the prediction from Eq. (3).

the number of category a SNe incorrectly classified by the tracer as b increases (blue circles in the left part of the figure); similarly, the fraction of category b SNe misclassified as a also increases

(orange circles in the right part of the figure). As a consequence of misclassifications, the measured distributions of q for each of the inferred populations a and b broaden and their means converge, so that the step measured using an inaccurate tracer is systematically smaller than the true step.

2.2. Notations and definitions

For clarity, we set here our definitions and nomenclature in the large-number limit. N is the number of targets, the subscript x denotes true conditions ($x = \{a; b\}$) and the superscript x denotes actual classifications by a tracer. Hence, N_a^b is the number of targets that are truly a but classified as b . Accordingly, $N^b = N_a^b + N_b^b$ is the number of targets classified as b by a tracer, and $N_b = N_b^b + N_a^b$ is the number of targets that are intrinsically b .

We call contamination the fraction of targets for which the tracer classification differs from the truth. It could either be defined as the fraction $c_a \equiv N_a^b/N_a$ of truly a targets classified as b (resp. $c_b \equiv N_b^a/N_b$) or as the fraction $c^a \equiv N_a^b/N^a$ of classified a targets that actually are b (resp. $c^b \equiv N_b^a/N^b$). The two definitions are related as $c^a = N_b/N^a c_b$ and $c^b = N_a/N^b c_a$. We also note that $1 - c_x \equiv N_x^x/N_x$ and $1 - c^x \equiv N_x^x/N^x$, where x is either a or b ¹.

The probability p^a of a target tracer t_i to be measured above the tracer threshold t_{cut} , thus classified as a , is the sum of (1) the probability that a target truly is a and properly identified as a , and (2) the probability that it truly is b but misclassified as a :

$$\begin{aligned} p^a &\equiv \frac{N^a}{N} = \frac{N_a^a}{N} + \frac{N_b^a}{N} = (1 - c_a) \times \frac{N_a}{N} + c_b \times \frac{N_b}{N} \\ &= (1 - c_a) \times p_a + c_b \times (1 - p_a) \end{aligned} \quad (1)$$

where $p_b \equiv N_b/N = 1 - N_a/N \equiv 1 - p_a$. Similarly:

$$p^b = c_a \times p_a + (1 - c_b) \times (1 - p_a). \quad (2)$$

While μ_a and μ_b are the true means of the SN distributions of categories a and b , respectively (see Sect. 2.1), μ^a and μ^b are the distribution means of the considered distributions for each group classified by the tracer (filled gray distributions in Fig. 1). Accordingly, the observed amplitude step $\gamma = \mu^a - \mu^b$, as measured by a tracer, is related to the true intrinsic step $\gamma_0 = \mu_a - \mu_b$ by

$$\begin{aligned} \gamma &= ((1 - c^a)\mu_a + c^a\mu_b) - (c^b\mu_a + (1 - c^b)\mu_b) \\ &= \gamma_0 \times [1 - (c^a + c^b)]. \end{aligned} \quad (3)$$

This prompts us to define the (total) contamination of a tracer as $c = c^a + c^b$. The linearly decreasing relation in the bottom panel of Fig. 1 illustrates this equation.

2.3. Reference, comparison tracers, and measurement errors

It is unlikely to have access to the true population classification; instead it is necessary to rely on a reference tracer with respect to which the other tracers can be compared. This reference tracer

¹ Following the standard binary classification terminology, if a is the positive condition and b is the negative condition, then N_a are the real positive cases (P), N_b the real negative ones (N), N_a^a are the true positives (TP) and N_b^b are the true negatives (TN). Thus, N_b^a are the false positives (FP) and N_a^b the false negatives (FN). Finally c_a is the false negative rate (FNR) and c_b is the false positive rate (FPR); c^a is the false discovery rate (FDR) and c^b is the false omission rate (FOR).

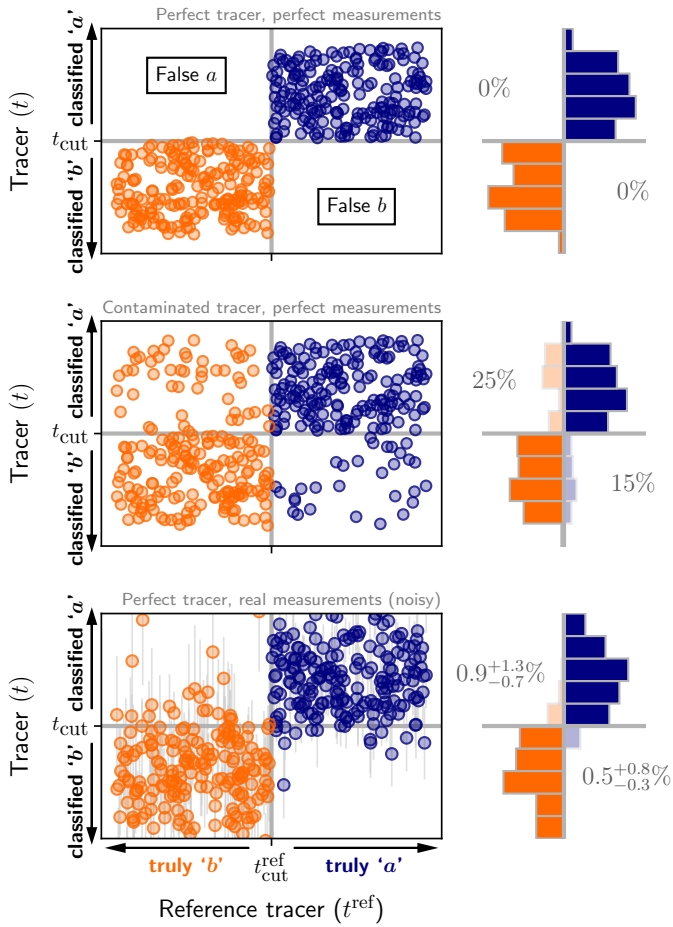


Fig. 2. Correlation between uniform mock reference and comparison tracers. The color indicates the true classification following the color-coding in Fig. 1: blue for a , orange for b . The right panel of each row shows the marginalized distribution of each quadrant, following the color-coding of the circles. The light histograms correspond to the off-diagonal terms: false a in light orange, false b in light blue. *Top panels:* perfect and noise-free tracer, i.e., without any off-diagonal points; *middle panels:* imperfect ($c_a = 15\%$ and $c_b = 25\%$) yet noise-free tracer, the sum of its off-diagonal terms corresponds to its contamination parameters; *bottom panels:* perfect but noisy tracer. In this case, plain off-diagonal fractions are 11% in false a and 7.5% in false b , while both should be 0% for a perfect tracer; the likelihood estimator (Eq. (11), see Sect. 2.4), properly accounting for the fraction of off-diagonal terms caused by measurement errors, provides contamination estimates $c_a = 0.5^{+0.8}_{-0.3}\%$ and $c_b = 0.9^{+1.3}_{-0.7}\%$, compatible with zero.

is itself an observable associated with its own contamination parameters c_a^{ref} and c_b^{ref} ; however, in this analysis, we will generally consider a perfectly accurate reference tracer: $c_a^{\text{ref}} = c_b^{\text{ref}} = 0$. The correlation of any other tracer with respect to this reference tracer enables the contamination of this comparison tracer to be derived.

This is illustrated in Fig. 2. In the case of error-free measurements (top two panels), the contamination will simply be the fractions of off-diagonal terms of the correlation plot between the reference tracer and the comparison tracer. However, measurement uncertainties complicate the picture as they randomly scatter points into the off-diagonal parts of the plot, even in the case of a perfect tracer, as illustrated in the bottom panel of the figure.

We denote f_i as the probability that a given tracer measurement $t_i \pm \delta t_i$ is below the cutoff value t_{cut} (and therefore the target is classified as b). Then f_i is expressed (for normally distributed errors) as

$$f_i = \mathcal{P}(t_i < t_{\text{cut}}) = \int_{-\infty}^{t_{\text{cut}}} \mathcal{P}(\hat{t}_i | t_i) d\hat{t}_i = \int_{-\infty}^{t_{\text{cut}}} \mathcal{N}(\hat{t}_i; t_i, \delta t_i) d\hat{t}_i \quad (4)$$

where \hat{t}_i correspond to the true value of t_i .

By definition, we then have $N^b \equiv \sum f_i$ and $N^a \equiv \sum (1 - f_i)$; similarly, the number of off-diagonal elements are $N_b^a = \sum_{i \in a} f_i$ and $N_a^b = \sum_{i \in b} (1 - f_i)$. Therefore, even in the context of a perfect tracer (for which we expect $c_a = c_b = c^a = c^b = 0$), the previous simple contamination estimates of $c^a = N_b^a/N^a$ or $c^b = N_a^b/N^b$ can appear to be non-zero due to measurement uncertainties, and can degrade the intrinsic tracer contamination estimates.

Instead, c^a and c^b should be estimated as the fractions of off-diagonal terms that are not caused by measurement errors. This will be done by defining the probability function, and comparing it to observations.

2.4. Building the probability function to estimate c^a and c^b

In order to get the contamination of a tracer ($c = c^a + c^b$; see Sect. 2.2), we first express the probability function of the intrinsic parameters c_a and c_b .

Following the derivations presented in Appendix A, we can express the probability of measuring t_i when the target i belongs to population b as

$$\mathcal{P}(i \in b, t_i | c_b) \propto (1 - p_a) \left((1 - c_b) \times f_i + c_b \times (1 - f_i) \right). \quad (5)$$

This can be understood as the probability that a target is b (i.e., $(1 - p_a)$) times the chance that a tracer is measured below a cut (f_i), and thus classifying the target as b , while accounting for the fraction of false negatives ($1 - c_b$), plus the chance that the tracer is measured above the cut ($1 - f_i$) times the fraction of false positives (c_b).

Similarly for the probability of measuring t_i when the target i belongs to population a ,

$$\mathcal{P}(i \in a, t_i | c_a) \propto p_a \left(c_a \times f_i + (1 - c_a) \times (1 - f_i) \right), \quad (6)$$

and therefore

$$\begin{aligned} \mathcal{P}(t_i | c_a, c_b) &= \mathcal{P}(i \in a, t_i | c_a) + \mathcal{P}(i \in b, t_i | c_b) \\ &\propto p_a \times \left((1 - c_a) \times (1 - f_i) + c_a \times f_i \right) \\ &\quad + (1 - p_a) \times \left(c_b \times (1 - f_i) + (1 - c_b) \times f_i \right). \end{aligned} \quad (7)$$

As mentioned in Sect. 2.3, in reality, we do not know if an individual target belongs to class a or b : the contamination cannot be directly tied to the truth, but only to another tracer used as a reference, plagued by its own contamination and measurement errors. Using $^{\text{ref}}$ to denote the parameters for the reference tracer, p_a and $p_b = 1 - p_a$, which cannot be estimated directly anymore, can be derived from the reference tracer as

$$\begin{aligned} p_a &\equiv \mathcal{P}(i \in a, t_i^{\text{ref}} | c_a^{\text{ref}}) \\ &= p_a^{\text{ref}} \times \left((1 - c_a^{\text{ref}}) \times (1 - f_i^{\text{ref}}) + c_a^{\text{ref}} \times f_i^{\text{ref}} \right) \end{aligned} \quad (8)$$

and

$$\begin{aligned} p_b &\equiv \mathcal{P}(i \in b, t_i^{\text{ref}} | c_b^{\text{ref}}) \\ &= (1 - p_a^{\text{ref}}) \times \left(c_b^{\text{ref}} \times (1 - f_i^{\text{ref}}) + (1 - c_b^{\text{ref}}) \times f_i^{\text{ref}} \right). \end{aligned} \quad (9)$$

where p_a^{ref} , the fraction of truly a targets, have to be assumed a priori since the reference tracer is noisy (see Sect. 4).

Assuming the reference tracer to be perfect (i.e., $c_a^{\text{ref}} = c_b^{\text{ref}} = 0$), Eq. (7) becomes

$$\mathcal{P}(t_i, t_i^{\text{ref}} | c_a, c_b) \propto p_a^{\text{ref}} \times (1 - f_i^{\text{ref}}) \times \left((1 - c_a) \times (1 - f_i) + c_a \times f_i \right) + (1 - p_a^{\text{ref}}) \times f_i^{\text{ref}} \times \left(c_b \times (1 - f_i) + (1 - c_b) \times f_i \right). \quad (10)$$

Finally, the estimation of the tracer's parameters c_a and c_b with respect to the reference tracer is made by minimizing

$$\mathcal{L} = -2 \sum_i \ln \mathcal{P}(t_i, t_i^{\text{ref}} | c_a, c_b), \quad (11)$$

and the total contamination $c = c^a + c^b$ for a comparison tracer can be computed from

$$c^a = \frac{N_b}{N^a} c_b = \frac{(1 - p_a^{\text{ref}}) N}{\sum (1 - f_i)} c_b$$

$$c^b = \frac{N_a}{N^b} c_a = \frac{p_a^{\text{ref}} N}{\sum f_i} c_a. \quad (12)$$

We tested and validated our model and our code using simulations. We generated mock datasets of various contaminations and sizes, which were fitted with our implementation of the likelihood described in this section. The results confirm that our implementation of the algorithm is correct.

3. Data

We work with the Nearby Supernova Factory (SNfactory, Aldering et al. 2002) SNe Ia dataset published in Rigault et al. (2020) (see also Aldering et al. 2020). This dataset has two advantages for this analysis: it is at low redshifts ($0.02 < z < 0.08$), so the local environment is measurable, and it contains spectrophotometric integral field units (IFU) environmental data, necessary to accurately estimate the local specific star formation rate (lsSFR). We use the publicly available catalog and images from SDSS and PanStarrs for the photometric measurements or their derived quantities.

This section briefly summarizes the methodology used to extract the different tracers considered in this analysis, which follows those developed in the literature, namely: the spectroscopically derived lsSFR (Sect. 3.1), photometrically derived lsSFR (Sect. 3.6), local colors (Sect. 3.4), local and global host stellar masses (Sect. 3.5), and the global host morphologies (Sect. 3.7).

3.1. Spectroscopic lsSFR

The spectroscopically derived lsSFR is detailed in Sect. 3 of Rigault et al. (2020). We use their measurements, which are generated in two stages: first, the star formation rate (SFR) is derived from the $H\alpha$ emission line luminosity (Calzetti 2013), spectroscopically measured within the local 1 kpc aperture radius, after subtraction of the stellar continuum background. The second step is the measurement of the local stellar mass, as described in Sect. 3.5. For both quantities, a full posterior distribution is derived such that their ratio sets the posterior distribution of the lsSFR measurements. Hereafter, we will refer to this tracer as the spectroscopic lsSFR.

3.2. Photometric measurements

We used flux-calibrated optical images from SDSS (DR12, Alam et al. 2015) to derive the photometric environmental tracers.

We measured *ugriz* SDSS local fluxes and their uncertainties in projected circular apertures centered on the SN location, using the SUM_CIRCLE method of SEP² (Barbary 2016). To compare our results with those of literature studies, we used aperture radii of $X = 1, 1.5,$ and 3 kpc. Counts were converted to flux assuming a zero point of 22.5 mag for the *gri* bands, and 22.46 and 22.52 for the *u* and *z* bands, respectively.

To test the accuracy of the sky background subtraction, we drew 500 random source-free apertures around each target. Presumably, the histogram of the error-normalized background levels should be a standard $\mathcal{N}(0, 1)$ pull distribution. However, we regularly observed that the pull mean was slightly too high, corresponding to an inaccurate background correction, and that the pull dispersion was larger than unity, meaning that the error on the background level had been under-estimated. We thus further corrected each aperture photometric measurement by the median of the 500 random sky apertures, and scaled the quoted error by the normalized median absolute deviation of the sky levels.

To derive global tracers, we first associated the SN with its host employing the directional light radius method (Sullivan et al. 2006; Gupta et al. 2016); ellipses used to determine the directional light radius were obtained by the SUM_ELLIPSE method of SEP. Then we used the *ugriz* global galaxy model magnitudes and fluxes from the corresponding SDSS catalog entries (see details in Rigault et al. 2020).

Both local and global photometric measurements were then corrected for Milky Way dust absorption using the EXTINCTION³ library assuming a Fitzpatrick (1999) extinction curve with $R_V = 3.1$ and the dust extinction map from Schlegel et al. (1998).

3.3. SED fitting and k -correction

From each *ugriz* photometric dataset, we used LEPHARE⁴ (Arnouts et al. 1999; Ilbert et al. 2006; Arnouts & Ilbert 2011) to fit for the associated spectral energy distribution (SED) using templates from Bruzual & Charlot (2003) (hereafter BC03) as did Jones et al. (2018), among others. For the reproducibility, our configuration file is available online⁵. It contains the following assumptions:

Dust: We used 41 bins of $E(B - V)$ extinction values ranging from 0 to 1 (in steps of 0.01 from 0 to 0.2, of 0.03 from 0.2 to 0.5, and then of 0.05 up to 1), and we used the Fitzpatrick (1999) extinction curve (extracted from the hyper- z program with $R_V = 3.1$; Bolzonella et al. 2000);

Redshift range: We used a redshift range from 0 to 0.1 with a bin size of $\Delta z = 0.002$;

Emission lines: We included this contribution (Kennicutt 1998);

Cosmology: We used $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$, $\Omega_m = 0.3$, and $\Omega_\Lambda = 0.7$.

The SED fit was made at the fixed (known) redshift of the host, the stellar mass M_* was bounded between 10^6 and 10^{13} solar masses, and the *r*-band absolute magnitude between -10 and -26 . We included the SDSS suggested error

² github.com/kbarbary/sep v1.10.

³ extinction.readthedocs.io

⁴ v2.2 see [LePhare website](https://lephare.org/)

⁵ github.com/MartinBriday/pylephare

floor (0.05, 0.02, 0.02, 0.02, 0.03 for the *ugriz* bands (see, e.g., [kcorrect.org](#) and [Childress et al. 2013a](#)).

We estimated the posterior distribution of each SED fitted parameter and spectrum using Monte Carlo simulations. For each *ugriz* flux measurement (i.e., for each local radius of each SN) we randomly drew 500 realizations assuming that the bands are independent and that flux errors are normally distributed. We ran the SED fitting procedure for these 500 realizations and the best fitting parameters (`_BEST`) and spectral distributions were used to fix the respective posteriors. We used the median rest-frame magnitudes measured on each of the 500 realizations to estimate the *ugriz* *k*-corrected magnitudes used in this analysis; the 16% and 84% percentiles indicate the corresponding errors.

3.4. Colors

The colors were estimated from the *k*-corrected magnitudes (see Sect. 3.3): the *u* – *r* color is the difference of the median of the *k*-corrected *u*- and *r*-band magnitudes, and the color error is the quadratic sum of the individual standard deviations.

3.5. Stellar masses

Local and global masses were derived using the procedure described in Sect. 3.3 of [Rigault et al. \(2020\)](#) (see also [Jones et al. 2018](#)). In brief, we use the relation from [Taylor et al. \(2011\)](#) to convert *g* and *i* *k*-corrected magnitudes into stellar masses. This relation has a 0.1 dex intrinsic dispersion that is added in quadrature to the stellar mass uncertainties derived from photometric uncertainties alone; this scatter dominates the error budget, especially for global measurements (see [Smith et al. 2020](#) for a discussion about the consistency of stellar mass estimators in the context of SN host analyses).

3.6. Photometric lsSFR

[Jones et al. \(2018\)](#) use photometry-based sSFR estimations to assess the lsSFR parameter in place of the $H\alpha$ -based measurements as they do not have local spectroscopy. They employ LEPHARE in a similar fashion to that described in Sect. 3.3 and estimate their sSFR posterior (and its errors) from the 50% ([16%, 84%]) of the individual sSFR values from the Monte Carlo realizations. They use the SDSS *u*-band plus *grizy* from PanSTARRS DR1 (PS1, [Chambers et al. 2016](#)) to do so. To be consistent, when deriving the sSFR, we also used these data, applying the same photometric measurement on PS1 data as we did for SDSS (see Sect. 3.2). Calibrated PS1 images were downloaded from the cutout service⁶. Again for consistency, we also used the [Jones et al. \(2018\)](#)'s LEPHARE configuration file when measuring the sSFR this way (D. Jones, priv. comm.). The measurements were finally normalized by the surface area. Hereafter we refer to this tracer as the photometric lsSFR, in contrast to the spectroscopic lsSFR.

3.7. Morphology

The inverse concentration index (i.c.i., [Shimasaku et al. 2001](#); [Strateva et al. 2001](#), see also the [SDSS web site](#)) is a commonly used morphological tracer. It is the ratio of radii containing 50% and 90% of the Petrosian flux in *r*-band. With this tracer,

⁶ PanSTARRS website.

early-type galaxies typically have an *i.c.i.* ~ 0.3 , while late-type galaxies have an *i.c.i.* closer to 0.45. Following [Kauffmann et al. \(2003\)](#), we used $i.c.i._{cut} = 1/2.7 = 0.37$ to distinguish between late- and early-type galaxies (also see, e.g., [Choi et al. 2010](#)). Using slightly lower boundaries, such as 0.35 suggested by [Banerji et al. \(2010\)](#), among others, has marginal influence on our results.

Galaxy classification based on the Petrosian flux is one approach, but it is not the only one. Among a few others, the *i.c.i.* is the most convenient for this analysis as it discriminates between two galaxy morphology populations (early and late types) separated by a cutoff value, which motivates our choice.

4. Contamination with respect to the spectroscopic lsSFR

Because the sSFR is usually used as a reference age tracer when available (see, e.g., [Yoshikawa et al. 2010](#); [Labbé et al. 2013](#); [Casado et al. 2015](#); [Karman et al. 2017](#)), we used the spectroscopic lsSFR measurement from [Rigault et al. \(2020\)](#) as a reference tracer, assuming that this quantity is a perfectly accurate (yet imprecise) progenitor age tracer: $c_{spec-lsSFR}^a = c_{spec-lsSFR}^b = 0$ (see Sect. 2). We test this hypothesis in Sect. 6.2.

Following most of the preceding host environmental studies, which split their samples at the median value, we also assume throughout the paper that $p_a^{ref} = 50\%$. We have found by simulations that the derivation of the c_a and c_b tracer parameters are unaffected by this choice (bias lower than the 1σ error), as long as the actual true parameter $25\% \lesssim p_a^{ref} \lesssim 75\%$, which is to be expected simply based on rate analyses (e.g., [Mannucci et al. 2006](#); [Rodney et al. 2014](#); [Wiseman et al. 2021](#)); the bias becomes significant (at more than 3σ) if $p_a < 10\%$ or $p_a > 90\%$.

4.1. Comparing the environmental tracers with the spectroscopic lsSFR.

Following [Rigault et al. \(2020\)](#), we classify as population *a* every SNe Ia with $\log(spec - lsSFR) > -10.82$ dex as this value corresponds to the median of this tracer for the SNfactory sample. Given the measurement errors, each SN Ia therefore has a probability $p(\text{young})$ of being observed in population *a* (referring to $1 - f_i^{ref}$ in Sect. 2.4). As highlighted in Sect. 2.3, the sum of the off-diagonal terms are fully captured by the measurement errors since we assumed $c_{spec-lsSFR}^a = c_{spec-lsSFR}^b = 0$.

Each tracer has its own cutoff value to classify a given SN Ia in population *a*. We apply the assumption typically used in the literature when available and the median value otherwise, which is usually also the literature assumption. The well-studied mass step usually has the threshold boundary set at $\log(M_*/M_\odot)_{cut} = 10$ dex (e.g., [Kelly et al. 2010](#); [Sullivan et al. 2010](#); [Betoule et al. 2014](#); [Scolnic et al. 2018](#)) and we use this value in this analysis. We note that since we have $\sim 60\%$ of SNe Ia with a host-mass greater than $10^{10} M_\odot$ (in agreement with, e.g., [Roman et al. 2018](#)), and since N_a/N is given by the reference tracer (here $p_a^{ref} = 50\%$), this means that $c_{gmass}^a \neq c_{gmass}^b$. Concerning the local mass, [Jones et al. \(2018\)](#) and [Rigault et al. \(2020\)](#) use the median to divide their respective samples. For the sample here we find the median to be $\log(M_*/M_\odot)_{cut} = 8.37$ dex; to compare with [Jones et al. \(2018\)](#), we used a 1.5 kpc ‘‘local’’ aperture for the local mass. Following [Roman et al. \(2018\)](#), [Jones et al. \(2018\)](#), and [Kelsey et al. \(2021\)](#), we did the same for the local

(3 kpc) color and the photometric local (1.5 kpc) sSFR finding medians at $u - r = 1.74$ mag and $\log(\text{phot} - \text{lsSFR}) = -10.32$ dex, respectively. Finally, as explained in Sect. 3.7, we used *i.c.i.* = 0.37 to divide the morphologies of our SN Ia host galaxies into category *a* or *b* when respectively above or below this value.

In Fig. 3 we show the correlation between the spectroscopic lsSFR and the other environmental tracers. We first note that all tracers correlate relatively well with the spectroscopic lsSFR. Based on the Spearman rank coefficient, the most correlated tracer is the local $u - r$ color ($|\rho| = 0.71$) closely followed by the global mass ($|\rho| = 0.64$); the host galaxy morphology and the local stellar mass show weaker correlations, with $|\rho| = 0.45$ and 0.32, respectively.

The population of off-diagonal terms appear to be consistent with these Spearman-ranked correlations: the higher the fraction of off-diagonal terms, the lower the Spearman coefficient value. However, as detailed in Sect. 2.3, only the fraction of off-diagonal terms not caused by measurement errors has to be accounted for to measure the accuracy of an indicator to trace the reference, here the spectroscopic lsSFR.

To do that, we fit the parameters c_a and c_b of the tracers, assuming that the spectroscopic lsSFR measurements are (noisy) perfectly accurate indicators of *a* and *b*, by minimizing Eq. (11). Since the measurements are noisy, we need to set the fraction of truly *a*, that we assume to be 50% (i.e., we fix $N_a = N_b$). The fit is made using Markov chain Monte Carlo routines using the EMCEE (Foreman-Mackey et al. 2013) package to sample the full posterior distributions of the parameters c_a and c_b of the tracers. Each tracer is fitted independently. The resulting median parameters c^a and c^b are displayed in Fig. 3 with their 1σ scale (16%/84%), and the aggregated tracer contamination $c = c^a + c^b$ is summarized in Table 1.

4.2. Measuring the environmental magnitude step

To derive the SNe Ia magnitude steps, we follow the procedure detailed in Sect. 4.2.2 of Rigault et al. (2020). Given $p_i^f = 1 - f_i^f$ the probability that the tracer measurement is above the tracer's cutoff value (see Sect. 2.3), thus classified as *a* by this tracer, we fit the magnitude offset γ between the two populations (i.e., the step) together with the stretch and color standardization coefficients α and β . This is done by χ^2 minimization between $\mu_{\Lambda\text{CDM}}$ and the standardized SN Ia distance modulus:

$$\mu = m - M + \alpha x_1 - \beta c + \gamma p^f. \quad (13)$$

When doing the fit, we fix the cosmology (Planck Collaboration XIII 2016), and the covariances between m , x_1 , and c are taken into account; p^f is an independent measurement. When fitting each environmental tracer independently to derive its associated step γ , the α and β coefficients are free to vary and might therefore differ between tracers. However, if all tracers are probing the same underlying effect, α and β should be the same, since the stretch and color standardization should not depend on the accuracy with which we are able to probe this underlying effect. When fitting α , β , and γ simultaneously, because stretch (and color) are connected to the host properties, the recovered value of α and β will be unbiased if the true underlying tracer is used to determine γ (Dixon 2021). We further investigate this issue in Sect. 6.1, where the value of α and β will be fixed to those derived together with the step of the reference tracer.

5. Results

The derived tracer contaminations with respect to the spectroscopic lsSFR and their associated magnitude steps γ , using the SNfactory dataset, are summarized in Table 1 and shown in Fig. 4.

The step amplitudes as a function of the tracer contaminations follow the expected trend given by Eq. (3), shown by the straight black line in the figure, remarkably well. As detailed in Sect. 2, this diagonal, which goes from γ_0 at $c = 0\%$ (the reference tracer step value) to 0 at $c = 100\%$ (the black diamond in the figure), is expected if two SN Ia populations exist with different mean magnitudes, and if we are using tracers that are not perfectly able to discriminate between the two populations ($c > 0$) to measure their magnitude offset. We note that $c = 100\%$ corresponds to randomly distributed SNe Ia between the two classes, and the expected magnitude difference between the two resulting groups is thus 0 by definition (as seen in Fig. 1).

We added to Fig. 4 the recent results from the literature with the step measurements that were made using the same techniques as the ones we used. Specifically, we added the lsSFR from Rigault et al. (2020); the local (3 kpc) $U - V$ (similar to $u - r$) and global host stellar mass (split at $10^{10} M_\odot$) from Roman et al. (2018); the local (1.5 kpc) stellar mass and photometric sSFR from Jones et al. (2018); and the morphology from Pruzhinskaya et al. (2020). We used the global host mass-step from Roman et al. (2018) for it is derived using the state-of-the-art Betoule et al. (2014) + SNLS five-year sample and the Malmquist bias correction were not made using the 5D implementation of the Beams with Bias Correction (BBC; Scolnic & Kessler 2016; Kessler & Scolnic 2017). Smith et al. (2020) showed that this implementation can bias the reported step if intrinsic SN–host correlations are not accounted for. We also used the global host mass step from Smith et al. (2020). Finally, we plotted in this figure the local (4 kpc) $U - R$ and stellar mass from Kelsey et al. (2021) with a transparent marker, as γ is fitted after the standardization in that paper (while we fit it as a third standardization parameter, see Sect. 4.2). For these literature data points we use the reported steps while using our derived tracer contaminations. We note that if the steps are considered to be SN sample dependent (e.g., due to the light curve extraction pipeline), this implies that the contaminations are purely galaxy properties that are unrelated to the SNe Ia.

Figure 4 shows that step amplitudes measured using the SNfactory data and measured using any of the literature environmental tracers are in remarkable agreement with the corresponding independent literature measurements. For instance the Jones et al. (2018) local mass step is 0.067 ± 0.017 mag, while we measure 0.053 ± 0.031 mag and the local $U - V$ color step from Roman et al. (2018) is 0.091 ± 0.013 mag and we find 0.096 ± 0.035 mag. The SNfactory SNe Ia data thus seem to be representative of that of the literature.

Under the assumption of multiple populations, implicitly implied by the environmental step functional form, the fact that the relationship between observed environmental step and tracer contamination is compatible with our two-populations model, for both SNfactory and the literature data points, suggests that two populations are enough to explain the observations and that both populations differ in standardized brightness by $\gamma_0 \sim 0.13$ mag. Fitting for γ_0 using literature data points, as shown in Fig. 4, we find 0.121 ± 0.010 mag. This claim assumes that spectroscopic lsSFR is a perfect tracer. We study the use of the other tracers as the reference tracer in Sect. 6.2.

In reality, no tracer is perfect and if the spectroscopic lsSFR contamination were to be a small percentage, then γ_0

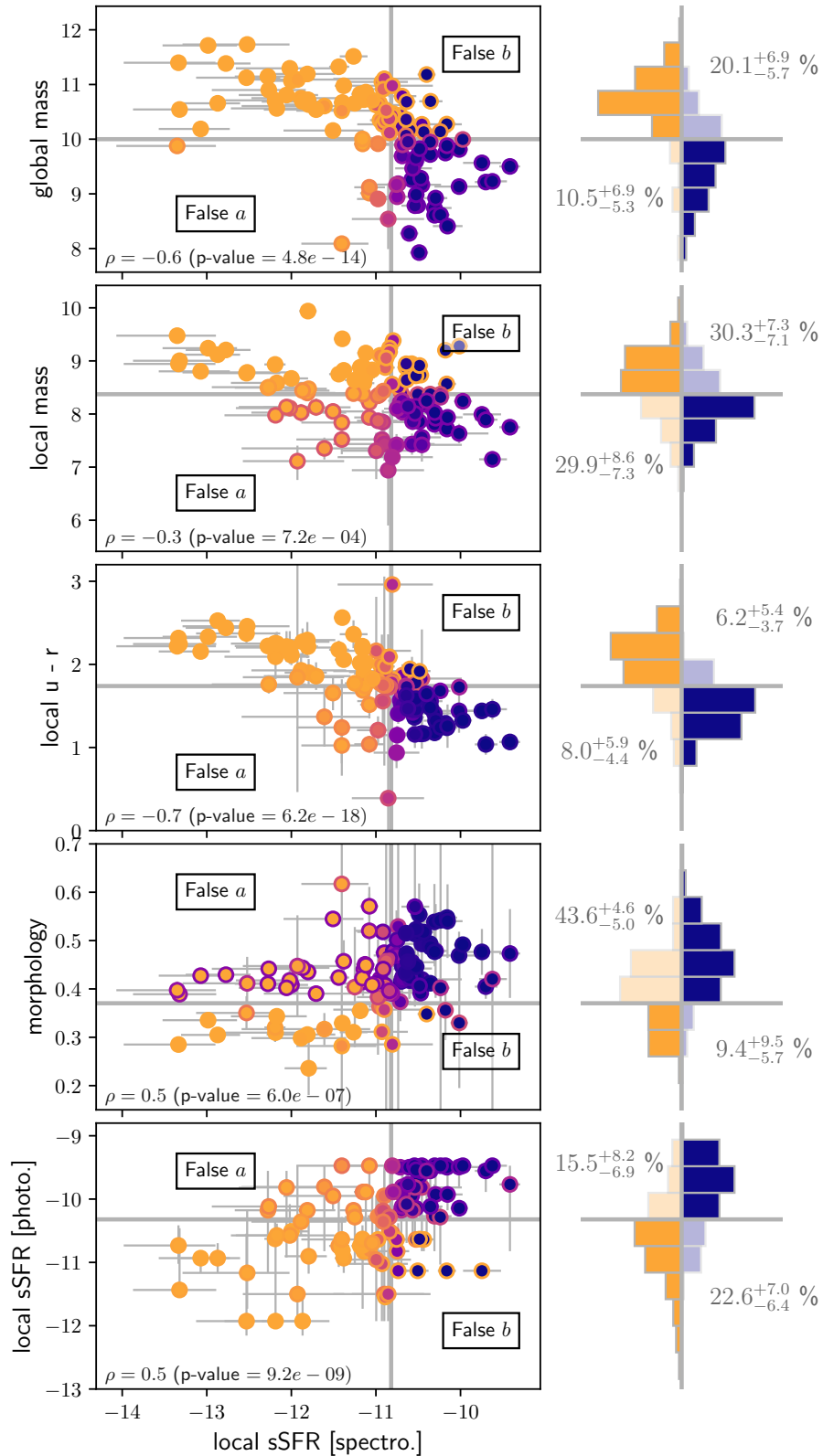


Fig. 3. Correlation between each environmental tracer and the spectroscopic lsSFR, as mentioned in Sect. 4. The vertical gray line is the spectroscopic lsSFR cutoff, set at $\log(\text{lsSFR}) = -10.82$ dex. The horizontal gray lines are the other tracer cutoffs, as defined in Sect. 4. Each figure indicates the false a and false b classification quadrants. The color of the circles, which vary from orange to blue, represents the probability for a SN Ia to be in population a (blue) or b (orange) from the spectroscopic lsSFR point of view, whereas the edge color represents the same probability, but from the comparison tracer point of view. In the right column, the histograms plot the distributions in each quadrant. Orange (resp. blue) bars correspond to the truly b (resp. a) spectroscopic lsSFR classification. The estimated c^a and c^b parameters are given in percentages within the corresponding quadrants where the histogram is transparent.

Table 1. Comparison of the standardization coefficients α , β , and γ (the stretch and color coefficients and the magnitude step value) and the tracer intrinsic contamination ($c = c^a + c^b$) with respect to the spectroscopic lsSFR used as a reference tracer.

Tracer	t_{cut}	$c^a + c^b$	α	β	γ	γ^*
local sSFR [spectro.]	-10.82	0.0	-0.156 ± 0.013	3.042 ± 0.128	0.127 ± 0.032	0.127 ± 0.032
local sSFR [photo.]	-10.32	$36.9^{+10.8}_{-9.3}$	-0.138 ± 0.012	3.059 ± 0.128	0.071 ± 0.032	0.078 ± 0.032
local $u - r$	1.74	$12.9^{+7.2}_{-5.7}$	-0.150 ± 0.013	3.115 ± 0.132	0.092 ± 0.035	0.095 ± 0.030
local mass	8.37	$60.2^{+11.2}_{-10.6}$	-0.137 ± 0.012	3.142 ± 0.138	0.053 ± 0.030	0.051 ± 0.028
global mass	10.00	$30.8^{+8.3}_{-7.7}$	-0.151 ± 0.013	3.121 ± 0.132	0.116 ± 0.030	0.116 ± 0.027
morphology	0.37	$53.3^{+8.9}_{-7.5}$	-0.138 ± 0.013	3.044 ± 0.139	0.030 ± 0.039	0.051 ± 0.036

Notes. γ^* is the magnitude step value when fixing α and β to those obtained by the reference tracer standardization (see Sect. 6.1); t_{cut} is the tracer cutoff.

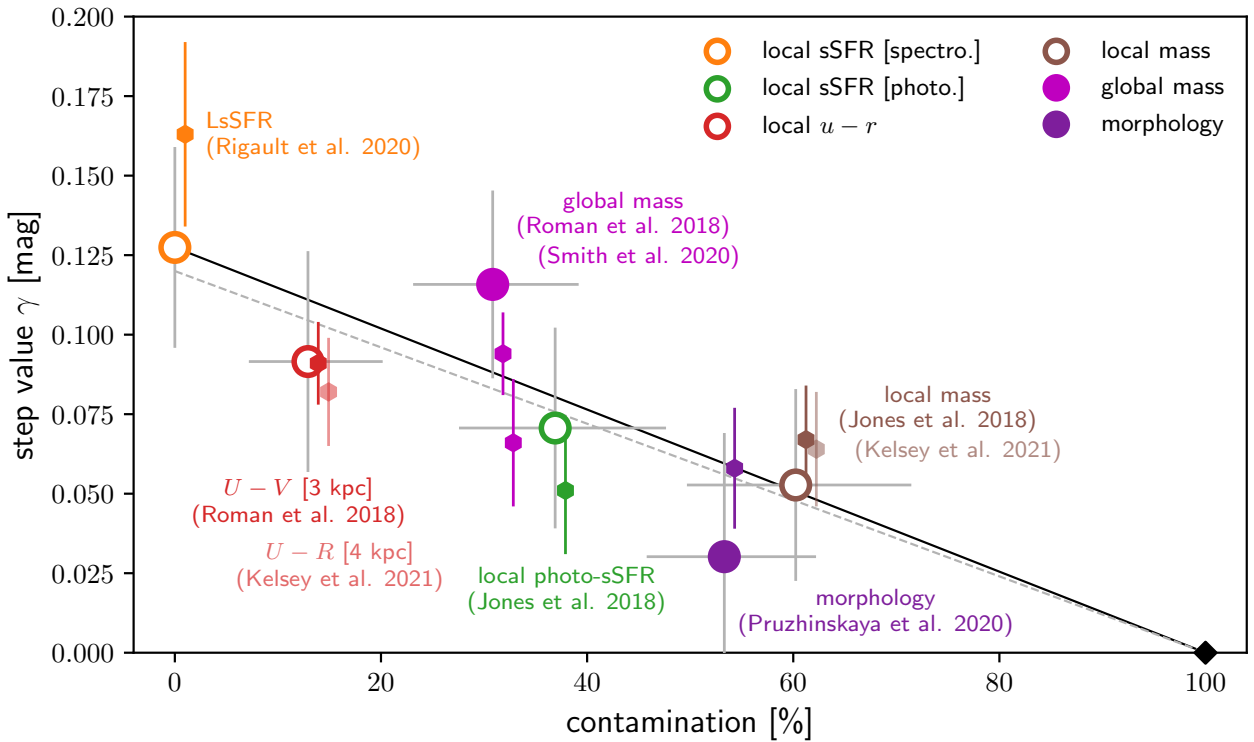


Fig. 4. SNe Ia magnitude steps, γ , as a function of the tracer contamination ($c = c^a + c^b$), using the spectroscopic lsSFR as reference ($c_{\text{spec-lsSFR}} \equiv 0$). The large circles are the steps derived using the SNfactory dataset (this work), while the small hexagons are the literature results (see Sect. 5). The large open (resp. full) circles are local (resp. global) measurements (see legend); the SNfactory-based and literature results have the same colors. The literature contaminations are those derived using the SNfactory sample and have been shifted by +1% for visibility. The Kelsey et al. (2021) related hexagonal markers are transparent as they measure the magnitude step after standardization, while all other data points (both with our SNfactory sample and the literature results) fit γ as a third standardization parameter (see Sect. 4.2). The full black diamond indicates the 0 mag step associated by definition with 100% contamination, corresponding to a random SN population classification. The straight black line shows our model of the measured step as a function of the contamination, linking the reference tracer point to the black diamond (see Sect. 2). The dashed gray line is a fit to the literature measurements, constrained to pass through the random classification value (black diamond).

would actually be higher. For instance if $c_{\text{spec-lsSFR}} = 10\%$ and $\gamma_{\text{spec-lsSFR}} = 0.13$ mag then $\gamma_0 = \gamma_{\text{spec-lsSFR}}(1 - c)^{-1} = 0.145$ mag. Consequently, the reported γ_0 measurements made in this analysis assuming we have a noisy but perfect tracer are, in fact, lower limits on the actual SN Ia population difference in magnitude means. The true spectroscopic lsSFR contamination is beyond the scope of this work.

6. Discussion

In this section we present variations to the main analysis, and then discuss the consequences of our findings. We first study the

impact of fixing the stretch and color standardization coefficients to that of the reference tracer. We then change which tracer is used as a reference and compare their ability to describe the data.

6.1. Fixing α and β

In Sect. 5, we fit the standardization coefficients α , β , and γ for each environmental tracer independently to find that the two-populations model detailed in Sect. 2 seems to explain the observed variations between the tracer γ parameters. Their apparent inconsistency is due to the ability of a tracer to accurately distinguish between the two underlying populations. In

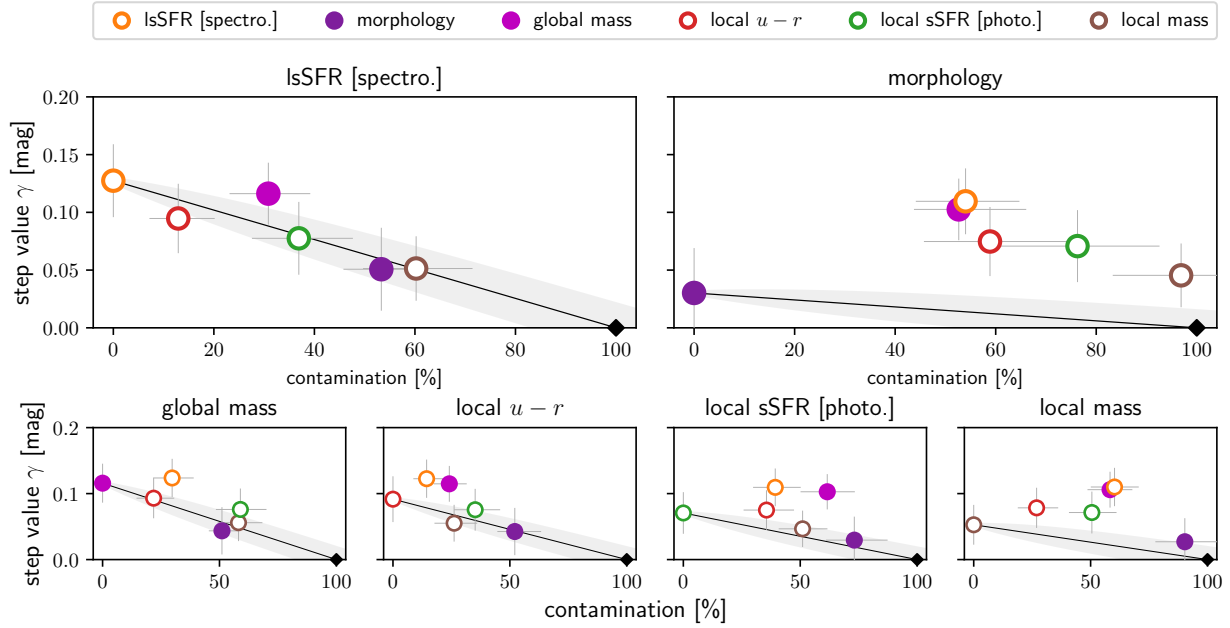


Fig. 5. Similar to Fig. 4, but changing the reference tracer choice by (from left to right, top to bottom) the spectroscopic lsSFR, the host galaxy morphology, the global host galaxy stellar mass, the local $u - r$ color, the photometric lsSFR, and the local stellar mass. In this figure the α and β have been fixed to that of the reference tracer used in each subplot (see Sect. 6.1). The gray band around the black line is the expected scatter along the diagonal (see Sect. 6.2). The more the data deviate from the model expectation (black line), the less likely it is for the reference tracer to be closely connected to the actual underlying astrophysical origin.

that context, because the standardization coefficients are correlated, especially α and γ (see, e.g., Fig. 7 of Rigault et al. 2020), if one is not able to accurately measure γ since its environmental tracer is inaccurate, one will in turn bias the derivation of the other standardization parameters, as the fitter will use them to counterbalance the γ error.

The natural solution in the context of the two-populations model, which is implicitly assumed when doing a step analysis, is that the value of α and β must be fixed to those derived when using the reference tracer when fitting Eq. (13) for the comparison tracers. The results of this alternative and more accurate analysis is given in Table 1 (column γ^*) and illustrated in the top left plot in Fig. 5. We see, comparing this plot with Fig. 4, that the results converge on the model's expectations.

6.2. Testing the reference tracer

In this section, we vary which tracer is used as a reference tracer and we re-derive the resulting contamination terms, c^a and c^b , as well as the steps γ^* assuming the reference's α and β are as detailed in Sect. 6.1. If a tracer is a good reference, that is, if it can accurately discriminate between the true underlying two populations, then the other tracers should follow the diagonal line given by Eq. (3), anchored at the value of γ_0 of the reference tracer. If a reference tracer is bad, the contamination associated with this tracer does not probe the ability of a comparison tracer to discriminate between the two underlying populations. In that case the points are not expected to follow the diagonal model.

This is what we qualitatively observe in Fig. 5. Morphology is a bad reference tracer, as the other tracers lie far from its expected diagonal. This means that the morphology is not able to accurately discriminate between the two underlying populations causing the environmental steps observed by the different tracers. Conversely, the spectroscopic lsSFR, the global mass, and the local $u - r$ colors seem to be better reference tracers.

To quantify this observation we first need to model how much scatter we should expect along the diagonal if we had access to a perfect tracer. This is mandatory since the step measurements are not independent; they all are made from the same sample of SNe Ia, but using different galaxy property indicators.

We use the simulation tool from Sect. 2 to simulate a sample with the same characteristics as the SNfactory sample, $N = 110$, $p_a = 0.5$, $\sigma_a = \sigma_b = 0.1$, and the γ_0 corresponding to the reference tracer step γ in Table 1. We then assume a c_a and a c_b , which define the four N_i^j with $i = \{a, b\}$ and $j = \{a, b\}$. We randomly shuffle the sample to follow these N_i^j and we measure γ . This last step is repeated 5000 times to determine the scatter on γ caused by the randomness of which target belongs to the off-diagonal terms or not. If the two-populations model is correct, and if we measure the tracer γ parameters with a single dataset, then this scatter corresponds to the expected variations given tracer c_a and c_b parameters. The amplitude of this scatter as a function of the tracer contaminations is shown as a gray band along the model's diagonal in Fig. 5.

Once we have determined the scatter $\sigma(c)$ expected given the amount of contamination c , we can measure the χ^2 associated with the ability of each reference tracer to explain the data, such that

$$\chi_{\text{ref}}^2 = \sum_t \left(\frac{\gamma_t^* - \gamma(c_t)}{\sigma(c_t)} \right)^2, \quad (14)$$

where t refers to the comparison tracers, γ_t^* is the fitted step value fixing α and β to those of the reference (see Sect. 6.1), and $\gamma(c_t)$ is the expected step at contamination c_t following Eq. (3). Finally, since the c_t measurements are noisy, we compute the χ_{ref}^2 for each c_t chain walkers. We report in Fig. 6 (top panel) the median χ_{ref}^2 for each tracer used as reference together with the

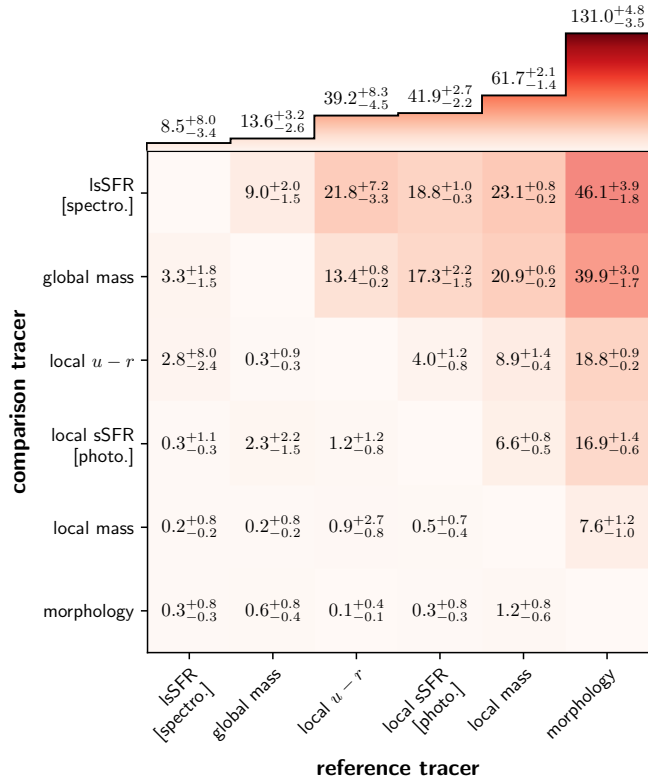


Fig. 6. Matrix reporting the median χ^2 for each reference tracer (columns) and each comparison tracer (rows) together with the 16% and 84% variations. The color red refers to the individual χ^2 values: the darker the color, the worse the value. At the top of the matrix is plotted, for each reference tracer choice, the median with the 16% and 84% variations when summing the tracer walkers following Eq. (14).

16% and 84% variations. This figure also displays the individual χ^2 contributions (main panel).

The χ^2 results confirm the qualitative observations. The spectroscopic IsSFR is the optimal reference tracer, followed by the global mass, the photometric IsSFR and the local $u-r$ colors. Local mass and morphology are the least suitable. Since the spectroscopic IsSFR is the best reference tracer choice, it means that it discriminates most accurately between the two underlying populations, which further strengthens the claim that there seems to be a prompt versus delayed age dichotomy.

Quantitatively, with $\chi^2 = 8.5$ for 5 degrees of freedom, the scatter along the contamination line for spectroscopic IsSFR as the reference is consistent with random scatter of 1.1σ . The scatter for global mass as the reference has $\chi^2 = 13.6$, corresponding to a scatter of 2.1σ . All other tracers are excluded from being accurate reference tracers at more than 5σ .

6.3. Scatter in the two-populations model

The two-populations model also has consequences for the observed scatter of the studied quantity q (see the introduction to the two-populations model in Sect. 2). If both populations differ by γ_0 on average in q , and keeping the assumption of 50% of the targets belonging to the a population, then we can show that marginalizing the populations results in an additional scatter in the dispersion of q by $0.5 \times \gamma_0$.

In the context of SNe Ia cosmology the studied quantity q is the standardized magnitude and interestingly the intrinsic scat-

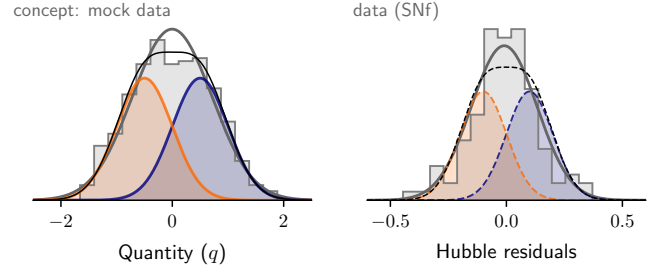


Fig. 7. Illustration of the scatter in the two-populations model. *Left panel:* distribution of quantity q for the same mock data as in the top panel of Fig. 1, represented by the gray histogram. The full thick gray line shows the Gaussian parameters associated with this distribution. The blue and orange curves represent the estimation of the q distribution from the a and b populations, respectively, while the black curve is their sum. *Right panel:* distribution of the Hubble residuals of the SNfactory dataset, represented by the gray histogram. The full thick gray line shows the Gaussian that parameterizes this distribution. The dashed orange and dashed blue curves show the expected distribution of the two underlying populations that would explain the full intrinsic distributions (see Sect. 6.3). The dashed gray line is their sum and should be compared to the histogram.

ter, corresponding to the part of the standardized magnitude dispersion along the Hubble diagram that cannot be explained by known sources of errors, typically is of 0.10 mag (Betoule et al. 2014; Scolnic et al. 2018).

Hence, assuming the two-populations model, if the 0.10 mag intrinsic SNe Ia scatter were fully caused by the existence of two underlying populations, the average standardized magnitude difference between these populations would be 0.2 mag. This is illustrated in the left panel of Fig. 7: in this mock example, the observed full distribution of q seems like a flattened distribution with a larger scatter than the underlying individual Gaussian distributions.

If we apply this concept to the SNfactory dataset, which also has an intrinsic dispersion of ~ 0.10 mag (Rigault et al. 2020), we can then guess the two underlying population distributions that would cause this effect; this is illustrated in the right panel of Fig. 7. In the case of the SNfactory standardization SNe Ia distribution, the central part (mag ~ 0) does not seem to qualitatively follow the expected distribution. This suggests that the entire intrinsic distribution might not be fully explained by the existence of a magnitude bias of ~ 0.2 mag between two underlying SN Ia populations.

Interestingly, this conceptual analysis provides key information on the upper limit of the astrophysical bias affecting SNe Ia standardized magnitudes in the context of the two populations model: it cannot be larger than twice the intrinsic dispersion and, consequently, it is smaller than ~ 0.2 mag.

In addition, if the magnitude step related to the spectroscopic IsSFR is ~ 0.16 mag, as claimed by Rigault et al. (2020), then the contamination of this tracer is lower than 25%.

7. Conclusion

We use a sample of 110 SNe Ia from the Nearby Supernovae Factory dataset to study the apparent inconsistencies in the literature between the different observed environmental dependencies of the standardized SNe Ia magnitudes. In the last ten years the SNe Ia luminosity has been shown to significantly depend on host properties, ranging from barely significant variations when split by host galaxy morphology (e.g., Pruzhinskaya et al. 2020)

to a very significant 15% luminosity difference when the SNe are split with respect to the spectroscopic specific star formation rate of their local environment (Rigault et al. 2020), leaving the 8% luminosity difference measured using the commonly used global host galaxy mass step in between (e.g., Sullivan et al. 2010; Roman et al. 2018).

To study these variations we first analyze the mathematical implications of assuming a step function (i.e., of comparing the SN Ia magnitude means when splitting the data into two bins). We show that doing so implicitly assumes two things: (1) that there exists two underlying populations that differ in standardized magnitudes and (2) that the environmental tracer used is somewhat able to distinguish them. Exploration of the implications of this implicit two-populations model enables us to demonstrate that the expected step observed by a tracer depends on its ability to accurately discriminate between the two underlying populations. In detail, if we call c the fraction of targets misclassified by the environmental tracer, and if we call γ_0 the true difference in mean magnitudes between the two populations, then the expected measured magnitude means offset γ is given by $\gamma = \gamma_0 \times (1 - c)$. The higher the contamination, hence the lower the tracer accuracy, the lower the expected measured step. In addition, the intrinsic magnitude dispersion caused by marginalizing the populations is half their magnitude offset. Since SNe Ia intrinsic dispersion is typically ~ 0.10 mag, the upper limit magnitude offset between the two SNe Ia populations would be ~ 0.20 mag, if the entire SNe Ia intrinsic scatter was caused by the existence of these two populations.

In light of that prediction, we derive the main literature environmental tracers for each of the 110 SNe Ia: spectroscopic and photometric measurements of the local specific star formation rate, the global and local stellar masses, the host galaxy morphology, and the local color. In this first analysis, we assume that one of these tracers is set as a reference. This provides a lower limit on the expected true amplitude of γ_0 . We draw from this analysis the following conclusions.

Tracer contamination model. Our model of a “two SN Ia populations model observed with tracers of various accuracy” explains well the observed variations. In Fig. 4 we show the expected versus measured magnitude step as a function of the derived tracer contaminations and we find good agreement, which supports a two-populations model. When applied to the steps reported in the literature, our model is able to explain the observed variations.

Spectroscopic *lsSFR* as reference tracer. When compared to the other tracers, using the spectroscopic local specific star formation rate as a reference tracer can explain all other observations with a scatter at 1.1σ . All other measurements are excluded as suitable reference tracers, with the possible exception of global mass, which shows a 2.1σ scatter, as we can see in Fig. 5 and quantified in Fig. 6.

The prompt versus delayed model. The spectroscopic *lsSFR* measures the fraction of young stars in the SNe Ia vicinity. As all observations are explained by using it as the reference tracer, the prompt versus delayed progenitor age model seems to best represent the behavior of the underlying populations. Nicolas et al. (2021) further show that this model also explains the observed redshift-drift of the SN Ia stretch distribution.

Origin of the mass-step. It seems that two populations related to progenitor age, combined with tracer accuracy, can explain all previous measurements of the mass-step. This con-

clusion is in agreement with former analyses (e.g., Rigault et al. 2013, 2020; Roman et al. 2018).

Standardization coefficients α, β . Because the standardizing parameters such as stretch and color are correlated with the two underlying populations, hence with their tracers, the use of an inaccurate tracer, such as morphology, biases the derivation of α and β , as would a non-simultaneous estimation of α, β , and γ .

The amplitude of γ . Under the assumption of the two-populations model, the amplitude of the astrophysical bias affecting the SNe Ia luminosity (i.e., the intercept of the $\gamma(c)$ plot) is close to the age-step reported in Rigault et al. (2020) (0.162 ± 0.029 mag) since the spectroscopic *lsSFR* is a good reference tracer. When we fit the intercept jointly on all literature data points using the derived contaminations from SNfactory sample, we find 0.121 ± 0.010 (see Fig. 4).

In light of the described two-populations model and the importance of tracer accuracy when assessing the amplitude of the astrophysical bias in SNe Ia cosmology, we highlight the importance of careful analyses of astrophysical biases when deriving cosmological parameters. This is true even when comparing two SN Ia samples at similar redshift ranges, if their selection function favored a given underlying population for any reason.

To avoid biases, one might want to probe as accurately as possible the underlying populations and to be careful when assessing them using only moderately good tracers such as global ones. In practice we urge caution with respect to current cosmological analyses that use the mass-step as the third standardization parameter to account for astrophysical dependencies in the SN Ia magnitude. The host stellar mass is not the underlying parameter affecting the SN Ia progenitor explosion mechanism; we see it, rather, as a tracer correlated with the true underlying physics. As astrophysical properties evolve significantly with cosmic time, it is critical to understand the relationship between SN Ia luminosity and the environment when doing SNe Ia cosmology.

Acknowledgements. We thank the anonymous referee for the constructive comments which helped to improve the conclusions of the paper. This project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation program (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. Additional support was provided by NASA under the Astrophysics Data 1095 Analysis Program grant 15-ADAP15- 0256 (PI: Aldering).

References

- Alam, S., Albareti, F. D., Allende Prieto, C., et al. 2015, *ApJS*, 219, 12
 Aldering, G., Adam, G., Antilogus, P., et al. 2002, *SPIE*, 4836, 61
 Aldering, G., Antilogus, P., Aragon, C., et al. 2020, *RNAAS*, 4, 63
 Arnouts, S., & Ilbert, O. 2011, *Astrophysics Source Code Library* [record ascl:1108.009]
 Arnouts, S., Cristiani, S., Moscardini, L., et al. 1999, *MNRAS*, 310, 540
 Aubourg, É., Tojeiro, R., Jimenez, R., et al. 2008, *A&A*, 492, 631
 Banerji, M., Lahav, O., Lintott, C. J., et al. 2010, *MNRAS*, 406, 342
 Barbary, K. 2016, *J. Open Source Software*, 1, 6
 Betoule, M., Kessler, R., Guy, J., et al. 2014, *A&A*, 568, A22
 Bolzonella, M., Miralles, J.-M., & Pelló, R. 2000, *A&A*, 363, 476
 Brout, D., & Scolnic, D. 2021, *ApJ*, 909, 26
 Bruzual, G., & Charlot, S. 2003, *MNRAS*, 344, 1000
 Calzetti, D. 2013, *Secular Evolution of Galaxies*, 419
 Casado, J., Ascasibar, Y., Gavilán, M., et al. 2015, *MNRAS*, 451, 888
 Chambers, K. C., Magnier, E. A., Metcalfe, N., et al. 2016, *ArXiv e-prints* [arXiv:1612.05560]
 Childress, M., Aldering, G., Antilogus, P., et al. 2013a, *ApJ*, 770, 108

- Childress, M., Aldering, G., Antilogus, P., et al. 2013b, *ApJ*, **770**, 107
- Choi, Y.-Y., Han, D.-H., & Kim, S. S. 2010, *J. Korean Astron. Soc.*, **43**, 191
- Dixon, S. 2021, *PASP*, **133**, 054501
- Fitzpatrick, E. L. 1999, *PASP*, **111**, 63
- Foreman-Mackey, D., Hogg, D. W., Lang, D., et al. 2013, *PASP*, **125**, 306
- Freedman, W. L., Madore, B. F., Gibson, B. K., et al. 2001, *ApJ*, **553**, 47
- Freedman, W. L., Madore, B. F., Hatt, D., et al. 2019, *ApJ*, **882**, 34
- Gupta, R. R., D'Andrea, C. B., Sako, M., et al. 2011, *ApJ*, **740**, 92
- Gupta, R. R., Kuhlmann, S., Kovacs, E., et al. 2016, *AJ*, **152**, 154
- Henne, V., Pruzhinskaya, M. V., Rosnet, P., et al. 2017, *New Astron.*, **51**, 43
- Ilbert, O., Arnouts, S., McCracken, H. J., et al. 2006, *A&A*, **457**, 841
- Jones, D. O., Riess, A. G., & Scolnic, D. M. 2015, *ApJ*, **812**, 31
- Jones, D. O., Riess, A. G., Scolnic, D. M., et al. 2018, *ApJ*, **867**, 108
- Karman, W., Caputi, K. I., Caminha, G. B., et al. 2017, *A&A*, **599**, A28
- Kauffmann, G., Heckman, T. M., White, S. D. M., et al. 2003, *MNRAS*, **341**, 54
- Kelly, P. L., Hicken, M., Burke, D. L., et al. 2010, *ApJ*, **715**, 743
- Kelsey, L., Sullivan, M., Smith, W., et al. 2021, *MNRAS*, **501**, 4861
- Kennicutt, R. 1998, *LIA Colloq. 34: The Next Generation Space Telescope: Science Drivers and Technological Challenges*, 429, 81
- 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, *J. Korean Astron. Soc.*, **52**, 141
- Knox, L., & Millea, M. 2020, *Phys. Rev. D*, **101**, 043533
- Labbé, I., Oesch, P. A., Bouwens, R. J., et al. 2013, *ApJ*, **777**, L19
- 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
- Nicolas, N., Rigault, M., Copin, Y., et al. 2021, *A&A*, **649**, A74
- Perlmutter, S., Aldering, G., Goldhaber, G., et al. 1999, *ApJ*, **517**, 565
- Planck Collaboration XIII. 2016, *A&A*, **594**, A13
- Planck Collaboration VI. 2020, *A&A*, **641**, A6
- Ponder, K. A., Wood-Vasey, W. M., Weyant, A., et al. 2020, *ApJ*, submitted, [arXiv:2006.13803]
- Pruzhinskaya, M. V., Novinskaya, A. K., Pauna, N., et al. 2020, *MNRAS*, **499**, 5121
- Reid, M. J., Pesce, D. W., & Riess, A. G. 2019, *ApJ*, **886**, L27
- 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. 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. 2020, *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
- Scannapieco, E., & Bildsten, L. 2005, *ApJ*, **629**, L85
- Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, *ApJ*, **500**, 525
- Scolnic, D., & Kessler, R. 2016, *ApJ*, **822**, L35
- Scolnic, D. M., Jones, D. O., Rest, A., et al. 2018, *ApJ*, **859**, 101
- Scolnic, D., Perlmutter, S., Aldering, G., et al. 2019, *Astro 2020: Decadal Survey on Astronomy and Astrophysics*, 2020, 270
- Shimasaku, K., Fukugita, M., Doi, M., et al. 2001, *AJ*, **122**, 1238
- Smith, M., Nichol, R. C., Dilday, B., et al. 2012, *ApJ*, **755**, 61
- Smith, M., Sullivan, M., Wiseman, P., et al. 2020, *MNRAS*, **494**, 4426
- Strateva, I., Ivezić, Ž., Knapp, G. R., et al. 2001, *AJ*, **122**, 1861
- Sullivan, M., Le Borgne, D., Pritchett, C. J., et al. 2006, *ApJ*, **648**, 868
- Sullivan, M., Conley, A., Howell, D. A., et al. 2010, *MNRAS*, **406**, 782
- Taylor, E. N., Hopkins, A. M., Baldry, I. K., et al. 2011, *MNRAS*, **418**, 1587
- Uddin, S. A., Mould, J., Lidman, C., et al. 2017, *ApJ*, **848**, 56
- Wiseman, P., Sullivan, M., Smith, M., et al. 2021, *MNRAS*, **506**, 3330
- Wong, K. C., Suyu, S. H., Chen, G. C.-F., et al. 2020, *MNRAS*, **498**, 1420
- Yoshikawa, T., Akiyama, M., Kajisawa, M., et al. 2010, *ApJ*, **718**, 112

Appendix A: Mathematical derivation of the modeling for two populations of SNe Ia

A.1. Two-populations model

The two-populations model estimates the probability of measuring certain fractions of false positives and false negatives, given known fractions of intrinsic false positives and false negatives. Closely following the notations presented in Section 2, we introduce the following variables:

- $k_i = \{a; b\}$ is a discrete indicator describing the true type of target i given the two SN Ia populations;
- c_a and c_b are, respectively, fractions of intrinsic false a and false b targets;
- \hat{t}_i is the true tracer value for i target;
- t_i is the measurement of the tracer for i target;
- δt_i is the measurement uncertainty of the tracer for i target;
- t_{cut} is the cutoff value, discriminating between the two categories, for tracer t .

Within this context, the probability of measuring t_i while the target i truly belongs to b population, given c_b , is expressed by

$$\begin{aligned}
\mathcal{P}(k_i, t_i | c_a, c_b) &= \int_{-\infty}^{+\infty} d\hat{t}_i \int d\theta \mathcal{P}(k_i, t_i, \hat{t}_i, \theta | c_a, c_b) \\
&= \int_{-\infty}^{+\infty} d\hat{t}_i \int d\theta \mathcal{P}(t_i | k_i, \hat{t}_i, \theta, c_a, c_b) \mathcal{P}(k_i, \hat{t}_i, \theta | c_a, c_b) \\
&= \int_{-\infty}^{+\infty} d\hat{t}_i \int d\theta \mathcal{P}(t_i | k_i, \hat{t}_i, \theta, c_a, c_b) \mathcal{P}(\hat{t}_i | k_i, \theta, c_a, c_b) \mathcal{P}(k_i, \theta | c_a, c_b) \\
&= \int_{-\infty}^{+\infty} d\hat{t}_i \int d\theta \mathcal{P}(t_i | k_i, \hat{t}_i, \theta, c_a, c_b) \mathcal{P}(\hat{t}_i | k_i, \theta, c_a, c_b) \mathcal{P}(k_i | \theta, c_a, c_b) \mathcal{P}(\theta | c_a, c_b)
\end{aligned} \tag{A.1}$$

where θ captures all of the other model parameters that may exist (i.e., the nuisance parameters). The first of the four terms in the last integral is related to measurement uncertainties, the second is the tracer probability, the third is the type probability, the fourth is the probability of drawing a target of class k , and the rest is how nuisance parameters are related to c_a and c_b .

In this paper, we make the following assumptions:

1. Knowledge of \hat{t}_i is all that is needed to obtain t_i , so $\mathcal{P}(t_i | k_i, \hat{t}_i, \theta, c_a, c_b) = \mathcal{P}(t_i | \hat{t}_i) = \mathcal{N}(t_i; \hat{t}_i, \delta t_i)$.
2. The unknown underlying distribution of \hat{t}_i only depends on the SN Ia population type k_i and the fraction of false a or false b targets, so $\mathcal{P}(\hat{t}_i | k_i = a, \theta, c_a) = \mathcal{P}(\hat{t}_i | k_i = a, c_a)$ and $\mathcal{P}(\hat{t}_i | k_i = b, \theta, c_b) = \mathcal{P}(\hat{t}_i | k_i = b, c_b)$.
3. Because we are only interested in knowing if a target is measured above or below a given cut, we use simple normalized top hats (\mathcal{U}) to build the t probability distribution functions, such that $\mathcal{P}(\hat{t}_i | k_i = b, c_b) = (1 - c_b) \mathcal{U}(t_{\text{min}}, t_{\text{cut}}) + c_b \mathcal{U}(t_{\text{cut}}, t_{\text{max}})$ and $\mathcal{P}(\hat{t}_i | k_i = a, c_a) = c_a \mathcal{U}(t_{\text{min}}, t_{\text{cut}}) + (1 - c_a) \mathcal{U}(t_{\text{cut}}, t_{\text{max}})$, where $(t_{\text{min}}, t_{\text{max}})$ corresponds to the boundaries for the parameter \hat{t}_i , though their values do not affect the inference (we use $t_{\text{cut}} - t_{\text{min}} = t_{\text{max}} - t_{\text{cut}} \gg \delta t_i$). The \mathcal{U} are normalized such that $\mathcal{U}(x; \text{min}, \text{max}) = (\text{max} - \text{min})^{-1}$ if x within min and max and 0 otherwise.

We tested the impact on this hypothesis on our results by simulating many mock samples with non-top hat functions, namely Gaussians or Gaussian mixtures with various parameters. When fitting these simulations with our baseline top hat model, we accurately recover the input c_a and c_b values for a large range of c value combinations.

4. In first approximation, the fraction of true a targets is constant. Notably, this requires that the fraction does not depend on redshift. While most likely overly simplistic for the general case, this assumption seems reasonable as we are studying data within a small redshift range ($0.03 < z < 0.08$). This results in $\mathcal{P}(k_i = a | \theta, c_a) = \mathcal{P}(k_i = a) = p_a$ and $\mathcal{P}(k_i = b) = (1 - p_a)$.
5. As a consequence of the given assumptions, there are no nuisance parameters (θ) in the model.

This way, for the case of $k_i = b$, equation A.1 simplifies to

$$\begin{aligned}
\mathcal{P}(k_i = b, t_i | c_b) &= \int_{-\infty}^{+\infty} d\hat{t}_i \int d\theta \mathcal{P}(t_i | \hat{t}_i) \mathcal{P}(\hat{t}_i | k_i, c_b) \mathcal{P}(k_i) \mathcal{P}(\theta) \\
&= \mathcal{P}(k_i = b) \int_{-\infty}^{+\infty} d\hat{t}_i \mathcal{P}(t_i | \hat{t}_i) \mathcal{P}(\hat{t}_i | k_i = b, c_b) \\
&= (1 - p_a) \int_{-\infty}^{+\infty} d\hat{t}_i \mathcal{N}(t_i; \hat{t}_i, \delta t) \left((1 - c_b) \mathcal{U}(t_{\text{min}}, t_{\text{cut}}) + c_b \mathcal{U}(t_{\text{cut}}, t_{\text{max}}) \right) \\
&= (1 - p_a) \left(\int_{-\infty}^{t_{\text{cut}}} d\hat{t}_i \mathcal{N}(\hat{t}_i; t_i, \delta t) (1 - c_b) \mathcal{U}(t_{\text{min}}, t_{\text{cut}}) + \int_{t_{\text{cut}}}^{+\infty} d\hat{t}_i \mathcal{N}(\hat{t}_i; t_i, \delta t) c_b \mathcal{U}(t_{\text{cut}}, t_{\text{max}}) \right) \\
&= \frac{1 - p_a}{\Delta} \left((1 - c_b) \int_{t_{\text{min}}}^{t_{\text{cut}}} d\hat{t}_i \mathcal{N}(\hat{t}_i; t_i, \delta t) + c_b \int_{t_{\text{cut}}}^{t_{\text{max}}} d\hat{t}_i \mathcal{N}(\hat{t}_i; t_i, \delta t) \right) \\
&= \frac{1 - p_a}{\Delta} \left((1 - c_b) \times f_i + c_b \times (1 - f_i) \right)
\end{aligned} \tag{A.2}$$

where $f_i = \int_{t_{\text{min}}}^{t_{\text{cut}}} d\hat{t}_i \mathcal{N}(\hat{t}_i; t_i, \delta t)$ is assimilated to the cumulative distribution function (see eq. 4 in the main text); $\Delta = (t_{\text{max}} - t_{\text{cut}}) = (t_{\text{cut}} - t_{\text{min}})$ is a constant normalization term.

A.2. Two-populations model with a reference tracer

Adding the reference tracer (and applying the same assumptions as above), we obtain

$$\begin{aligned}
\mathcal{P}(k_i = b, t_i, t_i^{\text{ref}} | c_b, c_b^{\text{ref}}) &= \int_{-\infty}^{+\infty} d\hat{t}_i \int_{-\infty}^{+\infty} d\hat{t}_i^{\text{ref}} \int d\theta \mathcal{P}(k_i, t_i, \hat{t}_i, t_i^{\text{ref}}, \hat{t}_i^{\text{ref}}, \theta | c_b, c_b^{\text{ref}}) \\
&= \int_{-\infty}^{+\infty} d\hat{t}_i \int_{-\infty}^{+\infty} d\hat{t}_i^{\text{ref}} \int d\theta \begin{aligned} &\times \mathcal{P}(t_i | k_i, \hat{t}_i, t_i^{\text{ref}}, \hat{t}_i^{\text{ref}}, \theta, c_b, c_b^{\text{ref}}) \\ &\times \mathcal{P}(t_i^{\text{ref}} | k_i, \hat{t}_i, \hat{t}_i^{\text{ref}}, \theta, c_b, c_b^{\text{ref}}) \\ &\times \mathcal{P}(\hat{t}_i | k_i, \hat{t}_i^{\text{ref}}, \theta, c_b, c_b^{\text{ref}}) \\ &\times \mathcal{P}(\hat{t}_i^{\text{ref}} | k_i, \theta, c_b, c_b^{\text{ref}}) \\ &\times \mathcal{P}(k_i | \theta, c_b, c_b^{\text{ref}}) \\ &\times \mathcal{P}(\theta | c_b, c_b^{\text{ref}}) \end{aligned} \\
&= \int_{-\infty}^{+\infty} d\hat{t}_i \int_{-\infty}^{+\infty} d\hat{t}_i^{\text{ref}} \int d\theta \mathcal{P}(t_i | \hat{t}_i) \mathcal{P}(t_i^{\text{ref}} | \hat{t}_i^{\text{ref}}) \mathcal{P}(\hat{t}_i | k_i, c_b) \mathcal{P}(\hat{t}_i^{\text{ref}} | k_i, c_b^{\text{ref}}) \mathcal{P}(k_i) \mathcal{P}(\theta) \\
&= \mathcal{P}(k_i = b) \int_{-\infty}^{+\infty} d\hat{t}_i \int_{-\infty}^{+\infty} d\hat{t}_i^{\text{ref}} \mathcal{P}(t_i | \hat{t}_i) \mathcal{P}(t_i^{\text{ref}} | \hat{t}_i^{\text{ref}}) \mathcal{P}(\hat{t}_i | k_i = b, c_b) \mathcal{P}(\hat{t}_i^{\text{ref}} | k_i = b, c_b^{\text{ref}}) \\
&= (1 - p_a) \begin{aligned} &\times \int_{-\infty}^{+\infty} d\hat{t}_i \mathcal{N}(t_i; \hat{t}_i, \delta t) \left((1 - c_b) \mathcal{U}(t_{\min}, t_{\text{cut}}) + c_b \mathcal{U}(t_{\text{cut}}, t_{\max}) \right) \\ &\times \int_{-\infty}^{+\infty} d\hat{t}_i^{\text{ref}} \mathcal{N}(t_i^{\text{ref}}; \hat{t}_i^{\text{ref}}, \delta t^{\text{ref}}) \left((1 - c_b^{\text{ref}}) \mathcal{U}(t_{\min}^{\text{ref}}, t_{\text{cut}}^{\text{ref}}) + c_b^{\text{ref}} \mathcal{U}(t_{\text{cut}}^{\text{ref}}, t_{\max}^{\text{ref}}) \right) \end{aligned} \\
&= \frac{1 - p_a}{\Delta'} \left((1 - c_b) \times f_i + c_b \times (1 - f_i) \right) \left((1 - c_b^{\text{ref}}) \times f_i^{\text{ref}} + c_b^{\text{ref}} \times (1 - f_i^{\text{ref}}) \right) \tag{A.3}
\end{aligned}$$

Similarly for $k_i = a$, we get

$$\mathcal{P}(k_i = a, t_i, t_i^{\text{ref}} | c_a, c_a^{\text{ref}}) = \frac{p_a}{\Delta'} \left(c_a \times f_i + (1 - c_a) \times (1 - f_i) \right) \left(c_a^{\text{ref}} \times f_i^{\text{ref}} + (1 - c_a^{\text{ref}}) \times (1 - f_i^{\text{ref}}) \right) \tag{A.4}$$

Finally, marginalizing over the population type, we find the general form of Eq. 10:

$$\begin{aligned}
\mathcal{P}(t_i, t_i^{\text{ref}} | c_a, c_b, c_a^{\text{ref}}, c_b^{\text{ref}}) &= \int dk_i \mathcal{P}(k_i, t_i, t_i^{\text{ref}} | c_a, c_b, c_a^{\text{ref}}, c_b^{\text{ref}}) \\
&= \mathcal{P}(k_i = a, t_i, t_i^{\text{ref}} | c_a, c_a^{\text{ref}}) + \mathcal{P}(k_i = b, t_i, t_i^{\text{ref}} | c_b, c_b^{\text{ref}}) \\
&\propto p_a \left(c_a^{\text{ref}} \times f_i^{\text{ref}} + (1 - c_a^{\text{ref}}) \times (1 - f_i^{\text{ref}}) \right) \left(c_a \times f_i + (1 - c_a) \times (1 - f_i) \right) \\
&\quad + (1 - p_a) \left((1 - c_b^{\text{ref}}) \times f_i^{\text{ref}} + c_b^{\text{ref}} \times (1 - f_i^{\text{ref}}) \right) \left((1 - c_b) \times f_i + c_b \times (1 - f_i) \right) \tag{A.5}
\end{aligned}$$

A.3. Testing the model robustness over the assumptions

We tested on simulations the two main assumptions made on this analysis, namely that the fraction of a class a target is a constant equal to 50% and that the probability distribution functions of the tracer values can be approximated by a combination of top hat distributions.

1. To test the impact of a fixed $p_a = 50\%$ term, we simulated many different models with varying p_a values ranging from 1 to 99%. Each time we built five different tracers with various c_a and c_b ranging from 5% to 50%, some symmetric ($c_a = c_b$) and some not. When fitting these simulated samples assuming our baseline model (hence with $p_a = 50\%$) we recovered the input c_a and c_b values with no bias as long as the input p_a is within 25 – 75% ; the bias become significant (at more than 3σ) only on the extreme ($p_a < 10\%$ or $p_a > 90\%$).
2. To test the assumption made on the tracer distribution in order to simplify the model, we built many simulations while varying the assumed distributions. We used Gaussian models or Gaussian mixture models with various parameter values to draw a perfect tracer prior to shuffling below (above) the tracer cutoff a fraction c_a (c_b) of targets to simulate the tracer's contamination. Random noise is added next, which we did for many combinations of c_a and c_b values. When fitting these simulated samples with our baseline model, each time we recovered the input c_a and c_b values with no bias.

We hence conclude that the assumptions made in the paper to simplify the likelihood have no consequences on our results.