A&A 662, A86 (2022)
https://doi.org/10.1051/0004-6361/202142577
© ESO 2022

Astronomy & Astrophysics

Systematic errors on optical-SED stellar-mass estimates for galaxies across cosmic time and their impact on cosmology

Ana Paulino-Afonso[1], Santiago González-Gaitán[1], Lluís Galbany[2,3], Ana Maria Mourão[1], Charlotte R. Angus[4], Mathew Smith[5], Joseph P. Anderson[6], Joseph D. Lyman[7], Hanindyo Kuncarayakti[8,9], and Myriam Rodrigues[10]

[1] CENTRA, Instituto Superior Técnico, Universidade de Lisboa, Av. Rovisco Pais 1, 1049-001 Lisboa, Portugal
e-mail: apaulinoafonso@tecnico.ulisboa.pt
[2] Institute of Space Sciences (ICE, CSIC), Campus UAB, Carrer de Can Magrans, s/n, 08193 Barcelona, Spain
[3] Institut d'Estudis Espacials de Catalunya (IEEC), 08034 Barcelona, Spain
[4] DARK, Niels Bohr Institute, University of Copenhagen, Jagtvej 128, 2200 Copenhagen, Denmark
[5] School of Physics and Astronomy, University of Southampton, Southampton SO17 1BJ, UK
[6] European Southern Observatory, Alonso de Córdova 3107, Casilla 19, Santiago, Chile
[7] Department of Physics, University of Warwick, Coventry CV4 7AL, UK
[8] Tuorla Observatory, Department of Physics and Astronomy, 20014 University of Turku, Finland
[9] Finnish Centre for Astronomy with ESO (FINCA), 20014 University of Turku, Finland
[10] GEPI, Observatoire de Paris, PSL University, CNRS, 5 Place Jules Janssen, 92190 Meudon, France

Received 3 November 2021 / Accepted 25 January 2022

ABSTRACT

Studying galaxies at different cosmic epochs entails several observational effects that need to be taken into account to compare populations across a large time-span in a consistent manner. We use a sample of 166 nearby galaxies that hosted type Ia supernovae (SNe Ia) and have been observed with the integral field spectrograph MUSE as part of the AMUSING survey. Here, we present a study of the systematic errors and bias on the host stellar mass with increasing redshift, which are generally overlooked in SNe Ia cosmological analyses. We simulate observations at different redshifts ($0.1 < z < 2.0$) using four photometric bands (*griz*, similar to the Dark Energy Survey-SN program) to then estimate the host galaxy properties across cosmic time. We find that stellar masses are systematically underestimated as we move towards higher redshifts, due mostly to different rest-frame wavelength coverage, with differences reaching 0.3 dex at $z \sim 1$. We used the newly derived corrections as a function of redshift to correct the stellar masses of a known sample of SN Ia hosts and derive cosmological parameters. We show that these corrections have a small impact on the derived cosmological parameters. The most affected is the value of the mass step Δ_M, which is reduced by ~0.004 (6% lower). The dark energy equation of state parameter w changes by $\Delta w \sim 0.006$ (0.6% higher) and the value of Ω_m increases at most by 0.001 (~0.3%), all within the derived uncertainties of the model. While the systematic error found in the estimate of the host stellar mass does not significantly affect the derived cosmological parameters, it is an important source of systematic error that needs to be corrected for as we enter a new era of precision cosmology.

Key words. cosmology: observations – cosmological parameters – supernovae: general – galaxies: fundamental parameters

1. Introduction

Type Ia supernovae (SNe Ia) have been successful as standard candles in probing the expansion history of our Universe over the last few decades (see e.g. Riess et al. 1998, 2018; Perlmutter et al. 1999; Betoule et al. 2014; Scolnic et al. 2018; Des 2019). However, SNe Ia are not perfect standard candles, and several empirical corrections are used to estimate their intrinsic luminosity. For example, light-curve shapes (Phillips 1993) and colours (Riess et al. 1996; Tripp 1998) have been used to reduce the scatter of their peak magnitudes by 50% and improve distance errors down to ~7%. With increasing samples of spectroscopically confirmed (e.g. Scolnic et al. 2018; Smith et al. 2020) and photometrically classified SNe Ia (Jones et al. 2018a), we are now in a phase where understanding the origin of these empirical corrections will improve our constraints and provide better corrections. This has potential implications for the determination of the equation of state of the Universe.

The observed scatter of SNe Ia distance residuals for the best-fit cosmological model is close to the 0.1 mag level (see e.g. Brout et al. 2019). This indicates that either there is a limit to which one can standardise SNe Ia, or there are additional correlations to their peak brightness that are not yet known because of limits on the quality of existing samples. These additional correlations are thought to arise from uncertainties related to the progenitor properties, the physics of SNe Ia explosions, and/or the environment in which they occur (see e.g. Scannapieco & Bildsten 2005; Mannucci et al. 2006; Maoz et al. 2014; Livio & Mazzali 2018). The drive to obtain ever more accurate standardisations of SNe Ia has motivated the search for additional empirical corrections based on the properties of the host galaxy used as tracers of the SNe Ia progenitors (e.g. Hicken et al. 2009; Sullivan et al. 2010; Kelly et al. 2010; Lampeitl et al. 2010; Gupta et al. 2011; D'Andrea et al. 2011; Hayden et al. 2013; Rigault et al. 2013; Childress et al. 2013; Johansson et al. 2013; Pan et al. 2014; Uddin et al. 2017, 2020; Ponder et al. 2021; Smith et al. 2020).

Article published by EDP Sciences

One of the most commonly used empirical corrections is based on the host stellar mass, with studies finding that SNe Ia occurring in galaxies with $M_\star > 10^{10}\,M_\odot$ require additional brightness corrections compared to those found in galaxies of lower stellar mass (e.g. Sullivan et al. 2010; Kelly et al. 2010; Lampeitl et al. 2010). Such a correction has been found in multiple studies and at various degrees of confidence (3–6σ) using multiple samples in the low- and high-redshift Universe (e.g. Sullivan et al. 2010; Kelly et al. 2010; Lampeitl et al. 2010; Childress et al. 2013; Johansson et al. 2013; Pan et al. 2014; Uddin et al. 2017, 2020; Ponder et al. 2021). However, it has been shown that more recent fitting frameworks lead to reduced corrections (e.g. Brout et al. 2019; Smith et al. 2020). There is currently no consensus on the physical motivation for this correction, as the stellar mass of galaxies is found to correlate with other global properties of the host galaxy: star-formation rate (SFR; e.g. Speagle et al. 2014), metallicity (e.g. Tremonti et al. 2004; Curti et al. 2020), and dust (e.g. Garn & Best 2010). Thus, it has also been found that the excess scatter could be corrected using other physical parameters of the host galaxy such as their metallicity and stellar age (Gupta et al. 2011; D'Andrea et al. 2011; Hayden et al. 2013; Pan et al. 2014; Moreno-Raya et al. 2016), SFR (Sullivan et al. 2010) or dust (Brout & Scolnic 2021).

The studies mentioned above focused on the global properties of the host galaxy because, for large cosmological distances, these are the only possible measurements with current instrumentation. Nonetheless, the progenitors of SNe Ia might reside in a particular region of the galaxy that is not well traced by their global properties. Recent studies on nearby galaxies have traced the empirical corrections to the local environment in which the SNe Ia occur Stanishev et al. (2012), Rigault et al. (2013, 2015, 2020), Galbany et al. (2014, 2016a), Jones et al. (2015, 2018b), Moreno-Raya et al. (2016), Roman et al. (2018), Kim et al. (2018, 2019), Rose et al. (2019, 2021), Kelsey et al. (2021). In these studies, the authors focused on the local SFR (traced by Hα emission or local $U - V/u - g$ colours) to find that SNe Ia in actively star-forming environments are fainter than those found in more passive environments. However, Jones et al. (2015, 2018b) find no conclusive evidence that correlations built from the local properties are better than those found with global properties.

Despite the existence of different empirical corrections, the correction based on the global host stellar mass has been the mostly used in cosmological analyses using SNe Ia (e.g. Sullivan et al. 2011; Betoule et al. 2014; Scolnic et al. 2018; Popovic et al. 2021). This is a consequence of the stellar mass being a more straightforward measurement to obtain, as it is the most robust parameter that can be estimated from photometry alone (e.g. Pforr et al. 2012). Nonetheless, care should be taken when estimating stellar masses and comparing estimates across a large redshift range, especially when using a small number of photometric bands as is typical in photometric studies of SNe. In this scenario, we need to account for observational effects (cosmological dimming and rest-frame coverage) that can impact the derived parameters. We aim to quantify the systematic errors on the estimates of stellar masses from the same photometric bands across a large redshift range, and test their impact on the derived cosmological parameters from SN studies.

In this paper, we use a sample of 166 nearby galaxies with integral field spectroscopic (IFS) data from the All-weather MUse Supernova Integral field Nearby Galaxies (AMUSING) survey (Galbany et al. 2016b) to simulate photometric observations of the same galaxies in the redshift range $0.1 < z < 2.0$. Using our host galaxy IFS data, we simulated *griz* observations and derived the host galaxy properties with commonly used spectral energy distribution (SED) fitting codes. We then took the observed differences between the new simulated properties and those derived in the local Universe to estimate a redshift-dependent stellar-mass correction. We used this new correction in our cosmological analysis and show the impact on the derived cosmological parameters.

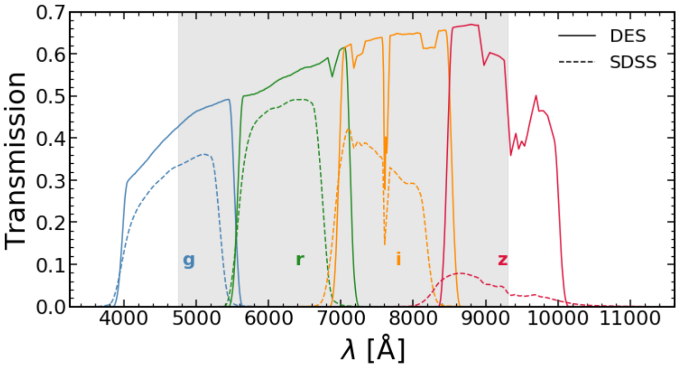

Fig. 1. Coverage of the MUSE spectroscopic data (shaded region) in comparison to the coverage of the DECam and SDSS *griz* filters.

This manuscript is organised as follows: in Sect. 2 we briefly explain the AMUSING survey on which our manuscript is based. In Sect. 3 we explain our novel method for simulating galaxy observations at higher redshift. Section 4 details the different stellar mass estimates that are used throughout the paper. We show our results regarding systematic errors on stellar mass estimation and their impact on the derivation of cosmological parameters, and we discuss our findings within the current ΛCDM paradigm in Sect. 5. We summarise our main conclusions in Sect. 6. We use AB magnitudes (Oke & Gunn 1983), a Chabrier (Chabrier 2003) initial mass function (IMF) unless otherwise explicitly stated, and assume a ΛCDM cosmology with $H_0 = 70\,\mathrm{km\,s^{-1}Mpc^{-1}}$, $\Omega_M = 0.3$, and $\Omega_\Lambda = 0.7$.

2. The AMUSING survey

In this work we use a sample of SN host galaxies drawn from the AMUSING survey[1] Galbany et al. (in prep.). Data were obtained with the Wide Field Mode of the MUSE instrument (Bacon et al. 2010) installed at the UT4 of the Very Large Telescope in Chile. Each pointing has an approximately 1′–1′ field of view (FoV) taken at a scale of 0.2″ pixel^{-1}. The spectra have a wavelength coverage in the optical range (4750 Å–9300 Å, see Fig. 1 for a comparison with the DECam and SDSS *griz* filter sets) with a fixed spectral sampling of 1.25 Å (spectral resolution of around 1800 at the blue edge and 3600 at the red edge). Our observations have a median seeing of ∼1″ which corresponds to a physical resolution of around 600 pc at the median redshift of our sample, $\langle z \rangle = 0.03$ (with 75% of the sample below $z = 0.05$), corresponding to a distance of ∼124 Mpc.

The data used in our work have been reduced using the MUSE pipeline (v1.2.1, Weilbacher et al. 2014) and the Reflex environment (Freudling et al. 2013). Tasks performed by the

[1] Based on observations made with ESO Telescopes at the Paranal Observatory (programmes 95.D-0091(A/B), 96.D-0296(A), 97.D-0408(A), 98.D-0115(A), P99 – 099.D-0022(A), P100 – 100.D-0341(A), P101 – 101.D-0748(A/B), P102 – 102.D-0095(A), P103 – 103.D-0440(A/B) and P104 – 104.D-0503(A/B).

pipeline include standard reduction such as subtracting bias, flat fielding, galactic extinction corrections, and flux/wavelength calibrations. For removal of the sky background, we use either an offset pointing to an empty region or blank sky regions within the science frames themselves (for smaller targets) and use the Zurich Atmosphere Purge package (ZAP, Soto et al. 2016) to perform this task. To reconstruct the final data product we applied a geometrical transformation of the individual slices to align them in a datacube. For more information on this procedure, we refer to Galbany et al. (2016b) and Krühler et al. (2017). We further corrected the fluxes of the observed spectra by matching the flux of the integrated galaxy light in the r-band to, by order of priority of available data, Pan-STARRS, DES, and SDSS photometry (Galbany et al., in prep.).

Our study is based on a subsample of the AMUSING survey that selects only SNe Ia host galaxies for which the FoV covers the entire galaxy, and no significant foreground contamination by bright stars or background contamination by distant galaxies is found in the MUSE datacubes. No galaxies with $z \geq 0.1$ are selected, with the great majority (~75%) having $z < 0.05$. There is no additional cut on any other property within the sample. Moreover, as the existence of foreground stars and/or background galaxies does not depend on either the host galaxy or the SNIa, the resulting subsample is akin to a random sampling of the parent sample. This process was conducted through visual inspection of each object and its corresponding segmentation map. This map is defined as the selection of all pixels belonging to the flagged object of interest, and this was done as a combination of two steps.

First, we searched for *Gaia* matches within the field of view of the MUSE datacube with a 1′ radial search around the cube centre using the `astroquery` package (Ginsburg et al. 2019). We then selected as foreground stars all objects with good parallax (π) measurements (i.e. $\pi > 2\pi_{\rm err}$ with $\pi_{\rm err}$ being the error on the parallax). With the final list of foreground stars, we built individual circular masks centred on each and with a radius containing 95% of the flux measured within a 3″ radius. As the final object map, we then selected all connected spaxels with a $S/N > 3$ that belong to the target object and do not overlap with the circular masks defined in the previous step. A similar S/N cut is applied when measuring photometry in the simulated observations (see Sect. 3).

All segmentation maps were individually inspected to select only objects without clear interlopers and with no other nearby objects (either bright foreground stars or background galaxies) that may contaminate the light of the galaxy of interest. After this inspection, a total of 166 galaxies were selected to be included in our study. To establish a comparison with other host galaxy samples in the literature, we computed the physical properties (stellar masses and SFRs) of our AMUSING subsample using MAGPHYS, as described in Sect. 4, and a *griz* magnitude set (using any of the other codes described below does not change the results significantly). This is comparable to the stellar mass estimates of the SDSS (Sako et al. 2018, $\langle z \rangle = 0.17$) and DES-SN program (Smith et al. 2020, $\langle z \rangle = 0.36$) samples[2]. As we show in Fig. 2, the AMUSING subsample spans similar stellar mass ranges as the samples from SDSS and has more massive galaxies on average than the sample from DES-SN program. This latter difference could be naturally explained by different cosmic epochs probed by the two samples. Our AMUSING lower redshift sample galaxies would have had more time to build up their stellar masses. In terms of SFRs, we find that our sample is slightly less star forming on average than the other two programs, but that can easily be explained by the median redshift of the sample, as one expects galaxies to increase their star formation as they move from $z \sim 0$ to $z \sim 2$ (see e.g. Madau & Dickinson 2014; Speagle et al. 2014). This shows that our population of galaxies is not particularly biased, and the differences among different surveys can be attributed to the different redshift ranges that are being targeted.

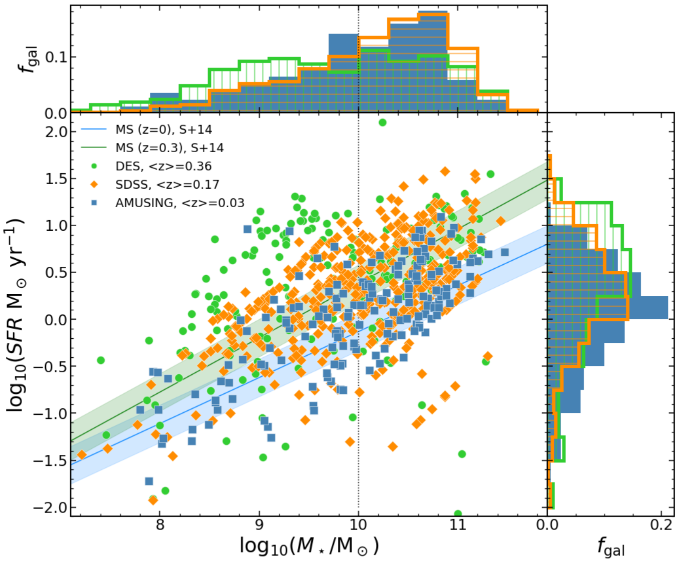

Fig. 2. Comparison of the AMUSING sample stellar masses and SFRs computed using MAGPHYS, see Sect. 4 (in blue), with the sample from Smith et al. (2020) (in green) and the one from Sako et al. (2018) (in orange). We show as lines with shaded regions the expected relation between stellar mass and SFR (commonly referred to as the main sequence) for the population of star-forming galaxies at different redshifts (adapted from Speagle et al. 2014). We show in the *upper panel* the stellar mass distributions and in the *right panel* the SFR distributions for the three samples. We highlight the $10^{10}\ M_\odot$ threshold for SNe Ia brightness corrections (see e.g. Sullivan et al. 2010) as the vertical dotted line.

3. Reconstructing data cubes at higher redshift

3.1. Extending the MUSE datacubes

We are interested in testing the impact of the observed wavelength range on estimated stellar masses. To simulate observations in a large redshift range we need to have an extended wavelength baseline. However, we note that for galaxies in our AMUSING subsample the differences in rest-frame coverage from one galaxy to another are negligible ($\Delta z \approx 0.03$) and much smaller than what we aim to simulate in our work.

To perform a simulation of the galaxy spectral energy distribution (SED) using a broad range of filters, we therefore artificially extended the data available in the MUSE datacubes (which cover the region 4750–9000 Å) to span a larger rest-frame wavelength coverage: 1200 Å–20 000 Å. To do so we use STARLIGHT (Cid Fernandes et al. 2005) to perform a spaxel-by-spaxel fit of the local spectra and then use the best-fit model to get the extended wavelength coverage (see Fig. 3). Prior to the fit with STARLIGHT, all major emission lines are masked as none of the models include them (the blue line in Fig. 3). We expect that

[2] We computed stellar masses using both their published catalog photometry and MAGPHYS and find negligible differences to their published values (smaller than 0.05 dex).

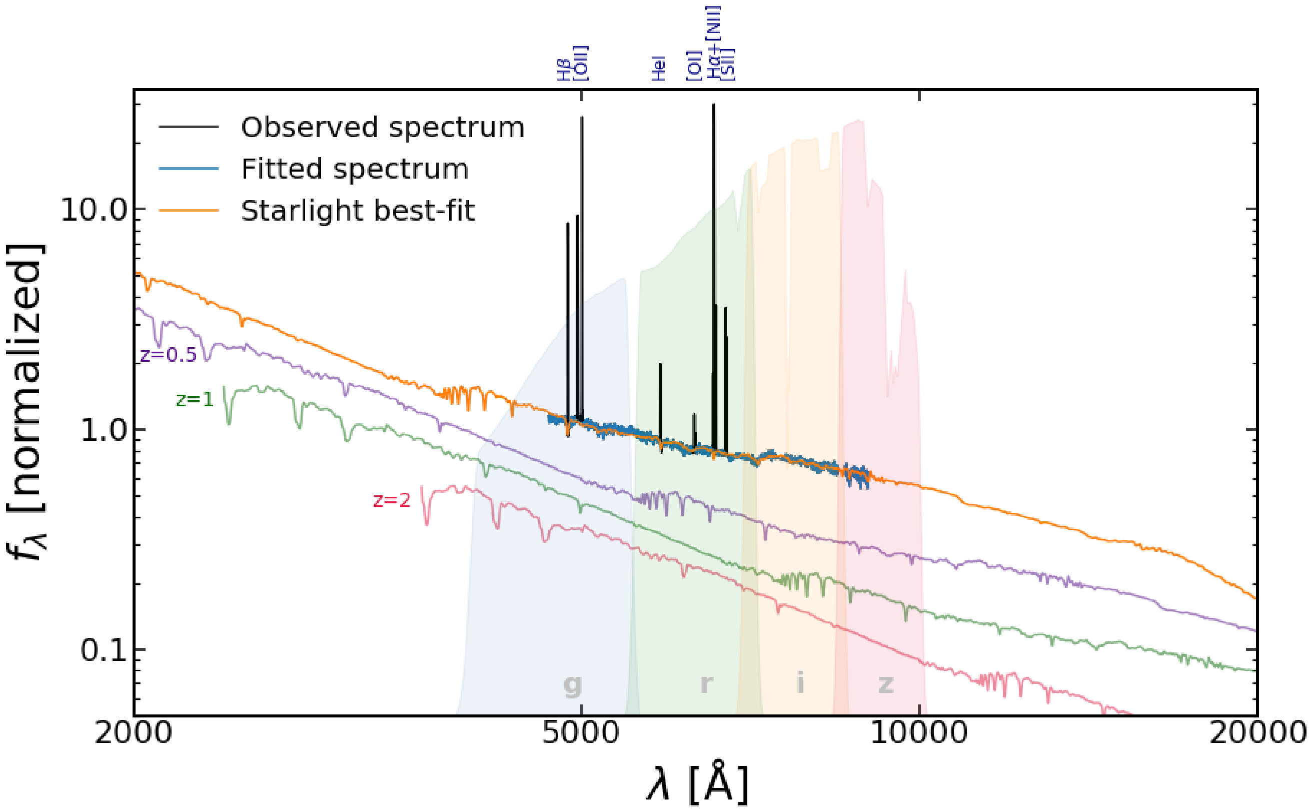

Fig. 3. Example spectrum with an extended wavelength coverage obtained from the best-fit STARLIGHT model (in orange) compared to the original MUSE data (in black). The best fit is done on the masked spectra (in blue). The transmission curves of the DES filters are shown as shaded regions. We also show the observed wavelengths of the redshifted spectra at z = 0.5, 1, and 2 in this figure as the purple, green, and red lines, respectively. Vertical offsets were applied for better visualisation. On top of the plot we identify the observed strong emission lines.

the masking of emission lines will have negligible impact on the derived stellar masses, which is the main goal in our work (e.g Whitaker et al. 2014). This fit is done for all spaxels belonging to each object map as defined in the previous section. The choice of the extended coverage takes into account that simulated galaxies will be used with optical and near-infrared (NIR) filters across a large redshift range ($z \lesssim 2$).

We use a combination of 45 base spectra built with the Bruzual & Charlot (2003) library and a Chabrier (2003) IMF. The base spectra span 15 stellar ages from 1 Myr to 13 Gyr and three metallicities (Z = 0.004, 0.02 and 0.05). The best-fit SSP template is then constructed as a linear combination of these base spectra that best approximates the observed spectra.

3.2. Artificial redshifting of galaxies

To estimate how the perceived properties of galaxies change across cosmic time, we wrote an algorithm (hereafter referred to as ARGAS) to simulate observations of how galaxies in the local Universe would look if they were at higher redshift. This is done by artificially redshifting galaxies, closely following the method described in Paulino-Afonso et al. (2017; see also Barden et al. 2008).

The core of the algorithm consists of three separate transformations. First, we apply a flux correction to the datacube (the dimming factor) that scales as the inverse of the luminosity distance to the galaxy. We then re-scale each wavelength slice of the cube (i.e. a 2D image at that wavelength) to match the pixel scale of the high-redshift observations whilst preserving the physical scale and flux of the galaxy. Finally, we redshift the extended galaxy spectra of each spaxel to match the observed frame at the requested redshift.

We show in Fig. 4 the effects of the scaling and dimming on images of a galaxy for the four different filters. The same method was applied to all slices of the extended datacube to re-create a MUSE observation at higher redshifts. From this extended and redshifted datacube, one can then extract photometry from filters within the observed wavelength interval between $1200\times(1+z)$ Å and $20\,000\times(1+z)$ Å for assessing possible biases in the estimation of physical galaxy parameters from photometric data (e.g. stellar mass or SFRs).

We applied each of these effects (dimming, scaling, and redshifting) separately and find that cosmological dimming is counteracted by the reduced physical resolution of higher redshift images. This experiment nicely confirms the concept of surface brightness which is independent of distance for instruments with the same resolution. While the flux observed at higher redshift is lower due to the cosmological dimming effect, this occurs because each pixel also covers a larger physical area of the galaxy which naturally corresponds to higher emitted flux per pixel. Furthermore, as both the luminosity distance and angular diameter distance scale similarly with redshift, they tend to cancel each other. We find less flux in the outskirts of galaxies as we move towards higher redshifts. Nonetheless, the different rest-wavelength coverage of the photometric filters has the most significant impact on the derived physical parameters. The rest-frame coverage changes with redshift, that is, towards bluer wavelengths as we move to higher redshifts when using the same

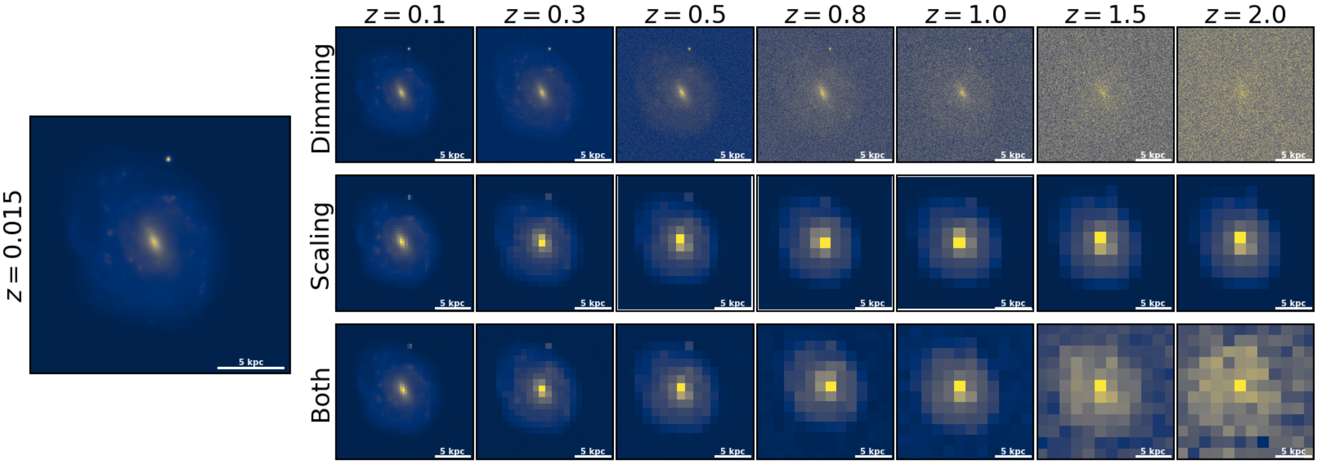

Fig. 4. Example of the different observational effects as simulated for a single scale instrument (same as used in our simulations, with a pixel scale of 0.2″) showing a galaxy (PGC 128348, host of SN Ia ASASSN-14jg) simulated at various redshifts (no SED redshifting is applied here). The original image is shown in the *single panel on the left*. The simulated images are shown in the grid with redshifts increasing *from left to right* and, *from top to bottom*, the effects of dimming, scaling and both applied to the galaxy. Each square has the exact same physical scale of ≈20 × 20 kpc.

filter set, leading to the major contribution to the observed differences. The use of 3D data from MUSE allows a more accurate depiction of observational effects than simply simulating integrated SEDs, as it allows us to measure the impact of flux loss in galaxy outskirts due to surface brightness dimming, and to get a good idea of the observed wavelength dependence of the flux.

3.3. Noise addition

To simulate realistic observation conditions, we need to add noise to the simulated high-redshift images. We assume that the noise is well described by a Gaussian distribution with a width defined by $\sigma_{\rm rms}$. We tested two approaches to simulating noise.

One approach is to scale the noise of the original MUSE datacube to the desired exposure time of the simulated observations. In doing this, we assume that the RMS is inversely proportional to the exposure time. In practice, we build a 2D noise map matching each of the filters we want to test. For each exposure time, we have that for an exposure time t the noise is described by a $\sigma_{\rm rms}(t) = \sigma_{\rm rms,0} \times t_0/t$, with t_0 and $\sigma_{\rm rms,0}$ being the exposure time and noise properties of the original datacube.

A second approach is to define a magnitude limit for each set of observed filters. To do this, we simulate a point-like object as a 2D Gaussian profile with an $FWHM = 3$ pixels (which is the typical sampling of a PSF, depending only on the instrument) with a flat spectrum with a constant value $f_\star$. We determine $f_\star$ to be the value for which the integrated magnitude in the observed filter and within a 3″ aperture is equal to the desired magnitude limit. We then compute the $\sigma_{\rm rms}$ that allows the simulated star to be detected with an $S/N = 5$ in the 3″ aperture. This helps simulate the conditions of typical surveys, for which the limiting depth is similar across the observed fields. The value of $\sigma_{\rm rms}$ is estimated by exploring a fixed list of values, computing the magnitude of the star at each value of $\sigma_{\rm rms}$ and comparing that to the real magnitude of the star. Once the difference between magnitudes exceeds 0.2 mag[3], we select that value of $\sigma_{\rm rms}$ to fix our simulated survey depth. To remove the bias of having a particular realisation of a 2D Gaussian noise distribution to define our final value of $\sigma_{\rm rms}$, we repeated this procedure 200 times and defined our final value of $\sigma_{\rm rms}$ as the median of those 200 realisations.

In the remainder of the paper, we use simulations with noise added using the latter of the two approaches described above. Our choice was based on the fact that this approach is the one that can most easily be matched to existing survey designs given the publicly available information. The conclusions from our work do not change if we choose the first approach to add noise to the images. We simulated galaxies with four different limiting magnitudes, $m_{\rm lim}$: 25, 27, 29, and 31. The results in this work are all based on a value of $m_{\rm lim} = 27$ (akin to the wide COSMOS survey, Scoville et al. 2007; Koekemoer et al. 2007) used for all redshifts. The conclusions from this work remain similar if we use any of the other three values, with the exception that we fail to detect most of the sources at $z > 1.5$ when simulating with $m_{\rm lim} = 25$. This implies that to observe galaxies at $z < 1$ using an instrument with the simulated plate scale of 0.2″ pix^{-1}, it suffices to have a depth of ~25 mag across all photometric bands. For $m_{\rm lim} = 27$, we detect ≳90% of the sample in all photometric bands at all redshifts.

4. Estimating stellar masses

Estimating a galaxy stellar mass from photometric data has been a powerful driver of extragalactic studies over the past decades. In particular, SED-fitting codes have often been used to this end (Le Borgne & Rocca-Volmerange 2010; Burgarella et al. 2005; Ilbert et al. 2006; da Cunha et al. 2008; Kriek et al. 2009; Carnall et al. 2018; Johnson et al. 2021). However, getting the right stellar mass estimate is not yet a well-posed problem due to the large number of model choices available prior to fitting data (e.g. Pforr et al. 2012; Mitchell et al. 2013; Acquaviva et al. 2015; Mobasher et al. 2015; Lower et al. 2020). To estimate the stellar masses and SFRs for the galaxies in our sample, we performed our SED fitting using several publicly available SED-fitting codes, which we describe below. We tried to use the same set of templates and configurations among different SED-fitting codes, although this is not always possible because of individual code design choices. We detail below the set of templates and choices used with each code. All our fitting was done using the photometric data derived for DES *griz* filter set, as seen in Fig. 1.

[3] A $S/N = 5$ means a 20% error on the flux, which translates to $-2.5\log_{10}(1.2) \approx 0.2$ mag.

When fitting the SEDs, the redshift of the galaxies is known from the spectra and fixed.

4.1. ZPEG

In ZPEG (Le Borgne & Rocca-Volmerange 2010) the template library for the stellar populations is built from PEGASE.2 (Fioc & Rocca-Volmerange 1997) from a set of nine exponentially declining star-formation histories, where

$$\mathrm{SFR}(t) \propto \frac{\exp\left(-t/\tau\right)}{\tau}, \tag{1}$$

with t being the age of the galaxy and τ the e-folding time, a parameter with the following possible values: 0.1, 0.2, 0.3, 0.4, 0.5,0. 75, 1, 1.5, or 2 Gyr. The SED is computed for 201 time-steps from 0 to 14 Gyr and the standard nebular emission prescription is used (see Fioc & Rocca-Volmerange 1997, for details on this). For each template, the initial metallicity has a value of 0.004 and evolves with time (with new stars having the metallicity of the ISM). We use the Kroupa (2001) IMF for this set of templates as PEGASE does not include base templates derived using Chabrier (2003). Nonetheless, we expect the differences in stellar masses from using these two IMFs to be small ($M_{\star,\mathrm{Kroupa}} = 1.06 M_{\star,\mathrm{Chabrier}}$, e.g. Speagle et al. 2014). We assume a uniform dust screen model using the Calzetti et al. (2000) law and with $E(B-V) = 0$–0.3 with 0.05 mag steps.

4.2. LePhare

LEPHARE was originally a photometric redshift code (Arnouts et al. 1999; Ilbert et al. 2006), but it can also be used to estimate a number of physical parameters of galaxies from the best-fit templates. This code is one of the most flexible of those used in this paper, and to minimise differences among different codes we use LePhare with the same templates as those described in the previous section (ZPEG templates). The only difference is the addition or absence of emission lines on top of the original templates created following the prescription described by Ilbert et al. (2006).

4.3. MAGPHYS

MAGPHYS (da Cunha et al. 2008) uses stellar templates constructed from the stellar libraries by Bruzual & Charlot (2003) and the dust absorption model follows Charlot & Fall (2000). The adopted IMF is that defined by Chabrier (2003). In this code, the star-formation histories are derived from an exponentially declining model and superimposed random bursts. Stellar metallicities are uniformly sampled between 0.02 and 2 times solar metallicity. Although there is no freedom to change the underlying templates, the code compares the data to the entire library and builds the probability distributions for each physical parameter (e.g. stellar mass, SFR, dust, among others). Moreover, while this constraint limits our ability to compare directly with other codes, we use a set of libraries that are commonly used in the community and can serve as a standard reference.

4.4. CIGALE

CIGALE is a code that was used to build optical-to-infrared SED models with and without AGN contributions, and can also be used to estimate the physical parameters of galaxies with no AGN contribution and limited wavelength coverage, as is our case Burgarella et al. (2005), Noll et al. (2009), Boquien et al. (2019). This code allows us a few degrees of freedom, and we try to match the set of available templates to those prescribed by MAGPHYS. The major difference is that we cannot replicate the same star-formation histories, and we use an exponentially delayed τ model ($\mathrm{SFR}(t) \propto t \times \exp\left(-t/\tau\right)/\tau$) with τ having the same values between 0.1 and 2 Gyr as described in Sect. 4.1.

4.5. PROSPECTOR

Finally, we use PROSPECTOR (Johnson et al. 2020, 2021), which allows a Bayesian exploration of the parameter space based on a set of template libraries of choice. We try our best to mimic the template configuration of MAGPHYS. We allow for the variation of three parameters: stellar-mass (with a top-hat prior $8 < \log_{10}(M_\star/M_\odot) < 12$); metallicity (with a top-hat prior $-1.7 < \log(Z/Z_\odot) < 0.3$); and an exponentially declining star formation history with a log-uniform prior $0.1 < \tau < 30$ Gyr). The IMF is fixed to that of Chabrier (2003), and we use the dust law defined by Charlot & Fall (2000) with the dust index fixed at −0.7 (the same as assumed in MAGPHYS).

5. Results and discussion

The goal of our work is to study the impact of observational strategies on the derived stellar masses of galaxies. To test this, we applied our artificial redshifting code (ARGAS) to 166 galaxies from the AMUSING survey and simulated observations at seven different redshifts $z = 0.1, 0.3, 0.5, 0.8, 1.0, 1.5$, and 2.0. At each redshift, we compute the photometric data in the four *griz* bands from DECam and use the SED fitting codes described in Sect. 4 to get the best stellar mass of the galaxy. For each code, we use the stellar mass computed at the original redshift of the galaxy ($z \sim 0.03$) – using the same filters and templates – as a frame of reference.

5.1. Underestimation of stellar masses

After obtaining our stellar mass estimates, we compare the one obtained at each simulated redshift with that obtained locally using the same filter set and library templates. The median differences for our 166 galaxy sample between the simulated and local values are shown in Fig. 5 (see also Table 1).

One of the first findings is that, despite the observed differences among the different codes used, there is a systematic underestimation of the stellar mass that depends on the redshift. This has implications for the implementation of the mass-step correction, as it implies that galaxies which are observed to be below the $10^{10}\,M_\odot$ threshold for correction may actually lie above it. This effect becomes more prominent as we move towards higher redshifts as more galaxies are affected (larger median offset from the true value). This is an important aspect that needs to be considered when estimating stellar masses for a singular dataset (i.e. observed with the same photometric bands) across an extensive redshift range, as is the case of large surveys such as DES. Given our defined set of filters and our choice of stellar population templates, we find that the LEPHARE (excluding emission lines) code is the overall best code in estimating stellar masses for galaxies at $z \leq 0.5$. Interestingly, ZPEG performs better for galaxies $0.5 < z \leq 1$. This is likely due to a combination of the nebular emission prescription included in the templates used and the filters where the emission lines are expected to fall.

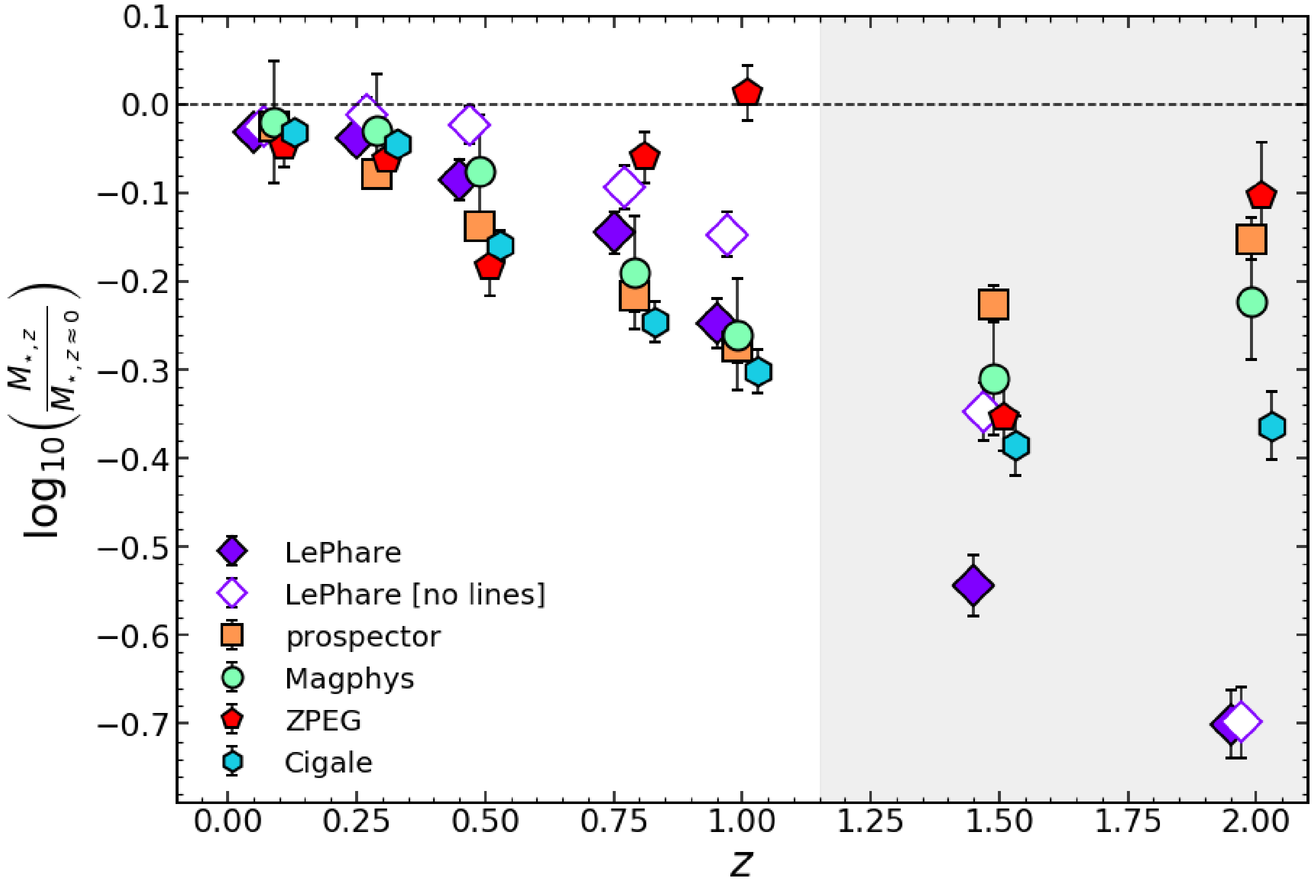

Fig. 5. Median stellar mass difference for our sample of galaxies as a function of redshift in the case of different SED fitting codes. The stellar mass reference at $z \approx 0$ is computed for each individual code using the same photometric filters. This difference can reach 0.2–0.3 dex by $z \sim 1$. We find that LePhare (with no added emission lines) gives the best overall results for $z < 1$ of all used codes, but it still underestimates the stellar masses at $z \gtrsim 0.5$. The shaded region indicates the redshifts for which the SED fitting codes are not well calibrated as we are mostly probing regions in the rest-frame UV.

Table 1. Median difference [in dex] on estimated stellar masses for all simulated redshifts (one per column) and different codes (one per row) used in this work.

Code	$z = 0.1$	$z = 0.3$	$z = 0.5$	$z = 0.8$	$z = 1.0$	$z = 1.5$	$z = 2.0$	All
LePhare	−0.03 ± 0.01	−0.04 ± 0.02	−0.09 ± 0.02	−0.15 ± 0.02	−0.25 ± 0.03	−0.54 ± 0.03	−0.70 ± 0.04	−0.11 ± 0.01
LePhare [nolines]	−0.03 ± 0.01	−0.01 ± 0.02	−0.02 ± 0.02	−0.09 ± 0.02	−0.15 ± 0.03	−0.35 ± 0.03	−0.70 ± 0.04	−0.08 ± 0.01
magphys	−0.02 ± 0.07	−0.03 ± 0.06	−0.08 ± 0.06	−0.19 ± 0.06	−0.26 ± 0.06	−0.31 ± 0.06	−0.22 ± 0.07	−0.10 ± 0.03
ZPEG	−0.05 ± 0.02	−0.06 ± 0.03	−0.18 ± 0.03	−0.06 ± 0.03	0.01 ± 0.03	−0.35 ± 0.04	−0.10 ± 0.06	−0.12 ± 0.01
Cigale	−0.03 ± 0.00	−0.05 ± 0.01	−0.16 ± 0.02	−0.25 ± 0.02	−0.30 ± 0.02	−0.39 ± 0.03	−0.36 ± 0.04	−0.11 ± 0.01
prospector	−0.02 ± 0.00	−0.08 ± 0.01	−0.14 ± 0.02	−0.22 ± 0.02	−0.27 ± 0.02	−0.23 ± 0.02	−0.15 ± 0.02	−0.10 ± 0.01

Notes. Errors are computed as $\sigma/\sqrt{N}$, with N being the number of galaxies in the bin. In the last column, we show the overall performance across all redshifts.

Although there are several studies in the literature that tackle a similar issue of estimating physical parameters, they present results using a much broader filter set. For instance, Pforr et al. (2012), Mitchell et al. (2013), and Mobasher et al. (2015) use optical, NIR, and mid-infrared (MIR; IRAC photmetry), Acquaviva et al. (2015) use additional UV photometry and more recently Lower et al. (2020) use FIR data from *Herschel* to constrain physical parameters. This extended set of photometric points is what is usually required for accurately constraining SED fitting models, given the number of available variables that need constraining (Acquaviva et al. 2015; Mobasher et al. 2015). Additionally, none of these studies evaluate the same galaxy simulated at different redshifts. They either consider exclusively mock galaxies (Pforr et al. 2012; Mitchell et al. 2013; Lower et al. 2020), real data (Acquaviva et al. 2015), or a mix of both (Mobasher et al. 2015). Nevertheless, the differences among different codes are consistent with results from Mobasher et al. (2015), who found an average spread of 0.136 dex in stellar mass differences estimated from different SED fitting codes using a similar set of assumptions in the model templates. The maximum scatter on the estimation of stellar masses was found to be due to contamination from nebular emission, reaching values of up to 0.5 dex (Mobasher et al. 2015). With respect to stellar mass estimation bias as a function of redshift, both Pforr et al. (2012) and Mitchell et al. (2013) find no significant differences. However, in their test case, they were using a much larger filter set, and estimating stellar masses for mock galaxies simulated to be at the redshift at which they were being observed.

Interestingly, Pforr et al. (2012) tested the impact of assuming different filter sets on photometry of mock galaxies, which includes two sets close to the one we study (*ugriz* and *UBVRI*). Contrary to our results, they find no significant difference with redshift, even for these smaller filter sets. However, we note that galaxies in their study are derived from simulations at the redshift at which they are being observed, and only include star-forming galaxies with young stars dominating the SED at optical wavelengths. We suppose that it is the fact that we are observing two different types of SEDs at each redshift that is driving the difference between our works. Namely, in our work we use a more evolved population that is the same at all redshifts, whereas Pforr et al. (2012) simulate star-forming galaxies that evolve with the redshifts they are testing. This tends to counterbalance the effect of filter shifting (when applied over the same population), which is likely due to a combination of the added u-band coverage and a population of younger galaxies. These are two complementary approaches to a similar problem that nicely test different aspects of stellar mass estimates across large redshift ranges.

In our experimental setup, we are attempting to fit the same galaxies using different rest-frame coverage (here corresponding to the different simulated redshifts) and a small filter set for SED fitting to mimic the conditions for large sky surveys where most SNe are found. We find that the one feature that most affects the measured stellar mass is the possibility to constrain the 4000 Å break that allows one to have an idea of the fraction of young and old stars in the galaxy and better constrains the average stellar population age. As we move towards higher redshifts, we are sampling increasingly bluer wavelengths, and thus giving more weight to the younger stellar population (e.g. Pforr et al. 2012; Mobasher et al. 2015) that can outshine the older stellar populations which add up to most of the galaxy stellar mass (especially in star-forming galaxies, see e.g. Sorba & Sawicki 2018). Furthermore, as these younger stars have lower mass-to-light ratios, estimates of stellar masses based on these wavelengths tend to be lower than the true value (Pforr et al. 2012).

5.2. Impact on cosmology

We find that galaxy stellar mass corrections depend strongly on the observed redshift. This can be a problem for cosmological fits based on SN data that span a sizeable cosmic time and use SN host galaxy stellar masses as the third empirical correction to their brightness. Our findings imply that some galaxies observed at stellar masses lower than $10^{10}\,M_{\odot}$ are more likely to actually be above that correction threshold. This is increasingly critical as we move towards higher redshifts. In this subsection, we use our derived corrections to estimate their impact on the derived best-fit cosmological models.

We use the median stellar mass difference to re-estimate the best-fit cosmological parameters for the Betoule et al. (2014) sample. We do this using two different approaches. The first uses the best approximation we derive from the set of SED-fitting codes that were tested in our paper (i.e. LEPHARE [no-lines] as shown in Fig. 5). In the second approach, we combine all individual corrections using a weighted average to produce a global correction curve for the estimated stellar masses. To derive the stellar mass correction curve as a function of redshift ($\Delta M(z)$), we interpolate linearly between the simulated redshifts. We show these correction curves in Fig. 6. We restrict our stellar mass corrections to be valid only at $z \leq 1$. This has negligible impact on our tests for cosmological parameters because the majority of SNe are below that redshift limit.

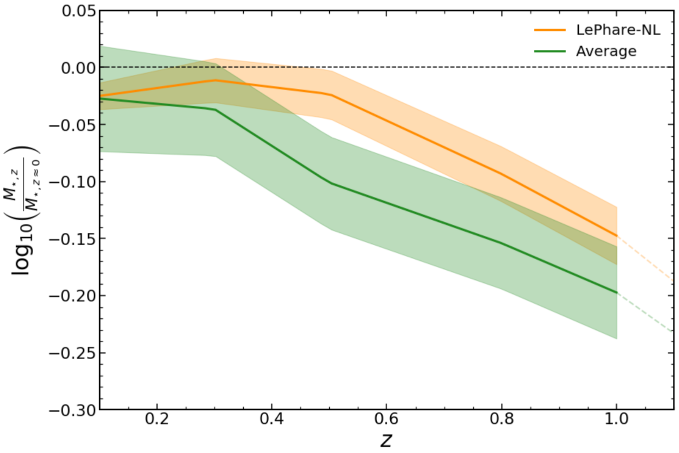

Fig. 6. Mass correction function for each redshift to be applied to the estimated observed stellar masses. The orange line represents the global mass correction using our best approximation (LEPHARE [no-lines]). The green line shows the correction to be applied using the weighted average correction derived from all SED-fitting codes. The shaded regions represent the uncertainty on the correction at each redshift.

The distance modulus to each SN can be modelled as (e.g. Betoule et al. 2014)

$$\mu(z) = M_B + 5\log_{10}(H_0[z, \Omega_m, w]) - \alpha \times (s - \bar{s}) + \beta \times c, \tag{2}$$

where s is the stretch term and c is the colour term. The SN luminosity is parameterized by

$$M_B = \begin{cases} M_{B,1} + \Delta_M, & \text{if } M_\star \geq 10^{10}\,M_\odot \\ M_{B,1}, & \text{otherwise} \end{cases}, \tag{3}$$

with $M_{B,1}$ being a free parameter, and Δ_M the magnitude difference to be applied for SNe Ia in massive hosts.

We estimate the best-fit parameters and corresponding probability density distributions using an MCMC approach with the package MONTEPYTHON (Brinckmann & Lesgourgues 2019; Audren et al. 2013). Our analysis is conducted using the "Joint Light-Curve Analysis" sample (Betoule et al. 2014, hereafter referred to as JLA). This sample combines data from 740 SNe Ia up to redshift $z \sim 1.3$ and data from the cosmic microwave background (CMB, Planck Collaboration VI 2020). We use a likelihood defined as (see e.g. González-Gaitán et al. 2021)

$$-2\ln(\mathcal{L}) = \sum_{\rm SN} \left\{ \frac{[\mu(z) - \mu_{\rm obs}]^2}{\sigma^2_{\rm tot}} \right\}, \tag{4}$$

where the uncertainty is defined as the diagonal of the covariance matrix:

$$\sigma^2_{\rm tot} = \sigma^2_{m_B} + (\alpha\sigma_S)^2 + (\beta\sigma_C)^2 + \sigma^2_{\rm int}. \tag{5}$$

We assume that $\sigma_{\rm int} = 0.105$, which is the average value for the JLA sample. We use the constraints of CMB data as a prior in our model in the same functional form as in Eq. (18) by Betoule et al. (2014), only updating the values with the latest release from the *Planck survey* (Planck Collaboration VI 2020).

To incorporate the uncertainty on the stellar mass correction models (see Fig. 6), we create 50 different correction curves

Table 2. Best-fit parameters of the cosmological model based on the three different configurations described in Sect. 5.2.

Parameter	Fiducial	Mass-corrected [LePhare-NL]	Mass-corrected [average]
$w0_{\rm fld}$	$-1.029^{+0.069}_{-0.043}$	$-1.022^{+0.053}_{-0.056}$	$-1.023^{+0.053}_{-0.055}$
α	$0.142^{+0.006}_{-0.007}$	$0.141^{+0.007}_{-0.006}$	$0.141^{+0.007}_{-0.006}$
β	$3.080^{+0.076}_{-0.083}$	$3.070^{+0.087}_{-0.074}$	$3.068^{+0.087}_{-0.075}$
M	$-19.108^{+0.036}_{-0.043}$	$-19.110^{+0.040}_{-0.037}$	$-19.111^{+0.041}_{-0.037}$
Δ_M	$-0.064^{+0.016}_{-0.030}$	$-0.063^{+0.022}_{-0.023}$	$-0.060^{+0.020}_{-0.022}$
Ω_m	$0.309^{+0.019}_{-0.012}$	$0.310^{+0.016}_{-0.015}$	$0.310^{+0.015}_{-0.015}$
$H0$	$68.249^{+1.248}_{-2.034}$	$68.075^{+1.594}_{-1.600}$	$68.043^{+1.652}_{-1.546}$

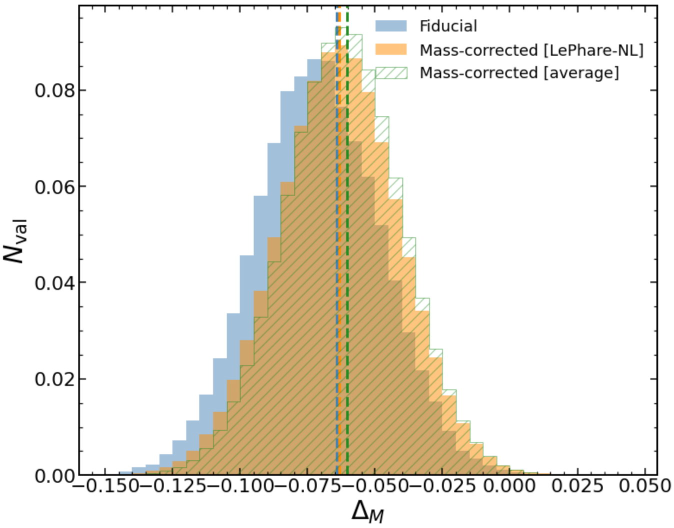

Fig. 7. Resulting posterior distributions on Δ_M for different runs. We show the LEPHARE-NL stellar mass correction results in orange and average stellar mass correction results in green, compared to the fit using the original stellar masses from Betoule et al. (2014). Vertical lines indicate the best-fit value for each configuration. We find that the fiducial model has a slightly larger value for Δ_M than either of the mass-corrected models, being very close to the best-fit value for the LEPHARE-NL correction model. Nevertheless, the resulting distributions for both mass-corrected models are similar.

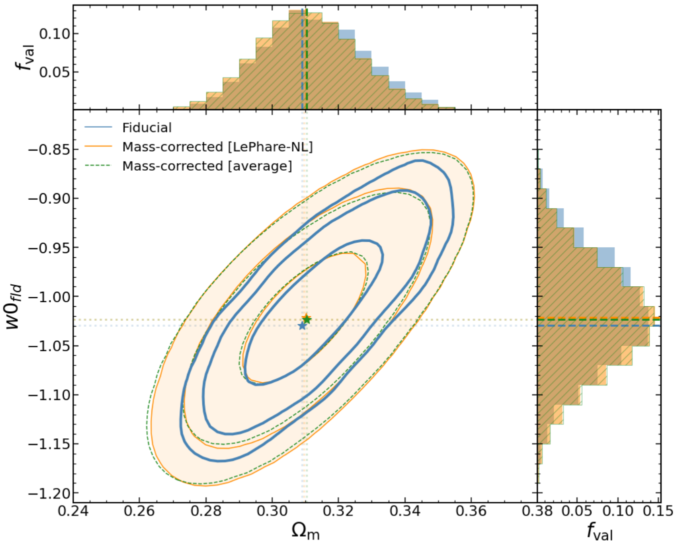

Fig. 8. Resulting posterior distributions on ω and Ω_m for different stellar mass corrections (LEPHARE-NL stellar mass correction in orange and average stellar mass correction in green) compared to the fit using the original stellar masses from Betoule et al. (2014) in blue. The contour levels correspond, from inside out, to 68%, 95%, and 99% of the posterior distribution. We show as stars (same colours as contours) the best-fit value for each configuration. Vertical and horizontal lines indicate the best-fit value of each parameter for the three different setups. There is no significant difference on the constraints when using the mass-corrected dataset with respect to the fiducial model. We find small differences in the best-fit values (star symbols), with the LEPHARE-NL stellar mass correction configuration showing the largest difference with respect to the fiducial model.

that are randomly perturbed around the median correction, and within the shown uncertainty region. We then run our cosmological fits for each of the 50 individual corrections. Finally, we combine the results into a single posterior distribution for each parameter, which is marginalised over the uncertainty on the stellar mass correction.

We do this exercise in three different configurations: one using the original stellar masses from the JLA sample, which is our fiducial model; and the two other MCMC runs use the derived stellar mass corrections shown in Fig. 6 applied to the measured stellar masses of the JLA sample. The best-fit values and corresponding uncertainties for each of these configurations is shown in Table 2. We also show all the posterior distributions for the fitted parameters in Fig. A.1.

We find that the parameter that changes the most when applying our stellar mass corrections is Δ_M. In our LEPHARE-corrected model, the best-fit parameter value decreases by ~2%, while when we apply our average correction, the difference with respect to the fiducial model is ~6%. As Δ_M is the parameter that is linked to the host stellar mass (Eq. (3)), it is expected to be the most affected by applying corrections to the original stellar masses (see Fig 7).

With respect to the constraints on the cosmological parameters, we find smaller differences when comparing to the fiducial model (<1%, see Figs. 7 and 8). The value of Ω_m increases by δ ~ 0.001 (~0.3%) in the LEPHARE-corrected and average-corrected fits. The value of w_0 increases by δ ~ 0.007 (~0.7%) when using the LEPHARE stellar mass corrections. To put this in context, this possible systematic bias corresponds to approximately one-tenth of the expected error budget in w_0 from Euclid (Amendola et al. 2018). The difference in w is slightly smaller when we use the average correction, with it increasing only by δ ~ 0.006 (~0.6%). Finally, we also find small changes in the H_0 parameter: δ ~ −0.2 (~−0.3%) for both the LEPHARE and the average stellar mass corrections. However, we note that these differences are all within the fitting uncertainties.

6. Conclusions

We study the impact of observational effects (namely cosmological dimming and rest-frame coverage) on estimations of the physical parameters of galaxies. In particular, we aim to assess the possible systematic bias on the estimation of stellar masses when analysing galaxy samples across a large redshift range. To achieve this goal, we used a sample of 166 SNe Ia host galaxies with IFS from the AMUSING survey. With these galaxies it was possible to simulate observations of galaxies at redshifts $0.1 < z < 2$ using a *griz* filter set to mimic the DES-SN program. Five different codes – CIGALE, LEPHARE, MAGPHYS, PROSPECTOR, ZPEG – were used to estimate stellar masses allowing a better identification of possible bias associated with the choice of SED-fitting models. We studied the implications of our results on the determination of cosmological parameters using mass step correction. Our main conclusions can be summarised as follows:

- Regardless of the code used to estimate stellar mass, this parameter is systematically underestimated, with the degree of underestimation increasing towards higher redshift. Depending on the individual code, this difference reaches around 0.2–0.3 dex by $z \sim 1$.
- We find that when correcting the observed stellar masses for a public SNe Ia sample, there is a small impact on the best-fit parameters of the cosmological model. The impact is of the same order of magnitude whether we use the LEPHARE-NL or the average stellar mass corrections.
- The cosmological parameters have the greatest impact when deriving the best-fit value of the magnitude correction Δ_M, which is reduced by ~2% and ~6% for the LEPHARE-NL and average stellar mass corrections, respectively. The cosmological parameters show deviations from the fiducial value below 1%: Ω_m increases by 0.3% ($\delta \sim 0.001$); w is reduced by 0.6% ($\delta \sim 0.006$); and H_0 decreases by 0.3% ($\delta \sim 0.2$). These differences are all within the fitting uncertainties, but could be a non-negligible source of systematic errors in the coming decade.

Our main conclusion is that stellar mass is systematically underestimated across a large redshift range, and that the extent of this underestimation depends on the redshift of the observed host galaxy. Forthcoming surveys, such as *Euclid* and/or *Nancy Grace Roman* Space Telescope, can help minimise these effects by providing a more significant baseline of rest-frame coverage (with added filters in the NIR regime), which will help to minimise the error budget. By doubling the number of filters into the NIR regime, one can hope to better constrain the region around 4000 Å rest-frame to higher redshifts, helping quantify the numbers of old and young stars in the galaxy, which are crucial for accurate stellar mass estimates.

Acknowledgements. This work was supported by Fundação para a Ciência e a Tecnologia (FCT) through the research grant UIDB/00099/2020 and the CRISP project PTDC/FIS-AST-31546/2017. L.G. acknowledges financial support from the Spanish Ministry of Science, Innovation and Universities (MICIU) under the 2019 Ramón y Cajal program RYC2019-027683 and from the Spanish MICIU project HOSTFLOWS PID2020-115253GA-I00. CRA was supported by grants from VILLUM FONDEN (project numbers 16599 and 25501). J.D.L. acknowledges support from a UK Research and Innovation Future Leaders Fellowship (MR/T020784/1). Computations were performed at the cluster "Baltasar-Sete-Sóis", supported by the H2020 ERC Grant "Matter and strong field gravity: New frontiers in Einstein's theory" grant agreement no. MaGRaTh-646597, and at COIN, the CosmoStatistics Initiative, whose purchase was made possible due to a CNRS MOMENTUM 2018–2020 under the project "Active Learning for large scale sky surveys". This work was only possible by the use of the following PYTHON packages: NumPy & SciPy (Walt et al. 2011; Jones et al. 2001), Matplotlib (Hunter 2007), and Astropy (Astropy Collaboration et al. 2013).

References

Acquaviva, V., Raichoor, A., & Gawiser, E. 2015, ApJ, 804, 8
Amendola, L., Appleby, S., Avgoustidis, A., et al. 2018, Liv. Rev. Rel., 21, 2
Arnouts, S., Cristiani, S., Moscardini, L., et al. 1999, MNRAS, 310, 540
Astropy Collaboration (Robitaille, T. P., et al.) 2013, A&A, 558, A33
Audren, B., Lesgourgues, J., Benabed, K., & Prunet, S. 2013, JCAP, 1302, 001
Bacon, R., Accardo, M., Adjali, L., et al. 2010, SPIE Conf. Ser., 7735, 773508
Barden, M., Jahnke, K., & Häußler, B. 2008, ApJS, 175, 105
Betoule, M., Kessler, R., Guy, J., et al. 2014, A&A, 568, A22
Boquien, M., Burgarella, D., Roehlly, Y., et al. 2019, A&A, 622, A103
Brinckmann, T., & Lesgourgues, J. 2019, Physics of the Dark Universe, 24, 100260
Brout, D., & Scolnic, D. 2021, ApJ, 909, 26
Brout, D., Scolnic, D., Kessler, R., et al. 2019, ApJ, 874, 150
Bruzual, G., & Charlot, S. 2003, MNRAS, 344, 1000
Burgarella, D., Buat, V., & Iglesias-Páramo, J. 2005, MNRAS, 360, 1413
Calzetti, D., Armus, L., Bohlin, R. C., et al. 2000, ApJ, 533, 682
Carnall, A. C., McLure, R. J., Dunlop, J. S., & Davé, R. 2018, MNRAS, 480, 4379
Chabrier, G. 2003, ApJ, 586, L133
Charlot, S., & Fall, S. M. 2000, ApJ, 539, 718
Childress, M., Aldering, G., Antilogus, P., et al. 2013, ApJ, 770, 108
Cid Fernandes, R., Mateus, A., Sodré, L., Stasińska, G., & Gomes, J. M. 2005, MNRAS, 358, 363
Curti, M., Mannucci, F., Cresci, G., & Maiolino, R. 2020, MNRAS, 491, 944
da Cunha, E., Charlot, S., & Elbaz, D. 2008, MNRAS, 388, 1595
D'Andrea, C. B., Gupta, R. R., Sako, M., et al. 2011, ApJ, 743, 172
Des, C. 2019, ApJ, 872, L30
Fioc, M., & Rocca-Volmerange, B. 1997, A&A, 500, 507
Freudling, W., Romaniello, M., Bramich, D. M., et al. 2013, A&A, 559, A96
Galbany, L., Stanishev, V., Mourão, A. M., et al. 2014, A&A, 572, A38
Galbany, L., Stanishev, V., Mourão, A. M., et al. 2016a, A&A, 591, A48
Galbany, L., Anderson, J. P., Rosales-Ortega, F. F., et al. 2016b, MNRAS, 455, 4087
Garn, T., & Best, P. N. 2010, MNRAS, 409, 421
Ginsburg, A., Sipőcz, B. M., Brasseur, C. E., et al. 2019, AJ, 157, 98
González-Gaitán, S., de Jaeger, T., Galbany, L., et al. 2021, MNRAS, 508, 4656
Gupta, R. R., D'Andrea, C. B., Sako, M., et al. 2011, ApJ, 740, 92
Hayden, B. T., Gupta, R. R., Garnavich, P. M., et al. 2013, ApJ, 764, 191
Hicken, M., Wood-Vasey, W. M., Blondin, S., et al. 2009, ApJ, 700, 1097
Hunter, J. D. 2007, Comput. Sci. Eng., 9, 90
Ilbert, O., Arnouts, S., McCracken, H. J., et al. 2006, A&A, 457, 841
Johansson, J., Thomas, D., Pforr, J., et al. 2013, MNRAS, 435, 1680
Johnson, B. D., Leja, J., Conroy, C., & Speagle, J. S. 2020, ApJS, 254, 21
Johnson, B. D., Leja, J., Conroy, C., & Speagle, J. S. 2021, ApJS, 254, 22
Jones, D. O., Riess, A. G., & Scolnic, D. M. 2015, ApJ, 812, 31
Jones, D. O., Scolnic, D. M., Riess, A. G., et al. 2018a, ApJ, 857, 51
Jones, D. O., Riess, A. G., Scolnic, D. M., et al. 2018b, ApJ, 867, 108
Jones, E., Oliphant, T., Peterson, P., et al. 2001, SciPy: Open Source Scientific Tools for Python [Online; Accessed 2016–03-23]
Kelly, P. L., Hicken, M., Burke, D. L., Mandel, K. S., & Kirshner, R. P. 2010, ApJ, 715, 743
Kelsey, L., Sullivan, M., Smith, M., et al. 2021, MNRAS, 501, 4861
Kim, Y.-L., Smith, M., Sullivan, M., & Lee, Y.-W. 2018, ApJ, 854, 24
Kim, Y.-L., Kang, Y., & Lee, Y.-W. 2019, J. Korean Astron. Soc., 52, 181
Koekemoer, A. M., Aussel, H., Calzetti, D., et al. 2007, ApJS, 172, 196
Kriek, M., van Dokkum, P. G., Labbé, I., et al. 2009, ApJ, 700, 221
Kroupa, P. 2001, MNRAS, 322, 231
Krühler, T., Kuncarayakti, H., Schady, P., et al. 2017, A&A, 602, A85
Lampeitl, H., Smith, M., Nichol, R. C., et al. 2010, ApJ, 722, 566
Le Borgne, D., & Rocca-Volmerange, B. 2010, ZPEG: An Extension of the Galaxy Evolution Model PEGASE.2
Livio, M., & Mazzali, P. 2018, Phys. Rep., 736, 1
Lower, S., Narayanan, D., Leja, J., et al. 2020, ApJ, 904, 33
Madau, P., & Dickinson, M. 2014, ARA&A, 52, 415
Mannucci, F., Della Valle, M., & Panagia, N. 2006, MNRAS, 370, 773
Maoz, D., Mannucci, F., & Nelemans, G. 2014, ARA&A, 52, 107
Mitchell, P. D., Lacey, C. G., Baugh, C. M., & Cole, S. 2013, MNRAS, 435, 87
Mobasher, B., Dahlen, T., Ferguson, H. C., et al. 2015, ApJ, 808, 101
Moreno-Raya, M. E., Mollá, M., López-Sánchez, Á. R., et al. 2016, ApJ, 818, L19

Noll, S., Burgarella, D., Giovannoli, E., et al. 2009, A&A, 507, 1793
Oke, J. B., & Gunn, J. E. 1983, ApJ, 266, 713
Pan, Y. C., Sullivan, M., Maguire, K., et al. 2014, MNRAS, 438, 1391
Paulino-Afonso, A., Sobral, D., Buitrago, F., & Afonso, J. 2017, MNRAS, 465, 2717
Perlmutter, S., Aldering, G., Goldhaber, G., et al. 1999, ApJ, 517, 565
Pforr, J., Maraston, C., & Tonini, C. 2012, MNRAS, 422, 3285
Phillips, M. M. 1993, ApJ, 413, L105
Planck Collaboration VI. 2020, A&A, 641, A6
Ponder, K. A., Wood-Vasey, W. M., Weyant, A., et al. 2021, ApJ, 923, 197
Popovic, B., Brout, D., Kessler, R., Scolnic, D., & Lu, L. 2021, ApJ, 913, 49
Riess, A. G., Press, W. H., & Kirshner, R. P. 1996, ApJ, 473, 88
Riess, A. G., Filippenko, A. V., Challis, P., et al. 1998, AJ, 116, 1009
Riess, A. G., Rodney, S. A., Scolnic, D. M., et al. 2018, ApJ, 853, 126
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
Roman, M., Hardin, D., Betoule, M., et al. 2018, A&A, 615, A68
Rose, B. M., Garnavich, P. M., & Berg, M. A. 2019, ApJ, 874, 32
Rose, B. M., Rubin, D., Strolger, L., & Garnavich, P. M. 2021, ApJ, 909, 28
Sako, M., Bassett, B., Becker, A. C., et al. 2018, PASP, 130, 064002
Scannapieco, E., & Bildsten, L. 2005, ApJ, 629, L85
Scolnic, D. M., Jones, D. O., Rest, A., et al. 2018, ApJ, 859, 101
Scoville, N., Abraham, R. G., Aussel, H., et al. 2007, ApJS, 172, 38
Smith, M., Sullivan, M., Wiseman, P., et al. 2020, MNRAS, 494, 4426
Sorba, R., & Sawicki, M. 2018, MNRAS, 476, 1532
Soto, K. T., Lilly, S. J., Bacon, R., Richard, J., & Conseil, S. 2016, ZAP: Zurich Atmosphere Purge [record ascl:1602.003]
Speagle, J. S., Steinhardt, C. L., Capak, P. L., & Silverman, J. D. 2014, ApJS, 214, 15
Stanishev, V., Rodrigues, M., Mourão, A., & Flores, H. 2012, A&A, 545, A58
Sullivan, M., Conley, A., Howell, D. A., et al. 2010, MNRAS, 406, 782
Sullivan, M., Guy, J., Conley, A., et al. 2011, ApJ, 737, 102
Tremonti, C. A., Heckman, T. M., Kauffmann, G., et al. 2004, ApJ, 613, 898
Tripp, R. 1998, A&A, 331, 815
Uddin, S. A., Mould, J., Lidman, C., Ruhlmann-Kleider, V., & Zhang, B. R. 2017, ApJ, 848, 56
Uddin, S. A., Burns, C. R., Phillips, M. M., et al. 2020, ApJ, 901, 143
Walt, S. V. D., Colbert, S. C., & Varoquaux, G. 2011, Comput. Sci. Eng., 13, 22
Weilbacher, P. M., Streicher, O., Urrutia, T., et al. 2014, in Astronomical Data Analysis Software and Systems XXIII, eds. N. Manset, & P. Forshay, ASP Conf. Ser., 485, 451
Whitaker, K. E., Franx, M., Leja, J., et al. 2014, ApJ, 795, 104

Appendix A: Full results from cosmological fits

In this section we show the posterior distributions for all the fitted parameters in our MontePython model (see description in Sect. 5.2). In Fig. A.1 we show that, for most parameters, the distributions are similar, with Δ_M being the variable that benefits the most from correcting stellar masses, with the magnitude correction to be applied depending on the host stellar mass (see Eq. 3).

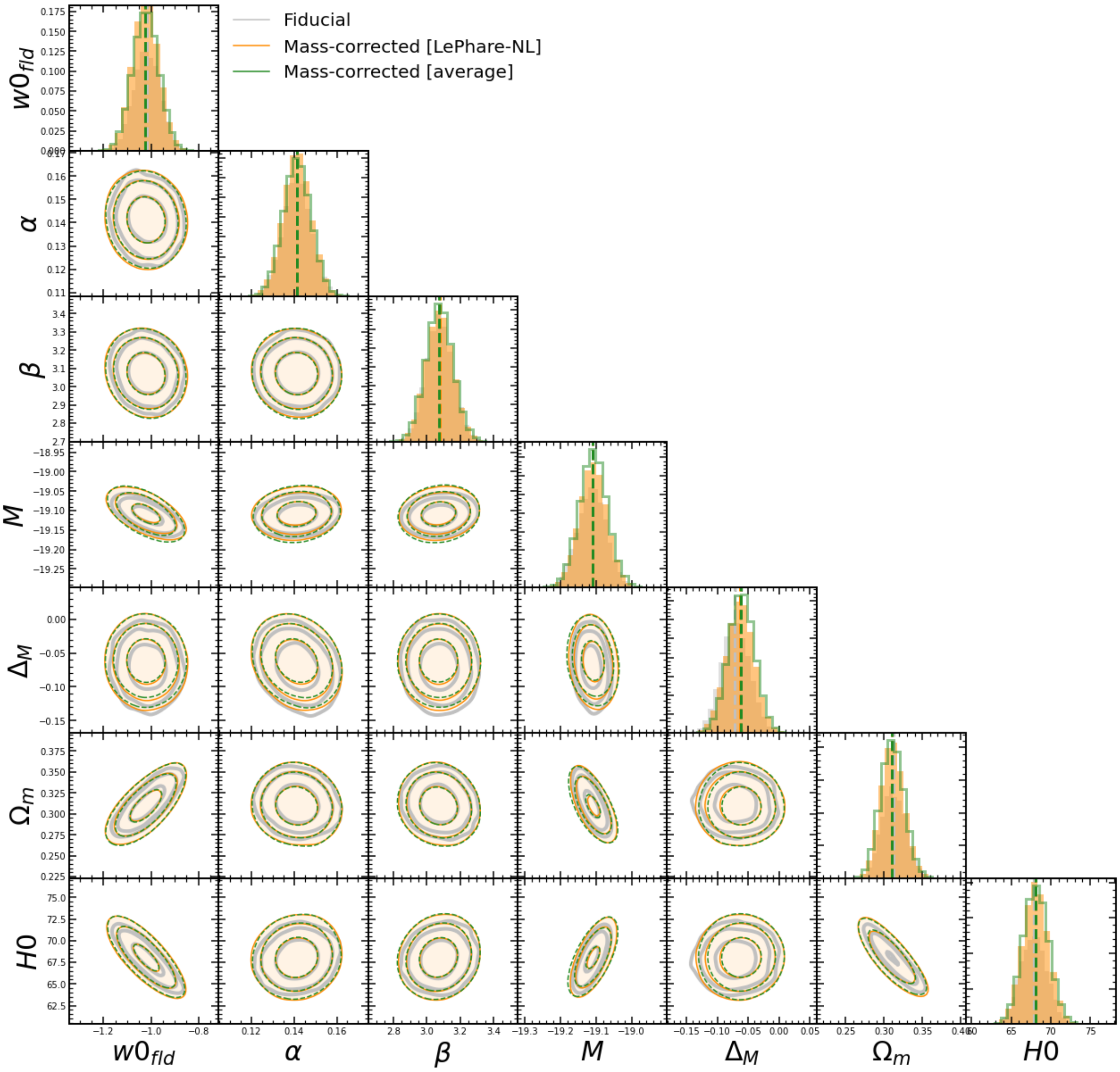

Fig. A.1. Resulting posterior distributions on all the free parameters for our cosmological model using the three different configurations: fiducial model (in grey), stellar masses corrected using the best approximation with LePhare-NL (in orange), and stellar masses corrected with the average difference among different codes (in green). The contour levels correspond, from inside out, to 68%, 95%, and 99% of the posterior distribution.