Horizon-AGN virtual observatory – 1. SED-fitting performance and forecasts for future imaging surveys

C. Laigle,[1]⋆ I. Davidzon,[2] O. Ilbert,[3] J. Devriendt,[1] D. Kashino [4] C. Pichon,[5,6,7]
P. Capak,[2] S. Arnouts,[3] S. de la Torre,[3] Y. Dubois,[5] G. Gozaliasl [8,9,10] D. Le Borgne,[5]
S. Lilly,[4] H. J. McCracken,[5] M. Salvato [11] and A. Slyz[1]

[1]*Sub-department of Astrophysics, University of Oxford, Keble Road, Oxford OX1 3RH*
[2]*IPAC, Mail Code 314-6, California Institute of Technology, 1200 East California Boulevard, Pasadena, CA 91125, USA*
[3]*Aix Marseille Université, CNRS, CNES, LAM (Laboratoire d'Astrophysique de Marseille) UMR7326, F-13388, Marseille, France*
[4]*Department of Physics, ETH Zürich, Wolfgang-Pauli-strasse 27, CH-8093 Zürich, Switzerland*
[5]*Institut d'Astrophysique de Paris, Sorbonne Universités, CNRS, UMR 7095, 98 bis bd Arago, F-75014 Paris, France*
[6]*Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK*
[7]*Korea Institute for Advanced Study (KIAS), 85 Hoegiro, Dongdaemun-gu, Seoul 02455, Republic of Korea*
[8]*Finnish Centre for Astronomy with ESO (FINCA), Quantum, Vesilinnantie 5, University of Turku, FI-20014 Turku, Finland*
[9]*Department of Physics, University of Helsinki, PO Box 64, FI-00014 Helsinki, Finland*
[10]*Helsinki Institute of Physics, University of Helsinki, P.O. Box 64, FI-00014, Helsinki, Finland*
[11]*Max-Planck-Institut für extraterrestrische Physik, PO Box 1312, Giessenbachstr. 1., D-85741 Garching, Germany*

ABSTRACT

Using the light-cone from the cosmological hydrodynamical simulation HORIZON-AGN, we produced a photometric catalogue over $0 < z < 4$ with apparent magnitudes in COSMOS, Dark Energy Survey, Large Synoptic Survey Telescope (LSST)-like, and *Euclid*-like filters at depths comparable to these surveys. The virtual photometry accounts for the complex star formation history (SFH) and metal enrichment of HORIZON-AGN galaxies, and consistently includes magnitude errors, dust attenuation, and absorption by intergalactic medium. The COSMOS-like photometry is fitted in the same configuration as the COSMOS2015 catalogue. We then quantify random and systematic errors of photometric redshifts, stellar masses, and star formation rates (SFR). Photometric redshifts and redshift errors capture the same dependencies on magnitude and redshift as found in COSMOS2015, excluding the impact of source extraction. COSMOS-like stellar masses are well recovered with a dispersion typically lower than 0.1 dex. The simple SFHs and metallicities of the templates induce a systematic underestimation of stellar masses at $z < 1.5$ by at most 0.12 dex. SFR estimates exhibit a dust-induced bimodality combined with a larger scatter (typically between 0.2 and 0.6 dex). We also use our mock catalogue to predict photometric redshifts and stellar masses in future imaging surveys. We stress that adding *Euclid* near-infrared photometry to the LSST-like baseline improves redshift accuracy especially at the faint end and decreases the outlier fraction by a factor ∼2. It also considerably improves stellar masses, reducing the scatter up to a factor 3. It would therefore be mutually beneficial for LSST and *Euclid* to work in synergy.

Key words: methods: observational – techniques: photometric – galaxies: evolution – galaxies: formation.

1 INTRODUCTION

Our understanding of galaxy formation, evolution, and of their distribution in the large-scale structure has taken a giant step forward in the last decade, owing to large multiwavelength data sets. Properties of different galaxy populations, and their evolution across cosmic time, can be constrained by measuring one-point statistics, such that the luminosity and stellar mass functions (e.g. Ilbert et al. 2006, 2013; Davidzon et al. 2017; Bundy et al. 2017). Two-point statistics, i.e. measuring the spatial correlation of galaxies, make it possible to investigate the role of the local environment (e.g.

⋆ E-mail: clotilde.laigle@physics.ox.ac.uk

Abbas & Sheth 2006; de la Torre et al. 2010; Hatfield & Jarvis 2017) and to infer halo properties, via simplifying assumptions such as the so-called halo model (e.g. McCracken et al. 2015; Coupon et al. 2015; Legrand et al. 2018). More generally, higher order statistics (see Moresco et al. 2017) as well as topological tools such as filament tracers, can help disentangle complex environmental effects, distinct from the isotropic influence of local density peaks (e.g. Malavasi et al. 2017; Laigle et al. 2018; Kraljic et al. 2018). When implementing such statistics, one must assess the impact of observational biases on inferring the underlying properties of the population.

In particular, investigations focusing on galaxy stellar mass assembly rely on three fundamental quantities: photometric redshifts, stellar masses, and SFRs. Large-area surveys can significantly reduce statistical errors in these kinds of measurements, and probe a wide variety of galaxy types and environments. Therefore, the dominant source of uncertainties in state-of-the-art studies became the selection biases of the surveys, the source extraction techniques, and the physical models assumed in the analysis (when needed). Even when a high-resolution galaxy spectral energy distribution (SED) is available, inferring physical properties from it is an ill-conditioned problem (Moultaka & Pelat 2000; Moultaka et al. 2004), which prevents a complete inversion approach to be successful (see e.g. Ocvirk et al. 2006). Difficulties are even more severe when only apparent magnitudes in broad-band filters are available. In that case, SED-fitting codes are routinely used because of their versatility. These codes fit pre-computed libraries of galaxy templates to the photometry of observed objects (see a review in Walcher et al. 2011; Conroy 2013). Some very promising alternative techniques are also being developed (see Salvato, Ilbert & Hoyle 2019, for a review), including 'clustering redshift' (Newman 2008; Ménard et al. 2013), 'photo-web' (Aragon-Calvo et al. 2015) and more recently machine-learning (see e.g. Masters et al. 2015; Beck et al. 2017; Pasquet et al. 2019; Gomes et al. 2018; Hemmati et al. 2018). However, these alternative techniques generally require large and representative spectroscopic samples, which is not the case of SED-fitting algorithms. In order to build a template library for the SED-fitting procedure, one relies none the less on several assumptions, mainly concerning star formation histories (SFHs), metal enrichment, and dust extinction and spatial distribution (Conroy 2013). These priors inevitably introduce systematics in the recovered physical quantities, which in turn may impair the statistical measurements and bias conclusions on galaxy mass assembly scenarii. For instance, Bundy et al. (2017) find that depending on the assumed SFH in the SED-fitting estimates, massive ($> 3 \times 10^{11}\,M_\odot$) galaxies between $z = 0$ and 0.8 may show either a mild stellar mass growth or a lack of evolution; this systematic uncertainty is dominant, considering that their extremely large sample of galaxies ($>41\,000$), collected across $\sim 140\,\mathrm{deg}^2$, makes shot noise and cosmic variance almost negligible.

Furthermore, in order to understand the physical processes regulating galaxy mass assembly, it is important to compare observational measurements to semi-analytical and hydrodynamical simulations, where different theoretical models of galaxy evolution have been implemented (e.g. De Lucia & Blaizot 2007; Vogelsberger et al. 2013; Dubois et al. 2014; Schaye et al. 2015). At present, such a task is not straightforward: a fair comparison should take into account biases and uncertainties affecting the observational analysis before comparing to simulated galaxies.

Therefore, it is of pivotal importance to assess the performances of photometric extraction and SED-fitting codes when recovering redshift and stellar mass in order to understand their impact on the statistical analyses of the galaxy population. Broadly speaking,

observational biases can occur because of image confusion (i.e. blending between two nearby galaxies), the choice of algorithm used to extract galaxy flux, and the assumptions made in the SED-fitting procedure. Previous works have already explored some of these effects. As an example, Mobasher et al. (2015) have quantified the global performances of an exhaustive list of existing SED-fitting codes, while relying on a large observed and semi-analytical mock catalogue (see also Hildebrandt et al. 2010). Focusing on mass and age estimates, Pforr, Maraston & Tonini (2012) and Pacifici et al. (2012) investigated the impact of the chosen template SFH, while the effect of dust and metallicity has been studied in Mitchell et al. (2013) and Hayward & Smith (2015).

Beyond the impact of simplistic SFHs (like the τ-model defined in Bruzual 1983), metallicity or dust distribution (see also Guidi et al. 2016), the performance of SED fitting is extremely sensitive to the choice of photometric filters, the depth of the survey, and flux measurements (see Bernardi et al. 2013). Hydrodynamical simulations have already been widely used to test the impact of the photometry extraction, as they allow to work on – often high resolution – mock images of realistic galaxies. Amongst the tested effects, the choice of the apertures (Price et al. 2017) and the lack of resolution (integrated photometry versus pixel-by-pixel fitting; Sorba & Sawicki 2015, 2018) have been found to systematically underestimate stellar masses (see also Sanderson et al. 2018) or to impair morphological estimators (Bottrell et al. 2017). All these past investigations underline the importance of understanding and quantifying biases when recovering physical parameters from surveys, which could be as large as two orders of magnitude in some particular mass and redshift ranges (see e.g. the effect of dust on stellar mass computation, Mitchell et al. 2013). However, most of the literature is based either on simple phenomenological prescriptions or semi-analytical models (SAMs), or when the sample is based on hydrodynamical simulation, it consists in no more than a handful of galaxies (e.g. Guidi et al. 2016). Hence we still lack a study relying on a sample that combines highly realistic baryon physics with a large cosmological volume (in order to minimize statistical uncertainties), capturing both galaxies' internal properties and environment. Moreover, this study must be an end-to-end analysis, i.e. including the same limitations introduced by the observational strategy and data reduction pipeline in current or future surveys. The present work aims to remedy this gap.

To this end, we exploit the light-cone from the HORIZON-AGN cosmological hydrodynamical simulation (Dubois et al. 2014). From this simulation, a mock catalogue of about 750 000 galaxies was extracted between $z = 0$ and 4, down to $M_* = 10^9\,M_\odot$. Hence, our sample combines large number statistics over a wide redshift range with a wealth of information on galaxy properties. Our aim is to carefully understand possible systematics arising when fitting the complex photometry of the galaxies with simplified templates. For this purpose, galaxy photometry has to be as realistic as possible. One advantage of using hydrodynamical simulations over SAMs for this work is to better resolve galaxies in space and time. Fluxes spatially vary across the simulated galaxies (in the limit of the resolution of the simulation) depending on metallicity enrichment and dust attenuation, and therefore the integrated photometry will present a complexity similar to the real galaxies. In addition, SFHs in the hydrodynamical simulation vary on a fine time grid and depend not only on the merger history of their host halo but also on stellar and active galactic nucleus (AGN) feedback, and on the detail of the gas accretion history. As emphasized in e.g. Mitchell et al. (2018), several quantities (e.g. the gas return time-scale) are naturally constrained by gravitational forces and hydrodynamics, whereas

Table 1. A summary of the configurations envisaged in this study. Depths are quoted in AB magnitudes. The depths in all bands are summarized in online Table A1. A complete list of the COSMOS bands is provided in table 1 of L16. A HORIZON-AGN photometric catalogue is built for each configuration.

Name	Bands	i-band depth	NIR depth	References
COSMOS-like	26 bands from u to 4.5 μm	26.2 ± 0.1 (3σ)	24.7 ± 0.1 (K_s, 3σ)	Laigle et al. (2016)
LSST-like	u, g, r, i, z, y	27.0 (5σ)	NA	LSST Science Collaboration et al. (2009)
Euclid + DES	g, r, i, z, riz, Y, J, H	24.5 (riz, 10σ) and 24.3 (i, 10σ)	24.0 (H band, 5σ)	Abbott et al. (2018); Laureijs et al. (2011)
Euclid + LSST	$u, g, r, i, z, y, riz, Y, J, H$	27.0 (5σ)	24.0 (H band, 5σ)	Rhodes et al. (2017)

they would conversely need to be globally tuned in SAMs. Finally, the light-cone geometry mimics that of observed surveys, and allows us for instance to implement the attenuation by the intergalactic medium (IGM) for each galaxy by drawing individually lines of sight through the foreground gas distribution.

The goal of this first study is to assess the photometric redshift (z_{phot}), stellar mass (M_*), and star formation rate (SFR) uncertainties caused by the choice of the filters, the signal-to-noise ratio (S/N) of the photometry and the SED-fitting recipe used for analysing the real galaxies.

For this purpose, observed-frame photometry is post-processed with COSMOS-like S/N for each galaxy of the HORIZON-AGN light-cone (as described in Section 2). Then photometric redshifts and physical properties (M_* and SFR) of mock galaxies are measured by applying the same pipeline used in the COSMOS field (Laigle et al. 2016, hereafter L16). This procedure allows us to identify which source of uncertainty dominate the error budget (Section 3). After validation on the COSMOS2015 data, we mimic (in Section 4) the expected photometry for the *Euclid* mission (Laureijs et al. 2011), along with the Dark Energy Survey (DES, Abbott et al. 2018) and the Large Synoptic Survey Telescope (LSST, LSST Science Collaboration et al. 2009), to predict the expected z_{phot} and M_* accuracy they should provide at completion. The possible synergy between these surveys is also explored. We then summarize our analysis and draw conclusions in Section 5. Additional material can be found in the online Appendices, where we provides more details about how the virtual photometry has been computed (online Appendix A); we further discuss dust and IGM absorption (online Appendix B), zero-point magnitude offsets (online Appendix C), and redshift errors (online Appendix D). These virtual catalogs are going to be made publicly available at https://www.horizon-simulation.org/data.html.

Throughout this study, we use a flat Λ-cold dark matter cosmology with $H_0 = 70.4 \, \text{km} \, \text{s}^{-1} \, \text{Mpc}^{-1}$, $\Omega_{\text{m}} = 0.272$, $\Omega_\Lambda = 0.728$, and $n_{\text{s}} = 0.967$ (Komatsu et al. 2011, *WMAP*-7). All magnitudes are in the AB (Oke 1974) system. The initial mass function (IMF) follows Chabrier (2003). Quantities are said 'observed' when they include observational noise (for magnitudes) or when they are measured through SED fitting (redshift, stellar mass, and SFR). If directly derived from the simulation, they are defined as 'intrinsic'.

2 DATA AND METHODS

2.1 Description of the observational surveys

The virtual photometric catalogue from the HORIZON-AGN simulation is built to mimic the COSMOS2015 catalogue. It also includes the photometry expected from the *Euclid* space-based telescope[1] (Abbott et al. 2018), the LSST (LSST Science Collaboration et al.

[1]https://www.euclid-ec.org

2009) and the DES (Laureijs et al. 2011) in terms of filter passbands and depths. We briefly describe hereafter the different configurations investigated in this work to quantify the performances of galaxy redshift and physical property computation. Table 1 provides a summary of these surveys.

2.1.1 The COSMOS field

The COSMOS deep optical and near-infrared (NIR) catalogue (COSMOS2015) described in Laigle et al. (2016, hereafter L16) is used as a reference to test the performances of our estimation of galaxy properties. The catalogue includes more than 1 million objects detected within the 2 deg^2 of the COSMOS field, observed in 30 bands from ultraviolet (UV) to IR (0.25–8 μm). Here, the analysis is restricted to the 'ultradeep' stripes, i.e. four rectangular regions that in COSMOS2015 have been covered with higher NIR sensitivity (in the UltraVISTA-DR2 survey, $K_s < 24.7$, 3σ) than the rest of the area.

COSMOS2015 contains far- and near-UV photometry (FUV and NUV, respectively) from *GALEX* (Zamojski et al. 2007), but only NUV was used for the estimation of photometric redshifts and masses. In the optical, it includes the same u, B, V, r, i, and z data as previous releases from the Canada–Hawaii–France and Subaru telescopes (Capak et al. 2007; Ilbert et al. 2009). This baseline is complemented with Subaru medium- and narrow-band images between 4000 and 8500 Å. In the NIR, Y, J, H, K_s images come from the second data release (DR2) of the UltraVISTA survey (McCracken et al. 2012), and the Y-band image from Subaru/Hyper-Suprime-Cam (HSC, Miyazaki et al. 2012). The *Spitzer* Large Area Survey with HSC (Capak et al. in preparation) provides mid-IR (MIR) coverage with the four IRAC channels centred at 3.6, 4.5, 5.8, and 8.0 μm.

In order to derive photometry coherently across different bands, the point spread function (PSF) in each filter has been rescaled using a Moffat profile modelling. After the PSF homogenization, fluxes are extracted within fixed apertures of 3 arcsec using SEXTRACTOR (Bertin & Arnouts 1996) in dual-image mode. Following a reduction procedure similar to that described in McCracken et al. (2012), the detection image is a χ-squared sum of the four NIR images of UltraVISTA DR2 and the z^{++} band. *Spitzer* sources are extracted by means of the code IRACLEAN (Hsieh et al. 2012).

Estimates obtained through SED fitting (photometric redshift, stellar mass, and other physical quantities) are also provided for each entry of the catalogue. The method adopted to obtain these estimates is described in Section 2.4, where the same technique is applied to simulated galaxies. Further details about the COSMOS2015 catalogue can be found in L16.

2.1.2 Future surveys: Euclid *and* LSST

To compute galaxy properties from SED fitting in comparable conditions to *Euclid* and LSST, HORIZON-AGN galaxies are also

post-processed to get the photometry in *Euclid*, LSST, and DES filters with depths similar to the ones expected for these surveys. We note that sometimes the expected depths from the literature are provided for point sources, and as a consequence give generally too optimistic estimators of the limiting magnitudes of extended sources. Our adopted limiting magnitudes are therefore probably not exactly the ones which will be obtained in the future, but they none the less reflect the relative depths of these upcoming surveys. The photometric baselines are detailed below, and the adopted limiting magnitudes in all bands are summarized in online Table A1.

Euclid + DES configuration: Euclid will provide photometry in one broad-band optical (*riz* filter) and three NIR filters (*Y, J, H*) with expected depths at completion of 24.5 (10σ, extended sources) in the optical and 24.0 (5σ, point sources) in the NIR bands. This broad-band baseline alone is not sufficient to compute photometric redshifts with a high enough accuracy (especially to constrain the Balmer break in the optical), therefore it has to be complemented with ground-based optical photometry (see e.g. Sorba & Sawicki 2011). In particular, DES provides photometry over 5000 deg² in the Southern sky in *g, r, i,* and *z* with depth of 24.33, 24.08, 23.44, and 22.69 (10σ, extended sources, Abbott et al. 2018), which matches the *Euclid* requirements (Laureijs et al. 2011). DES photometry provides a finer sampling of the optical range than the *Euclid riz* filter alone. Collaboration between *Euclid* and DES is planned. Therefore in the current work, we explore what would be the expected performance of such a configuration.

LSST configuration: the survey conducted on LSST will provide photometry in the optical over 30 000 deg². LSST single visit depth should reach 24.5 in *r* (5σ, point sources), and the co-added survey depth should reach 26.3, 27.5, 27.7, 27.0, 26.2, and 24.9 (5σ, point sources) in *u, g, r, i, z,* and *y* bands, respectively (LSST Science Collaboration et al. 2009). For weak-lensing studies, the 'gold' sample of LSST galaxies with a high *S/N* is defined with a magnitude cut *i* < 25.3. We use this cut in the present work when studying the LSST-like configuration.

In the Southern sky, *Euclid* and LSST will overlap over at least 7000 deg². It is therefore natural to explore the possible gain to combine both data sets. To this end, we also analyse HORIZON-AGN galaxies in the *Euclid* + LSST-like configuration.

2.2 The HORIZON-AGN simulation

This study relies on HORIZON-AGN[2] (Dubois et al. 2014), a cosmological hydrodynamical simulation in overall fairly good agreement with observations, in the redshift and mass regime of the present analysis (see Kaviraj et al. 2017).

The simulation box, run with the RAMSES code (Teyssier 2002), is $L_{box} = 100\,h^{-1}$ Mpc on a side, and the volume contains 1024^3 dark matter (DM) particles, corresponding to a DM mass resolution of $8 \times 10^7\,M_\odot$. The initially coarse 1024^3 grid is adaptively refined down to 1 physical kpc. The refinement procedure leads to a typical number of 6.5×10^9 gas resolution elements (leaf cells) in the HORIZON-AGN simulation at $z = 1$.

Heating of the gas from a uniform UV background takes place after redshift $z_{reion} = 10$, following Haardt & Madau (1996). Gas can cool down to 10^4 K through H and He collision and with a contribution from metals that follows the rates tabulated in Sutherland & Dopita (1993). Star formation occurs in regions

where gas number density is above $n_0 = 0.1\,\mathrm{H\,cm^{-3}}$, following a Schmidt law: $\dot\rho_* = \epsilon_* \rho_g / t_{ff}$, where $\dot\rho_*$ is the SFR mass density, ρ_g the gas mass density, $\epsilon_* = 0.02$ the constant star formation efficiency, and t_{ff} the gas local free-fall time. Feedback from stellar winds and supernova (both type Ia and II) are included into the simulation with mass, energy, and metal releases. Galactic black hole formation is also implemented in HORIZON-AGN, with accretion efficiency tuned to match the black hole–galaxy scaling relations at $z = 0$. Black hole energy is released in either quasar or radio mode depending on the accretion rate (see Dubois et al. 2012, for more details).

The light-cone has been extracted on-the-fly as described in Pichon et al. (2010). For the light-cone extraction, gas leaf cells were replaced by gas particles, and treated as the stars and DM particles. All particles were extracted at each coarse time-step according to their proper distance to the observer at the origin. In total, the light-cone contains about 22 000 portions of concentric shells. The light-cone projected area is 5 deg² below $z = 1$, and 1 deg² above. However, we restrict ourselves to 1 deg² over the whole redshift range considered in this study. The full light-cone up to $z = 4$ contains about 19 replica of the HORIZON-AGN box.

2.3 Generating a mock photometric catalogue

2.3.1 Galaxy extraction

The ADAPTAHOP halo finder (Aubert, Pichon & Colombi 2004) is run on the light-cone over $0 < z < 4$ to identify galaxies from the stellar particles distribution. Local stellar particle density is computed from the 20 nearest neighbours, and structures are selected with a density threshold equal to 178 times the average matter density at that redshift. Galaxies resulting in less than 50 particles ($\simeq 10^8\,M_\odot$) are not included in the catalogue. Since the identification technique is redshift dependent, ADAPTAHOP is run iteratively on thin light-cone slices (about 4000 slices up to $z = 4$) of few comoving Mpc (cMpc). Slices are overlapping to avoid edge effects (i.e. cutting galaxies in the extraction) and duplicates are removed.

2.3.2 Galaxy SED computation and dust attenuation

Although the simulation assumes a Salpeter IMF (Salpeter 1955) to model stellar mass losses, we have decided to post-process it with a Chabrier IMF (Chabrier 2003). The choice of the IMF is significant, as it controls both the stellar mass-loss prescription and the overall mass-to-light ratio. The Chabrier IMF brings the simulated galaxy counts in much better agreement with the COSMOS2015 galaxy counts (see online Appendix A. Magnitudes are ~0.4 mag fainter with a Salpeter IMF compared to Chabrier's). For each galaxy, each stellar particle is linked to a single stellar population (SSP) obtained with the stellar population synthesis model of Bruzual & Charlot (2003, hereafter BC03). Because stellar particle ages and metallicities vary on a much finer grid than the BC03 models, an interpolation is carried out between SSPs to reproduce the desired values. In addition, each SSP is also rescaled to match the initial stellar mass of the particle, in order to follow the same mass loss fraction for the simulated galaxies as in BC03 (see online Appendix A4). This rescaling is essential to avoid discrepancies between the intrinsic and computed galaxy properties coming from the different SSP prescriptions, which is out of topic for the present work. It should be noted that the metallicity of stellar particles in HORIZON-AGN has been boosted by a empirically computed factor, to match the observed mass–metallicity relation. This factor f_Z is redshift dependent as follows: $f_Z = 4.08430 - 0.213574z - 0.111197z^2$ (see Kaviraj et al. 2017, for more details).

Dust attenuation is also modelled for each star particle using the gas metal mass distribution as a proxy for the dust distribution. Gas metal mass is evaluated in a cube of 138 comoving kpc³ around each galaxy and a constant dust-to-metal mass ratio is adopted. Although a mesh-based code is used to run HORIZON-AGN and in particular to follow the gas distribution, gas cells have been turned in particles when extracting the light-cone. In order to get a smooth metal field around the galaxy from this gas particle distribution, a Delaunay tessellation is computed on the particles to avoid cells with null values in underdense regions, and then interpolated on a regular grid with a resolution of ∼1 ckpc. The dust column density and the optical depth along the line of sight are computed for each stellar particle in the galaxy using the $R_V = 3.1$ Milky Way dust grain model by Weingartner & Draine (2001). Further details on the dust computation and dust-to-metal mass ratio calibration are given in online Appendix A. This dust attenuation model only takes into account absorption and does not include scattering. While this is not a problem in the rest-frame optical and NIR, the impact in the rest-frame UV is non-negligible (Kaviraj et al. 2017). This effect is not corrected in our catalogue and as a result, our galaxies are up to 0.8 mag brighter without scattering in the UV part of the spectrum. In fact, including scattering would have a similar impact as a steepening of the dust attenuation law in the UV. A dust-free version of the catalogue is also produced in order to isolate the impact of dust attenuation on the computation of galaxy physical properties.

Flux contamination by nebular emission lines is not included in our virtual photometry. Consistently, emission-line parametrization is also turned off in the SED-fitting computation. In real surveys, emission lines can help determining the photometric redshifts. On the other hand, the dispersion in the emission-line ratios is poorly modelled by the SED-fitting code and can bias the redshift estimation.

Finally, we stress that we do not model extinction by the Milky way. Photometry from observed survey is generally corrected from Galactic extinction using galactic reddening maps (e.g. Schlegel, Finkbeiner & Davis 1998). However, the amount of absorption in a given band will depend on the source SED. As shown in Galametz et al. (2017), the bandpass extinction can vary by up to 20 per cent from the average correction depending on the SED. The discrepancy between the effective extinction and the average correction of the photometry can potentially lead to additional systematics which are therefore not accounted for here.

Note that HORIZON-AGN only reproduces global mass assembly up to some point. Although the simulation broadly matches the mass function evolution with redshift and the SFR main sequence (Kaviraj et al. 2017), at low mass ($\log M_*/M_\odot > 9.5$), the mass function is systematically overestimated given the present sub-grid stellar feedback and star formation recipes. Conversely, at high redshift ($z > 4$), it is also underestimated because of limited spatial resolution. To be conservative, forecasts were therefore limited within the redshift and mass range where the simulation is reliable. Online Appendix A5 lists the limitations of our modelling.

2.3.3 IGM absorption

Prior to convolving the galaxy spectrum with photometric filters, attenuation by the IGM must be implemented. The knowledge of the gas distribution in the light-cone allows us to consistently implement IGM absorption, while accounting for variation from one sight-line to the other. Conversely, it is globally accounted for with an analytical prescription at the SED-fitting stage. Therefore,

our photometric catalogue allows to test the effect of the inhomogeneous IGM attenuation on photometric redshifts and galaxy properties. The focus is on HI absorption in the Lyman-series (hence neglecting metal lines, such as e.g. the CIV forest). Details about the implementation of the attenuation on the line of sight of each galaxy is given in online Appendix A. An IGM-free version of the catalogue is also produced in order to isolate the impact of IGM absorption.

2.3.4 Photometry extraction and error implementation

Eventually, the integrated galaxy spectra are computed between 91 and 16×10^5 Å (with 1221 wavelength points) by adding all the SSPs of a given galaxy together. Note that IR dust emission is not computed because not used here. Apparent total magnitudes are obtained by convolving the spectrum (redshifted to the intrinsic redshifts of the galaxies) with the same filter set as used in COSMOS2015 (u, B, V, r, i^+, z^{++}, Y, J, H, K_s, [3.6 μm], [4.5 μm], and the 14 intermediate and narrow bands[3]), *Euclid*, DES, and LSST (see Section 2.1).

Photometric errors are added in each band to reproduce the S/N distribution and sensitivity limit of COSMOS2015 and DES, and the ones expected for *Euclid* and LSST (see Table 1 and online Appendix A). It should be emphasized here that the photometry has been derived from the entire distribution of star particles and not through a realistic flux extraction from images (as it is done in real observations via tools like SEXTRACTOR, Bertin & Arnouts 1996). Consequently, our mock catalogue does not include all the associated photometric issues including potential systematic effects like blending, object fragmentation, imperfect background sky subtraction, and PSF homogenization and offsets due to the rescaling of fixed aperture to total fluxes, which will be the topic of a future work. Photometric errors as implemented in the current catalogue simply correspond to Gaussian noise with standard deviation depending on the galaxy flux and the depth of the surveys.

2.4 SED fitting: method

2.4.1 Photometric redshifts

Photometric redshifts (z_{phot}) are computed using the code LEPHARE (Arnouts et al. 2002; Ilbert et al. 2006) with a configuration similar to Ilbert et al. (2013). The SED library includes spiral and elliptical galaxies from Polletta et al. (2007), along with bluer templates of young star-forming galaxies built by means of the BC03 model. Dust extinction is added to the templates according to one of the following attenuation curves: Prevot et al. (1984), Calzetti et al. (2000), or a modified version of Calzetti et al. (2000) with the addition of the 'graphite bump' at ∼2 175 Å (e.g. Fischera & Dopita 2011; Ilbert et al. 2009). The $E(B - V)$ values range from 0 to 0.5. IGM absorption is implemented following the analytical correction of Madau (1995).

As strong nebular emission (such as [OII] or Hα lines) can significantly increase the flux measured in a photometric filter, nebular emission lines were considered in the z_{phot} computation of real data (L16). On the other hand, such options are disabled

---

[3]At lower redshift, *GALEX* FUV and NUV filters are relevant, as they bring information about young stellar populations. However, it is more difficult to reproduce their flux extraction and the related uncertainties. Therefore, NUV and FUV are excluded from the mock catalogue.

when LEPHARE is run on HORIZON-AGN galaxies, since nebular emission lines are not modelled for them (Section 2.3).

Each template in the LEPHARE library is fit to the virtual photometry of the HORIZON-AGN galaxies. The code computes the goodness of fit (χ^2) for each redshift solution and their likelihood ($\mathcal{L}$). The z_{phot} estimate for a given galaxy is defined as the median of the $\mathcal{L}(z)$ distribution, and the 1σ uncertainty (σ_z) is the interval enclosing 68 per cent of its area (see Section 3.1.2). In order to improve the SED-fitting performance, a mild luminosity prior is also applied to reduce the fraction of catastrophic outliers. This prior is based on the observed luminosity function in the rest-frame *B* band (e.g. Ilbert et al. 2006; Zucca et al. 2009; López-Sanjuan et al. 2017). Given the extremely low number density expected in the (extrapolated) bright end of the luminosity function, our codes excludes solutions with absolute magnitude $M_B > -24.2$. No other prior on the redshift distribution is applied in LEPHARE.

As done in L16, fluxes rather than magnitudes are used when running LEPHARE. This allows to deal robustly with faint or non-detected objects.

Systematic offsets: with real data sets, an important aspect of the z_{phot} computation is the derivation of systematic offsets which are applied to match the predicted magnitudes and the observed ones (Ilbert et al. 2006) based on the spectroscopic sub-sample. This calibration is designed to empirically correct both for the incomplete template library and possible systematics in the galaxy magnitude extraction. However, this calibration might be biased because it relies on a spectroscopic subsample. As described in online Appendix C, we test the computation of systematic offsets in the COSMOS-like catalogue by using a sub-sample of galaxies matching the spectroscopic catalogue on COSMOS. We find that there is no need for this calibration in the simulated catalogue. The offsets introduced by an imperfect knowledge of the templates are therefore negligible in HORIZON-AGN.[4]

2.4.2 Stellar mass and star formation rate

Stellar mass and SFR are then derived using another template library built by using the BC03 model, in the same way as for the COSMOS2015 catalogue. In this second run, similarly to what was done in COSMOS2015 for computational reasons, only two extinction laws are used (Arnouts et al. 2013 and Calzetti et al. 2000). The IMF is assumed to be Chabrier's (2003), and the stellar metallicity of each template to be either solar ($Z_\odot$) or subsolar ($0.4 Z_\odot$).

In our library, the SFHs used to build the SEDs are parametrized with an analytic equation. It can be exponentially declining, i.e. $SFR(t) \propto e^{-t/\tau}$. An alternative definition is the 'delayed' SFH, more suitable to model a galaxy with gas infall: $SFR(t) \propto e^{-t/\tau} t/\tau$. In both cases, τ is the *e*-folding time, varying between one-tenth and several Gyr.[5] In the latter case, τ also represents the galaxy age (since its formation) at which the SFR peaks. In particular, such SFHs cannot reproduce multiple bursts of star formation. The SFH models start forming stars from $t = 0$. When building the library of

templates, *t* is sampled from a few hundreds Myr to the age of the Universe at a given redshift, with up to 44 steps at $z = 0$.

3 SED-FITTING PERFORMANCE: PRESENT SURVEYS

3.1 Photometric redshifts

Let us first investigate our ability to recover photometric redshifts from the photometry, using the simulated COSMOS-like catalogue.

3.1.1 Comparison between z_{sim} and z_{phot}

The accuracy of our SED-fitting method can be first tested by comparing galaxy redshifts in the light-cone (z_{sim}) to those obtained by LEPHARE.[6] Fig. 1 (left-hand panel) presents such a comparison for galaxies with $K_s < 24.7$. In addition to the magnitude cut, pathological cases are also excluded with reduced chi-square values $\chi^2_{red} > 10$, which represent <0.1 per cent of the whole sample (namely, 387 out of 541 555 K_s-selected galaxies). This kind of χ^2 selection is also applied in real surveys and removes similar fractions of problematic objects (Davidzon et al. 2017; Caputi et al. 2015). Overall, we do not find significant systematics affecting LEPHARE redshift estimates. Despite the simplistic implementation of dust extinction, the limited number of templates and the fact that they have been calibrated to represent the *observed* universe, our recipe captures the main features of the simulated galaxy SEDs, and recovers their redshifts with a precision comparable to what is achieved with the COSMOS2015 catalogue. In our simulation, the normalized median absolute deviation (NMAD, Hoaglin, Mosteller & Tukey 1983) for the entire sample is $1.48 \times \text{median}(|\Delta z|)/(1 + z_{sim}) = 0.031$. The fraction of outliers, defined as objects with $|\Delta z| > 0.15 (1 + z_{sim})$, is $\eta = 4.6$ per cent (Fig. 1, left-hand panel). In the real survey, with the same cut at $K_s < 24.7$, a comparison between photometric and spectroscopic redshifts results yields $NMAD = 0.013$ and $\eta = 3.1$ per cent, respectively. Such difference between simulation and observation is explained because the spectroscopic samples are not representative (as detailed below).

Specific sub-samples of galaxies may show a smaller scatter than the global one. Galaxies' z_{phot} precision depends on their S/N, which in turn correlates with apparent magnitude, therefore it is useful to estimate the NMAD as a function of the latter. Fig. 1 (right-hand panel) presents the z_{phot} versus z_{spec} diagram after dividing HORIZON-AGN galaxies in four i^+-band magnitude bins. NMAD and η of each sub-sample are reported in Table 2, along with the corresponding metrics obtained in L16 for COSMOS2015. The precision of HORIZON-AGN z_{phot} is comparable to that found in the real survey. However the fraction of catastrophic failures is systematically larger in COSMOS2015. Such a discrepancy can be explained by observational uncertainties related to image source extraction (e.g. confusion noise) not considered in HORIZON-AGN. In addition, it should be remembered that the observed η and NMAD are provided only for a sub-sample of high-confidence spectroscopic galaxies, which are generally biased towards bright objects.

Comparison with a zCOSMOS-like sub-sample: the COSMOS2015 z_{phot} precision was assessed by using data collected during various spectroscopic campaigns (see table 5 of L16). One of the most important is the zCOSMOS survey, accounting for almost half of

---

[4]In fact, the virtual photometry of the simulated galaxies presents less diversity than in real data sets, because it is built by the mean of BC03 SSP models for all the galaxies, which might be a reason why these offsets are negligible.

[5]The following time-steps are used: $\tau = 0.1, 0.3, 1, 3, 4,$ and 30 Gyr for the exponentially declining SFHs, and $\tau = 1$ and 3 Gyr for the delayed SFHs.
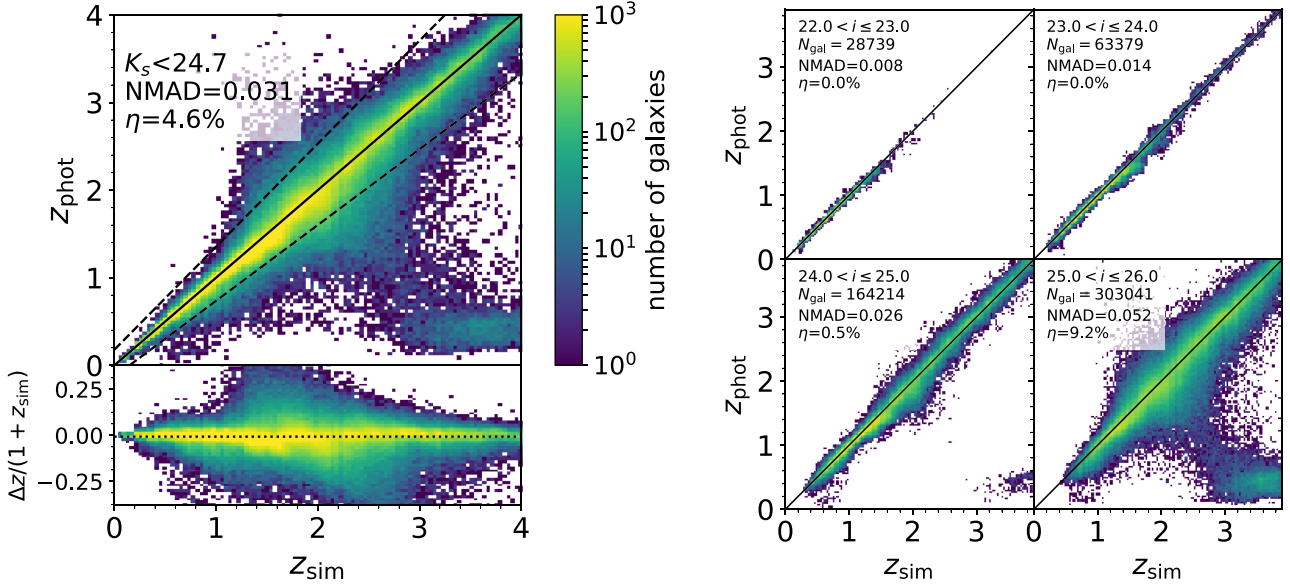
[6]z_{sim} includes galaxy peculiar velocity.

Figure 1. Left: comparison between galaxy redshifts in the HORIZON-AGN light-cone (z_{sim}) and the photometric redshifts recovered by LEPHARE (z_{phot}) for the whole sample of 506 361 objects selected at $K_s < 24.7$. The z_{phot} uncertainty computed as NMAD is shown along with the catastrophic error fraction (η, see Section 3.1). Solid line is the 1:1 bisector, while dashed lines mark the $\pm 0.15 (1 + z)$ threshold used to compute η. In the bottom panel, Δz is defined as $z_{phot} - z_{spec}$. Right: comparison between z_{sim} and z_{phot} as a function of apparent magnitude in the $i*$ band (same colour scale as in the left-hand panel). The number of galaxies per magnitude bin is shown in each panel along with NMAD uncertainty and catastrophic error fraction, and is also reported in Table 2.

Table 2. Statistical errors (NMAD) andpercentage of catastrophic errors (η) in different i^+ magnitude bins. Results for COSMOS2015 galaxies (L16) are compared to the outcome of our simulation.

i^+	COSMOS2015		Hz-AGN with IB		Hz-AGN without IB	
mag	NMAD	η (per cent)	NMAD	η (per cent)	NMAD	η (per cent)
(22,23]	0.010	1.7	0.008	0.0	0.023	0.0
(23,24]	0.022	6.7	0.014	0.0	0.028	0.1
(24,25]	0.034	10.2	0.026	0.5	0.037	0.8
(25,26]	0.057	22.0	0.052	9.2	0.065	11.7

high-quality galaxy spectra at $z < 3$ (Lilly et al. 2007).[7] To reproduce a similar subset in HORIZON-AGN, a sub-sample of galaxies in the light-cone is identified using selection criteria similar to zCOSMOS. We randomly extract mock galaxies at $i^+ < 22.5$ until we match both the magnitude and redshift distributions of the zCOSMOS-Bright sample (Lilly et al. 2007), and do the same at $B < 25$ to mimic the zCOSMOS-Deep (Lilly et al. in preparation). The zCOSMOS-Deep selection aimed at enforcing a target selection at $z > 1.5$, but some faint galaxies at lower redshift were also observed. Yet, those interlopers are not included in the zCOSMOS-Deep mock sample by applying a sharp $1.5 < z < 2.5$ cut. It has to be noted that a large fraction of the catastrophic failures seen at $z_{spec} < 1$ in the real data correspond indeed to the selection that is not replicated in the simulation. The result is a set of about 5000 galaxies for which

pseudo-spectroscopic measurements are created by perturbing z_{sim} with a Gaussian random error having $\sigma = 0.0004 (1 + z)$.[8]

Fig. 2 shows the z_{phot} versus z_{spec} comparison in HORIZON-AGN (upper panel) and COSMOS2015 (lower panel). Since stars are not included in our simulation the analysis is restricted to $z > 0.1$ to avoid both stellar interlopers present in the real survey and bright COSMOS2015 galaxies erroneously classified as stars. With such a zCOSMOS-like selection, the NMAD measured in HORIZON-AGN is in excellent agreement with COSMOS2015 (0.0080 and 0.0081, respectively). On the other hand, catastrophic errors are more numerous in the real sample, as also found in the previous test (see Table 2). Online Appendix D2, focuses on the 22 outliers with $|\Delta z| > 0.3 (1 + z_{spec})$ to understand this difference. In most of the case, the failure arises either because of uncertain photometry (fragmented or blended objects) or spectroscopic misidentification.

Impact of IGM absorption: the HORIZON-AGN light-cone allows us to quantify the importance of correctly accounting for IGM absorption, by comparing the z_{phot} estimate computed from the IGM and IGM-free (i.e. turning off IGM both in the photometry computation and at the SED-fitting stage) versions of the catalogue. This test is carried out with the dust-free version of the catalogue. In HORIZON-AGN, IGM absorption is implemented along the line of sight of each galaxy knowing the foreground HI distribution, while in LEPHARE, absorption due to the intervening IGM between the galaxy and the observer is taken into account by applying an *average* correction as a function of redshift based on an analytical relation

---

[7]The full description of the zCOSMOS-Deep sample characteristics and the evaluation of the redshift estimation performance will be published in a future paper (Lilly et al. in preparation).

[8]The standard deviation of the Gaussian random error corresponds to the 1σ uncertainty of zCOSMOS-Bright galaxy estimated by repeated measurements (Lilly et al. 2007). Given the order of magnitude of z_{phot} error, that pseudo-spectroscopic perturbation is negligible in the following analysis.
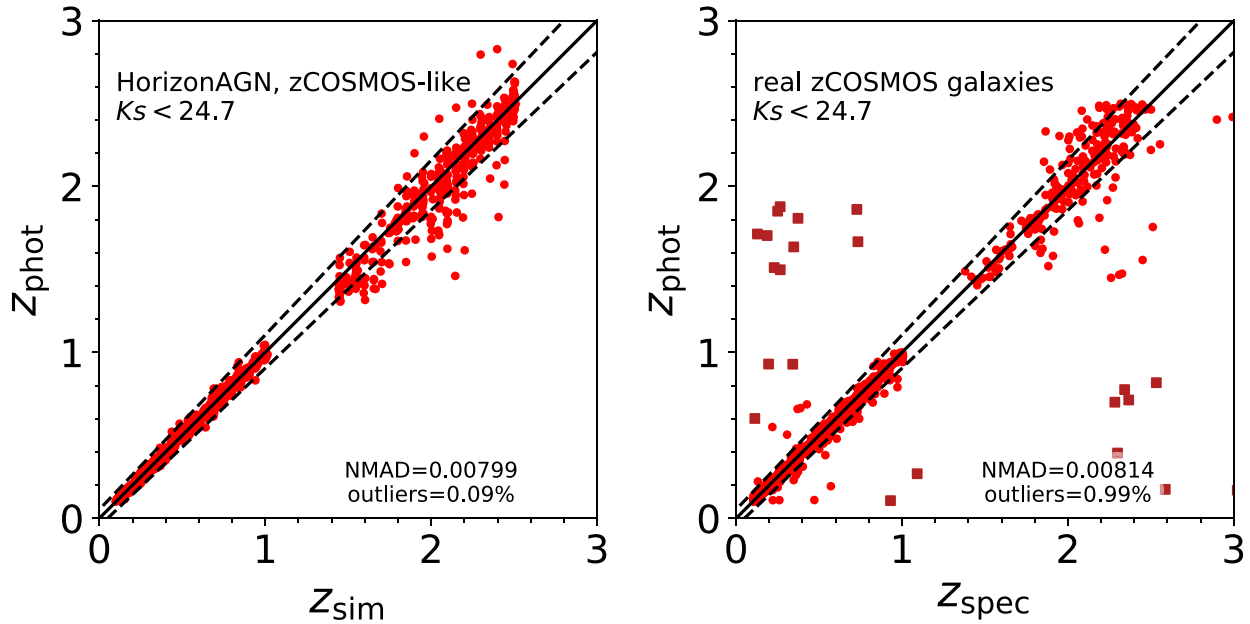
Figure 2. Left: comparison between redshift of HORIZON-AGN galaxies (red dots) versus photometric redshifts from LEPHARE. Solid and dashed lines show 1:1 relation and $\pm0.05(1+z)$ deviation. The gap at $z \sim 1.5$ is due to the different selection functions of zCOSMOS-Bright and Deep. Right: same as left-hand panel, but for real COSMOS2015 galaxies having photometric redshift computed by LEPHARE and spectroscopic redshifts from zCOSMOS-Bright and Deep survey (red circles and squares). Squares indicate catastrophic failures with $|\Delta z| > 0.3\,(1 + z_{spec})$, a class of objects that is not present in the comparison using the mock sample (see Section 3.1.1).

(Madau 1995)[9]. Overestimating the IGM correction (or neglecting the line-of-sight variability of IGM opacity) might impact the performance of z_{phot} estimate for distant galaxies (as suggested in Thomas et al. 2017). IGM absorption plays a role mostly at $z > 2$, with a more dramatic attenuation of galaxy photometry from $z \sim 3$ as illustrated by online Fig. A2.

The global redshift accuracy estimated by the NMAD is impacted at a below per cent level, only at the very faint end of the galaxy population. Interestingly, implementing IGM absorption slightly helps constraining the redshift of faint galaxies, when averaging it over the entire redshift range (In the bin $25 < i < 26$, NMAD = 0.049 and $\eta = 8.7$ per cent in the IGM-free version, while NMAD = 0.045 and $\eta = 7.6$ per cent with IGM). However, at $z > 3$, the fraction of outliers populating the clump located at $[z_{sim} > 2.5] \cap [z_{obs} < 1.5]$ strongly increases when IGM is included: at $24 < i < 25$, only 7 per cent of the existing outliers populate this region in the no-IGM case, while they are 58 per cent with IGM. At $25 < i < 26$, in this region, they are 16 per cent and 47 per cent in the no-IGM and IGM cases, respectively. This population of outliers also occurs when fitting observed population of galaxies (see fig. 11 in L16) and our work suggests therefore that the way IGM is accounted for is important to mitigate it.

Impact of medium bands: let us now quantify the improvement of z_{phot} estimates due to medium-band photometry. The inclusion of these bands in a deep extragalactic survey is very useful to better constrain the redshift from spectral features occurring in the optical wavelength range (Lyman and Balmer breaks depending on the

redshift, nebular emission lines in real data sets) but expensive in exposure time. It is therefore important to check whether it is worthwhile. Even though the major advantage of those filters is to find contribution from nebular emission lines (not implemented in our virtual magnitudes), they should also in principle help to constrain the galaxy continuum. Indeed, when medium-band filters are removed, the z_{phot} precision degrades considerably (see Table 2). For instance, the NMAD of bright objects ($22 < i < 23$) is degraded by almost a factor 3, going from NMAD = 0.008 (when medium bands are included) to 0.023. At the fainter magnitudes ($24 < i < 26$), the difference is less remarkable, but none the less the absence of medium-bands results in a z_{phot} scatter larger by 30–40 per cent. The outlier fraction, especially at $i > 25$, also increases (see Table 1).

3.1.2 Photometric redshift errors from marginalized likelihood

When working on extragalactic surveys, an accurate knowledge of the redshift probability distribution function (PDF) is instrumental to remove observational uncertainties in galaxy statistics. For instance, the PDFs of a certain set of galaxies can be used to correct their luminosity function for the so-called Eddington bias (see Schmidt et al. 2014). Therefore, it is crucial to verify whether the SED fitting produces a reliable PDF(z) for a given galaxy. In a Bayesian framework, this PDF is the *posterior* probability distribution, proportional to the product of the *prior distribution* and the marginalized *likelihood* $\mathcal{L}(z)$. From $\mathcal{L}(z)$,[10] LEPHARE computes $\sigma_{z,\,fit}$, namely the photometric redshift 1σ error. This is defined as the redshift interval, centred at z_{phot}, that encloses 68.27 per cent of the $\mathcal{L}(z)$ area.

---

[9]As explained in online Appendix A, there is a slight discrepancy between the average IGM absorption in HORIZON-AGN and the correction implemented in LEPHARE, the latter being stronger than the former. However the present work does not aim at correcting this discrepancy as it is likely to also happen when fitting the photometry of real galaxies.

[10]$\mathcal{L} \propto \exp(-^{1/2}\chi^2)$, with $\chi^2(z) = \sum_{filters\,i}(F_{obs\,i} - F_{SED\,i}(z, T))^2/\sigma_{i,obs}^2$, where $F_{SED\,i}(z, T)$ is the flux predicted for a template T in the filter i at z, $F_{obs\,i}$ is the observed flux in the filter i and σ_{obs} is the associated uncertainty.
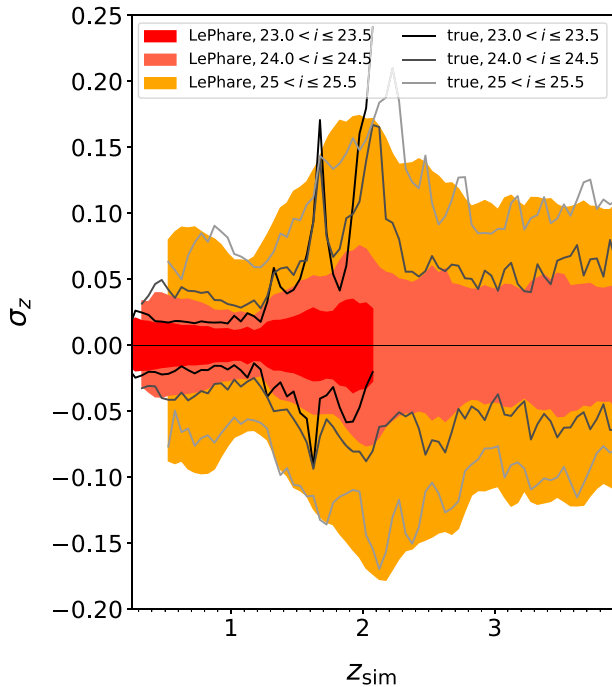
Figure 3. Photometric redshift 1σ uncertainties in the HORIZON-AGN simulation, as a function of z_{sim} and divided in three bins of i^+-band magnitude. For each of these bins, the shaded area is the average $\sigma_z(z)$ error interval as it results from the SED-fitting likelihood analysis in LEPHARE ($\sigma_{z,\,fit}$ described in Section 3.1.2). Solid lines shows an alternate estimate of σ_z, directly retrieved from the scatter between photometric and true redshifts ($\sigma_{z,\,true}$). Both computations stop at redshifts where the statistics becomes too low (<20 galaxies).

Reliability of redshift 1σ errors from SED fitting: the simulation provides the true z_{phot} uncertainties directly from the difference between SED-fitting estimates and z_{sim}. Hence, it can be checked whether the error bars provided by LEPHARE actually represent the redshift 1σ uncertainty. Previous tests with spectroscopic data suggest that they could be underestimated (see e.g. Dahlen et al. 2013, L16).

Fig. 3 shows the evolution of the median of $\sigma_{z,\,fit}$ as a function of z_{sim}, for galaxies in three different bins of i^+-band magnitude. From these bins galaxies with degenerate redshift solutions are excluded, i.e. whose $\mathcal{L}(z)$ function shows two peaks.[11] In such a case, the integrated 68.27 per cent of the $\mathcal{L}(z)$ area is strongly skewed towards the secondary solution and the $\sigma_{z,\,fit}$ does not represent the pure statistical error but also includes systematics. Nevertheless, if those galaxies are re-introduced in the sample, results shown in Fig. 3 remain the same at $i^+ < 25$ and change less than 20 per cent in the faintest bin.

We first find that $\sigma_{z,\,fit}$ values increases with magnitude, as shown in Fig. 3. Such behaviour is expected since fainter galaxies have a lower S/N ratio and a lower constraint on the SED fit. We also find a redshift dependency of the uncertainties, mainly due to the different efficiency of optical and NIR photometry: the former being deeper, including medium bands and can tightly constrain the Balmer break at $z < 1.3$; whereas at $z > 1.5$, the break is entirely shifted in the

NIR regime, which is sampled with fewer, less sensitive bands. At $z > 2.5$ the $\sigma_{z,\,fit}$ amplitude slightly decreases as optical blue bands start to constrain the Lyman break position. Overall, LEPHARE predicts a symmetric scatter, with upper and lower errors such that $\sigma_{z,\,fit}^+ \simeq -\sigma_{z,\,fit}^-$.

To establish whether the uncertainties derived by LEPHARE are reliable, they are compared to the 1σ errors directly retrieved from the simulation ($\sigma_{z,\,true}$). Using the same i^+-selected galaxies for which $\sigma_{z,\,fit}$ was computed, we measure $\sigma_{z,\,true}$ by means of their $\Delta z = z_{sim} - z_{phot}$ distribution, finding the interval that includes 68.27 per cent of it. Despite some noise, relatively good agreement is found between $\sigma_{z,\,fit}$ and $\sigma_{z,\,true}$, an indication that the uncertainty provided by LEPHARE is generally a good proxy of the actual 1σ error dispersion (Fig. 3). However at bright magnitudes ($i^+ < 24.5$) and for $1 < z_{sim} < 2.5$, $\sigma_{z,\,fit}$ is generally underestimated compared to $\sigma_{z,\,true}$. As shown in online Appendix D1, this underestimation might be due either to the underestimation of photometric errors, or to a lack of representativeness of the set of templates for this galaxy population, making $\mathcal{L}(z)$ too spiky around the median z. A similar trend was already discussed in L16. From a comparison with spectroscopic redshifts (their fig. 13), L16 suggested indeed that the 1σ uncertainties produced by LEPHARE were underestimated. As a consequence, the authors proposed a magnitude-dependent boosting factor (f_σ) that would enlarge $\sigma_{z,\,L16}$ so that $\sim$68 per cent of the COSMOS2015 z_{spec} would fall within $\sigma_{z,\,L16} \times f_\sigma$ from z_{phot}. This boosting factor[12] was however constant with redshift and increasing with magnitudes. From our analysis, it appears that this factor would generally overcorrect the z_{phot} errors at faint magnitudes, but particular care should be given to the errors of bright galaxies at $1 < z < 2.5$.

Finally, note that the global behaviour of $\sigma_{z,\,fit}$ as a function of magnitude and redshift depends on the photometric baseline available, and does not necessarily hold for different configurations (see e.g. Section 4).

z_{phot} error comparison between COSMOS-like and COSMOS2015: let us now compare the HORIZON-AGN redshift uncertainties with those of the COSMOS2015 galaxies ($\sigma_{z,\,L16}$) as calculated in L16 by means of LEPHARE. The method is the same as that applied to simulated galaxies, based on the marginalized $\mathcal{L}(z)$. Fig. 4 shows the median $\sigma_{z,\,L16}$ as a function of redshift, for galaxies with $23.0 < i^+ < 23.5$ and $25.0 < i^+ < 25.5$. In the figure $\sigma_{z,\,fit}$ is also reported, in bins of z_{phot} instead of z_{sim} to allow the comparison with real data.

The trend of $\sigma_{z,\,L16}$ and $\sigma_{z,\,fit}$ are remarkably similar, both showing an increase between $z = 1.5$ and 2, as discussed above. However, COSMOS2015 galaxies have a median σ_z that is about 50 per cent larger than HORIZON-AGN. This difference is likely to be driven either by simplifications in the modelling of the photometry itself, or by failures in the photometry extraction of real data which are not modelled in the simulated catalogue. On the one hand, the simulated photometry includes indeed less variety than realistic galaxies: we use a single and constant IMF, a constant dust-to-metal ratio, single SSP model for stellar mass losses, and emission lines are not modelled. These simplifications can naturally reduce the scatter of the z_{phot} estimate compared to the observed catalogue. On the other hand, although magnitude errors in the simulated catalogue are implemented consistently with COSMOS2015, 'catastrophic detections' (such that blended or

---

[11]LEPHARE automatically identifies a galaxy with two acceptable redshift solutions when the secondary peak includes >2 per cent of the integrated $\mathcal{L}(z)$ distribution.

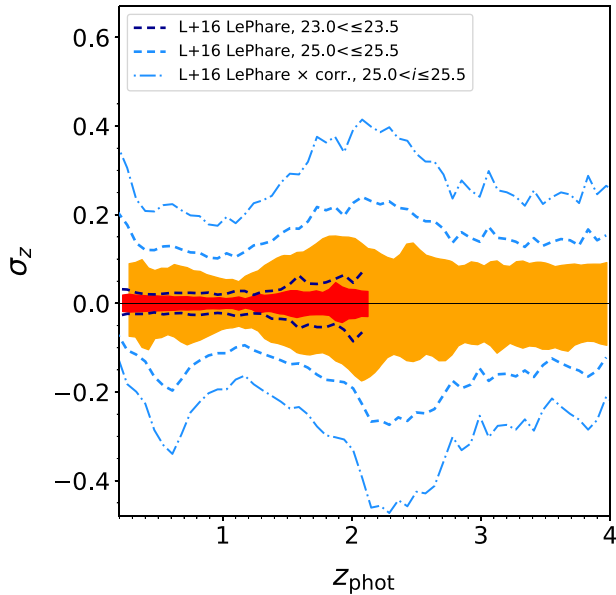[12]For galaxies with $i^+ > 20$, $f_\sigma = 0.1 \times i^+ - 0.8$; $f_\sigma = 1.2$ otherwise.

Figure 4. Median photometric redshift 1σ uncertainty of HORIZON-AGN galaxies at $23 < i^+ \leq 23.5$ and $25 < i^+ \leq 25.5$, as a function of z_{phot} (shaded areas, colours as in Fig. 3). Dashed lines show the median $\sigma_{z,\mathrm{L16}}$ of real COSMOS2015 galaxies in the same magnitude bins (dark blue: $23 < i^+ \leq 23.5$, and light blue: $25 < i^+ \leq 25.5$). Both HORIZON-AGN and COSMOS2015 errors are computed by LEPHARE using the z_{phot} likelihood function. The dotted–dashed lines embrace the median of COSMOS2015 *enhanced* errors at $25 < i^+ \leq 25.5$, i.e. the original $\sigma_{z,\mathrm{L16}}$ values have been increased by a multiplicative 'boosting factor' as prescribed in L16. See Section 3.1.2 for more details. Note that the y-axis range is different from Fig. 3.

fragmented objects) and systematics (astrometry calibration issues, background removal, lack of modelling of the PSF variation within the field, miscentring of the galaxies, etc.) are not considered in the virtual photometry. These errors will propagate in the z_{phot} errors. Exploring these effects is out of the scope of this paper and will be the topic of a future work.

3.2 Physical quantities: mass and star formation rate

3.2.1 Stellar mass estimate

The overall comparison between the intrinsic stellar masses (M_{sim}) and those retrieved via SED fitting (M_{phot}) is shown in Fig. 5 in photometric redshift bins up to $z \sim 4$. The observed stellar masses are in very good agreement with the intrinsic ones. The left- and right-hand panels of Fig. 6 present, respectively, the median and the dispersion around the median σ_M of $\log M_{\mathrm{phot}}/M_{\mathrm{sim}}$ as a function of $\log M_{\mathrm{phot}}$ in different redshift bins. The dispersion around the median value being potentially asymmetric, we measure σ_M^+ and σ_M^- as the value which encloses 34 per cent of the full population respectively above and below the median. The values σ_M^+ and σ_M^- are displayed as the upper and lower lines on the right-hand panels of Fig. 6.

Impact of z_{phot} uncertainties: in order to determine how much of the trend is driven by the propagation of uncertainties from the photometric redshift estimation, stellar mass computation through SED fitting is reproduced in a second step while fixing the redshift at z_{sim} instead of z_{phot}. The dashed lines in the top panels of

Fig. 6 correspond to the median and dispersion of $\log M_{\mathrm{phot}}/M_{\mathrm{sim}}$ using the photometric redshifts in the computation of M_{phot}, while the solid lines correspond to the same quantities but using the intrinsic redshifts from the simulation in the computation of M_{phot}. Comparing the solid and dashed lines therefore allows to quantify the impact of the photometric redshift uncertainty propagation in the stellar mass computation, which is very limited. Finally, it should be noted that the dashed lines provide a direct comparison with observations, as the galaxy population is split in bins of z_{phot}, and the computation includes both dust and redshift uncertainties. As a complement, the top panel of online Fig. B1 can also be compared with Fig. 5. Overall, the propagation of z_{phot} uncertainties has only a small impact on retrieving stellar mass. The scatter is relatively stable over the redshift and mass ranges and is generally smaller than 0.1 dex. M_{phot} is preferentially underestimated up to $z \sim 2$ by at most ~ 0.12 dex. At $z > 2$ and $\log M_{\mathrm{phot}} > 9.5$, the trend tends to reverse and M_{phot} ends up slightly overestimated.

Impact of dust attenuation: in order to isolate the role played by attenuation in driving this behaviour, the same computation is performed on the attenuation-free catalogue, while using the intrinsic redshift from the simulation. The median and dispersion of $\log M_{\mathrm{phot}}/M_{\mathrm{sim}}$ in the dust-free case are shown in the bottom panels of Fig. 6. Without attenuation, one is left with a weak systematic underestimation of the stellar mass especially at low redshift, which can be driven either by the too simplistic (single-burst) SFHs (see e.g. Leja et al. 2018) or the discretization of the metallicity in galaxy template (Mitchell et al. 2013). At higher redshift, these assumptions are more likely to correctly represent the actual SFHs and metallicity distributions. It should be noted that the impact of attenuation (overestimation) and of these simplified assumptions (underestimation) tends to compensate each other. For example, at $1.5 < z < 2$, M_{phot} is closer to M_{sim} when dust is included.

3.2.2 Star formation rate estimate

It is known that SFR derived from SED fitting has to be considered with caution, given the simplistic shape of the SFHs assumed in the templates, which for instance cannot account for recent bursts of star formation. Ilbert et al. (2015) predicted an overall offset of 0.25 dex and a scatter up to 0.35 dex, from a comparison of SFR derived on the one hand from SED fittnig and on the other hand from IR + UV flux.

The bottom panel of Fig. 5 illustrates this point, as it presents the overall comparison between the intrinsic SFR (SFR$_{\mathrm{sim}}$) and that derived from SED fitting (SFR$_{\mathrm{phot}}$) in photometric redshift bins up to $z \sim 4$. Up to $z \sim 3$, SFR$_{\mathrm{phot}}$ presents a bimodal behaviour, with a systematic underestimation for a large fraction of the galaxy population up to $z \sim 1.5$ and an overestimation of low-mass galaxies above. Moving towards high-z, the bimodality tends to disappear but the scatter remains very large. The left- and right-hand panels of Fig. 7 present the median of $\log \mathrm{SFR}_{\mathrm{phot}}/\mathrm{SFR}_{\mathrm{sim}}$ and the dispersion around the median σ_{SFR}. The dispersion around the median value being potentially asymetric, we measure σ_{SFR}^+ and σ_{SFR}^- as the value which encloses 34 per cent of the population respectively above and below the median. The values σ_{SFR}^+ and σ_{SFR}^- are displayed as the upper and lower lines on the right-hand panels of Fig. 7. The median evolves between -0.6 at $\log M_*/M_\odot > 10.5$ and $z < 1.5$ and 0.3 at $\log M_*/M_\odot < 10.$, while the scatter varies from ~ 0.6 at $0.8 < z < 1.1$ to ~ 0.15 at high-z.
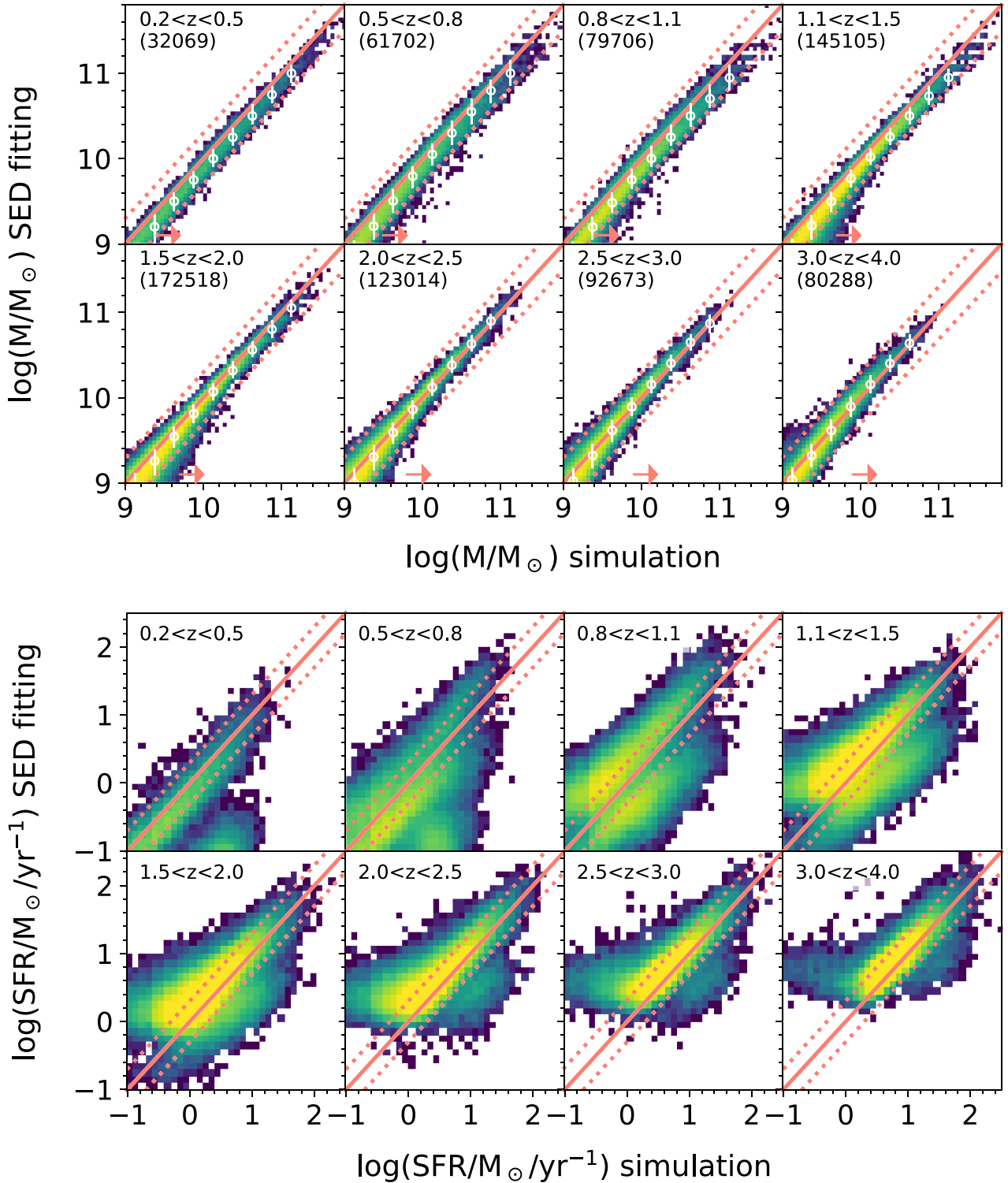
Figure 5. Top: comparison between stellar mass estimates obtained through SED fitting and masses directly derived from the sum of stellar particles, for HORIZON-AGN galaxies with $K_s < 24.7$ and $0.2 < z_{phot} < 4$. Dust and IGM attenuation are included in the photometry. The redshift is taken as being z_{phot} in the mass and SFR computation. Redshift bins and the number of objects is indicated in the upper left corner of each panel. Solid line is the 1:1 relation and dotted lines show ± 0.3 dex offset from it. White circles are the median of the SED-fitting estimates in running bins and the error on it. The density map has the same colour scale as in Fig. 1. Bottom: comparison between SFR estimates obtained through SED fitting and the SFR directly derived by adding the mass of the stellar particles formed over the last 100 Myr, for the same HORIZON-AGN galaxies. The z_{phot} range of each panel is indicated in the upper left corner (number of galaxies in each redshift bin is the same for both mass and SFR comparison).
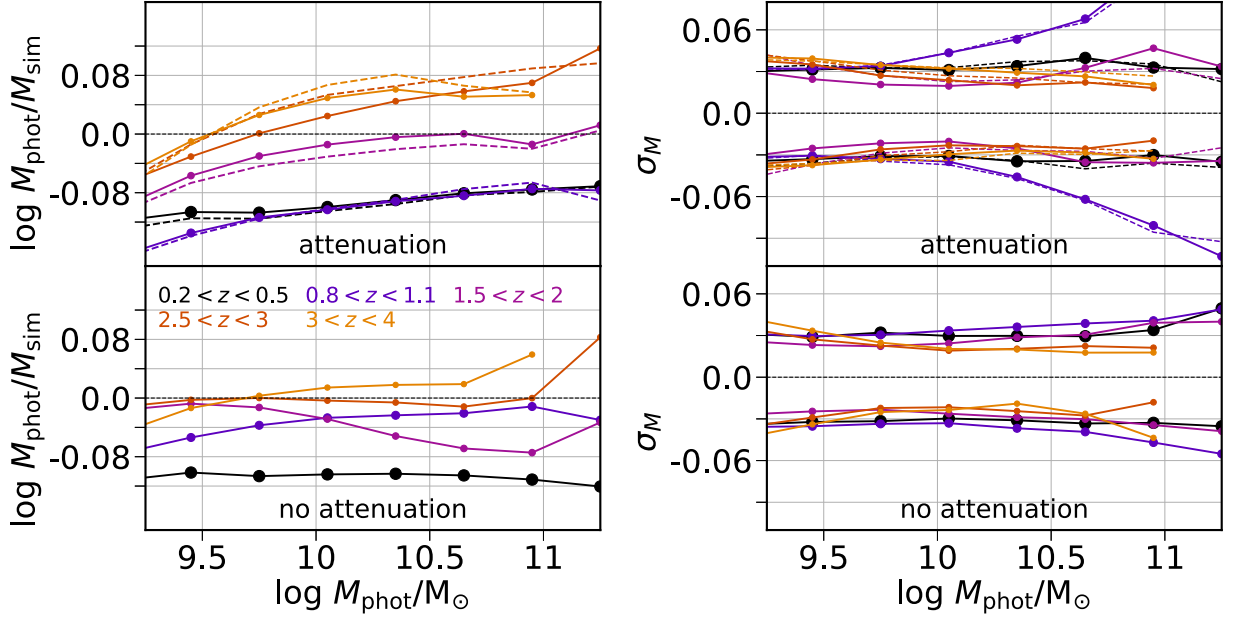
Figure 6. Left: median offset $\log M_{phot}/M_{sim}$ as a function of M_{phot}. Right: dispersion around the median values. The upper and lower lines correspond respectively to σ_M^+ and σ_M^- and enclose 34 per cent of the population above and below the median. Increasing marker sizes correspond to decreasing redshifts. Top panels are the results for HORIZON-AGN galaxies including attenuation by dust and the IGM in their virtual photometry while fixing the redshift at z_{phot} (dashed line) or z_{sim} (solid line) in the computation of M_{phot}. Bottom panels correspond to the computation of M_{phot} from attenuation-free version of the same sample with the redshift fixed to z_{sim}. For the sake of clarity only the results for half of the redshift bins are displayed.
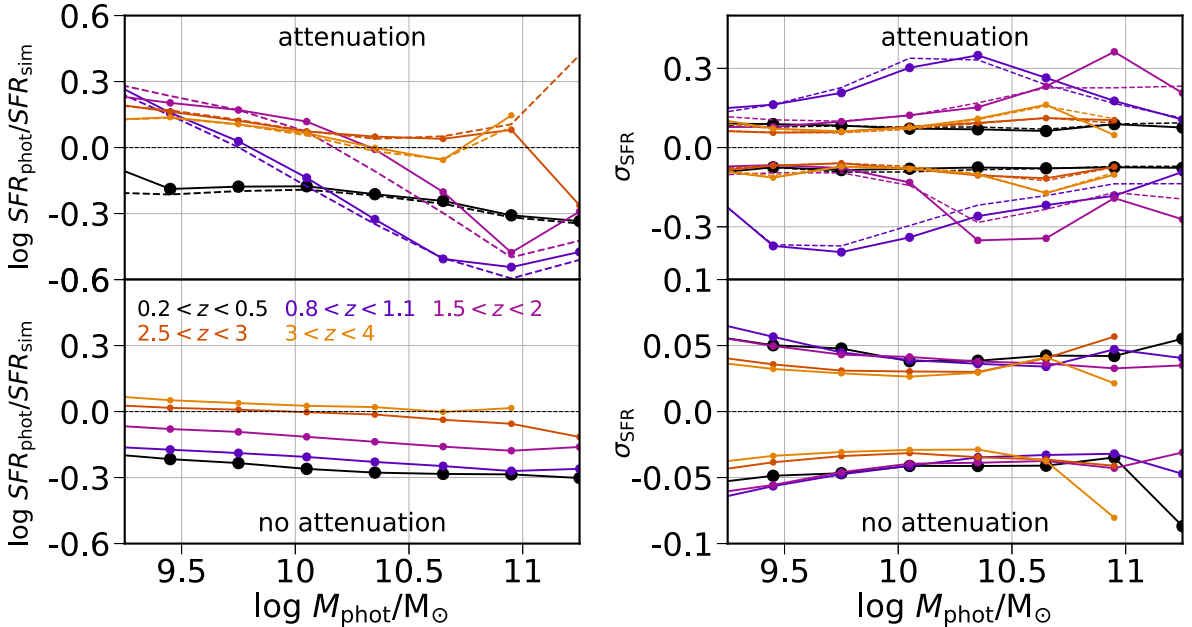
Figure 7. Left: median offset $\log SFR_{phot}/SFR_{sim}$, as a function of M_{phot}. Right: dispersion around the median values. The upper and lower lines correspond respectively to σ_{SFR}^+ and σ_{SFR}^- and enclose 34 per cent of the population above and below the median. Increasing marker sizes correspond to decreasing redshifts. Top panels are the results for HORIZON-AGN galaxies including attenuation by dust and IGM in their virtual photometry while fixing the redshift at z_{phot} (dashed line) or z_{sim} (solid line) in the computation of SFR_{phot}. Bottom panels correspond to the attenuation-free version of the same sample with the redshift fixed to z_{sim}. Note that on the left-hand panel, the y-axis range is not the same on the top and the bottom panels.

Impact of z_{phot} uncertainties: in order to determine the role of redshift uncertainties in driving the trend, the SFR is also computed while fixing the galaxy redshift at their intrinsic values z_{sim} instead of z_{phot}. The dashed lines in the *top* panels of Fig. 7 correspond to the median and dispersion of $\log SFR_{phot}/SFR_{sim}$ using the photometric

redshifts in the computation of SFR_{phot}, while the solid lines correspond to the same quantities but using the intrinsic redshifts from the simulation in the computation of SFR_{phot}. Comparing the solid and dashed lines therefore allows to quantify the impact of the photometric redshift uncertainty propagation in the SFR

computation. As a complement, the top panel of online Fig. B2 can also be compared with the bottom panel of Fig. 5. This comparison highlights that working with the simulated redshift removes the bimodality in the lowest redshift range $0.2 < z < 0.5$, which therefore is driven by redshift degeneracies. However, the bimodality remains in all the other redshift bins.

Impact of dust attenuation: one expects the impact of dust on the precision of SFR to be much stronger than on the mass. Indeed, it attenuates preferentially blue bands, which are a tracer of recently formed stars, hence directly connected to the SFR. The comparison of the SFR in the run with and without attenuation confirms this fact. The median and dispersion of $\log SFR_{phot}/SFR_{sim}$ in the dust-free case are shown in the bottom panels of Fig. 7, and the overall comparison of SFR_{phot} versus SFR_{sim} without dust is shown in the bottom panel of online Fig. B2. At $z > 2$ and in a dust-free Universe, the SFR is very well recovered without any bimodal behaviour, while it tends to be slightly underestimated at $z < 2$. As for the mass, this remaining underestimation is likely to be driven by the oversimplified SFH and metallicity underlying models, which cannot fully render the complexity of low redshift galaxy SEDs. On the contrary, when dust is accounted for in the virtual Universe, a bimodal behaviour appears, due to the SFR-dust degeneracy. The cause of this trend is investigated in online Appendix B3. In particular, there is a direct correlation between the attenuation in the rest-frame NUV and the SFR. Overestimating the attenuation A_{NUV} at the SED-fitting stage yields an overestimation of the SFR. None of the two extinction curves used in LEPHARE are a good fit for the one used in HORIZON-AGN, and this discrepancy is likely to be the main driver of this bimodality.

3.3 Performance of current surveys: summary

The virtual photometric catalogue, calibrated to mimic COS-MOS2015, has allowed a fully consistent test of the performance of LEPHARE when computing galaxy redshifts, masses, and SFR from broad-band photometry. We summarize below our main findings:

(i) In the same configuration as COSMOS2015, photometric redshifts are retrieved with the same overall precision (as estimated from NMAD and 1σ uncertainties) in the virtual data set as in the observed one. When binning the data sets in apparent magnitudes, the simulation yields as precise estimates as the observations. In particular, the 1σ uncertainties measured from $\mathcal{L}(z)$ represent on overall a good estimate of the intrinsic errors (as measured from the difference between z_{sim} and z_{zphot}), except for the bright galaxies at $1. < z < 2.5$ which have in general their errors underestimated. However, the averaged correcting factor for redshift errors proposed in earlier works (see e.g. Ilbert et al. 2013; Laigle et al. 2016) would generally overcorrect the errors of faint galaxies. Redshift errors are generally smaller in HORIZON-AGN as in COSMOS, as the mock catalogue does not include systematics in the extraction of photometry from noisy images, and presents less diversity in terms of photometry. For the same reasons, although the simulated catalogue allows us to retrieve the overall redshift distribution of the catastrophic population of outliers, it systematically underestimates their fraction. Intermediate bands allows to improve redshift accuracy.

(ii) Stellar masses are very well recovered, despite the use of single-burst SFH model and discrete metallicity in the SED-fitting templates, which do not a priori represent the complex SFHs of simulated galaxies. Only a small underestimation of at most ~ 0.12 dex persists at low redshift, and an overall scatter of the order of 0.1 dex. Conversely, dust induces a slight overestimation of the mass
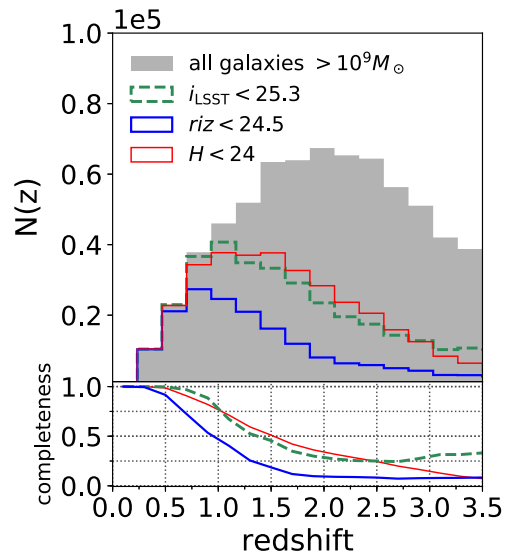

Figure 8. Redshift distribution $N(z_{sim})$ of HORIZON-AGN simulation (grey histogram, $M > 10^9 M_{\odot}$) along with the $N(z)$ of different sub-samples selected at $H < 24$, $riz < 24.5$, and $i < 25.3$ (upper red, lower blue, and green dashed histograms). For each sub-sample, the completeness fraction as a function of z_{sim} is shown in the bottom panel using same colours.

at high redshift. The impact of redshift uncertainties in driving the scatter is very limited.

(iii) Unsurprisingly, the SFR directly derived from the SFH are a quite poor proxy of the intrinsic SFR. The simplistic SFH and metallicity enrichment induce an underestimation, while the dust modelling (mainly the choice of the attenuation curve) induces a bimodality. As a result, the dispersion around the median values evolve between 0.2 and 0.6 dex.

4 SED-FITTING PERFORMANCE: FORECASTS

In this section, we use the mock catalogues reproducing *Euclid*-, LSST-, and DES-like photometry in order to predict the performance of these surveys in the three configurations presented in Table 1.

4.1 Redshift and mass completeness in *Euclid* and LSST

Let us first present the expected completeness of each survey, estimated from virtual catalogues using the intrinsic redshift, intrinsic stellar masses, and total unperturbed magnitudes. The fraction of 'detected' galaxies (i.e. those brighter than the magnitude limit of the survey) is measured as a function of redshift and stellar mass. The magnitude cuts correspond to those used for weak lensing galaxy selection in *Euclid* ($riz < 24.5$) and LSST ($i < 25.3$); the completeness at $H < 24$ is also computed, namely the 5σ detection limit expected at the completion of the *Euclid* mission. It is argued in the following that the latter threshold is the most suited for galaxy evolution science cases.

Fig. 8 compares the intrinsic redshift distribution of all the objects in the light-cone with $M_{sim} > 10^9 M_{\odot}$ to the sub-sample of galaxies detected in the *Euclid*-like catalogue, in the case of either H band or optical selection. The i-band sample expected to be detected in LSST is also included, showing a redshift distribution similar to the *Euclid* H-band-selected sample. The cut applied in the riz band results in a lower completeness, below 50 per cent already at $z_{sim} = 1$. Both the $H < 24$ and $i < 25.3$ completeness drop below 50 per cent at $z_{sim} = 1.5$. The fact that Fig. 8 does not include galaxies below

Table 3. Mass completeness limits (90 and 50 per cent threshold) in the *Euclid* (*H* < 24 or *riz* < 24.5) and LSST (*i* < 25.3) configurations as a function of redshift, as estimated from intrinsic quantities.

z_{sim} range	z_{sim} median	90 per cent mass completeness [log $M_\odot$]			50 per cent mass completeness [log $M_\odot$]		
		H < 24	*riz* < 24.5	*i* < 25.3	*H* < 24	*riz* < 24.5	*i* < 25.3
(0.00,0.25]	0.22	9.00	9.00	9.00	9.00	9.00	9.00
0.25,0.50]	0.41	9.00	9.02	9.00	9.00	9.00	9.00
0.50,0.75]	0.65	9.08	9.43	9.00	9.00	9.05	9.00
0.75,1.00]	0.89	9.31	9.81	9.27	9.05	9.42	9.00
1.00,1.25]	1.14	9.48	10.12	9.67	9.26	9.70	9.30
1.25,1.50]	1.38	9.64	10.52	9.77	9.38	9.92	9.44
1.50,1.75]	1.63	9.76	10.65	10.01	9.51	10.22	9.60
1.75,2.00]	1.87	9.83	10.89	10.13	9.60	10.31	9.65
2.00,2.25]	2.11	9.88	10.78	10.23	9.66	10.32	9.70
2.25,2.50]	2.37	9.94	10.76	10.14	9.70	10.25	9.72
2.50,2.75]	2.62	10.01	10.72	10.06	9.75	10.26	9.66
2.75,3.00]	2.87	10.04	10.68	9.98	9.82	10.22	9.58
3.00,3.25]	3.12	10.17	10.53	9.93	9.87	10.22	9.54
3.25,3.50]	3.36	10.21	10.34	9.81	9.94	10.15	9.48
3.50,3.75]	3.61	10.27	10.33	9.82	9.96	9.98	9.37
3.75,4.00]	3.87	10.28	10.26	9.71	9.99	10.07	9.32
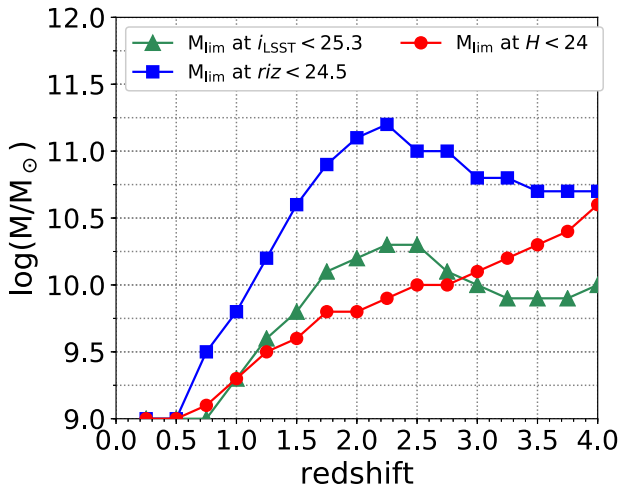
Figure 9. Redshift evolution of the stellar mass limits defined as the 90 per cent completeness threshold for galaxy samples selected in *Euclid H* (*red* line and circles), *riz* (*blue* line and squares), and LSST *i* filter (*green* line and triangles).

$10^9 M_\odot$ has a negligible impact at $z > 0.5$ because such a low-mass population is generally fainter than the magnitude limits considered here (see L16). Even at $z < 0.5$, where all selections are complete, one does not expect the addition of $M_{sim} < 10^9 M_\odot$ galaxies to significantly impact our results.

The stellar mass completeness (M_{lim}) is shown in Fig. 9. This is a lower limit, as a function of redshift, above which >90 per cent of galaxies are detected in the selection band (*i*, *riz*, or *H*). The results are summarized in Table 3, along with a less conservative $M_{lim}(z)$ threshold (50 per cent completeness). As shown in Fig. 9, the 90 per cent stellar mass completeness of LSST galaxies is below $10^9 M_\odot$ at $z < 0.5$. It increases at $z > 0.5$, because of dimming, reaching $M_{lim} = 1.5 \times 10^{10} M_\odot$ at $z \sim 2$. This threshold decreases at higher redshift as galaxies within our mass range in the early universe have higher SFRs (see e.g. Speagle et al. 2014) so they become brighter in the rest-frame UV probed by the *i* band. Conversely, at $z \lesssim 2$ the SFR starts to decline while more stellar mass is assembled, allowing an easier detection in the *H* band (see

a similar discussion in Davidzon et al. 2017, for a *Spitzer*/IRAC-selected sample).

For galaxy evolution studies relying on the *Euclid* photometry one should prefer an *H*-band selection, instead of the nominal *riz* < 24.5, if the scientific goals require a sample highly complete in stellar mass. Given an average galaxy SED, the *Euclid* optical selection would correspond to a cut at $H \sim 23$, while the survey will go deeper in the NIR by about 1 mag. Therefore, in the following sections, the analysis is carried on with the *H* < 24 selected sample. Modulo observational uncertainties, the nominal mass completeness for the *Euclid H* < 24 sample is well fitted by the function $M_{lim}(z) = 4.5 \times 10^8 (1 + z)^{2.4} M_\odot$, reaching a maximum of $2 \times 10^{10} M_\odot$ at $z = 4$ in the studied redshift range (see Fig. 9). Although the limit based on *riz* detections is generally higher (e.g. $8 \times 10^{10} M_\odot$ at $z \sim 2$) it is enlightening to note that it starts to decline at $z > 2$, as already discussed for the optical selection in LSST.[13]

4.2 Forecasts for *Euclid* and LSST photometric redshifts

Photometric redshifts are computed in the same way as in the COSMOS-like case (Section 2.4). The performance of SED fitting in the *Euclid* + DES, LSST-only, and *Euclid* + LSST configurations for z_{phot} estimation is presented in Fig. 10. The sample is split in different *i*-band bins (*i* taken from either LSST or DES photometry). The results in terms of NMAD and catastrophic outliers are presented in Table 4, for both *H* and *i* magnitude bins.

In all configurations, the usual population of outliers at high redshift and faint magnitudes $i > 24$ is found (see the discussion in Section 3.1 about the impact of IGM absorption). *Euclid*

---

[13]Recall that the star-forming main sequence in HORIZON-AGN reproduces well the observed one at $z \sim 4$, but at lower redshifts the simulation underestimates it by $\lesssim 0.3$ dex (see fig. 3 in Kaviraj et al. 2017). Consequently, observed galaxies at $2 < z < 4$ are expected to be brighter in the rest-frame UV, which would therefore enhance their *riz* magnitude by at most ~ 0.5 mag. Such a magnitude offset corresponds to a stellar mass limit approximately ~ 0.2 dex lower than that displayed in Fig. 9. A *H* < 24 selection would still provide a higher completeness. The underestimation of the SFR in the virtual catalogue being independent of mass, this remark is also valid for the LSST catalogue.
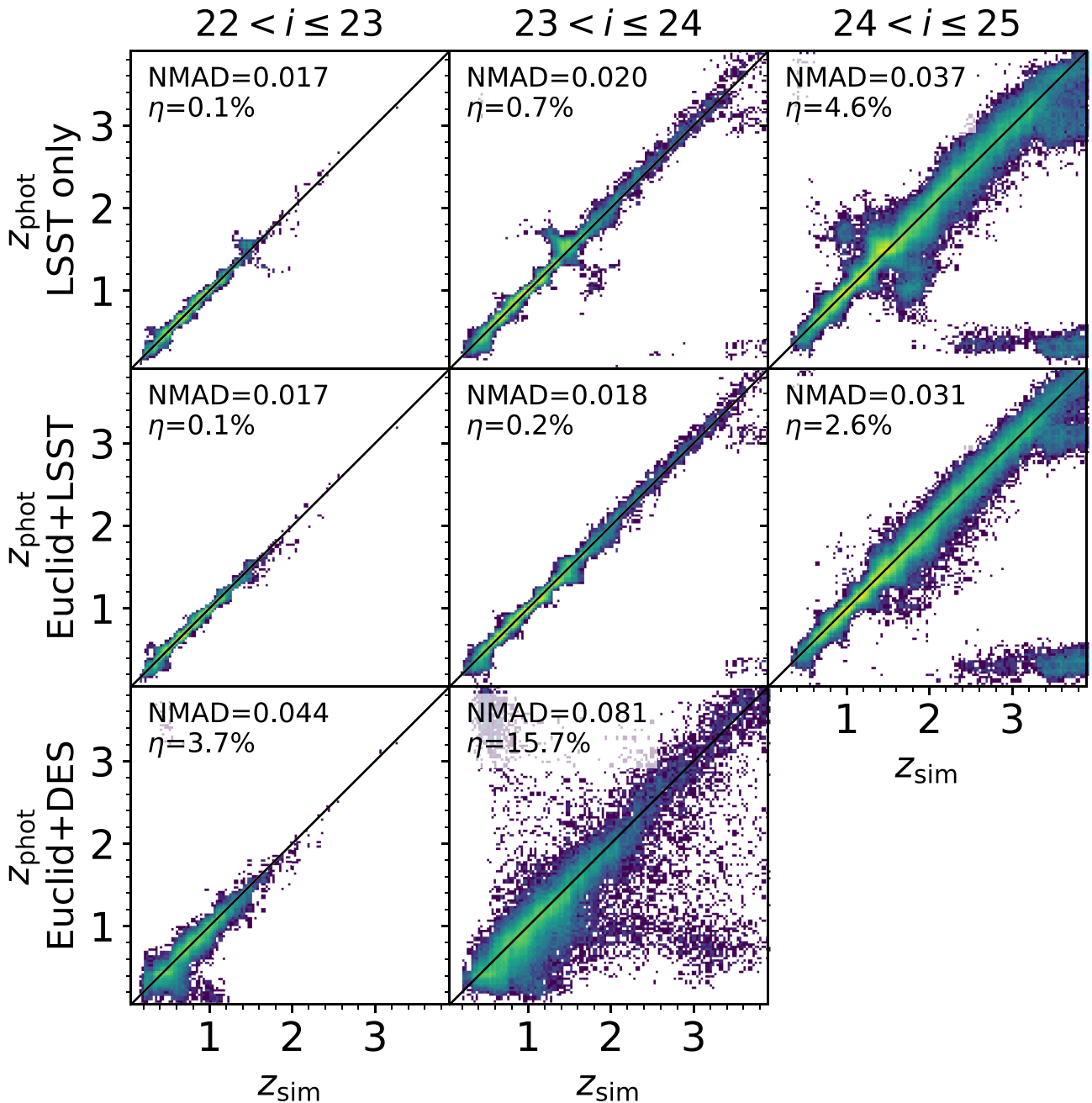
Figure 10. Predictions of z_{phot} quality in future galaxy surveys (see Section 4.2); symbols and colours as in Fig. 1. Results for three different baselines are shown: LSST (top), *Euclid* + LSST (middle), and *Euclid* + DES (bottom); Table 4 summarizes these results. Each column shows galaxies in a different *i*-band magnitude bin, with the *i*-band virtual observations coming from either LSST or DES. In the latter case, the comparison is limited to $i < 24$ because of the shallower DES sensitivity.

(combined with optical DES photometry) performs relatively well at bright magnitudes ($i < 23$). However, because of the lack of blue optical bands to constrain the Balmer break, the accuracy at very low redshift ($z < 0.5$) is lower than in COSMOS, even for bright galaxies. At fainter magnitudes, the main limitation of this configuration is the shallow depth of the survey.

In the LSST configuration, without NIR photometry, a large fraction of catastrophic outliers is present at $1.2 < z < 2.5$ at all magnitudes with a relatively symmetric patterns. The reason is the same as for the increase of z_{phot} uncertainties in this redshift range (see e.g. Fig. 4). At this redshift, the Balmer break is not constrained anymore by the optical bands and enters NIR. Without NIR bands to properly constrain its position, determining the redshift is challenging (see also Fotopoulou & Paltani 2018; Gomes et al. 2018). The situation improves from $z > 2.5$ when the Lyman break enters the optical bands.

The LSST-like catalogue alone performs therefore less well than the LSST + *Euclid* one. While the NMAD improves only by ∼1 per cent in the faintest magnitude bin, adding the *Euclid* NIR bands allows to reduce the fraction of outliers by a factor 2.

Table 4. Statistical errors (defined as NMAD) and percentage of catastrophic errors (η) in different i and H magnitude bins, in the different configurations studied here.

i band	Euclid + DES		Euclid + LSST		LSST only	
mag	NMAD	η (per cent)	NMAD	η (per cent)	NMAD	η (per cent)
(22,23]	0.044	3.7	0.017	0.1	0.017	0.1
(23,24]	0.081	15.7	0.018	0.2	0.020	0.7
(24,25]	–	>40.0	0.031	2.6	0.037	4.6
riz band	Euclid + DES		Euclid + LSST		LSST only	
mag	NMAD	η (per cent)	NMAD	η (per cent)	NMAD	η (per cent)
(22,23]	0.039	3.0	0.017	0.1	0.017	0.1
(23,24]	0.065	10.7	0.018	0.3	0.020	0.8
(24,24.5]	0.119	28.8	0.029	3.2	0.034	4.6
H band	Euclid + DES		Euclid + LSST		LSST only	
mag	NMAD	η (per cent)	NMAD	η (per cent)	NMAD	η (per cent)
(21,22]	0.045	2.6	0.018	0.0	0.018	1.2
(22,23]	0.080	12.4	0.022	0.1	0.026	2.4
(23,24]	0.153	35.4	0.040	3.6	0.049	7.1
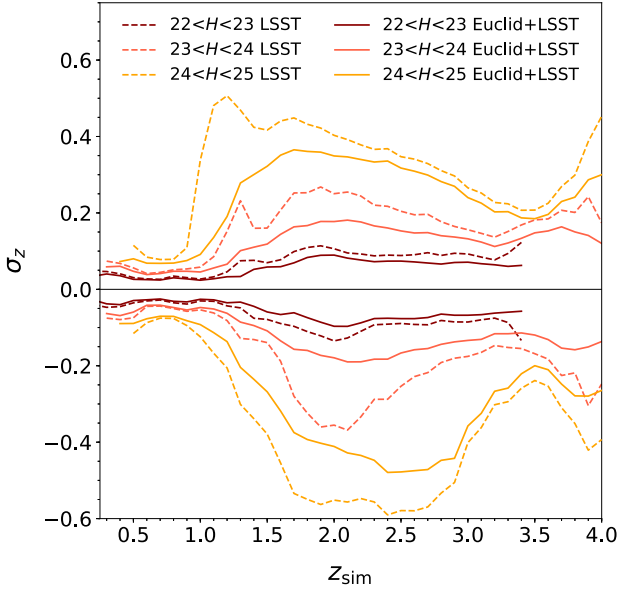
Figure 11. Photometric errors computed by LEPHARE in the LSST-only (dashed line) and *Euclid* + LSST (solid line) configurations in three different H bins.

The benefit to combine *Euclid* and LSST is further illustrated in Fig. 11. It presents the photometric errors computed by LEPHARE in the LSST-only and *Euclid* + LSST configurations in different H bins. The z_{phot} errors dramatically improve in the redshift range $1.5 < z < 2.5$, especially in the faintest magnitude bins. This is true as long as galaxies are bright enough in the NIR. At fainter magnitudes $H \gtrsim 25$, *Euclid* is not deep enough to properly constrain the Balmer break. From the *Euclid* perspective, adding the LSST optical bands to the *Euclid* + DES baseline considerably decreases redshift uncertainties and fraction of outliers (see Table 4) especially at faint magnitudes, as it provides deeper photometry in the g, r, i, and z band. Furthermore the addition of the u band is considerably useful from $z \sim 2.5$, when the Lyman-break enters the u band.

4.3 Forecast for *Euclid* and LSST stellar masses

Fig. 12 presents the overall comparison between intrinsic and reconstructed stellar masses in five photometric redshifts bins between 0.2 and 3, in the *Euclid*+DES, LSST-only and LSST + *Euclid* configurations. The number of objects in each redshift bin varies as the performance of z_{phot} estimation varies from one configuration to the other.

With the LSST-like catalogue, the performance is much poorer than with the *Euclid* + DES or *Euclid* + LSST configurations, with a very large scatter from $z > 0.7$. Indeed, without NIR photometry, the stellar mass will be determined on the basis of the photometric filters which trace the young stellar populations. For example, the massively star-forming galaxies at high redshift (e.g. the massive galaxies in the bin $1.7 < z_{phot} < 2.2$) will generally get their mass overestimated, which drives the very large scatter above the median. On the contrary, passive galaxies generally get their mass underestimated (e.g. in the bin $0.7 < z_{phot} < 1.2$), which drives the very large scatter below the median. The resulting scatter (as defined from the rms of $\log M_{phot}/M_{sim}$) can be as large as 0.5 at $z > 2$ (see e.g. Fig. 13).

It can also be noted that the *Euclid* + LSST configuration performs in general better than the *Euclid* + DES one at $z < 2$: the additional LSST optical bands help to constrain the mass reconstruction. However, more bands do not always yield a better fit. At $z > 2$ and $\log M_{sim} > 10.5$, the scatter is larger in the *Euclid* + LSST configuration than in the *Euclid* + DES one (namely with NIR photometry only). Although counterintuitive, this discrepancy might be a consequence of the very different depth in optical and NIR. Flux errors being much smaller in the optical, the blue part of the spectrum will provide a stronger constraint to the fit compared to the NIR. When no template can fit well both the optical and NIR photometry, the preference will be given to the optical since the error bars are smaller. As a result, the error on the mass might be higher, because the optical part of the spectrum is a poorer proxy for stellar mass than the NIR. In the case of star-forming galaxies, it can lead to an overestimation of the mass. Removing the LSST u band is in general not sufficient to bring better agreement, and the other optical bands still contribute a lot to this discrepancy.

As a summary, Fig. 13 presents the evolution of completeness with redshift and mass, along with the evolution of the median and rms of $\log M_{sim} - \log M_{phot}$.

4.4 Performance of future surveys: summary

We can draw the following conclusion concerning the expected performance of future surveys.

(i) With the depth of the surveys for weak lensing galaxy selection in *Euclid* ($H < 24$) and LSST ($i < 25.3$), one can expect at $z = 2$ a 90 per cent completeness at $\log M_*/M_\odot > 9.9$ and 10.2 respectively.

(ii) The *Euclid* + DES z_{phot} accuracy is of the order of several per cent even at low redshift and for bright objects, due to the absence of deep optical photometry to constrain the Balmer break position ($z < 1.5$); the fraction of catastrophic outliers dramatically increases with fainter objects. The LSST z_{phot} accuracy is of the order of 2 per cent at bright magnitudes ($i < 24$). The absence of NIR photometry does not allow to properly constrain the Balmer break at $1.5 < z < 2.5$, leading to a significant fraction of catastrophic outliers in this redshift range. A data set which would include *Euclid* NIR photometry in addition to the LSST optical photometry would
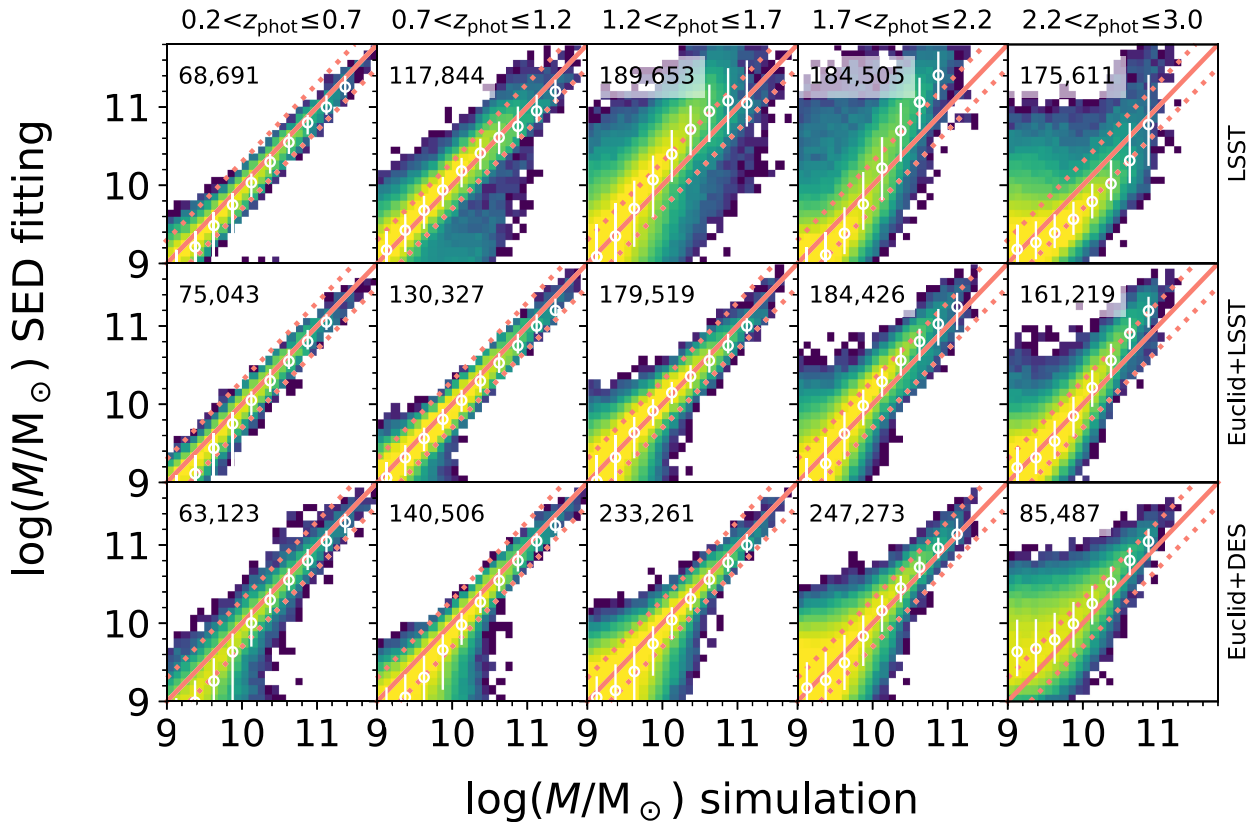
Figure 12. Stellar mass comparison for the three photometric baselines: LSST (top), *Euclid* + LSST (middle), and *Euclid* + DES (bottom). Panels in a given column include galaxies in the photometric redshift bin indicated on the top (the number of galaxies in each bin is quoted in the upper left corner of each panel). Colours and symbols are the same as in Fig. 5.

provide a better z_{phot} accuracy than LSST alone or *Euclid* alone; it would also decrease the fraction of outliers by 2.

(iii) Our LSST-like catalogue benefits a lot from NIR *Euclid*-like photometry for stellar mass reconstruction, reducing the scatter up to a factor 3. There is therefore a mutual benefit for LSST and *Euclid* to work in synergy.

5 GENERAL SUMMARY AND CONCLUSION

Using the realistic photometric catalogue extracted from the HORIZON-AGN hydrodynamical simulation, we investigated the performance of SED-fitting algorithms to compute galaxy properties. Compared to previous studies, the additional value of the present modelling relies on the use of an hydrodynamical light-cone from which the photometry has been consistently post-processed. This light-cone contains a large diversity of galaxies in terms of masses and star formation activity (but also orientation with respect to the line of sight), over a representative cosmological volume. Galaxy photometry therefore naturally accounts for the diversity of SFHs, metallicity enrichment and dust distribution which result from the complex history of their formation, driven by the combination of pristine gas infall, stellar and AGN feedback, mergers, etc., which is consistently followed in the simulation. This light-cone also allowed for the self-consistent implementation of in-homogeneous IGM absorption within each galaxy spectrum in order to test its impact on z_{phot} estimation.

We used the well-calibrated COSMOS2015 data set to assess the performance of the SED-fitting software LEPHARE in extracting galaxy z_{phot} from our mock catalogue. We also quantified our ability

to estimate the corresponding z_{phot} uncertainties. We then estimated the biases in galaxy masses and SFR estimation through SED-fitting relying on our ability to turn on and off various physical processes in the mocks (see Table 2 and Section 3.3 for a detailed summary). Finally, we quantified the expected performance for the upcoming imaging surveys *Euclid* and LSST, given the available photometric baseline and expected depths (see Tables 3 and 4 and Section 4.4 for a detailed summary).

In addition to these findings specific to some survey configurations, this work has allowed to draw the following general conclusions on the process of measuring galaxy properties from their photometry:

Choice of the photometric baseline: the added value of having medium bands in the photometric baseline to improve redshift precision is obvious when comparing z_{phot} computed with and without medium bands (Table 2). At the faint end of the galaxy population, better estimating the galactic continuum with these bands improves NMAD and η by ∼50 per cent. One can expect that the gain is even larger in the real universe, when nebular line emission can be used to constrain redshift more efficiently;

Deriving z_{phot} without deep optical photometry is challenging below $z < 1.5$ (bottom panel in Fig. 10).

NIR photometry is mainly driving the performance of z_{phot} at $1.5 < z < 2.5$ and of stellar mass computation (e.g. compare top and bottom lines in Fig. 12). Adding optical bands to the NIR photometry helps reducing the scatter (compare middle and bottom lines in Fig. 12).
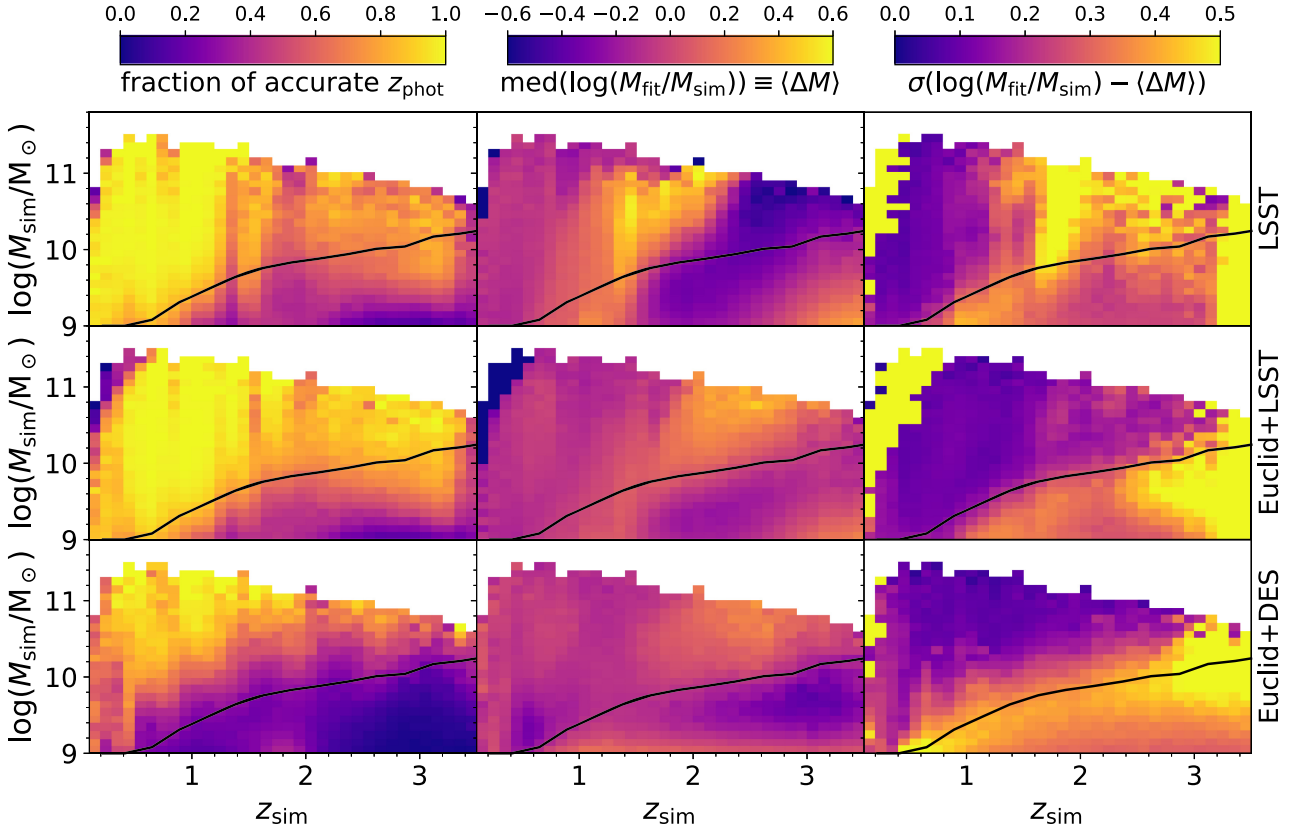
Figure 13. SED-fitting properties of galaxies in bins of z_{sim} and M_{sim} from the LSST catalogue (top) and *Euclid* (middle and bottom, combining with LSST and DES photometry, respectively). Left: fraction of accurate photometric redshift, defined as the fraction of galaxies within the given pixel having $|\Delta z|/(1 + z_{sim}) < 0.05$; middle: median logarithmic offset between intrinsic stellar masses (M_{sim}) and SED-fitting estimates (M_{fit}); and right: standard deviation of the scatter between M_{sim} and M_{fit} after removing the systematic offset. In all the panels, a black solid line delimits the 90 per cent stellar mass completeness expected for a galaxy sample selected at $H < 24$ (see Fig. 9).

Impact of dust and IGM attenuation on the z_{phot} estimates: while the impact of dust is significant, it globally does not bias much the z_{phot} reconstruction performance with the current method to include it in SED-fitting. Overestimating the IGM absorption at the SED-fitting stage also impacts the z_{phot} estimate and can explain a large fraction of the population of catastrophic outliers at $z_{sim} > 2.5$ and $z_{phot} < 1.5$, as seen in real data. There are however other possible explanations for this observed population of outliers, including systematics at the stage of photometry extraction, which we do not test in the present work.

Uncertainties on z_{phot} estimates: 1σ uncertainties estimated from the SED-fitting code LEPHARE are a good representation of the intrinsic z_{phot} errors, except for the bright galaxies at $1 < z < 2.5$ for which the errors are generally underestimated. The remaining discrepancies can be understood in the limits of our end-to-end pipeline (e.g. no emission lines, no modelling of the possible failures in the extraction of the photometry).

Uncertainties on M_{phot} and SFR estimates: the scatter and systematics in the stellar mass and SFR computation are a combination of three effects: (1) the inherently limited SFH and metallicity enrichment pattern used to build the template library, which usually drives a global underestimation of stellar mass and SFR; (2) the way dust is accounted for (in particular the choice of the dust extinction curves at the SED-fitting stage) and the degeneracy between dust

and SFR in the blue bands; and (3) the propagation of z_{phot} errors in the mass and SFR estimates, which increase the scatter. The net result exhibits a complex trend, which also depends on the photometric baseline available (e.g. compare Figs 5 and 12). As mentioned before, the impact of these effects on SFR is more dramatic, with in particular a bimodal behaviour mainly driven by dust.

Amongst possible actions to improve the redshift, mass, and SFR estimates, building a template library with an additional parametrizable double burst SFH could temper the remaining systematic offsets at low redshift. When computing masses and SFR, one could also try to build the template library in LEPHARE with additional dust extinction curves, which could mitigate the bimodal behaviour in the SFR computation. Finally, it would be worth testing if allowing the mean IGM absorption to slightly vary at $z > 3$ (in order to account for the line-of-sight variability of IGM opacity or the uncertainty on the model) can reduce the fraction of catastrophic outliers.

Our study does not account for systematics in the photometry extraction, and therefore our estimates of σ_z, η_z, and σ_M must be understood as lower limits. However in the light of our results, we can discuss if LSST and *Euclid* will, at face value, fulfil their requirements.

For LSST, the redshift errors quantified from the root-mean-square scatter ($\sigma_z^{rms} = rms(z_p - z_s)/(1 + z)$) and fraction of outliers $\eta_{3\sigma}$ (i.e the fraction of objects with $(z_p - z_s)/(1 + z_s) > 3\sigma_z^{rms}$)

must be respectively smaller than 0.05 (with a goal of 0.02) and 10 per cent at all redshifts, as specified by LSST Science Collaboration et al. (2009). For $z < 3$ and $i^+ < 24.5$ (respectively, $i^+ < 25.0$), we found $\sigma_z^{\rm rms} = 0.043$ (respectively, 0.060) and $\eta_{3\sigma} = 1.03$ per cent (respectively, 1.50 per cent). Note that $\sigma_z^{\rm rms}$ is quite sensitive to the presence of outliers, and computing it by excluding the outliers (as defined by $\eta_{3\sigma}$) yields $\sigma_z^{\rm rms} = 0.026$ (respectively, 0.032). At face value, the requirements are fulfilled, but as cautioned previously, these errors might be increased because of the systematics in the photometry extraction. When adding the *Euclid* photometric baseline to LSST, the errors decreases to $\sigma_z^{\rm rms} = 0.031$ (respectively, 0.044) (and $\sigma_z^{\rm rms} = 0.025$ (respectively, 0.030) when excluding the outliers).

As for *Euclid*, the expected redshift error $\sigma_z^{\rm rms}$ must be smaller than 0.05 (with a goal of 0.03) and the fraction of outliers $\eta_{0.15}$ (same definition as in this paper) is required to stay below 10 per cent, with a goal of 5 per cent (Laureijs et al. 2011). In the *Euclid* + DES configuration, at $riz < 23.5$ (respectively, 24.5), we get $\sigma_z^{\rm rms} = 0.09$ (resp. 0.17) and $\eta_{0.15} = 3.45$ per cent (respectively, 9.67 per cent). Excluding the outliers in the computation yields $\sigma_z^{\rm rms} = 0.46$ (respectively, 0.057). Photometry deeper than DES in bands narrower than the actual *Euclid riz* filter (e.g. the photometric baseline provided by LSST) will be required to improve these performances and extend them at fainter magnitudes.

Although measuring stellar mass is not pivotal for weak-lensing-based cosmology, it is of great interest for galaxy evolution science, i.e. to make the best of the legacy programs of *Euclid* and LSST, and therefore to fulfill their secondary science goals. In particular, the huge area of these surveys will allow to drastically decrease the statistical errors on mass functions, two-point correlation functions or any other environmental measurements (e.g. groups and clusters, cosmic web analysis). For these studies, the NIR coverage provided by *Euclid* will be of prime importance to extract accurate galaxy masses. On the other hand, without deep optical photometry in narrow optical filters, *Euclid* will be unable to separate galaxy populations from their colours, which is pivotal e.g. to study galaxy bimodality and the growth of the population of quiescent galaxies. To this end, combining *Euclid* and LSST would be a powerful configuration, which would benefit to both surveys and allows for the first time to address some of the most pressing questions in the field of galaxy formation today.

In following works, we will pursue this discussion by exploring how redshift and mass errors propagate into one and two-point statistics, and we will quantify the effect of imperfect photometry extraction from mock images.

REFERENCES

Abbas U., Sheth R. K., 2006, MNRAS, 372, 1749
Abbott T. M. C. et al., 2018, MNRAS, 480, 3879,
Aragon-Calvo M. A., van de Weygaert R., Jones B. J. T., Mobasher B., 2015, MNRAS, 454, 463
Arnouts S. et al., 2002, MNRAS, 329, 355
Arnouts S. et al., 2013, A&A, 558, A67
Aubert D., Pichon C., Colombi S., 2004, MNRAS, 352, 376
Beck R., Lin C.-A., Ishida E. E. O., Gieseke F., de Souza R. S., Costa-Duarte M. V., Hattab M. W., Krone-Martins A., 2017, MNRAS, 468, 4323
Bernardi M., Meert A., Sheth R. K., Vikram V., Huertas-Company M., Mei S., Shankar F., 2013, MNRAS, 436, 697
Bertin E., Arnouts S., 1996, A&AS, 117, 393
Bottrell C., Torrey P., Simard L., Ellison S. L., 2017, MNRAS, 467, 1033
Bruzual A. G., 1983, ApJ, 273, 105
Bruzual G., Charlot S., 2003, MNRAS, 344, 1000 (BC03)
Bundy K., Leauthaud A., Saito S., Maraston C., Wake D. A., Thomas D., 2017, ApJ, 851, 34
Calzetti D., Armus L., Bohlin R. C., Kinney A. L., Koornneef J., Storchi-Bergmann T., 2000, ApJ, 533, 682
Capak P. et al., 2007, ApJS, 172, 99
Caputi K. I. et al., 2015, ApJ, 810, 73
Chabrier G., 2003, PASP, 115, 763
Conroy C., 2013, ARA&A, 51, 393
Coupon J. et al., 2015, MNRAS, 449, 1352
Dahlen T. et al., 2013, ApJ, 775, 93
Davidzon I. et al., 2017, A&A, 605, A70
de la Torre S. et al., 2010, MNRAS, 409, 867
De Lucia G., Blaizot J., 2007, MNRAS, 375, 2
Dubois Y. et al., 2014, MNRAS, 444, 1453
Dubois Y., Devriendt J., Slyz A., Teyssier R., 2012, MNRAS, 420, 2662
Fischera J., Dopita M., 2011, A&A, 533, A117
Fotopoulou S., Paltani S., 2018, A&A, 619, A14
Galametz A., Saglia R., Paltani S., Apostolakos N., Dubath P., 2017, A&A, 598, A20
Gomes Z., Jarvis M. J., Almosallam I. A., Roberts S. J., 2018, MNRAS, 475, 331
Guidi G., Scannapieco C., Walcher J., Gallazzi A., 2016, MNRAS, 462, 2046
Haardt F., Madau P., 1996, ApJ, 461, 20
Hatfield P. W., Jarvis M. J., 2017, MNRAS, 472, 3570
Hayward C. C., Smith D. J. B., 2015, MNRAS, 446, 1512
Hemmati S. et al., 2018, preprint (arXiv:1808.10458)
Hildebrandt H. et al., 2010, A&A, 523, A31
Hoaglin D. C., Mosteller F., Tukey J. W., 1983, Understanding Robust and Exploratory Data Analysis. Wiley, New York
Hsieh B.-C., Wang W.-H., Hsieh C.-C., Lin L., Yan H., Lim J., Ho P. T. P., 2012, ApJS, 203, 23
Ilbert O. et al., 2006, A&A, 457, 841
Ilbert O. et al., 2009, ApJ, 690, 1236
Ilbert O. et al., 2015, A&A, 579, A2
Ilbert O. et al., 2013, A&A, 556, A55
Kaviraj S. et al., 2017, MNRAS, 467, 4739

Komatsu E. et al., 2011, ApJS, 192, 18

Kraljic K. et al., 2018, MNRAS, 474, 547

Laigle C. et al., 2016, ApJS, 224, 24 (L16)

Laigle C. et al., 2018, MNRAS, 474, 5437

Laureijs R. et al., 2011, preprint (arXiv:1110.3193)

Legrand L. et al., 2018, preprint (arXiv:1810.10557)

Leja J., Carnall A. C., Johnson B. D., Conroy C., Speagle J. S., 2018, preprint (arXiv:1811.03637)

Lilly S. J. et al., 2007, ApJS, 172, 70

López-Sanjuan C. et al., 2017, A&A, 599, A62

LSST Science Collaboration et al., 2009, preprint (arXiv:0912.0201)

Madau P., 1995, ApJ, 441, 18

Malavasi N. et al., 2017, MNRAS, 465, 3817

Masters D. et al., 2015, ApJ, 813, 53

McCracken H. J. et al., 2012, A&A, 544, A156

McCracken H. J. et al., 2015, MNRAS, 449, 901

Ménard B., Scranton R., Schmidt S., Morrison C., Jeong D., Budavari T., Rahman M., 2013, preprint (arXiv:1303.4722)

Mitchell P. D. et al., 2018, MNRAS, 474, 492

Mitchell P. D., Lacey C. G., Baugh C. M., Cole S., 2013, MNRAS, 435, 87

Miyazaki S. et al., 2012, in McLean I. S., Ramsay S. K., Takami H., eds, Proc. SPIE Conf. Ser. Vol. 8446, Ground-based and Airborne Instrumentation for Astronomy IV. SPIE, p. 84460z

Mobasher B. et al., 2015, ApJ, 808, 101

Moresco M. et al., 2017, A&A, 604, A133

Moultaka J., Pelat D., 2000, MNRAS, 314, 409

Moultaka J., Boisson C., Joly M., Pelat D., 2004, A&A, 420, 459

Newman J. A., 2008, ApJ, 684, 88

Ocvirk P., Pichon C., Lançon A., Thiébaut E., 2006, MNRAS, 365, 46

Oke J. B., 1974, ApJS, 27, 21

Pacifici C., Charlot S., Blaizot J., Brinchmann J., 2012, MNRAS, 421, 2002

Pasquet J., Bertin E., Treyer M., Arnouts S., Fouchez D., 2019, A&A, 621, A26

Pforr J., Maraston C., Tonini C., 2012, MNRAS, 422, 3285

Pichon C., Thiébaut E., Prunet S., Benabed K., Colombi S., Sousbie T., Teyssier R., 2010, MNRAS, 401, 705

Polletta M. et al., 2007, ApJ, 663, 81

Prevot M. L., Lequeux J., Prevot L., Maurice E., Rocca-Volmerange B., 1984, A&A (ISSN 0004-6361), 132, 389

Price S. H., Kriek M., Feldmann R., Quataert E., Hopkins P. F., Faucher-Giguère C.-A., Kereš D., Barro G., 2017, ApJ, 844, L6

Rhodes J. et al., 2017, ApJS, 233, 21

Salpeter E. E., 1955, ApJ, 121, 161

Salvato M., Ilbert O., Hoyle B., 2019, Nat. Astron., 3, 212

Sanderson R. E. et al., 2018, ApJ, 869, 12

Schaye J. et al., 2015, MNRAS, 446, 521

Schlegel D. J., Finkbeiner D. P., Davis M., 1998, ApJ, 500, 525

Schmidt K. B. et al., 2014, ApJ, 786, 57

Sorba R., Sawicki M., 2011, PASP, 123, 777

Sorba R., Sawicki M., 2015, MNRAS, 452, 235

Sorba R., Sawicki M., 2018, MNRAS, 476, 1532

Speagle J. S., Steinhardt C. L., Capak P. L., Silverman J. D., 2014, ApJS, 214, 15

Sutherland R. S., Dopita M. A., 1993, ApJS, 88, 253

Teyssier R., 2002, A&A, 385, 337

Thomas R. et al., 2017, A&A, 597, A88

Vogelsberger M., Genel S., Sijacki D., Torrey P., Springel V., Hernquist L., 2013, MNRAS, 436, 3031

Walcher J., Groves B., Budavári T., Dale D., 2011, Ap&SS, 331, 1

Weingartner J. C., Draine B. T., 2001, ApJ, 548, 296

Zamojski M. A. et al., 2007, ApJS, 172, 468

Zucca E. et al., 2009, A&A, 508, 1217

SUPPORTING INFORMATION

Supplementary data are available at *MNRAS* online.

Appendix A. Mocks additional features
Appendix B. Impact of absorption on SED
Appendix C. Zero-point magnitude offsets
Appendix D. Estimating the redshift errors

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

This paper has been typeset from a TeX/LaTeX file prepared by the author.