I. Apples to apples $A^2$: photometric redshift predictions for next-generation surveys

B. Ascaso$^1$ *; S. Mei$^{1,2}$, N. Benítez$^3$

$^1$GEPI, Observatoire de Paris, CNRS, Université Paris Diderot, 61, Avenue de l’Observatoire 75014, Paris France
$^2$University Denis Diderot, 4 rue Thomas Mann, 75205 Paris, France
$^3$Instituto de Astrofísica de Andalucía (IAA-CSIC), Glorieta de la Astronomía s/n, 18008, Granada, Spain

Accepted . Received

ABSTRACT
This is the first of a series of papers where we compare the expected performance of two of the largest stage IV next-generation surveys in the optical and infrared (LSST and Euclid), with a particular focus on cluster surveys. In this first paper, we introduce the mock catalogues we have utilized in this work, an N-body simulation+semi-analytical cone with a posterior modification with PhotReal, a technique which modifies the original photometry to make it more realistic by using an empirical library of spectral templates. We have confirmed the reliability of the mock catalogue by comparing the obtained color-magnitude relation, the luminosity and mass function and the angular correlation function with those of real data. We also analyze the behavior of the expected photometric redshifts for each different survey, in terms of photometric redshift resolution, photometric redshift bias and fraction of outliers. In addition, we discuss the benefits of using the BPZ odds photometric redshift quality parameter to select the best quality data of the sample. We provide consistent comparisons between the different surveys identifying the redshift and magnitude space where the performance of each survey becomes optimum. In particular, we find that very deep near infrared surveys such as Euclid will provide very good performance ($\Delta z/(1+z) \sim 0.025 - 0.053$) down to $H \sim 24$ AB mag and up to redshift $\sim 3$ depending on the optical observations available from the ground whereas extremely deep optical surveys such as LSST will obtain an overall lower photometric redshift resolution ($\Delta z/(1+z) \sim 0.045$) down to $i \sim 27.5$ AB mag, being substantially improved ($\Delta z/(1+z) \sim 0.035$) if we restrict the sample down to $i \sim 24$ AB mag. We highlight the fact that those numbers can be improved substantially by selecting a subsample of galaxies with the best quality photometric redshifts. We finally discuss the impact that these surveys will have for the community in terms of photometric redshift legacy once the data is available.

Key words: catalogues – surveys – galaxies: luminosity function, mass function – galaxies: photometry – galaxies: evolution – cosmology: observations

1 INTRODUCTION
In the next decade, several large-area cosmological surveys are expected to be completed in the optical: the Javalambre Physics of the Accelerated Universe Astrophysical Survey (J-PAS, Benítez et al. 2009a; Benitez et al. 2014), expected to start in 2015; the Dark Energy Spectroscopic Instrument (DESI, Levi et al. 2013) survey, expecting to start in 2018; the Large Synoptic Sky Telescope (LSST, Ivezic et al. 2008; LSST Science Collaboration et al. 2009) survey, also starting in 2018; the near infrared: the Euclid survey (Laureijs et al. 2011), which will start in 2020 and the Wide-Field Infrared Survey Telescope (WFIRST$^1$), which will start in 2020 as well, among others. One of the main goals of such surveys is the measurement of Baryonic Acoustic Oscillations (BAOs) up to redshift 1.3, therefore achieving a combined figure of merit (FoM) superior of 1000, making them to be classified as Stage IV experiments (Albrecht et al. 2004, 2009). Different strategies have been designed to achieve such goals: J-PAS will be sampling the optical spectrum with 54 narrow-bands, conse-
quenty obtaining a superb photometric redshift resolution $(\Delta z < 0.003(1 + z))$ for all galaxies down to flux limit of $i \sim 23.5$ mag (Benítez et al. 2014); DESI will target 20-30 million galaxies and quasars spectroscopically over 14000 degrees square; LSST will image a large portion of the sky with 6 broad-band optical bands down to very faint magnitudes $(r \sim 27.5)$, consequently obtaining very deep quality data; Euclid will obtain very deep infrared space data $(H \sim 24$ mag), providing an extremely deep and good resolution high-redshift data sample and WFIRST will obtain very deep photometry $(J \sim 27$ mag for 2400 square degrees and $H \sim 29$ mag for 3 square degrees) in six broad bands in the infrared, sampling also the high-redshift universe. All these surveys will be covering thousands of degrees of the northern and southern sky. This, together with the collection of external datasets (X-rays and Sunyaev-Zel'dovich effect maps, spectroscopic samples, etc) will result into an enormous quantity of available data, several order of magnitudes superior to the present data in terms of combined area, depth and photometric redshift resolution.

Up to date, different studies have been performed to estimate the photometric redshift performance of different surveys (Oyaizu et al. 2008; Banerji et al. 2008; Benítez et al. 2009a; Hildebrandt et al. 2010; Bellagamba et al. 2012; Dahlen et al. 2013; Gorecki et al. 2014; Molino et al. 2014) with a variety of different techniques. Two approaches have been considered in these works. The former uses empirical spectroscopic samples to calibrate the photometric sample while the latter extracts their results from simulations. The first methodology, while using real data, provides usually optimistic results since the spectroscopic samples are usually biased towards brightest samples. Also, these methods provide usually an average global accuracy for the whole population of galaxies at different redshift and magnitude ranges, making these results difficult to be compared consistently. The second methodology allows a treatment of spectroscopically complete mock samples, hence allowing to provide expected results as a function of magnitude or redshift.

However, it is well known that semi-analytic galaxy formation models are not fully representing the observational universe. Some of the inconsistencies between these models and the observations are not related to the inconsistency with the stellar mass function (Fontanot et al. 2009; Hansson et al. 2012; Skelton et al. 2012; Weinmann et al. 2012; Henriques et al. 2012, 2013; Mitchell et al. 2013), luminosity function (Fontanot et al. 2009; Cirasuolo et al. 2011; Somerville et al. 2012; Skelton et al. 2012; Hansson et al. 2012; Henriques et al. 2012; Guo et al. 2013; Gonzalez-Perez et al. 2014; Henriques et al. 2013), chemical abundances (Fontanot et al. 2009; De Lucia & Borgani 2012), and colors (Cohn et al. 2007; Weinmann et al. 2011; Skelton et al. 2012; Somerville et al. 2012; Hansson et al. 2012; Henriques et al. 2012), consequently producing a wrong tilt of the color-magnitude relation, ‘plume’ effects of redder galaxies belonging to the cluster spread within 1-2 bright magnitudes, absence of a smooth-transitory green valley and other related effects. These effects generally lead to an underestimation of the photometric redshift uncertainties. As shown by Arnalte-Mur et al. (2014), Ascaso et al. (2015b); obtaining photometric redshifts directly from mock photometry mimicking the ALHAMBRA survey (Moles et al. 2008) provided a mean photometric redshift dispersion three times higher than the expected for the real data [Molino et al. 2014]. After creating a new photometry set using a technique that uses a new empirical library of templates (see section § 2.2) to fit the original spectral distribution, the photometric redshift accuracy exactly matched that obtained from real data.

In order to create a solid basis for comparison between different methodologies and datasets, we start with this paper a series of works under the title: *Apples to apples, A²*, which intend to provide a handy well-characterized mock catalogue with realistic well-behaved photometric-redshifts that mimic two of the next generation surveys: LSST and Euclid. A separate work (Ascaso et al. in prep) will also be devoted to analyze the multiple narrow-band J-PAS survey.

In this work, we have applied a similar prescription as in Arnalte-Mur et al. (2014), Ascaso et al. (2015b). The whole point of applying this prescription is to produce a galaxy sample which resembles as much as possible the real universe in terms of galaxy properties, in particular their distribution of luminosities, redshift, and spectral types. The chosen mock catalogue is based on the publicly available light cone mock catalogues by Merson et al. (2013). These catalogs are extracted from an N-body simulation from the Millennium simulation, (Springel et al. 2003) and semi-analytic model of galaxy formation from GALFORM (Cole et al. 2000, Bower et al. 2006). We have used the Bayesian Photometric Redshift software 2.0 (BPZ2.0, Benítez 2000), Benítez et al 2015, in prep), to obtain photometric redshifts. This technique, called PhotReal (Benítez et al. in prep), is chosen to provide really accurate and realistic results at comparing with real data and avoid fake effects such as ‘plume’ or unrealistic color-magnitude relations.

At least two more works will continue these $A²$ series. In the second one (Ascaso et al., in prep), we will provide consistent clusters selection functions for all the surveys considered by using the Bayesian Cluster Finder (BCF, Ascaso et al. 2012, 2014a), a well behaved detector which is maximizing the advantages of many methods in the literature to detect not only red-sequence cluster and groups. The third $A²$ paper (Ascaso et al., in prep) will be devoted to forecast the cosmological parameters from the obtained cluster counts in the same consistent way. These mock catalogues will be publicly available and they will be useful for any consistent prediction between the considered surveys: LSST and Euclid in terms of galaxy evolution, galaxy clustering, large-scale structure or any other astrophysical purpose.

The structure of the paper is as follows. In section §2, we introduce the mock catalogue used in this work and we give an overview of the technique developed to provide reliable photometry and realistic photometric redshifts. Section §3, summarizes the main characteristics of the three next-generation surveys considered in terms of area, depth, number of bands and different technical properties. Section §4 provides an extended comparison with different observational dataset to check that their main properties such as stellar masses, magnitudes, colors and spectral types resemble those obtained from the PhotReal post-processed mock catalog. Section §5 describes the obtained results on photometric redshift resolutions and the fraction of outliers for each survey. Section §6 provides a way to improve the overall photometric redshift performance by obtaining the ‘best...
selection samples’ with an odds quality selection and finally, section 5 summarizes the results of this work and discusses the possible applications of the overall set of data to future studies on providing cluster selection functions and cluster cosmology constraints as well as the impact of this science.

The mock catalogue used in this paper has a fixed cosmology of $\Omega_M = 0.25$, $\Omega_L = 0.75$, $\Omega_b = 0.045$ and $h = 0.73$, chosen to match the cosmological parameters estimated from the first year results from the WMAP (Spergel et al. 2003) and which are not significantly different to those obtained by Planck (Planck Collaboration et al. 2013). Hence, throughout the whole paper and when necessary, we will use the same cosmology. All the magnitudes in the paper are given in the A$^2$ system.

2 MAIN SIMULATION

2.1 Original light mock catalogue

In this work, we based our simulations on the original mock catalogue by Merson et al. (2013). The spatial information of this mock catalogue comes from the Dark Matter halo merger trees extracted from the Millennium Simulation (Springel et al. 2005), a 2160$^3$ particle N-body simulation of the $\Lambda CDM$ cosmology. This simulation traces the cold dark matter structures from redshift $z=127$ throughout the present day in a cubic volume of $500h^{-1}$ Mpc$^3$. The minimum resolution of this simulation is 20 particles, corresponding to a halo resolution of $1.72 \times 10^3h^{-1}M_\odot$. These haloes have been populated with galaxies created with the variant of the GALFORM (Cole et al. 2000) semi-analytical model by Bower et al. (2006), which models many physical processes governing the formation and evolution of the galaxies such as the star formation and merger history, the radiative cooling of the gas, the feedback as a result of supernovae active galactic nuclei, etc. It also makes predictions for many galaxy properties including luminosity as a function of wavelength. However, these predictions do not always resemble those of the observed galaxies: the colors of the faint galaxies tend to be too red compared to the real ones, there exist a ‘plume’ of redder galaxies than the red main sequence at redshift lower than 1.4, etc. We have overcome these problems by running PhotReal (see section 2.2), an algorithm that creates more realistic photometry by using a set of empirical templates to fit the original spectral distribution. Finally, these galaxies have been associated a subhalo according to its status of central or satellite galaxy at each snapshot. More information on the construction of this mock can be found in Merson et al. (2013) and references herein.

We have chosen to use the Euclid 500 deg$^2$ wide mock catalogue since it is the deepest of the surveys considered and hence, we can always make it shallower. As for the surveyed area, we chose to use a 500 square degrees mock even if it is a factor of 20 to 40 times smaller than the total area of the targeted surveys as a compromise between having a easy-to-transfer, relatively small in size catalogue and an area large enough to have significant enough statistical results for our scientific purposes. Indeed, since the main purpose of $A^2$ is to analyze photometric redshift and cluster selection performance, none of these results should be substantially affected after applying a correction which increases the errors following a Poissonian approach to the results to account for this effect.

As for galaxy clusters and groups, the possible structures that might be missing due to the limitation of the area would be the rarest ones ($M > 10^{15} M_\odot$) with similar properties as those recently discovered such as, for instance, the SPT-CL J2344-4243 (The Phoenix, McDonald et al. 2012), ACT-CL J0102-4915 (El Gordo, Menanteau et al. 2012), SPT-CL J0546-5345 (Brodwin et al. 2010), SPT-CL J2106-5844 (Foley et al. 2011) at redshift > 1. These structures, which have been reported to be very unfrequent (Joe et al. 2009; Hoyle et al. 2012), are usually the easiest to detect among other structures at the same redshift almost with any technique (optical, X-rays, cosmic shear) so, we can ensure that if they exist, we will be detecting them. The other possible problem that might arise from using a smaller simulation would be related with those structures lying on the edges of the cone. However, when detecting galaxy clusters and groups in the next paper or these series, we will use the BCF which, automatically corrects the probabilities by the missing area of the cluster due to edge or masks issues (Ascaso et al. 2012, 2014a).

2.2 PhotReal: the photometry calibration technique

In order to obtain more realistic and reliable photometry and photometric redshifts, we have applied PhotReal, a similar algorithm as in previous works Arnalte-Mur et al. (2014), Zandivarez et al. (2014), Ascaso et al. (2015b) to the initial mock catalogue. We summarize here this technique developed to fit the spectral types to make the photometry more realistic. A more extended explanation will be given in a forthcoming work (Benítez et al. 2015, in prep).

This technique relies on a library of empirical templates developed for the well-known photometric redshift code, the Bayesian Photometric Redshift (BPZ2.0, Benítez 2000, Benítez in prep). This library consist of eight empirical templates with a very low outlier rate and bias in high quality photometric catalogues. Hence, it contains a complete (up to 1-2%), even if coarse-grained, representation of real galaxy colours for the galaxy populations sampled.

The PhotReal procedure starts by running BPZ2.0 in 'ONLY_TYPE yes’ mode to match the optical ugriz and infrared (if available) rest frame photometry of the initial mock catalogue to the best template from the empirical library. In other words, PhotReal forces the semi-analytical photometry to have a realistic template distribution by choosing the closest BPZ2.0 template. Once the SED is thus chosen, we generated galaxy fluxes through the survey set of filters in the same way as in the Benítez et al. (2009b, 2014) empirical mocks and add to them empirically calibrated photometric noise. This noise is a combination of the expected photometric noise from the observed relationship between magnitudes and errors in a similar characteristics survey filters (see §3), plus a systematic noise which is approximately constant with magnitude and most likely unavoidable when measuring galaxy colours in multiband photometry. This systematic is empirically cal-

2 http://community.dur.ac.uk/a.i.merson/lightcones.html
ibrated to be 8% for bluer objects and 6% for red galaxies (Benitez, in preparation).

Afterwards we run BPZ2.0 on those mock catalogues to obtain the final photometric redshift estimation and spectral type. This procedure of course implicitly introduces an error: there may be real galaxies which are present in the catalogues, but not covered by the BPZ2.0 templates or not adequately present in the spectroscopic redshift catalogues used to measure the outlier rate. Since we are not aware of any such substantial population within the depths for the surveys considered, we estimate that the procedure followed here does not introduce a significant bias.

The final PhotReal output includes photometry and photometric errors, spectral types, stellar masses, a single photometric redshift estimation and the redshift probability function \( P(z, T) \), condensing the probability as a function of redshift and spectral type, for all galaxies in the mock.

### 3 Considered Next Generation Surveys

Traditionally, the estimation of the accuracy of photometric redshifts has been given globally, without specifying their dependence with photometry or redshift (although see recent efforts by Hildebrandt et al. 2012; Dahlen et al. 2013; Raichoor et al. 2014; Molino et al. 2014); it has been estimated for different spectroscopic samples with very complicated selection functions to reproduce and different flux cuts have been performed. These issues carry a number of arduousness. First, it becomes complicated to have a clear idea about the best region of the parameters space to use if one wants to select the best quality photometric redshifts. Also, these assumptions make very hard the comparison between the photometric redshift accuracy of different surveys. Furthermore, a clean sample obtained by selecting the best quality photometric redshift has been proved to have many benefits for many scientific purposes but the previous points produce a non equivalent treatment for different surveys.

In the spirit of A², we pretend to perform a systematic comparison between the photometric redshift estimation for different next-generation stage IV surveys using the same mock catalogue and the same photometric redshift technique. In this section, we highlight the main characteristic of the two next-generation surveys we have considered in this work in terms of design, area coverage, wavelength range, photometry depth, complementary data and other technical issues.

#### 3.1 LSST

The Large Synoptic Survey Telescope (LSST; Ivezić et al. 2008; LSST Science Collaboration et al. 2009) is a stage IV experiment starting in 2018. The main objectives of this survey are also the constraints of the dark energy parameters, the weak lensing analysis and the study of transients. The survey is designed to be a large, wide-field ground-based survey imagine the whole visible southern sky from an 8.4m telescope Cerro Pachon (Chile) with six broad-band optical bands ugrizy down to \( r = 27.5 \) mag after coadding 10 years of operations. The depth of the survey will allow to produce very detailed shear and weak lensing maps in a similar way that was done with the Deep Lens Survey (DLS, Wittman et al. 2002; Schmidt & Thorman 2013; Ascaso et al. 2014). The expected depths can be found in Table 1 in Ivezić et al. (2008).

The photometric errors have been estimated according to the description given by the Survey Science Group: eLSST\(^4\), e.g. the expected error for a galaxy in a single observation is

\[
\sigma^2 = \sigma^2_{\text{sys}} + \sigma^2_{\text{photon}}
\]

with \( \sigma_{\text{sys}} \) is the systematic photometric error ranging between 0.005-0.01 mag for the LSST case, and \( \sigma_{\text{photon}} \) is the random photometric error, which can be written as

\[
\sigma^2_{\text{photon}} = (0.04 - \gamma)x + \gamma x^2,
\]

where \( x = 10^{0.4(m - m_5)} \)

where \( m_5 \) is the 5\( \sigma \) limiting magnitude depth and \( \gamma \) depends on the sky brightness and readout noise. Eventually, the final error obtained for \( N \) observations is

\[
\sigma_N = \sigma_1/\sqrt{N}.
\]

#### 3.2 Euclid

The Euclid survey (Laureijs et al. 2011) is an European space mission survey starting in 2020. Euclid will also be a Stage IV experiment and it is expected to provide very accurate measurements of the dark energy constrains from a number of different probes including BAOs, weak lensing, cluster counts and high-z supernova.

The survey will have a 'Euclid Wide Survey’ covering 15,000 square degrees of the sky in the near infrared (YJH), down to ~24 mag in H band providing also near-infrared spectroscopy and two ‘Euclid Deep Fields’, about 2 magnitude deeper than the wide survey which will cover around 20 square degrees each. The deep infrared imaging will map and systematically explore the mostly unknown-high redshift Universe.

Euclid space observations will be combined with other space and ground-based observations to obtain the source photometric redshifts and physical properties. Among the optical surveys that will be available form the ground, one is already available, the Sloan Digital Sky Survey (SSDS), and two are planned, the LSST and the DES (The Dark Energy Survey Collaboration 2009) surveys. Hence, we will consider two cases which will provide us a maximum and minimum limit in the photometric redshift and cluster survey predictions for the Euclid survey:

- **Euclid pessimistic case:** where the optical data counterpart will come from the five broad band photometry gryY from the DES down to the depth stated in Table 1 by Mohr et al. (2012). The photometric errors for the DES bands have been estimated from the mock catalogues by Chang et al. (2014) which mimic the properties of the DES data.
- **Euclid optimistic case:** where the optical counterpart will come from the combination of DES and LSST-like surveys.

---

3 http://www.lsst.org/lsst/

4 http://ssg.astro.washington.edu/lsst/magsfilters.shtml

5 http://www.euclid-ec.org/
In both cases the photometric errors for the 'Euclid Wide survey' have been calibrated from the Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey (CANDELS\footnote{http://candels.ucolick.org/} and normalized to the Euclid depth (e.g. H$\sim$24 at 5$\sigma$, Laureijs et al. (2011)).

4 PERFORMANCE OF THE MOCK GALAXIES

4.1 Model properties of the galaxies

With the double-fold objective of measuring the variation of the modeled properties of the galaxies used in this simulation and proving the non-dependence of the results shown here on the photometric redshift code used, we have used LePhare (Ilbert et al. 2006, 2009) to obtain an estimation of the stellar masses, ages and star formation rates (SFRs) of the galaxies in the original mock and the PhotReal post-processed catalogues.

To do so, we have used 27 stellar population synthesis models generated using Bruzual & Charlot (2003) models at 14 ages that range from 0.1 to 13.0 Gyr. These BC03 models use a Chabrier (2003) initial mass function with metallicities of $Z$=0.004, 0.008, and 0.2$Z_\odot$ and are characterized by exponentially declining star formation histories (SFHs) with timescales of $\tau$ = 0.1 – 30 Gyr. We adopt the Calzetti et al. (2000) reddening law with $0 \leq B - V \leq 0.5$. We first obtained the photometric redshift estimation and we later fixed it to provide model measurement of the galaxies.

Then, we have compared on one hand, the LePhare stellar masses estimated from the mock initial photometry with the original stellar masses in the mock and on the other, the LePhare stellar masses estimated from PhotReal post-processed photometry with the stellar masses obtained with BPZ2.0. In Fig. 1 we display the difference between the LePhare estimation to the original photometry and the initial mock catalogue masses (solid line); and the difference between the LePhare estimation to the PhotReal photometry and the stellar masses estimated with BPZ2.0 to this photometry (dashed line). The difference between the initial catalogue masses and LePhare estimations is a factor of two more biased towards negative values ($\langle SM(LePhare) - SM(Merson) \rangle = -0.028 \pm 0.494$) than the difference between the BPZ2.0 masses and their estimation with LePhare ($\langle SM(LePhare) - SM(BPZ2.0) \rangle = -0.018 \pm 0.432$). However, the dispersion obtained in the relation is similar.

Looking into this issue more in detail, we have plotted in Fig. 2 the density map of the logarithmic ratio between the stellar mass obtained from the fit to the PhotReal photometry and the estimation with LePhare as a function of the PhotReal mass as a function of redshift. In this comparison, we see that in the lower redshift bin, the bias of the stellar mass estimation is smaller than the previous comparison (-0.002). However, the dispersion is 3 times larger. As we increase in redshift, we notice a bimodal distribution of the difference, explained in Mitchell et al. (2013) for the use of different Initial Mass Function (IMF) and Stellar Population Synthesis (SPS) models. However, the dispersion do not increase, becoming comparable between $0.5 < z < 1.0$ and a factor of 1.5-2 smaller at $z>1.0$.

As a consequence, the PhotReal post-processed catalogue seems to be a factor of $\sim 1.5$ less sensible to the presence of dust in the model. In section 4.5 we will investigate more in detail the stellar mass function.

Similarly, we have plotted in Fig. 3 the logarithmic ratio between the original stellar masses provided in the original mock catalogue and the stellar masses fitted with LePhare to the original mock photometry. Dotted line: Distribution of the difference between the stellar masses obtained from the PhotReal post-processed mock catalogue and the stellar masses fitted with LePhare to the PhotReal mock photometry. The vertical dashed line indicates the zero value for reference.

In Fig. 4 we show the differences between stellar ages and star forming rates (SFR) obtained from the initial and transformed mock catalogue for different redshift bins. The stellar ages of the considered galaxies are very similar, particularly at high redshifts, where they fully agree. Overall, we find that more than $\sim 60\%$ of the sample has a difference in galaxy ages less than 0.5 dex.

As for the SFR, we find that the differences are smaller in the highest redshift bin ($1.5 < z < 3$), obtaining $<log(SFR/SFR_i)> = -0.281 \pm 0.509$. At lower redshift the dispersion increases a factor of 3-4 due to the presence of a secondary peak at all redshift bins with a systematic difference of $log(SFR) < -3$. These secondary peaks indi-
Figure 2. Density map of the logarithmic ratio between the stellar mass given in initial mock catalogue and its fit with LePhare as a function of the initial mock stellar mass for six different redshift bins.

Figure 3. Density map of the logarithmic ratio between the stellar mass obtained from the fit to the PhotReal photometry and the estimation with LePhare as a function of the PhotReal mass for six different redshift bins.

cicate that a substantial part of the galaxies in the PhotReal post-processed mock are better fitted by a spectrum with different star forming rates than originally defined, particularly at lower redshifts. This is compatible with the shift to bluer colors shown in the previous section where we show how the PhotReal post-processed catalogue makes the color-
magnitude relation (CMR) more realistic showing a blue cloud as expected in observational data.

While the objective of this test is not to obtain reliably model parameter of the galaxies in the mock but rather check the difference between their properties, we conclude that the main star formation history of the galaxies in the mock are not changing substantially, particularly at high redshift.

4.2 Spectral type recovery

The original mock catalogue by Merson et al. (2013) did not include the information regarding the spectral type of the galaxy or its spectrum due to disk space storage. We have instead performed a comparison of the spectral type obtained from the fit to the original photometry, \( t_i \), and the recovery spectral type obtained from the PhotReal photometry, \( t_f \). The spectral types are similar to the ones used in Benítez (2000); the first time refers to a typical Elliptical, type 2 to a Lenticular, type 3-4 to Early and late spirals respectively, and type 5-8 to starburst. In Fig. 4 we show these type distributions for the Euclid-Pes (top left panel), Euclid-Opt (bottom left panel) and LSST (top right panel) surveys respectively.

The difference between the original and recovered spectral type distribution is minimum for the Euclid-opt case (1.90%), as expected and it changes slightly for the Euclid-pes case (3.65%) and the LSST case (2.90%). While a higher departure of the final spectral distribution is observed for the LSST and Euclid-pes, we conclude that the spectral distribution is well-recovered for the three cases considered.
I. Apples to apples $A^2$: photometric redshift predictions for next-generation surveys

4.3 Color-magnitude relation

In the original semi-analytical mock catalogue, the first effect that we noticed al is the fact that the CMR usually exhibited a lack of blue cloud that it is not observed with observational data (e.g. Ascaso et al. 2008; Mei et al. 2009; Ascaso et al. 2014b) in agreement with Merson et al. (2015). PhotReal overcomes this problem by re-computing the photometry and therefore, shifting some galaxies from the red sequence to the blue cloud. In Fig. 6 we show the $z - J$ color versus $J$ magnitude diagram for a cluster in the mock catalogues as seen by the Euclid bands. The left panel refers to the original CMR from the mock catalogues whereas the right panel displays the CMR from the PhotReal catalogue. We have also overplotted the observational data from Papovich et al. (2010) at redshift 1.62. We observe that the PhotReal CMR fills well the blue valley, in agreement with the observational data.

We have also looked into the optical CMR obtained at low redshift. Fig. 7 shows the $i - z$ versus $z$ CMR for a cluster at $z \sim 0.52$ and Fig. 8 the $g - r$ versus $r$ for a cluster at $z \sim 0.29$, as seen by LSST. As before, the left panel refers to the original simulation while the right panel shows the PhotReal post-processed CMR. In all cases, the PhotReal post-processed CMR shows a well-defined blue cloud in addition to a defined CMR, resembling closely the observational CMR of the clusters in the literature.

A secondary effect related to colors, noticed in the original mock catalogues is the presence of a ‘plume’ of redder galaxies than the CMR. While this effect has been noticed in the infrared colors at high-redshift ($z \sim 1.4$), the effect continues appearing at lower redshift in the optical bands. In Fig 9 we show an example of the $z - J$ color-magnitude relation of a cluster at $z \sim 0.52$ and $M \sim 2.17 \times 10^{14}M_\odot$ as seen by the Euclid-Opt survey in the original mock catalogue (left panel) and in the PhotReal post-processed one (right panel). The blue squares indicate the galaxies considered to be in the ‘plume’ in the original mock catalogue and their position in the PhotReal post-processed mock catalogue. We observe that the presence of this systematic ‘plume’ disappears from the color-magnitude diagrams after the correction but they are still red galaxies, located in the main red sequence of the cluster. These galaxies have not changed substantially their colors and, in addition, they have also diluted their photometry within the photometric errors of the simulation. This result ensures us that, while we are making the mock cata-

Figure 5. Spectral type distribution obtained from the original Merson photometry (grey solid line) and from the PhotReal post-processed mock catalogue (red dashed line) for each of the considered surveys: Euclid-Pes (top left panel) and Euclid-Opt (bottom left panel) and LSST (top right panel). The Euc-opt shows an excellent recovery of the spectral type, whereas the difference between the original and the recovered simulation for the Euclid-pes and the LSST becomes a factor of 1.5-2 larger.

Figure 6. $z - J$ color versus $J$ magnitude diagram for a $7 \times 10^{13}M_\odot$ cluster at $z = 1.61$ in the mock catalogue as seen by Euclid with its bands. The left panel refers to the original mock CMR whereas the right panel represents the recovered CMR after introducing the realistic errors in the catalogs. The blue triangles refer to the CMR of the galaxies belonging to a cluster at $z \sim 1.62$ by Papovich et al. The PhotReal post-processed CMR samples the region with observational data, whereas the initial CMR only constraints to red colors.

Figure 7. $i - z$ color versus $z$ magnitude diagram for a $1.82 \times 10^{14}M_\odot$ cluster at $z = 0.29$ in the mock catalogue as seen by LSST with its bands. The left panel refers to the original mock CMR whereas the right panel represents the recovered CMR after introducing the real effects of the catalogs. The red colors refer to the galaxies identified as early ($t_3 < 3$) in the original mock catalogue and ($t_f < 3$) in the PhotReal post-processed mock catalogue.
Ascaso, Mei & Benítez

4.4 Luminosity Function

An additional check that we have performed is the comparison between the infrared and optical luminosity function (LF) at all redshift ranges. In Fig. 10, we plot the observational H-band rest-frame LF for three different redshift bins at $z > 1.5$ obtained from the fits by Stefanon & Marchesini (2013) from the Multiwavelength Survey by Yale-Chile (MUSYC, Gawiser et al. 2006), the Faint Infrared Extragalactic Survey (FIRES, Franx et al. 2000) and the GOODS Chandra Deep Field-South Survey (FIREWORKS, Wuets et al. 2008). We overplot the luminosity counts obtained for the Euclid-opt simulation in the original mock catalogue and in the PhotReal post-processed CMR. We find a remarkable good agreement between the observational SM function and both estimations in the mock catalogues, giving also support to the fact that we do not substantially alter the initial mock catalogue.

4.5 Stellar mass function

We have looked into the agreement of the observational stellar mass (SM) function with the one obtained from our simulations. In Fig. 12, we display the observational SM obtained by Muzzin et al. (2013) from the COSMOS/UltraVista field within $0 < z < 3$ for six different redshift bins. As before, we overplotted the SM counts for the original mock photometry and the PhotReal post-processed photometry. We find a remarkable good agreement between the observational SM function and both estimations in the mock catalogues, giving also support to the fact that we do not substantially alter the initial mock catalogue.

We have also compared the SM function for different galaxy types. In Figs. 13 and 14, we show observational SM functions by Muzzin et al. (2013) for quiescent galaxies and...
I. Apples to apples $\Delta^2$: photometric redshift predictions for next-generation surveys

Figure 11. I-band rest-frame luminosity function for three different redshift bins: $0. < z < 0.1$ (top panel), $0.45 < z < 0.85$ (model panel) and $0.85 < z < 1.3$ (bottom panel). The red diamonds refer to the magnitude counts of the initial mock catalogue and the blue triangles refer to the same counts for the PhotReal post-processed mock catalogue for the LSST survey. The solid line and the gray area are the observational luminosity function obtained by Montero-Dorta & Prada (2009) for the low redshift bin and Gabasch et al. 2006 with the Case 3 evolution for the other higher redshift bins.

Figure 12. Stellar mass function for six different redshift bins: $0.2 < z < 0.5$ (top left panel), $0.5 < z < 1.0$ (top right panel), $1.0 < z < 1.5$ (middle left panel), $1.5 < z < 2.0$ (middle right panel), $2.0 < z < 2.5$ (bottom left panel) and $2.5 < z < 3.0$ (bottom right panel). The red diamonds refer to the stellar mass counts of the initial mock catalogue and the blue triangles refer to the same counts for the PhotReal post-processed mock catalogue for the Euclid-opt survey. The solid line and the gray area are the observational SM function obtained by Muzzin et al. 2013.

star forming galaxies respectively, we also display in these plots the SM counts for these types of galaxies for the original and PhotReal post-processed catalogue. To do this, we have defined quiescent galaxies (star-forming galaxies) as those that the spectral classification obtained from the fit to the original $t_i$, or PhotReal post-processed photometry, $t_f$ has a spectral type $< 3$ ($\geq 3$).

In general, we find a very good agreement between the observational SM function for the star-forming galaxies and the initial and PhotReal post-processed SM distribution for any redshift bin. There exists also a good agreement between the observational SM function for the quiescent galaxies and both mock SM, at least for redshift higher than 1. Interestingly, the high mass end of the mass function obtained with PhotReal reproduces better the observational stellar mass function, compensating for the fact that the original mock catalogue shows lack of quiescent galaxies in this mass range.

4.6 Angular correlation function

As a final test for our mock catalogues, we have also compared the obtained angular correlation functions with observational data. For this purpose we used the public software athena\(^7\). Since the basis of our cosmological simulation is the Millennium simulation and we have not perturbed the galaxy positions, the overall correlation function should reflect reality. However, our aim is to test if the angular correlation function of passive and star-forming galaxies has unrealistically changed when applying PhotReal.

In Fig 15, we display the overall angular correlation function for the initial and the PhotReal post-processed mock limited to redshift 0.2. This is compared to the angular correlation functions obtained from the SDSS (Wang et al. 2013) split for different magnitude bins. For the sake of clarity, we have not split our sample in magnitude bins. However, we see that our estimation is well located within the different observational angular correlation functions.

Moreover, we have split the angular correlation functions for quiescent and star forming galaxies with the same criteria for classifying the initial and PhotReal post-processed photometry as in the previous section. In Figs. 16 and 17 we display the angular correlation function for the original and Euclid-Opt PhotReal mock catalogue for the quiescent and star-forming galaxies respectively for six different redshift bins. We observe that the angular correlation functions for the both types of galaxies are almost indistinguishable for the original and PhotReal post-processed catalogue, confirming the unalterability of the main properties of the mock galaxies.

\(^7\)http://www.cosmostat.org/athena.html
5 PHOTOMETRIC REDSHIFT PERFORMANCE

In this section, we investigate the behavior of the photometric redshifts obtained from the mock catalogues described in section §3 for each next generation survey in terms of photometric redshift resolution and catastrophic outliers rate.

5.1 Photometric Redshift Resolution

In order to obtain a robust measurement of the mean photometric redshift resolution, we have used the NMAD estimator following a similar approach as in Brammer et al. (2008); Molino et al. (2014). This estimator is known to be less sensitive to the behavior of the distribution. As such, the NMAD estimator is defined as:

$$\sigma_{NMAD} = 1.48 \times \text{median}\left(\frac{|\Delta z - \text{median}(\Delta z)|}{1 + z_s}\right)$$

where $\Delta z = z_{\text{phot}} - z_{\text{s}}$.

In Fig. 13 we show the density maps of the photometric redshift versus spectroscopic redshift for the three cases considered. We immediately notice several differences between the considered surveys. In the case of the LSST, the lack of infrared data gets compensated with the depth of the optical data, being able to reach higher redshift and magnitude limits with a photometric redshift precision of 0.045 for all the redshift range. We see some features at redshift $\sim$1.5, corresponding to the limit at which the spectrum of the objects is shifted to the infrared. The two cases considered for the Euclid survey sample well the optical and infrared spectrum providing a less biased and lower rates of catastrophic outliers than the other surveys. The expected accuracy for the Euclid pessimistic and optimistic case is $\sigma_{NMAD} = 0.053$ and 0.025 respectively. It is noticeable the higher dispersion found for the Euclid-Pes at lower redshift $<0.4$, which is probably due to the lack of UV data (i.e. Schmidt & Thornley 2013).

We have now investigated the photometric redshift performance as a function of the magnitude. The results are collected in Table I and in Fig. 14 we show the behavior of the photometric redshift bias as a function of magnitude. Clearly, the number of bands covering a wide range of the spectrum has a direct impact in the photometric redshift accuracy and photometric redshift bias. While the Euclid-Opt has photometric redshift precision of $<0.02$ down to $H = 24$ mag AB, we have an increase factor of 2 for the Euclid-Pes and the LSST cases for the same range of magnitudes. The same effect is observed in the photometric redshift bias.

We have also looked into the performance of the photometric redshifts for each different survey as a function of the redshift. The results are summarized in Table II and are displayed in Fig. 15. We find that the photometric redshift dispersion is always a factor of 1.5 higher for the Euclid-Pessimistic than for the LSST. This result remarks that the
presence of the infrared bands do not improve the photometric redshift resolution at $z < 1$. By contrast, the improvement in the photometric redshift performance found in using a deeper optical counterpart becomes remarkable when comparing the results of the two Euclid cases. The photometric redshift resolution of the Euclid-Optimistic is a factor of 2 to 5 better than in the Pessimistic case both in terms of photometric redshift accuracy and bias. In general, we find that the photometric redshift dispersion depends more strongly on magnitude than on redshift, in agreement with other works (Hildebrandt et al. 2012; Dahlen et al. 2013; Raichoor et al. 2014; Molino et al. 2014).

### Table 1. Photometric redshift comparison for different surveys as a function of magnitude

| $\Delta z$ | $\sigma_{\text{NMAD}}$ | $\eta$ | $\Delta z$ | $\sigma_{\text{NMAD}}$ | $\eta$ | $\Delta z$ | $\sigma_{\text{NMAD}}$ | $\eta$ |
|------------|-----------------|-------|------------|-----------------|-------|------------|-----------------|-------|
| 0.0339     | 0.0376          | 16.5  | 0.0313     | 0.0397          | 24.8292 | 0.0043     | 0.0145          | 0.1657 |
| 0.0292     | 0.0252          | 16.0  | 0.0326     | 0.0348          | 20.9461 | 0.0041     | 0.0140          | 0.2344 |
| 0.0191     | 0.0179          | 15.7  | 0.0312     | 0.0326          | 17.0296 | 0.0042     | 0.0136          | 0.1948 |
| 0.0161     | 0.0126          | 18.0  | 0.0126     | 0.0303          | 13.5556 | 0.0035     | 0.0141          | 0.1644 |
| 0.0275     | 0.0298          | 17.0  | 0.0228     | 0.0314          | 8.5210  | 0.0004     | 0.0152          | 0.1583 |
| 0.0186     | 0.0161          | 17.5  | 0.0126     | 0.0031          | 7.0722  | 0.0031     | 0.0153          | 0.1761 |
| 0.0269     | 0.0212          | 19.0  | 0.0298     | 0.0314          | 5.8210  | 0.0004     | 0.0151          | 0.1617 |
| 0.0265     | 0.0298          | 20.0  | 0.0306     | 0.0318          | 4.9948  | 0.0002     | 0.0150          | 0.1478 |
| 0.0254     | 0.0298          | 21.0  | 0.0292     | 0.0419          | 4.1289  | 0.0033     | 0.0150          | 0.1408 |
| 0.0245     | 0.0298          | 21.5  | 0.0280     | 7.4709          | 3.3720  | 0.0035     | 0.0150          | 0.1295 |
| 0.0238     | 0.0298          | 22.0  | 7.4808     | 2.7919          | 0.0039  | 0.0276     | 0.0150          | 0.1286 |
| 0.0235     | 0.0298          | 22.5  | 7.3433     | 2.3916          | 0.0043  | 0.0285     | 0.0152          | 0.1448 |
| 0.0235     | 0.0298          | 23.0  | 7.1155     | 2.1944          | 0.0054  | 0.0285     | 0.0158          | 0.1927 |
| 0.0237     | 0.0298          | 23.5  | 7.9340     | 2.2293          | 0.0078  | 0.0285     | 0.0173          | 0.3503 |
| 0.0240     | 0.0298          | 24.0  | 12.2873    | 2.5102          | 0.0110  | 0.0480     | 0.0201          | 0.8595 |
| 0.0247     | 0.0298          | 24.5  | 21.7991    | 3.1855          | 0.0155  | 0.0691     | 0.0251          | 2.4388 |
| 0.0265     | 0.0298          | 25.0  | 29.7845    | 4.4840          | 0.0282  | 0.0882     | 0.0302          | 4.6215 |
| 0.0308     | 0.0298          | 25.5  | --         | 6.7209          | --      | --         | --              | --     |
| 0.0383     | 0.0298          | 26.0  | --         | 10.0954         | --      | --         | --              | --     |
of each survey in Tables 1 and 2 respectively. In Fig 21, we show the outlier rate as a function of redshift for the different surveys.

While the Euclid-Opt survey has a very lower rate of catastrophic outliers as a function of magnitude, being less than 1% down to \( H = 23 \) mag, the LSST and Euclid-Pes show the highest rate of outliers between 5-12%. This separation illustrates that the multiple (\( > 12 \)) band surveys are much more effective in decreasing the rate of catastrophic outliers than surveys with fewer bands.

Looking at the same rate as a function of redshift, we see a milder dependence of the outlier rate and redshift. The rate of outliers remains close to \( \sim 1\%-2\% \) for the Eucl-Opt, \( \sim 1\%-2\% \) for the J-PAS, and \( \sim 1\%-2\% \) for the LSST.

The term 'outliers' refers to those galaxies that, due to the well-known color-redshift degeneracies, gets assigned a very similar color to a theoretical galaxy at a very different redshift. Even if the Bayesian probability alleviates this problem due to the introduction of a Bayesian prior (Benítez 2004, Bellagamba et al. 2012), the existence of such galaxies still remains and we need to quantify it.

One necessary requirement for good quality photometric redshifts is to have a small percentage of the total sample of galaxies becoming 'outliers'. We define fraction of outliers as in [Hildebrandt et al. (2010)]:

\[
\eta = \frac{\Delta z}{1 + z_s} > 0.15
\]  

The results obtained for different surveys as a function of magnitude and redshift are collected in the third column of each survey in Tables 1 and 2 respectively. In Fig 21, we define fraction of outliers as in [Hildebrandt et al. (2010)]:

\[
\eta = \frac{\Delta z}{1 + z_s} > 0.15
\]  

The term 'outliers' refers to those galaxies that, due to the well-known color-redshift degeneracies, gets assigned a very similar color to a theoretical galaxy at a very different redshift. Even if the Bayesian probability alleviates this problem due to the introduction of a Bayesian prior (Benítez 2004, Bellagamba et al. 2012), the existence of such galaxies still remains and we need to quantify it.

One necessary requirement for good quality photometric redshifts is to have a small percentage of the total sample of galaxies becoming 'outliers'. We define fraction of outliers as in [Hildebrandt et al. (2010)]:

\[
\eta = \frac{\Delta z}{1 + z_s} > 0.15
\]

The results obtained for different surveys as a function of magnitude and redshift are collected in the third column of each survey in Tables 1 and 2 respectively. In Fig 21, we define fraction of outliers as in [Hildebrandt et al. (2010)]:

\[
\eta = \frac{\Delta z}{1 + z_s} > 0.15
\]

The results obtained for different surveys as a function of magnitude and redshift are collected in the third column of each survey in Tables 1 and 2 respectively. In Fig 21, we define fraction of outliers as in [Hildebrandt et al. (2010)]:

\[
\eta = \frac{\Delta z}{1 + z_s} > 0.15
\]
I. Apples to apples $A^2$: photometric redshift predictions for next-generation surveys

Figure 19. Density maps of magnitude versus the photometric bias for the three different survey considered in this work (Euclid pessimistic, top left; Euclid optimistic, bottom left and LSST, top right). The color scale are the same as in Fig. 18. The axis scale are the same for all the panels. The behavior of the photometric redshifts for the LSST and both Euclid cases show larger dispersions and biases down to deeper magnitudes, having the Euclid Optimistic case a significantly smaller dispersion than the Pessimistic case and the LSST.

2-10% for the LSST and $\sim 10-20\%$ for the Euc-Pes survey. As before, the presence of the deep infrared data and the combination with deep optical data is a strong combination to decrease the rate of catastrophic outliers.

6 PHOTOMETRIC REDSHIFT QUALITY CUTS

As we have illustrated in the previous section, the existence of outliers in a photometric survey is unavoidable but can be quantified and keep under a moderate threshold. In addition, we can use one main advantage of some photometric redshift packages such as BPZ2.0, which is the fact that all the galaxies have assigned a redshift probability function. The study of this function provides a measure of the quality of the photometric redshift: very well behaved photometry redshifts will probably have a well defined narrow peak around the spectroscopic redshift of the galaxy, whereas galaxies with degeneracies will be assigned a multi peak probability distribution.

Benítez (2000) introduced the odds parameter: a measurement of the quality of a photometric redshift. This odds parameter is obtained by integrating the redshift probability function of the galaxy within $2\sigma$ of its more probable value. These parameters has been widely used to select the best quality photometric redshift sample to study a number of different features (e.g. Coe et al. 2006; Benítez et al. 2009; Bellagamba et al. 2012; Molino et al. 2014; Ascaso et al. 2015b) and allows to improve the overall photometric redshift precision of the sample by previously modeling the selection bias.

In this section, we discuss the benefit obtained in terms of the improvement of the photometric redshift accuracy and the decrement of the outlier rates when using this parameter to select the best quality photometric redshift sample for a particular survey.

6.1 Improvement in the photometric redshift resolution

In order to illustrate the increase in photometric redshift resolution by effectuating a cut in the odds parameter, we plot in Fig. 22 the photometric redshift accuracy as a function of the odds parameter for each of the three different surveys considered in this study. As expected, we see a decrease of the photo-z dispersion as the odds increase for all three surveys. Indeed, we can select a subsample of the accuracy desired by performing an accurate odds cut. For instance, while the overall accuracy of LSST is 0.045, we could obtain a subsample of twice this accuracy ($\sigma_{NMAD} \sim 0.02-0.4$) by selecting the 25% highest odds subsample.

We also illustrate the effect of this selection by plotting the analogous plot to the Fig. 18 in Fig. 23, the behavior of the photometric redshift as a function of spectroscopic redshift for the 50% best quality photometric redshifts for each of the surveys considered. The number quoted in the legend

Figure 20. Density maps of redshift versus the photometric bias for the J-PAS, upper left; LSST, upper right; Euclid pessimistic, bottom left and Euclid optimistic, bottom right. The color scale are the same as in Fig. 18. The axis scale are the same for all the panels. The performance of the photometric bias for the J-PAS is, in general terms, better than for the other surveys, but the redshift range is limited as well. The behavior for the LSST and Euclid-Pessimistic case becomes similar (less than a factor of 2) up to redshift 1.5. At higher redshift, the bias for the Euclid-Pessimistic case, than the Euclid Optimistic case, which obtains an overall good performance ($\sigma_{NMAD} \sim 0.02-0.4$) and negligible photometric redshift bias up to redshift 3.
Figure 21. Fraction of outliers as a function of the redshift for the three different survey considered in this work (LSST, red line; Euclid pessimistic, blue line and Euclid optimistic, green line). The solid lines refer to the overall population for each survey while the dotted lines refer to the 50% of the population with highest odds for each different survey. In general, the fraction of outliers increases gently with redshift for all the surveys considered here, obtaining the lower rates for Euclid Optimistic, Euclid Pessimistic, and LSST in increasing order. If we consider the best 50% quality odds sample, the outliers rate decreases more than a factor of 10.

Figure 22. Odds photometric redshift quality parameter versus photometric redshift dispersion for the three different survey considered in this work (LSST, red dotted line; Euclid pessimistic, blue dashed line and Euclid optimistic, green dotted-dashed line). Performing a safe quality odds parameter in each of these surveys can decrease the photometric redshift dispersion significantly.

Figure 23. Density maps of spectroscopic redshift versus photometric redshift for the 50% best quality photometric redshift for each different survey considered for all the simulation (LSST, upper right; Euclid pessimistic, bottom left and Euclid optimistic, bottom right). The color scale are the same as in Fig. 18. The overall photometric redshift precision for each survey is quoted in each panel. The photometric redshift resolution increases between a factor of 1.5 and 3 for all surveys by selecting a best quality photometric redshift subsample, being particularly affected the density regions of outliers.

6.2 Decrease of outlier rates

As mentioned in the previous section, in Fig. 21 we have illustrated the relation between the outlier rates as a function of redshift with solid lines for each of the different surveys. In this Figure, we have also plotted the outlier rates resulting for the 50% best photometric redshift quality subsample for every survey. We plainly observe a dramatic descend of the outlier rate for each survey. The new rates are decreased by a factor of 3 to 10, putting in evidence the virtues of this selection.

When working in a particular science topic, one needs to understand the nature of any possible selection based on the odds parameter and quantify the possible bias introduced on the final scientific results. Once this effect is taken into account, the improvement in photometric precision and the decrease of outliers rate can be fully exploited and we encourage researchers to exploit this feature.
7 DISCUSSION AND CONCLUSIONS

The advent of several large deep surveys is imminent and the need of precise tools to describe accurately their properties is pressing. In this work, we have developed mock catalogues mimicking realistically two of the upcoming generation IV surveys: LSST and Euclid (considering two cases for the optical counterpart, optimistic and pessimistic). We have confirmed that the main properties of this new mock reflect reality in terms of photometry, photometric errors, stellar masses, spectral types and photometric redshifts by comparing with observable relations. We have also provided predictions in terms of photometric redshift performance and selection cuts from the same basis for each survey. The mocks will become publicly available and they will become extremely useful for making internal comparisons between the different surveys.

Each of these surveys will provide output data observed with very different strategies. In this work, we have disengaged some of the advantages of using each of these different datasets in terms of photometric redshift resolution. First, the number of bands used to sample the optical and/or infrared spectrum is directly related to the level of photometric redshift accuracy and photometric redshift bias as confirmed by other works (e.g. Benítez et al. 2009a and references herein). On the other hand, deep broad-band optical surveys allow us to sample the luminosity function in the optical range even if with worse photometric redshift resolution. They are also an excellent complement for deep infrared surveys, particularly to dramatically decrease the resolution. They are also an excellent complement for deep infrared surveys, particularly to dramatically decrease the resolution. They are also an excellent complement for deep infrared surveys, particularly to dramatically decrease the resolution.

As stated in the introduction, a separate work (Ascaso et al. in prep) will be fully devoted to explore the photometric redshift expectations for J-PAS, a multiple narrow-band survey. We have then examined the performance of the photometric redshifts for a new ‘super-survey’ consisting in combining all the expected data for these surveys once they are made public and we are allowed to combine it. This ‘super-survey’ will consist then in the combination of J-PAS+LSST+DES+Euclid and the results should account for all the advantages described through this work. In Fig.24 we plot the density map of the photometric redshift versus the spectroscopic redshift for this ‘super-survey’.

The expected overall photometric redshift resolution for all the redshift range within $0 < z < 3$ and $i < 27.5$ AB will be of $\sigma_{NMAD} = 0.0135$, with almost negligible photometric redshift bias. As for the rate of catastrophic outliers, we plot in Fig.25 the expected outlier rate for this ‘super-survey’ as a function of magnitude and redshift. As we see, this survey will keep the outliers rate below 1% up to redshift $\sim 2$ and below 3% for all the redshift and magnitude range.

Hence, this future ‘super survey’ will support almost all the scientific cases for the community, becoming a superior machine comparable to a full spectroscopic survey ranging a very large range of redshift and magnitude.

In a forthcoming paper (Ascaso et al., in prep), we will continue this work within the frame of Apples to Apples, A$^2$ to explore and compare the expected selection function for cluster and group detections for the three surveys considered here. We will also publish a third paper providing cosmological constraints from these cluster counts for different surveys. These series of works will help setting tight on the expected predictions of these surveys and the mocks...
will be useful to set requirements to the science derived from them.

ACKNOWLEDGMENTS

BA acknowledges financial support for a postdoctoral fellowship from the Observatory of Paris. SM acknowledges financial support from the Institut Universitaire de France (IUF), of which she is senior member. We acknowledge Alex Merson and Carlton Baugh for useful discussions and comments to our work. BA acknowledges Rossella Licitra and Anand Raichoor for multiple discussions that helped to improve this paper.

REFERENCES

Abdalla, F., Annis, J., Bacon, D., et al. 2012, arXiv:1209.2451
Albrecht, A., Bernstein, G., Cahn, R., et al. 2006, arXiv:astro-ph/0609591
Albrecht, A., Amendola, L., Bernstein, G., et al. 2009, arXiv:0901.0721
Arnalte-Mur, P., Martínez, V. J., Norberg, P., et al. 2014, MNRAS, 441, 1783
Ascaso, B., Moles, M., Aguerri, J. A. L., Sánchez-Janssen, R., & Varela, J. 2008, A&A, 487, 453
Ascaso, B., Wittman, D., & Benítez, N. 2012, MNRAS, 420, 1167
Ascaso, B., Wittman, D., & Dawson, W. 2014a, MNRAS, 439, 1980
Ascaso, B., Lemaux, B. C., Lubin, L. M., et al. 2014b, MNRAS, 442, 589
Banerji, M., Abdalla, F. B., Lahav, O., & Lin, H. 2008, MNRAS, 386, 1219
Bellagamba, F., Meneghetti, M., Moscardini, L., & Bolzonella, M. 2012, MNRAS, 422, 553
Benítez, N. 2000, ApJ, 536, 571
Benítez, N., Gaztañaga, E., Miquel, R., et al. 2009, ApJ, 691, 241
Benítez, N., Moles, M., Aguerri, J. A. L., et al. 2009, ApJ, 692, L5
Benítez, N., Dupke, R., Moles, M., et al. 2014, arXiv:1403.5237
Bower, R. G., Benson, A. J., Malbon, R., et al. 2006, MNRAS, 370, 645
Brammer, G. B., van Dokkum, P. G., & Coppi, P. 2008, ApJ, 686, 1503
Brodwin, M., Ruel, J., Ade, P. A. R., et al. 2010, ApJ, 721, 90
Bruzual, G., & Charlot, S. 2003, MNRAS, 344, 1000
Calzetti, D., Armus, L., Bohlin, R. C., et al. 2000, ApJ, 533, 682
Chabrier, G. 2003, PASP, 115, 763
Chang, C., Busha, M. T., Wechsler, R. H., et al. 2014, arXiv:1411.0052
Cirasuolo, M., McLure, R. J., Dunlop, J. S., et al. 2010, MNRAS, 401, 11
Coe, D., Benítez, N., Sánchez, S. F., et al. 2006, AJ, 132, 926
Cohn, J. D., Evrard, A. E., White, M., Croton, D., & Ellingson, E. 2007, MNRAS, 382, 1738
Cole, S., Lacey, C. G., Baugh, C. M., & Frenk, C. S. 2000, MNRAS, 319, 168
Dahlen, T., Mobasher, B., Faber, S. M., et al. 2013, ApJ, 775, 93
De Lucia, G., & Borgani, S. 2012, MNRAS, 426, L61
Foley, R. J., Andersson, K., Bazin, G., et al. 2011, ApJ, 731, 86
Fontanot, F., De Lucia, G., Monaco, P., Somerville, R. S., & Santini, P. 2009, MNRAS, 397, 1776
Franx, M., Moorwood, A., Rix, H.-W., et al. 2000, The Messenger, 99, 20
Gabasch, A., Hopp, U., Feulner, G., et al. 2006, A&A, 448, 101
Gawiser, E., van Dokkum, P. G., Herrera, D., et al. 2006, ApJS, 162, 1
Gonzalez-Perez, V., Lacey, C. G., Baugh, C. M., et al. 2014, MNRAS, 439, 264
Gorecki, A., Abate, A., Ansari, R., et al. 2014, A&A, 561, A128
Guo, Q., Cole, S., Eke, V., Frenk, C., & Helly, J. 2013, MNRAS, 434, 1838
Guo, Y., Ferguson, H. C., Giavalisco, M., et al. 2013, ApJS, 207, 24
Hildebrandt, H., Arnouts, S., Capak, P., et al. 2010, A&A, 523, A31
Hansson, K. S. A., Lisker, T., & Grebel, E. K. 2012, MNRAS, 427, 2376
Henriques, B. M. B., White, S. D. M., Lemson, G., et al. 2012, MNRAS, 421, 2904
Henriques, B. M. B., White, S. D. M., Thomas, P. A., et al. 2013, MNRAS, 431, 3373
Hildebrandt, H., Erben, T., Kuijken, K., et al. 2012, MNRAS, 421, 2355
Hoyle, B., Jimenez, R., Verde, L., & Hotchkiss, S. 2012, J. Cosmology Astropart. Phys., 2, 9
Ilbert, O., Arnouts, S., McCracken, H. J., et al. 2006, A&A, 457, 841
Ilbert, O., Capak, P., Salvato, M., et al. 2009, ApJ, 690, 1236
Ivezic, Z., Tyson, J. A., Acosta, E., et al. 2008, arXiv:0805.2366
Jee, M. J., Rosati, P., Ford, H. C., et al. 2009, ApJ, 704, 672
Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, arXiv:1110.3193
Levi, M., Bebek, C., Beers, T., et al. 2013, arXiv:1308.0847
LSST Science Collaboration, Abell, P. A., Allison, J., et al. 2009, arXiv:0912.0201
McDonald, M., Bayliss, M., Benson, B. A., et al. 2012, Nature, 488, 349
Mei, S., Holden, B. P., Blakeslee, J. P., et al. 2009, ApJ, 690, 42
Menanteau, F., Hughes, J. P., Sifón, C., et al. 2012, ApJ, 748, 7
Merson, A. I., Baugh, C. M., Helly, J. C., et al. 2013, MNRAS, 429, 556
Merson, A. I., Baugh, C. M., Abdalla, F. B., et al. 2015, arXiv:1502.04984
Mitchell, P. D., Lacey, C. G., Baugh, C. M., & Cole, S. 2013, MNRAS, 435, 87
I. Apples to apples $A^2$: photometric redshift predictions for next-generation surveys

Moles, M., Benítez, N., Aguerri, J. A. L., et al. 2008, AJ, 136, 1325
Mohr, J. J., Armstrong, R., Bertin, E., et al. 2012, Proc. SPIE, 8451,
Molino, A., Benítez, N., Moles, M., et al. 2014, MNRAS, 441, 2891
Montero-Dorta, A. D., & Prada, F. 2009, MNRAS, 399, 1106
Muzzin, A., Marchesini, D., Stefanon, M., et al. 2013, ApJ, 777, 18
Oyaizu, H., Lima, M., Cunha, C. E., Lin, H., & Frieman, J. 2008, ApJ, 689, 709
Papovich, C., Momcheva, I., Willmer, C. N. A., et al. 2010, ApJ, 716, 1503
Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2013, arXiv:1303.5076
Raichoor, A., Mei, S., Erben, T., et al. 2014, arXiv:1410.2276
Schmidt, S. J., & Thorman, P. 2013, MNRAS, 431, 2766
Skelton, R. E., Bell, E. F., & Somerville, R. S. 2012, ApJ, 753, 44
Somerville, R. S., Gilmore, R. C., Primack, J. R., & Domínguez, A. 2012, MNRAS, 423, 1992
Spergel, D. N., Verde, L., Peiris, H. V., et al. 2003, ApJS, 148, 175
Springel, V., White, S. D. M., Jenkins, A., et al. 2005, Nature, 435, 629
Stefanon, M., & Marchesini, D. 2013, MNRAS, 429, 881
The Dark Energy Survey Collaboration 2005, arXiv:astro-ph/0510346
Wang, Y., Brunner, R. J., & Dolence, J. C. 2013, MNRAS, 432, 1961
Weinmann, S. M., Lisker, T., Guo, Q., Meyer, H. T., & Janz, J. 2011, MNRAS, 416, 1197
Weinmann, S. M., Pasquali, A., Oppenheimer, B. D., et al. 2012, MNRAS, 426, 2797
Wittman, D. M., Tyson, J. A., Dell’Antonio, I. P., et al. 2002, Proc. SPIE, 4836, 73
Wuyts, S., Labbé, I., Schreiber, N. M. F., et al. 2008, ApJ, 682, 985
Zandivarez, A., Díaz-Giménez, E., Mendes de Oliveira, C., et al. 2014, A&A, 561, A71