

# The impact from survey depth and resolution on the morphological classification of galaxies

M. Pović,<sup>1\*</sup> I. Márquez,<sup>1\*</sup> J. Masegosa,<sup>1</sup> J. Perea,<sup>1</sup> A. del Olmo,<sup>1</sup> C. Simpson,<sup>2</sup> J. A. L. Aguerra,<sup>3</sup> B. Ascaso,<sup>1,4</sup> Y. Jiménez-Teja,<sup>5</sup> C. López-Sanjuan,<sup>6</sup> A. Molino,<sup>1,7</sup> A. M. Pérez-García,<sup>3,8</sup> K. Viironen,<sup>6</sup> C. Husillos,<sup>1</sup> D. Cristóbal-Hornillos,<sup>6</sup> C. Caldwell,<sup>2</sup> N. Benítez,<sup>1</sup> E. Alfaro,<sup>1</sup> T. Aparicio-Villegas,<sup>1,5</sup> T. Broadhurst,<sup>9,10</sup> J. Cabrera-Caño,<sup>11</sup> F. J. Castander,<sup>12</sup> J. Cepa,<sup>3,8</sup> M. Cerviño,<sup>1,3</sup> A. Fernández-Soto,<sup>13,14</sup> R. M. González Delgado,<sup>1</sup> L. Infante,<sup>15</sup> V. J. Martínez,<sup>14,16</sup> M. Moles,<sup>1,6</sup> F. Prada<sup>1</sup> and J. M. Quintana<sup>1</sup>

*Affiliations are listed at the end of the paper*

Accepted 2015 July 21. Received 2015 July 21; in original form 2014 November 14

## ABSTRACT

We consistently analyse for the first time the impact of survey depth and spatial resolution on the most used morphological parameters for classifying galaxies through non-parametric methods: Abraham and Conselice–Bershady concentration indices, Gini, M20 moment of light, asymmetry, and smoothness. Three different non-local data sets are used, Advanced Large Homogeneous Area Medium Band Redshift Astronomical (ALHAMBRA) and Subaru/XMM-Newton Deep Survey (SXDS, examples of deep ground-based surveys), and Cosmos Evolution Survey (COSMOS, deep space-based survey). We used a sample of 3000 local, visually classified galaxies, measuring their morphological parameters at their real redshifts ( $z \sim 0$ ). Then we simulated them to match the redshift and magnitude distributions of galaxies in the non-local surveys. The comparisons of the two sets allow us to put constraints on the use of each parameter for morphological classification and evaluate the effectiveness of the commonly used morphological diagnostic diagrams. All analysed parameters suffer from biases related to spatial resolution and depth, the impact of the former being much stronger. When including asymmetry and smoothness in classification diagrams, the noise effects must be taken into account carefully, especially for ground-based surveys. M20 is significantly affected, changing both the shape and range of its distribution at all brightness levels. We suggest that diagnostic diagrams based on 2–3 parameters should be avoided when classifying galaxies in ground-based surveys, independently of their brightness; for COSMOS they should be avoided for galaxies fainter than  $F814 = 23.0$ . These results can be applied directly to surveys similar to ALHAMBRA, SXDS and COSMOS, and also can serve as an upper/lower limit for shallower/deeper ones.

**Key words:** surveys – galaxies: fundamental parameters.

## 1 INTRODUCTION

Morphology is one of the main characteristics of galaxies, and the morphological classification has been central to many advances in the picture of galaxy formation and evolution. Different correlations between morphology and other galaxy properties have been studied,

including the relation with stellar mass (e.g. Deng 2013), colour (e.g. Strateva et al. 2001; Bell et al. 2003; Hogg et al. 2003; Baldry et al. 2004; Weiner et al. 2005; Cirasuolo, Magliocchetti & Celotti 2005; Melbourne et al. 2007; Cassata et al. 2007; Pović et al. 2013), luminosity (e.g. Blanton et al. 2003; Kelm, Focardi & Sorrentino 2005), environment (e.g. Cassata et al. 2007), black hole mass (e.g. Kormendy & Richstone 1995; McLure, Dunlop & Kukula 2000; Graham et al. 2001b; Marconi & Hunt 2003), nuclear activity (e.g. Adams 1977; Heckman 1978; Ho, Filippenko & Sargent 1995;

\* E-mail: mpovic@iaa.es (MP); isabel@iaa.es (IM)

Kauffmann et al. 2003; Pierce et al. 2007; Choi, Woo & Park 2009; Gabor et al. 2009; Pović et al. 2012), and X-ray properties (e.g. Hickox et al. 2009; Pović et al. 2009b,a).

The methods of morphological classification of galaxies can be separated into three groups: (i) visual (e.g. Lintott et al. 2008, 2011; Nair & Abraham 2010; Baillard et al. 2011; Willett et al. 2013), for classifying the nearby and well-resolved galaxies, (ii) parametric, based on the galaxy physical (e.g. de Vaucouleurs 1948; Sérsic 1963; Peng et al. 2002, 2010; Simard et al. 2002, 2011; de Souza, Gadotti & dos Anjos 2004; Barden et al. 2012) and mathematical parameters, where assuming an analytic model for fitting the galaxy (e.g. Kelly & McKay 2005, 2004; Ngan et al. 2009; Andrae, Jahnke & Melchior 2011a; Andrae, Melchior & Jahnke 2011b; Jiménez-Teja & Benítez 2012), they perform the decomposition of well-resolved galaxies and provide the properties of different structures, and finally, (iii) non-parametric, which does not assume any particular analytic model.

The non-parametric methods are based on measuring the different galaxy quantities that correlate with the morphological types, i.e. colours (e.g. Strateva et al. 2001), spectral properties (e.g. Humason 1931; Morgan & Mayall 1957; Baldwin, Phillips & Terlevich 1981; Folkes, Lahav & Maddox 1996; Sánchez-Almeida et al. 2010), or light distribution (e.g. Doi, Fukugita & Okamura 1993; Abraham et al. 1994, 1996; Bershadly, Jangren & Conselice 2000; Conselice et al. 2000; Graham, Trujillo & Caon 2001a; Abraham, van den Bergh & Nair 2003; Lotz, Primack & Madau 2004; Yamauchi et al. 2005). The non-parametric methods are less time-consuming in comparison with other methods, and can provide an easy and fast separation between regular and irregular (or disturbed) sources and/or early- and late-type galaxies, down to intermediate redshifts ( $\sim 1.5$ ), or higher if dealing with space based, or good seeing ground-based data. Over the past years different morphological diagrams have been applied, relating galaxy light concentration and asymmetry or smoothness parameters either to classify galaxies (Abraham et al. 1994; Bershadly et al. 2000; Conselice et al. 2000; Graham et al. 2001a; Conselice 2006; Cassata et al. 2007; Pović et al. 2009b; Deng 2013), or to select merger candidates (Urrutia, Lacy & Becker 2008; Lotz et al. 2010a,b; Villforth et al. 2014). However, some of the previous works showed that these parameters can depend significantly on e.g. the aperture definition for measuring the galaxy flux/radius (Graham et al. 2001a; Strateva et al. 2001; Lisker 2008), and/or on the resolution and signal-to-noise ratio (hereafter S/N; Conselice et al. 2000; Graham et al. 2001a; Lisker 2008; Huertas-Company et al. 2009; Andrae et al. 2011a,b; Pović et al. 2012; Carollo et al. 2013; Cibinel et al. 2013; Petty et al. 2014). Different trends of these parameters can be observed in relation to the data quality, spatial resolution and depth. Since a reliable morphological classification is essential for galaxy formation and evolution studies, it is crucial to disentangle how strong is this impact<sup>1</sup> on each morphological parameter when dealing with different data sets. In most of previous works, this observational bias was analysed either for a particular parameter, either for a particular observational condition (survey), or for a particular sample of galaxies. In this work we present, for the first time, a systematic study, where in a consistent way the analysis was carried out for all parameters, on large sample of galaxies, and in relation with different observational conditions.

<sup>1</sup> Hereafter, when using the expression ‘observational bias’ we will refer to the impact of spatial resolution and data depth on analysed morphological parameters.

We analysed the observational bias that might affect each of the six commonly used morphological parameters: CABR, Gini coefficient, Conselice–Bershadly concentration index, M20 moment of light, asymmetry index, and smoothness. We studied how this bias depends on the spatial resolution and magnitude/redshift distributions of galaxies, and how it affects the diagnostic diagrams used to classify galaxies. We used a visually classified sample of local galaxies and simulated them to map the observational conditions of three different ground- and space-based deep surveys: the Advanced Large Homogeneous Area Medium Band Redshift Astronomical survey (Moles et al. 2008, ALHAMBRA), the Subaru/XMM-Newton Deep Survey (Furusawa 2008, SXDS), and the Cosmos Evolution Survey (Scoville et al. 2007, COSMOS). With this analysis our main goal is to measure a set of morphological parameters in the real (local) conditions and in the simulated conditions of non-local surveys in order to study how the spatial resolution and data depth affect each parameter and the commonly used morphological diagrams. We emphasize that is out of our aim to classify the galaxies in the three selected non-local surveys, neither reclassify local galaxies once they were scaled to map the conditions of ALHAMBRA, SXDS, and COSMOS. Comparing the parameters of local galaxies measured before and after shifting them in redshift and magnitude, we were able to put constraints on the main diagrams, observing how the position and the shape of the regions typical of early- and late-type galaxies change in the local and non-local conditions. Moreover, analysing the morphological diagrams in ground-based and space-based surveys, and at different magnitude cuts, we quantified how strong is the impact from spatial resolution and survey depth, respectively. The results obtained in this work can be applied to surveys similar to ALHAMBRA, SXDS, and COSMOS.

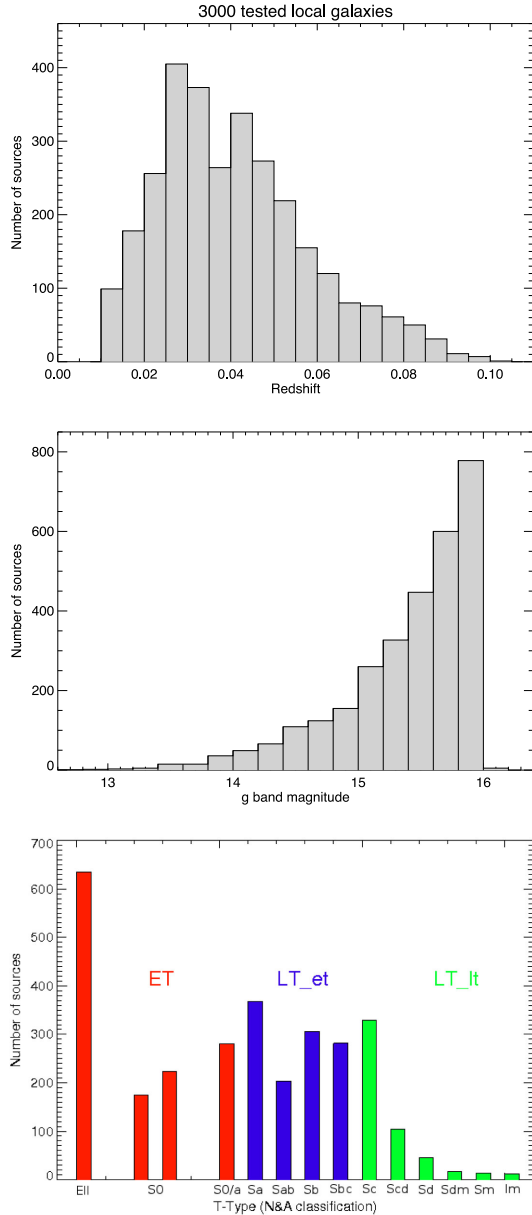
The paper is organized as follows: in Section 2, we described the used local and high-redshift samples and the corresponding surveys. The applied methodology is described in Section 3. The results obtained for local and simulated galaxies are showed in Section 4. In Section 5, we discussed how the impact from data depth and spatial resolution affects the morphological parameters and commonly used diagnostic diagrams for galaxy classification. Finally, we summarized our results in Section 6.

We assume the following cosmological parameters throughout the paper:  $\Omega_\Lambda = 0.7$ ,  $\Omega_M = 0.3$ , and  $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$ . Unless otherwise specified, all magnitudes are given in the AB system (Oke & Gunn 1983).

## 2 THE DATA

### 2.1 Local sample

To test how the observational bias affects the analysed morphological parameters, we used a sample of 3000 local galaxies at  $0.01 \leq z \leq 0.1$  (with a mean redshift of 0.04), observed in the Sloan Digital Sky Survey (SDSS) Data Release 4 (DR4) down to an apparent extinction-corrected magnitude of  $g < 16$ , and visually classified by Nair & Abraham (2010; hereafter **N&A**), using the  $g$  and  $r$  bands. The number of local galaxies is selected as a compromise between the computing time and the classification accuracy, since the computing time to train the Support Vector Machine (SVM) through the `GALSVM` code used in this work (see Section 3) is totally dependent and very sensitive to the size of the training data set (Huertas-Company et al. 2008, 2009). The galaxies were selected randomly out of  $\sim 14\,000$  sources contained in the **N&A** catalogue, making sure that the selected subsample is



**Figure 1.** Redshift (top),  $g$ -band magnitude (middle), and morphological type (bottom) distributions of the local sample used in this work.

representative in terms of the general properties of the whole data set (see Pović et al. 2013, and their fig. 3):  $g-r$  colour, morphological classification, and inclination in the case of the late-type galaxies. On the other hand, Fig. 1 shows the redshift,  $g$ -band magnitude, and N&A morphological classification (T-Type in their work) distributions of the selected local sample. As can be seen from the T-Type histogram, the selected sample occupies all range of morphologies, from elliptical to irregular galaxies. In this work, all analyses were performed dividing galaxies into three morphological groups:

- (i) *early-type galaxies (hereafter ET)* - with T-Type  $\leq 0$ , including elliptical (c0, E0, E+), lenticular (S0-, S0, S0+) and S0/a galaxies from the N&A classification,
- (ii) *early spirals (hereafter LT\_et)* - with  $0 < \text{T-Type} \leq 4$ , including Sa, Sab, Sb, and Sbc, and

- (iii) *late spirals and irregular galaxies (hereafter LT\_it)* - with T-Type  $> 4$ , including Sc, Scd, Sd, Sdm, Sm, and Im galaxies.

We stress that throughout this work, we do not re-classify local galaxies and every time we specify the morphological type, it refers to that from N&A classification. Here, we want to measure a set of the most used morphological parameters in real conditions of the local sample and in the simulated conditions of non-local surveys in order to evaluate how the spatial resolution and data depth affect each parameter. ET, LT\_et, and LT\_it galaxies represent 45, 38, and 17 per cent of the selected local sample, respectively, completely consistent with the whole N&A sample.

Irregular galaxies constitute only  $< 1$  per cent of all sources in the N&A catalogue. With such a small population, we were not able to provide reliable statistical analyses in comparison to the other two broader groups, Ell/S0 and spirals. This is the reason why we do not study irregulars separately, and divide instead all late-type galaxies into two subgroups: earlier and later. In our LT\_it defined group, galaxies classified as Irr by N&A form only 2 per cent of the whole population, being therefore insignificant for affecting the results represented in this paper.

## 2.2 Non-local samples

In this section, we describe the properties of non-local surveys used for studying the impact from spatial resolution and data depth on the morphological classification of galaxies. Since our work is statistical, to select the surveys we considered the following criteria:

- (i) a large number of detected galaxies, either a deep survey covering a smaller area, and/or a shallower survey covering a large area,
- (ii) surveys with accurate photometry as well as accurate photometric redshift measurements available, and
- (iii) surveys with different properties in terms of spatial resolution and depth.

Finally, as a result of these criteria, we selected three non-local surveys for our analysis: deep ground-based surveys ALHAMBRA and SXDS (ALHAMBRA being shallower than SXDS, but with a large covered area and accurate photometric and photometric redshift measurements, and SXDS as an example of the deepest available ground-based data), and COSMOS as an example of deep space-based surveys. We had to take into account the number of selected surveys, since for each of them the applied computational procedure is time consuming. In the following sections, we describe each of these surveys. Their basic properties are summarized in Table 1. We cannot compare directly the results obtained between the three analysed non-local surveys: first, because they have different depth, so the analysed magnitude cuts are different (see below), and secondly they have different photometric information available, so the bands that we used are different as well. We tested each survey separately, and tried to represent the obtained results depending on the properties of each one.

### 2.2.1 ALHAMBRA

The ALHAMBRA survey (Moles et al. 2008) is a photometric survey that imaged  $\sim 3.5 \text{ deg}^2$  of the sky, in eight different fields, through 20 optical and three near-infrared (NIR) filters. The average spatial resolution of the ALHAMBRA images is  $\sim 1$  arcsec (not higher than 1.6, and ranging mainly between 0.8 and 1.2 arcsec in the  $F613W$  band; see table 1 in Pović et al. 2013, for more details).

**Table 1.** Summary of the ground- and space-based high-redshift data selected for studying the observational bias of morphological parameters. Column description: Survey – selected high-redshift surveys; tot. A/used A – the total covered area of each survey, and the area that covers the data used in this work; Sp. Res. – the average survey spatial resolution; Lim. mag/Sel. sam. (band) – limiting survey magnitude, and limiting magnitude of the selected sample in the correspondent band; Num. of gal – the final number of selected galaxies used in this work to obtain the magnitude and redshift distributions in each survey, and move later the local galaxies to these conditions (see Sections 3 and 4); mag\_sim and z\_sim – magnitude cuts and the corresponding redshifts (for 95 per cent of the sample) used to simulate the local galaxies.

Survey	Tot. A/used A (deg <sup>2</sup> )	Sp. Res. (arcsec)	Lim_mag/Sel. sam. (band)	Num of gal	mag_sim	z_sim
ALHAMBRA	4.0 / ~ 3.0	1.1	25.0 / 23.0 (F613W)	>43 000	≤20.0, ≤21.5, ≤23.0	~0.45, ~0.62, ~1.0
SXDS	1.22 / ~0.65	0.8	27.7 / 24.5 ( <i>i'</i> )	>68 000	≤21.0, ≤23.0, ≤24.5	~0.65, ~1.2, ~2.2
COSMOS	2.0 / ~0.5	0.09	26.5 / 24.0 (F814W)	>107 000	≤21.0, ≤23.0, ≤24.0	~0.67, ~1.1, ~1.6

More than half million sources were detected down to the magnitude limit  $r = 25.0$ .

The Bayesian Photometric Redshift code (BPZ2.0) was used to measure the photometric redshifts (Benítez 2000; Molino et al. 2014), obtaining the accuracy for galaxies brighter than 22.5 and with magnitudes 22.5–24.5 in the constructed F814W band of  $\sim\delta z / (1 + z) = 0.011$  and  $\sim\delta z / (1 + z) = 0.014$ , respectively. All images and catalogues are available through the ALHAMBRA webpage.<sup>2</sup>

*ALHAMBRA sample selection.* In this work, we used the same sample as described in Pović et al. (2013), whereby we will provide here only a brief summary. We selected only sources with the BPZ quality parameter  $\text{ODDS} > 0.2$ , expecting no more than 3 per cent of photometric redshift outliers (Molino et al. 2014). For details about the galaxy–star separation see Pović et al. (2013). And finally, we selected only objects with magnitudes  $\leq 23.0$  in the F613W band, taking into account that above this magnitude limit S/N ratio decreases significantly, leading to less accurate photometric redshift estimations and geometrical galaxy–star classifications. The F613W photometric band was selected due to its higher S/N ratio in comparison with other filters (Aparicio-Villegas et al. 2010). The final selected sample has more than 43 000 galaxies.

### 2.2.2 SXDS

The SXDS (Sekiguchi et al. 2004) is a large, multiwavelength, and one of the deepest optical ground-based surveys, covering five continuous, rectangle subfields, and a total area of 1.22 deg<sup>2</sup>. All subfields were observed in five broad-band filters  $BVR_c i' z'$ , detecting in each more than 800 000 sources down to the limiting magnitudes  $B = 28.4$ ,  $V = 27.8$ ,  $R_c = 27.7$ ,  $i' = 27.7$ , and  $z' = 26.6$  at  $3\sigma$  and a typical seeing of 0.8. All images and photometric catalogues are described in Furusawa (2008), and can be downloaded from the SXDS webpage.<sup>3</sup>

The photometric redshift distribution of each sample of galaxies was determined using multiband photometry of objects in the  $K$ -band-selected catalogue of Hartley et al. (2013), having a depth of  $K = 24.3$ . In addition to the  $JHK$  photometry, photometric measurements were obtained for the  $uBVR_c i' z'$  and *Spitzer*/IRAC channels 1 and 2 using the method of Simpson et al. (2012). Photometric redshifts and object classifications were derived using the code EAZY (Brammer, van Dokkum & Coppi 2008), using a development of the method described in Simpson et al. (2013) and Caldwell et al. (in preparation). Measured photometric redshifts have accuracy of  $\delta z / (1 + z) = 0.031$ .

*SXDS sample selection.* The sample we selected in the SXDS field is based on the available photometric redshifts. Therefore, we used only those galaxies that overlap with the  $K$ -band survey, instead of the entire SXDS field. The survey covers 0.77 deg<sup>2</sup>, where the overlap region with the SXDS field is  $\sim 0.63$  deg<sup>2</sup>, covering the entire central subfield and fractions of the side ones. We discarded all galaxies with poor photometric redshift fits, having  $\chi^2 < 100$ . To separate between point-like and extended sources, we applied the classification described above (Simpson et al. 2013, and Caldwell et al., in preparation). Moreover, we selected all galaxies with  $i' \leq 24.5$ , which is the band that we will use in the following analysis, and the limit at which our sample is complete, taking into account the available redshifts and photometry. The final selected sample consists on more than 68 000 galaxies. Since the photometric redshift selection is based on the  $K$ -band, we might be missing the bluest sources detected in the SXDS. We did not perform a detailed analysis with respect to this, but we estimated by cross-matching  $i$ - and  $K$ -band catalogues an upper limit of  $< 9$  per cent of these sources, which should not affect significantly our results.

### 2.2.3 COSMOS

The COSMOS (Scoville et al. 2007) is a deep, 2 deg<sup>2</sup> multiwavelength survey. In this work, we use the observations from the *HST* ACS survey, where more than  $1.2 \times 10^6$  sources were detected down to a limiting magnitude 26.5 in the filter F814W. The ACS data reduction and images are described in Koekemoer et al. (2007), while the photometric catalogue is presented in Leauthaud et al. (2007).

The photometric redshifts used in this work are presented in Ilbert et al. (2009), and were measured using the LE PHARE code and photometric information from 30 broad, intermediate, and narrow-band filters from UV, optical, NIR and MIR bands. The obtained measurements show accuracy of  $\sim\delta z / (1 + z) = 0.007$  for galaxies brighter than  $i^+ = 22.5$ . At magnitudes fainter than  $i^+ < 24.0$ , the accuracy is  $\sim\delta z / (1 + z) = 0.012$ , while for the very faint sample ( $i^+ < 25.0$ ) the accuracy drops significantly with  $\sim 20$  per cent of outliers. All images and catalogues used in this work are available from the COSMOS webpage.<sup>4</sup>

*COSMOS sample selection.* The COSMOS 2 deg<sup>2</sup> field is covered with 81 *HST*/ACS images (Koekemoer et al. 2007). Taking into account the number and size (20 480 pixels square) of the images, and the computing time that we would need to process all the data with the GALSVN code (Section 3), we are finally using an area of 0.5 deg<sup>2</sup>. Considering the depth of the COSMOS data, for the study presented in this paper, the selected area provides

<sup>2</sup> <http://alhambrasurvey.com>

<sup>3</sup> <http://www.naoj.org/Science/SubaruProject/SXDS/index.html>

<sup>4</sup> <http://cosmos.astro.caltech.edu>

us with the statistical information completely comparable with the other two surveys. Moreover, we checked that the magnitude and redshift distributions of galaxies in the selected field correspond to the total ones. Our main criteria for the COSMOS sample selection is again based on the accuracy of the photometric-redshifts. We selected only sources with  $i^+ < 24.0$ , since the photo- $z$  accuracy decreases significantly at fainter magnitudes, as described above. For the galaxy–star separation, we used the classifications presented in both Leauthaud et al. (2007) and Ilbert et al. (2009). Finally, the selected sample has more than 107 000 galaxies.

### 3 METHODOLOGY

Non-parametric methods of galaxy classification are usually applied either when dealing with high-redshift and/or low-resolution data, where the galaxy decomposition and profile fitting becomes impossible, or in large surveys, where due to the high number of detected sources it is necessary to use more automated and statistically based approaches. We test how much the morphological parameters described in Section 3.1 (basically related with the distribution of light within the galaxy and its shape) are sensitive to depth and spatial resolution, and how this introduced observational bias might affect the final galaxy classification. To do this, we used a sample of local galaxies with available detailed visual classification (see Section 2.1), and measured their morphological parameters in two cases:

(i) *at their real redshifts (magnitudes), i.e. at  $z \sim 0$*  (see Section 3.2), and

(ii) *moving them to higher redshifts (therefore to fainter magnitudes) and lower resolution*, to simulate the conditions of galaxies on deep, ground- and space-based non-local surveys (see Section 3.3).

In both cases, all parameters are measured in a completely consistent way, using their definitions as described in the following section. Finally, we compared the results obtained in cases (1) and (2) to quantify the impact from data depth and spatial resolution on each morphological parameter, for each of the three analysed surveys (see Sections 4.2 and 4.3).

#### 3.1 Description of the tested morphological parameters

The six morphological parameters described here are by large, the most commonly used to distinguish between ET and LT galaxies, and to classify perturbed galaxies and interacting/merging systems. In the following definitions, when necessary, the galaxy centre is determined by minimizing the asymmetry index (Abraham et al. 1996, see below), while the total flux is defined as the one contained within 1.5 times the Petrosian radius ( $r_p$ , measured by SExtractor; see Bertin & Arnouts 1996; Huertas-Company et al. 2008).

(i) *Abraham concentration index* (hereafter CABR; Abraham et al. 1996) – measured as the ratio between the flux at 30 per cent of the Petrosian radius ( $F_{30}$ ) and the total flux ( $F_{\text{tot}}$ ):

$$\text{CABR} = F_{30}/F_{\text{tot}}. \quad (1)$$

It gives values from 0 to 1, where higher values correspond to higher central light fractions.

(ii) *Gini coefficient* (hereafter GINI; Abraham et al. 2003; Lotz et al. 2004) – presented as the cumulative distribution function of galaxy pixel  $i$  values:

$$\text{GINI} = \frac{1}{|\bar{X}|n(n-1)} \sum_i^n (2i - n - 1)|X_i|, \quad (2)$$

where  $n$  is the total number of pixels in a galaxy,  $X_i$  the pixel flux value, and  $\bar{X}$  the mean over all the pixel flux values, using as an aperture  $1.5 r_p$ . Usually, it correlates with CABR, having also values from 0 to 1, where higher ones correspond to galaxies with higher central concentrations. However, unlike CABR, it can distinguish between galaxies with shallow light profiles and those with the light concentrated in a few pixels, but outside the galaxy centre.

(iii) *Conselice–Bershady concentration index* (hereafter CCON; Bershady et al. 2000) – measured as the logarithm of the ratio of the circular radii containing 80 and 20 per cent of the total flux,

$$\text{CCON} = 5 \log(r_{80}/r_{20}). \quad (3)$$

In general, lower CCON values correspond to lower fractions of light in the central region.

(iv) *M20 moment of light* (hereafter M20; Lotz et al. 2004) – measured as the flux ( $f_i$ ) in each pixel ( $x_i, y_i$ ) multiplied by the squared distance to the centre ( $x_c, y_c$ ) of the galaxy, summed over the 20 per cent brightest pixels, and normalized by the total second-order moment ( $M_{\text{tot}}$ ):

$$\text{M}_{20} = \log\left(\frac{\sum_i M_i}{M_{\text{tot}}}\right), \quad \text{while } \sum_i f_i < 0.2 f_{\text{tot}}, \quad (4)$$

where

$$M_{\text{tot}} = \sum_i^n f_i ((x_i - x_c)^2 + (y_i - y_c)^2). \quad (5)$$

It shows the light distribution of central sources, off-centre star clusters, bars, spiral arms, etc.

(v) *Asymmetry* (hereafter ASYM; Abraham et al. 1996) – measured by subtracting the galaxy image rotated by  $180^\circ$  from the original image, taking also into account the background:

$$\text{ASYM} = \frac{\sum (|I(i, j) - I_{180}(i, j)| / 2)}{\sum I(i, j)} - \frac{\sum (|B(i, j) - B_{180}(i, j)| / 2)}{\sum I(i, j)}, \quad (6)$$

where  $I(i, j)$  is the flux in the  $(i, j)$  pixel position on the original image,  $I_{180}(i, j)$  is the flux in  $(i, j)$  on the image rotated by  $180^\circ$ , and  $B(i, j)$  and  $B_{180}(i, j)$  are fluxes of the background in the original and rotated image, respectively. Defined in this way, this parameter quantifies the degree to which the light of a galaxy is rotationally symmetric. It gives values from 0 to 1, where the higher ones correspond to more asymmetric galactic shapes.

(vi) *Smoothness or clumpiness* (hereafter SMOOTH; Conselice et al. 2000) – quantifies the degree of small-scale structure. To measure SMOOTH, the original galaxy image is smoothed out with a boxcar of a given width and then subtracted from the original image:

$$\text{SMOOTH} = \frac{\sum (|I(i, j) - I_S(i, j)| / 2)}{\sum I(i, j)} - \frac{\sum (|B(i, j) - B_S(i, j)| / 2)}{\sum I(i, j)}, \quad (7)$$

where  $I(i, j)$  ( $B(i, j)$ ) present again the flux (background) in the pixel position  $(i, j)$  on the original image, while  $I_S(i, j)$  and  $B_S(i, j)$  are the flux and the background, respectively, on the image smoothed by a boxcar of width  $0.25 r_p$ . The residual image provides with

information about possible regions of clumpiness, quantifying the level of small-scale structures.

### 3.2 Methodology applied on the local sample

We measured the reference morphological parameters of the local sample, at their real redshifts  $z \sim 0$ , in three different SDSS bands:  $g$ ,  $r$ , and  $i$ . This is needed in order to compare in a consistent way the morphologies of local and simulated galaxies in Sections 4.2 and 4.3, since the corresponding  $k$ -correction will imply a different filter for such a comparison. We used the function implemented in the GALSV<sup>5</sup> public code of galaxy classification (Huertas-Company et al. 2008, 2009) for measuring the parameters. The input information that the code needs, related with the source position, size, ellipticity parameters, and background, was obtained by running SEXTRACTOR in all three bands.

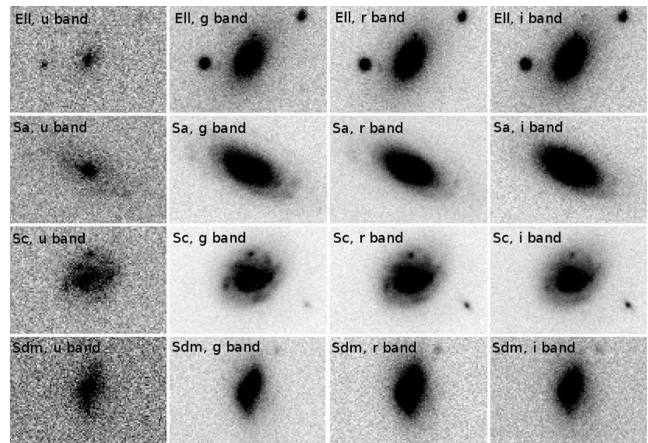
### 3.3 Methodology applied at higher redshifts

To simulate the conditions of non-local surveys (case 2 in Section 3), we used again the GALSV code (Huertas-Company et al. 2008, 2009), performing the following three steps for each of the three analysed surveys:

(i) First, we constructed the correspondent magnitude and redshift distributions of non-local galaxies (see Section 4.2 and Fig. 7). We randomly redshifted and scaled in luminosity the selected sample of local galaxies to match the distributions of non-local ones (see Table 1 for the used photometric band in each survey). In this way, when scaling the galaxy in flux, the surface brightness dimming was directly taken into account. Moreover, we re-sampled the local galaxies with the corresponding pixel-scale for each selected non-local survey, and convolved with its point spread function to match the spatial resolution.

(ii) Secondly, we dropped the simulated galaxies, obtained in the first step, into the real background of the high-redshift survey images. We made sure that the galaxies are dropped into empty regions of sky, minimizing the chance of superposition with high-redshift foreground/background sources. To detect the empty sky regions, the code makes use of corresponding SEXTRACTOR segmentation images. Moreover, although the probability of overlapping is related with the survey properties, we have been strict, dropping only 10 galaxies per image. This step is then repeated until dropping all simulated galaxies, generating every time new images. With all this, we make the probability of overlapping and merger confusion practically negligible (for more information see Huertas-Company et al. 2008). Finally, by dropping the local galaxy in a real background of the high-redshift sources, we expect to reproduce the noise from the real images.

(iii) Thirdly, we measured the morphological parameters described in Section 3.1. We took care of the  $k$ -correction effect introduced by cosmological redshift, and depending on the band selected in each survey, and also depending on the redshift to which the local galaxy was shifted, we measured the morphological parameters using the corresponding SDSS rest-frame band image. In the ALHAMBRA survey, since we are using the  $F613W$  AUTO magnitudes, the SDSS  $r$ -band images were used for galaxies simulated to  $z_{\text{sim}} \leq 0.13$ ,  $g$  band for the redshift range 0.13–0.5, and  $u$  band for higher redshifts. In the SXDS survey, using the  $i'$  magnitudes, the  $r$  band was used for redshifts  $z_{\text{sim}} \leq 0.4$ , the  $g$  band



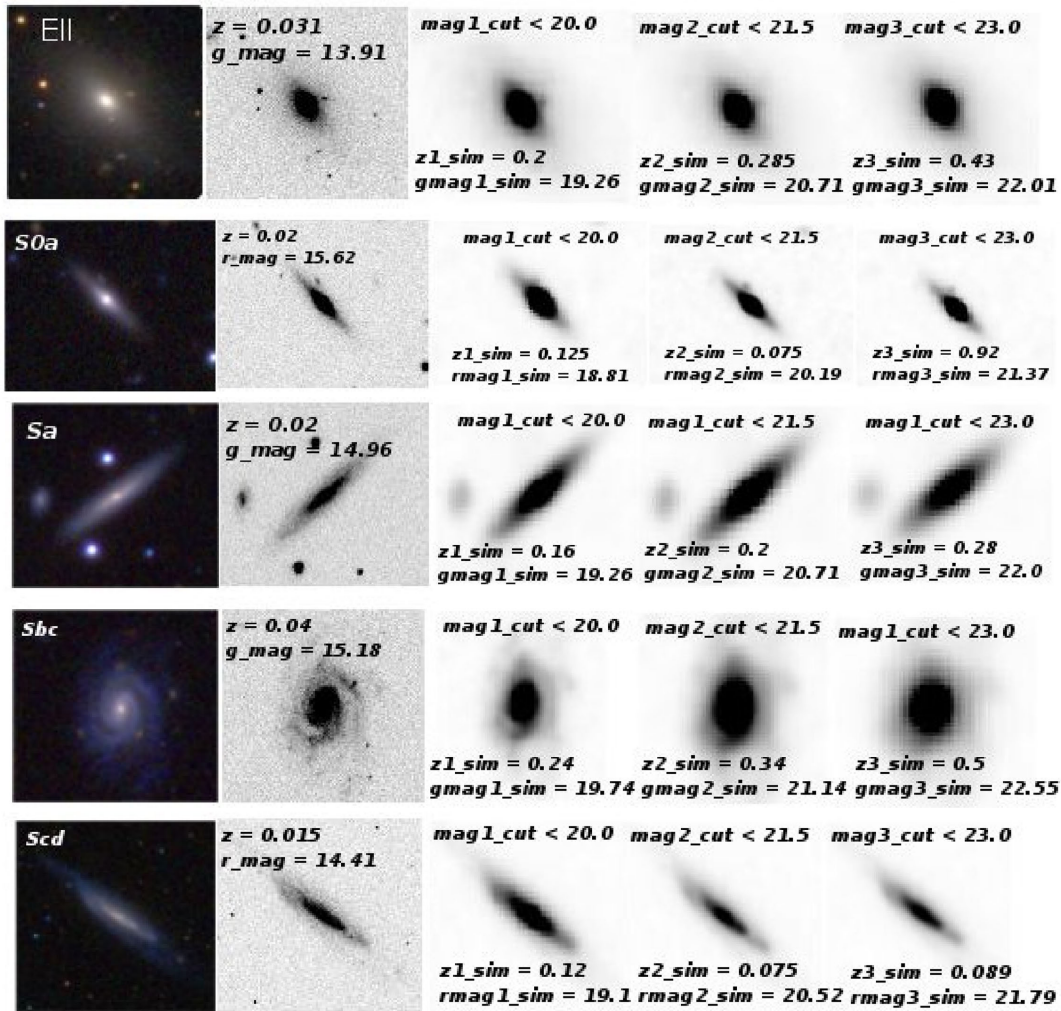
**Figure 2.** An example of Ell, Sa, Sc, and Sdm galaxies (from top to bottom) seen in four SDSS bands,  $u$ ,  $g$ ,  $r$ , and  $i$  (from left to right).

for  $0.4 < z_{\text{sim}} \leq 0.86$ , and the  $u$  band for  $z_{\text{sim}} > 0.86$ . Finally, in the COSMOS survey, using the  $F814W$  AUTO magnitudes, the  $i$  band was used for the lowest redshifts  $z_{\text{sim}} \leq 0.18$ , the  $r$  band for  $0.18 < z_{\text{sim}} \leq 0.5$ , the  $g$  band for  $0.5 < z_{\text{sim}} \leq 1.0$ , and the  $u$  band for  $z_{\text{sim}} > 1.0$ . Since the filter efficiency for the SDSS  $u$  band is much lower than that for  $g$ ,  $r$ , and  $i$  bands (Gunn et al. 1998), and taking into account that we are analysing three broad morphological groups (ET, LT<sub>et</sub>, and LT<sub>lt</sub>) as described in Section 2.1 (instead of studying a finer morphology), we used the  $g$ -band images instead of  $u$  when necessary. Fig. 2 shows an example of  $u$ ,  $g$ ,  $r$ , and  $i$  images of four galaxies from our local sample with different morphologies; note the poor information available in the  $u$  image in comparison to the other bands. Finally, since the morphological classification directly depends on the source brightness, in each survey we analysed the observational bias at the three different magnitude cuts:  $\text{mag1\_cut} \leq 20.0$ ,  $\text{mag2\_cut} \leq 21.5$ , and  $\text{mag3\_cut} \leq 23.0$  in the  $F613W$  band in ALHAMBRA;  $\leq 21.0$ ,  $\leq 23.0$ , and  $\leq 24.5$ , in the  $i'$  band in SXDS; and  $\leq 21.0$ ,  $\leq 23.0$ , and  $\leq 24.0$ , in the  $F814W$  band in COSMOS. In all surveys, we chose the first magnitude cut such that we have a sufficient number of sources to perform our analysis. The cut at which our selected photometric/photometric redshift sample is complete, is the last one. Finally, we chose a third cut intermediate to these two.

Figs 3–5 show examples of local galaxies, with different morphologies, after being scaled to the conditions of the ALHAMBRA, SXDS, and COSMOS surveys, respectively. In each survey, we show the galaxies being redshifted to the corresponding magnitude cuts, as explained above, providing for each cut the simulated values of magnitude and redshift. These images represent the output of the first step explained in this section, that we use in the second step to drop them in the real background of the non-local survey images, and to measure their morphological parameters in the third step. The colour images were gathered from the EFIGI<sup>6</sup> project, by combining the  $gri$  images (Baillard et al. 2011), while the original images were downloaded from the SDSS DR4 data base and correspond to the images used by N&A in their visual classification. We can observe in each survey how the galaxy information changes when going from brighter to fainter magnitudes (in general from lower to higher redshifts), but even more, how the role that the

<sup>5</sup> <http://gepicom04.obspm.fr/galSVM/Home.html>

<sup>6</sup> Extraction de Formes Idealizes de Galaxies en Imagerie; <http://www.astromatic.net/projects/efgi>



**Figure 3.** Example of simulated images of local galaxies (with different morphologies) after being scaled to map the conditions of the ALHAMBRA survey. For each galaxy (in each row), we present the following (from left to right): colour image, used rest-frame image(s), and simulated high-redshift images at three magnitude cuts (as written on the top of each redshifted image). The morphological type of each galaxy is noted on the colour image. The real redshift and magnitude of the galaxy are noted on the corresponding rest-frame band image(s), while the simulated high-redshift and magnitude are noted on the scaled images of each magnitude cut.

spatial resolution plays when classifying galaxies, where in COSMOS survey the galaxy information can be conserved up to much higher redshifts in comparison with the ground-based surveys. We studied this in detail in Sections 4 and 5.

#### Noise effect

With respect to the noise, since we are using very bright local galaxies, GALSVN assumes that the noise from the galaxies themselves is negligible. Therefore, the noise coming from the galaxy itself once it has been dropped into the non-local image is not treated by the code. We tested how strong this effect could be on the measured morphological parameters. We selected 30 galaxies, 10 from each morphological group defined in Section 2.1, and we measured the morphological parameters in the case of ALHAMBRA and COSMOS in two cases: (1) without the above noise added, like in our analysis, and (2) adding the random Poisson noise, typical of each survey, to the scaled local images (before dropping them into the background of non-local surveys). We then compared the results

obtained in these two cases for the whole samples, and in relation with the morphological types. We found insignificant differences of the above noise on the mean surface brightness ( $<1$  per cent), and on the parameters CABR, CCON, GINI, and M20 ( $<6$  and  $<5$  per cent in ALHAMBRA and COSMOS, respectively, at the faintest magnitude cuts), but more significant differences in the case of ASYM and SMOOTH. In the case of ALHAMBRA, the differences for these two parameters go up to  $\sim 40$  and  $\sim 50$  per cent, respectively. In the case of COSMOS, differences for ASYM are 13, 19, and 57 per cent at  $F814W \leq 21.0$ ,  $\leq 23.0$ , and  $\leq 24.0$ , respectively. For SMOOTH the corresponding values are 15, 35, and 50 per cent. From this result, we suggest that for the parameters ASYM and SMOOTH the noise effect should be taken into account when using them in morphological classification of galaxies, and especially when dealing with ground-based data. Therefore, in our analysis we discuss all results only for the four parameters CABR, GINI, CCON, and M20, that are more stable to noise and where our GALSVN measurements are reliable. The results based on ASYM and SMOOTH are provided only in the case of COSMOS, for the first two magnitude cuts, where the noise effect is lower.

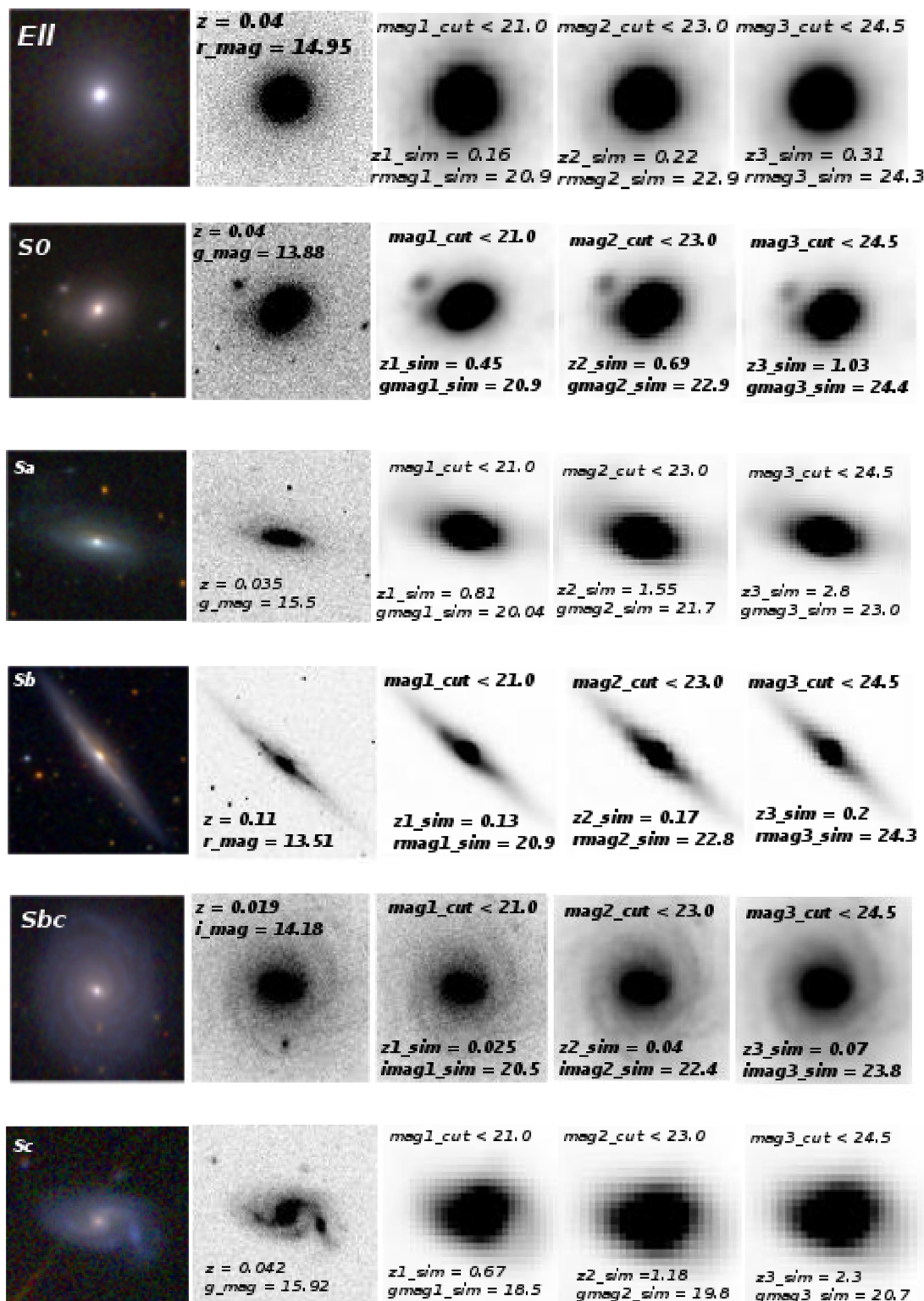


Figure 4. Same as Fig. 3, but in the SXDS survey.

## 4 RESULTS

### 4.1 Morphological parameters of the local sample at $z \sim 0$

For each galaxy, we measured the morphological parameters defined in Section 3.1 in the  $g$ ,  $r$ , and  $i$  bands, as was described in

Section 3.2. To test how sensitive are morphological parameters to the used wavelengths and if somehow it could affect our further analysis, for each parameter, we compared its measurements obtained in the three photometric bands, finding, in general, a good linear correlation in all cases. Fig. 6 shows the comparison of all six parameters between  $g$  and  $r$  (panel a), and  $g$  and  $i$  bands (panel b). In



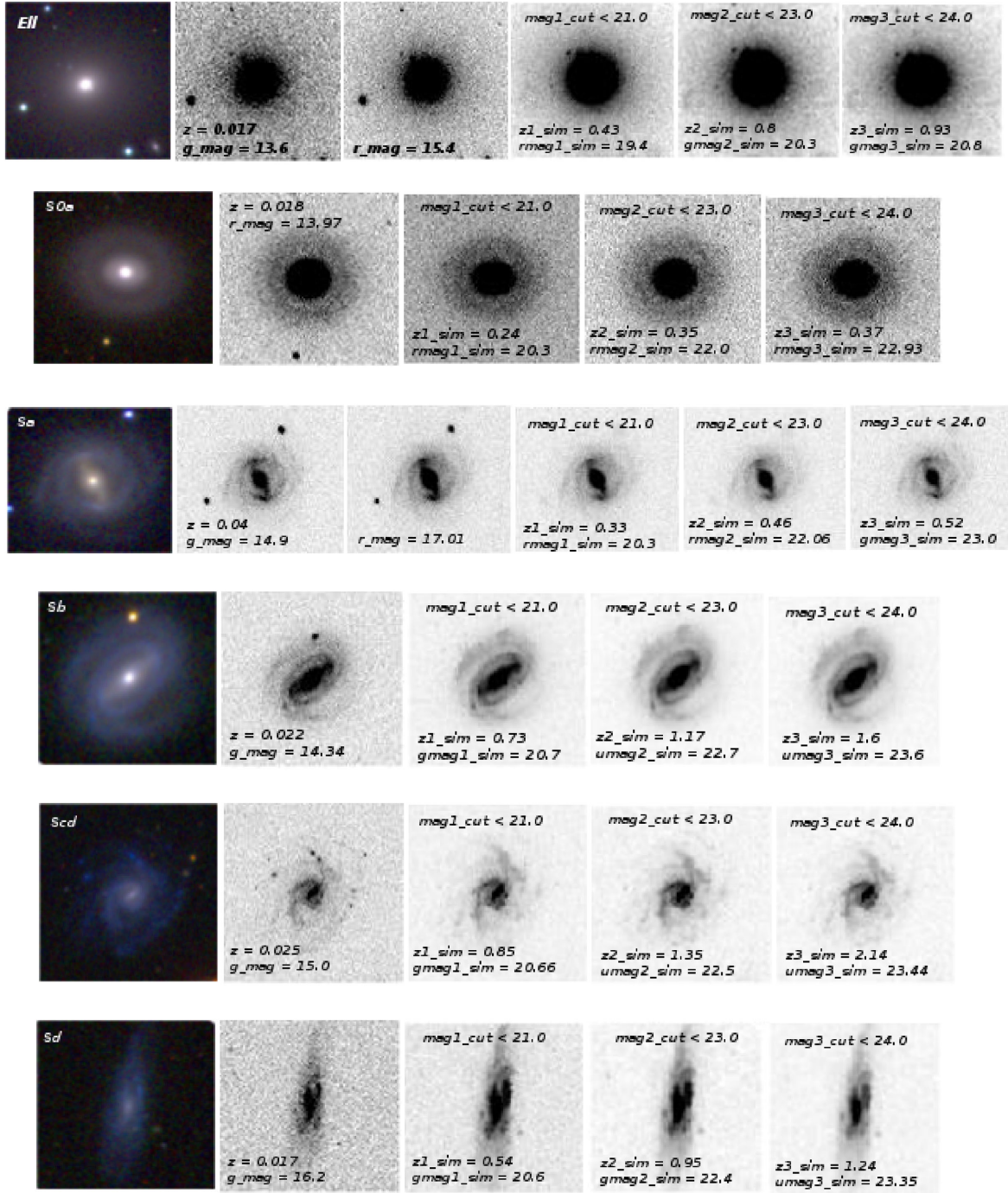
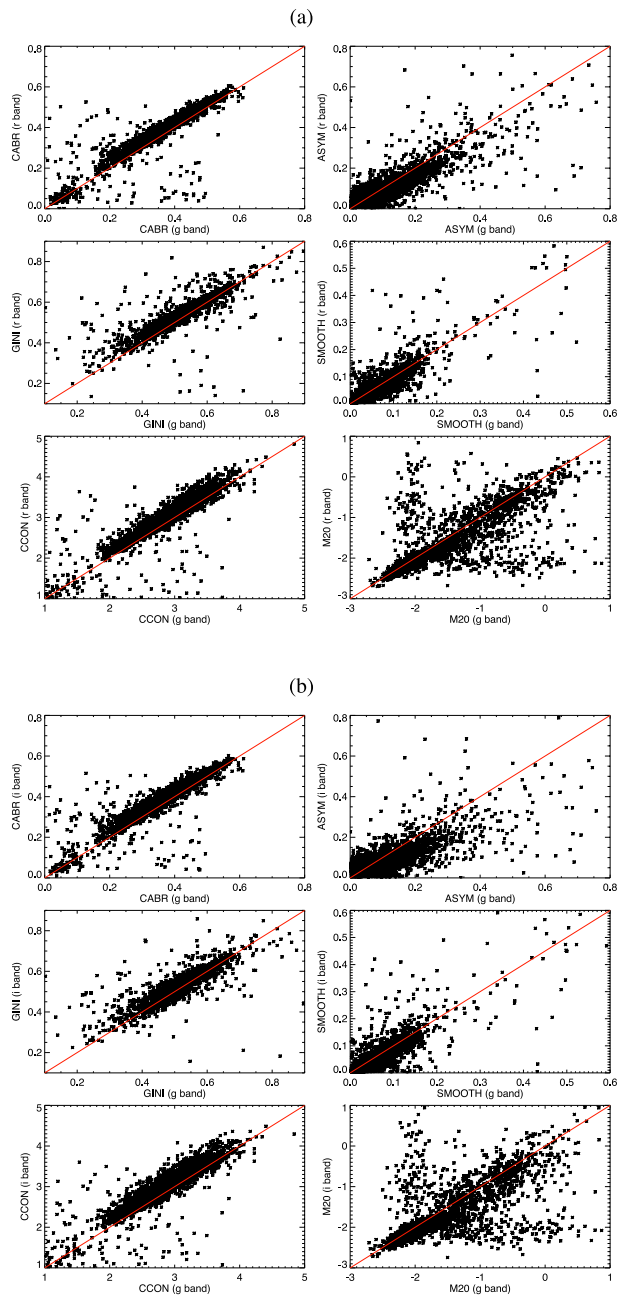


Figure 5. Same as Fig. 3, but in the COSMOS survey.

both cases, the concentration parameters, CABR, CCON, and GINI, show the best linear correlations, having similar Pearson Correlation Coefficient (PCC)<sup>7</sup>  $\sim 0.9$ . For M20, ASYM, and SMOOTH, the linear correlation is also conserved in both cases, however they are more sensitive to the selected band in comparison with the concentration indexes, and show lower PCC values (between  $\sim 0.7$  and

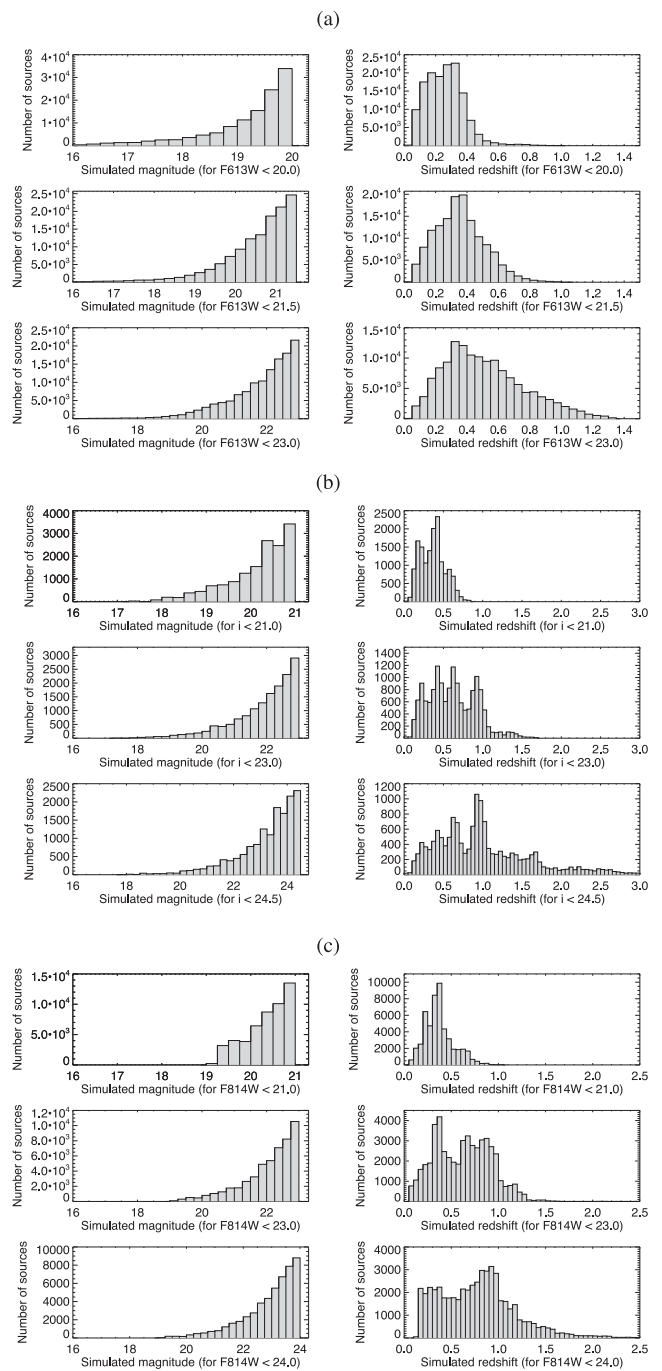
$\sim 0.85$ ). They also show a slightly higher dispersion in panel b, in comparison with panel a, with a difference in the PCC of  $\sim 0.05$ . M20 is the most sensitive to the selected band, with PCC = 0.733 between the  $g$  and  $r$  bands, and PCC = 0.688 between the  $g$  and  $i$  measurements. The number of catastrophic outliers, placed in the ‘wings’ above and below the linear correlations is  $\sim 10$  per cent in both panels. We checked the properties of these sources (magnitude, size, and redshift), without finding any significant difference in comparison with the whole sample. We only found a small difference in relation with the morphological type, with 6 per cent more LTs in the wings. To have an idea about how these outliers could affect

<sup>7</sup> PCC is a measure of the linear correlation between two variables giving a value  $-1 \leq \text{PCC} \leq +1$ , where  $+1$  shows total positive correlation,  $0$  no correlation, and  $-1$  total negative correlation (Pearson 1895).



**Figure 6.** Comparison between the six morphological parameters for the galaxies in the local sample measured in the  $g$  and  $r$  bands (panel a), and  $g$  and  $i$  bands (panel b). In each panel, we compared, from top to bottom, and from left to right: CABR, asymmetry parameter, GINI, SMOOTH, CCON, and M20. In all plots, the red solid line presents the perfect linear correlation between the two measurements.

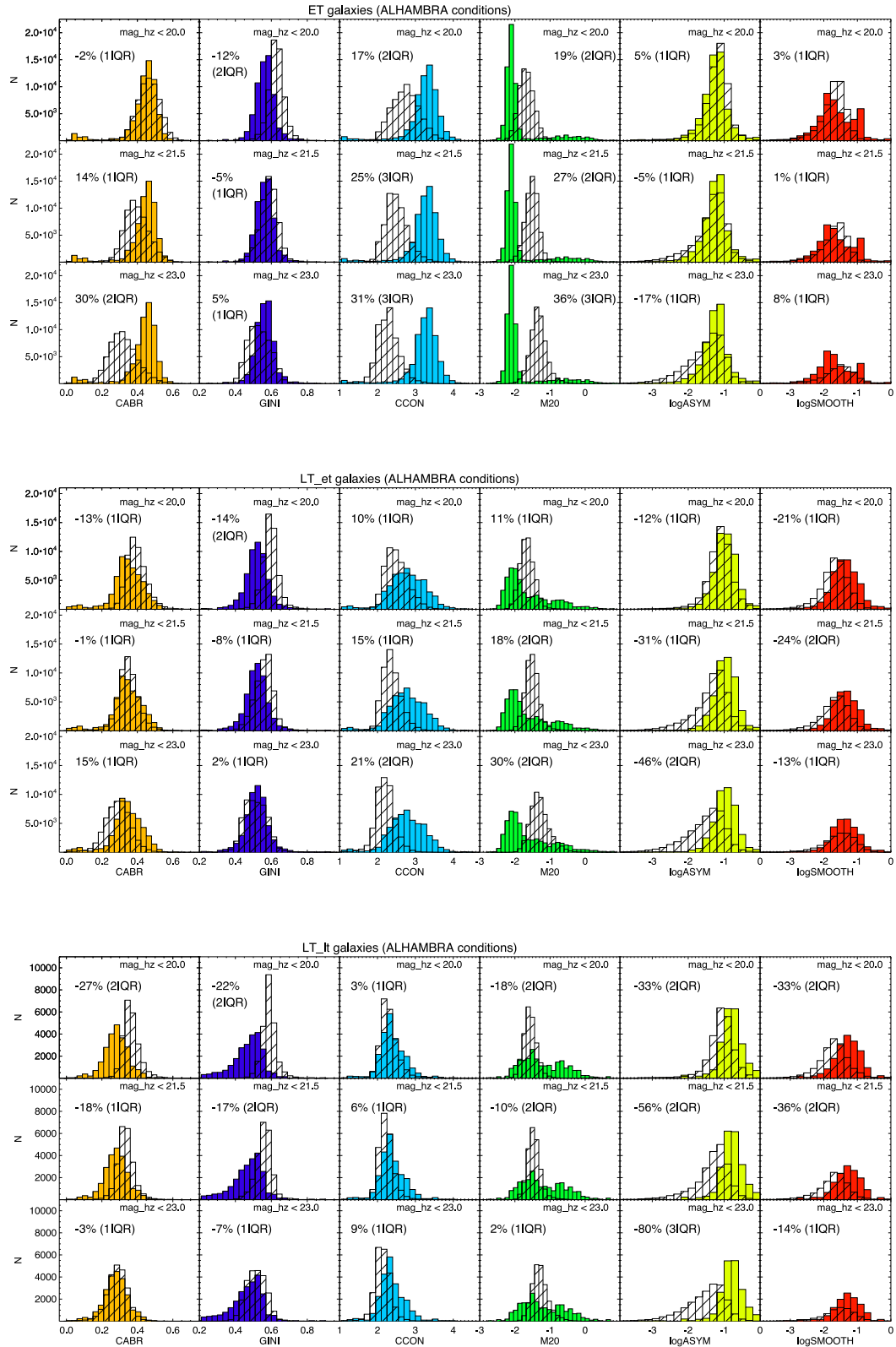
the results in our following analysis and if there are some important differences that we should study in more detail, we performed a test by excluding them from the whole sample and checking again the corresponding distributions of morphological parameters (see Section 4.2 and Figs 8–10), without finding any significant differences. However, this was done only to test the properties of outliers, but the all analysis and results presented in this paper were obtained using the whole sample, outliers included. For the comparison of morphological parameters between near-infrared and optical bands see Huertas-Company et al. (2014).



**Figure 7.** Magnitude (left plots) and redshift (right plots) distributions of simulated galaxies for the conditions in ALHAMBRA (panel a), SXDS (panel b), and COSMOS (panel c) surveys. In each survey, they are represented for three magnitude cuts, from top to bottom: for  $F613W \leq 20.0$ ,  $F613W \leq 21.5$ , and  $F613W \leq 23.0$  in the ALHAMBRA;  $i \leq 21.0$ ,  $i \leq 23.0$ , and  $i \leq 24.5$  in the SXDS; and  $F814 \leq 21.0$  (top),  $F814 \leq 23.0$  (middle), and  $F814 \leq 24.0$  in the COSMOS.

## 4.2 Morphological parameters of the simulated sample

To measure the morphological parameters of the simulated sample, we followed the methodology described in Section 3.3. We first moved randomly the local sample to map the magnitude and redshift distributions of galaxies in each high-redshift survey, at each selected magnitude cut. Fig. 7 shows the corresponding



**Figure 8.** Comparison between the morphological parameters of the sample of local galaxies measured at their real redshift (filled-colour histograms, representing the reference galaxy properties) and in conditions that correspond to the galaxies from the ALHAMBRA survey (filled-line histograms). Six parameters are compared, from left to right: CABR, GINI, CCON, M20, ASYM, and SMOOTH. (Top) three magnitude cuts of the ET galaxies, (middle) LT\_et galaxies, and (bottom) LT\_It galaxies. In all diagrams, the numbers represent the difference between the median value of the reference (local) and simulated distributions, normalized with the median of the local sample (expressed in per cent), and the corresponding IQR.

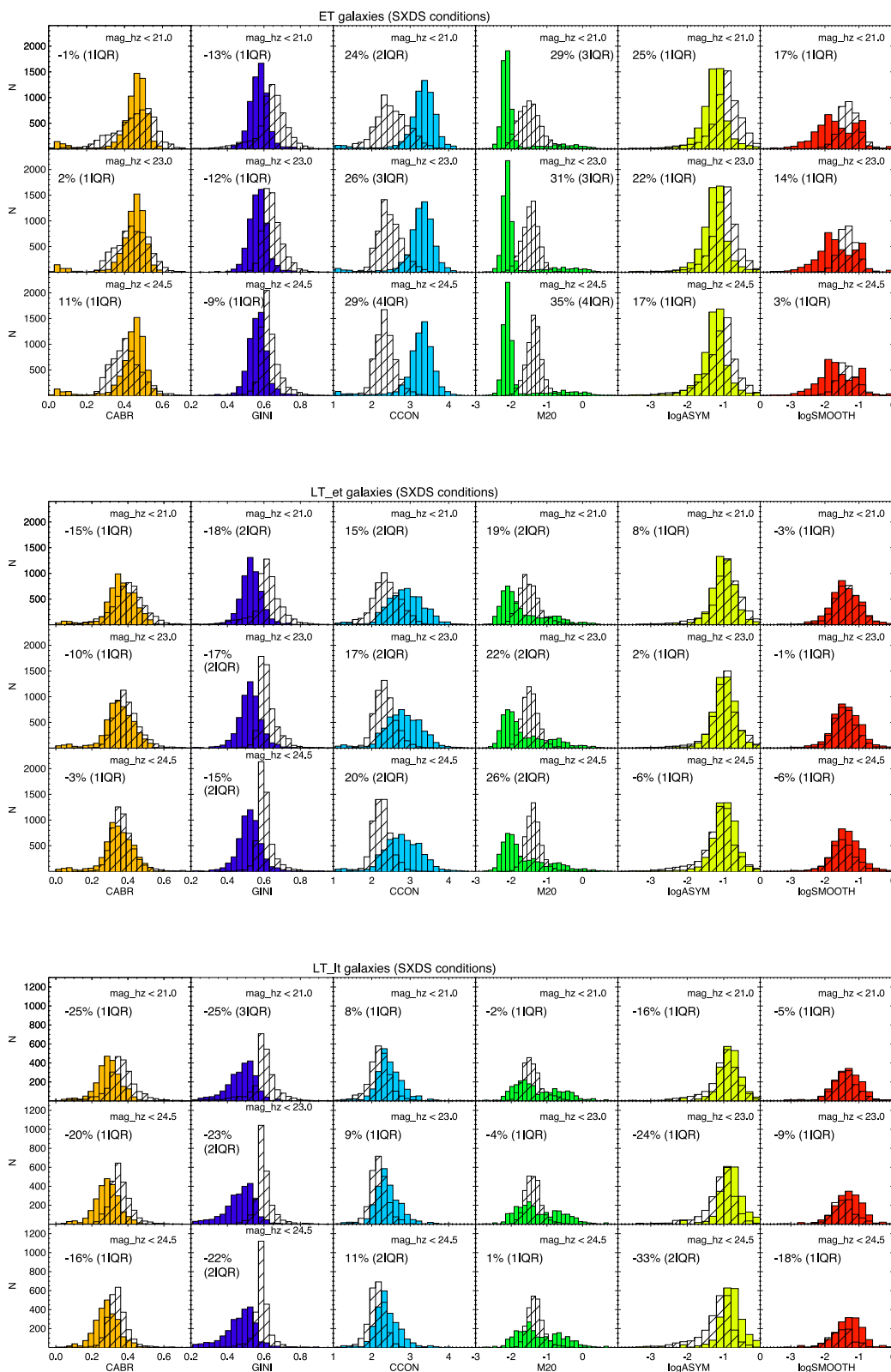


Figure 9. Same as Fig. 8, but moving the local sample to map the conditions of the SXDS survey.

magnitude and redshift distributions of the simulated galaxies in case of ALHAMBRA (panel a), SXDS (panel b), and COSMOS (panel c). Since the magnitude cut introduces also a redshift selection, we are dealing at the same time with different redshift ranges

in the three surveys (as summarized in Table 1). After simulating the local galaxies and dropping them into the real background of high-redshift surveys, we run again *SExtractor* to measure the same parameters as we did at  $z \sim 0$ . Finally, we measured the

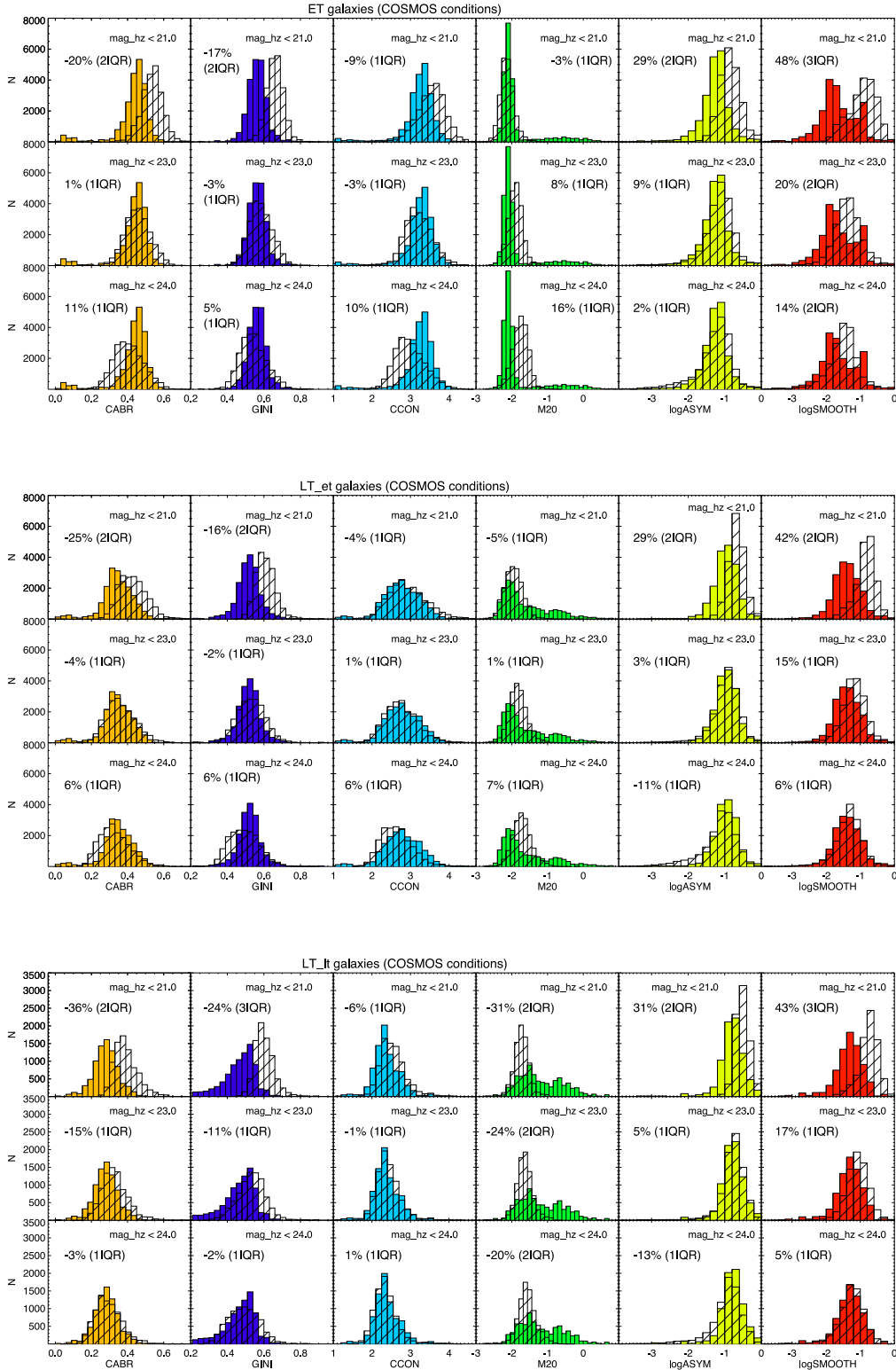


Figure 10. Same as Fig. 8, but moving the local sample to map the conditions of the COSMOS survey.

morphological parameters of simulated galaxies, again in the same way as we did at  $z \sim 0$ . As mentioned above, in each survey, depending on the used photometric band and the redshift to which the galaxy was moved, we used the corresponding rest-frame SDSS images to carry out all the measurements.

In each non-local survey, we had to repeat all the procedure a number of times, for each separate image, in order to cover the analysed survey areas (see Table 1). In case of ALHAMBRA, we were dealing with 48 subfields [4 CCD images per pointing, with 12 pointing in total (1 or 2 per each field); see Pović et al. 2013.

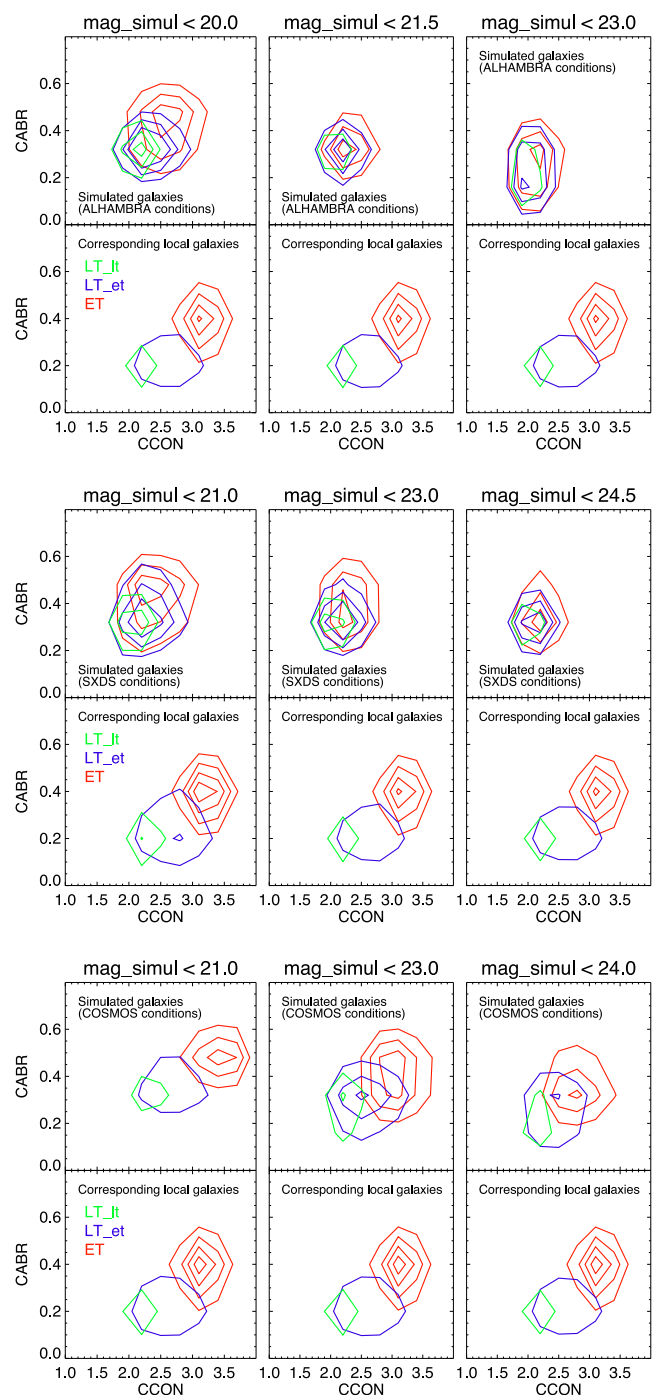
In SXDS, we repeated the procedure in each of five SXDS fields (see Section 2.2.2), while to cover the selected area of COSMOS we run the code on 17 images. Moreover, in all surveys we repeated the measurements three times, since having selected three magnitude cuts (see Table 1). Since each time we run `GALSVM` on one image, 3000 simulated galaxies were distributed randomly to the corresponding redshift and magnitude distributions of the non-local samples, for each galaxy we obtained a number of measurements that is equal to the number of used images. Therefore, the final catalogues contain 144 000, 15 000, and 51 000 simulated galaxies in the case of ALHAMBRA, SXDS, and COSMOS, respectively. Each catalogue contains the columns with morphological parameters measured at higher redshifts in the three magnitude cuts, besides the corresponding reference measurements (obtained at  $z \sim 0$ ). Once again, when assigning the reference value to the simulated one, we took care of the  $k$ -correction, assigning the value measured in the corresponding rest-frame band (that will depend on the photometric band we used for analysis in each non-local survey, and on the simulated redshift). The reader should keep this in mind for understanding easier the distributions of comparison sample of local galaxies, represented in Figs 11–17 in Section 4.3.

The obtained morphological parameters of simulated galaxies are represented with the filled-line histograms in Figs 8–10, that reproduce the conditions of the ALHAMBRA, SXDS, and COSMOS surveys, respectively. In each figure are shown the distributions of CABR, GINI, and CCON concentration indexes, M20, ASYM, and SMOOTH (from left-hand to right-hand panels) for three morphological groups: ET (top plots), LT\_et (middle plots), and LT\_lt (bottom plots), defined in Section 2.1. For each morphological group, we represent the distribution of parameters at the three analysed magnitude cuts. In these figures, we also represent the distributions of the reference morphological parameters of local galaxies, measured in Section 4.1, with the filled-colour histograms.

For all morphological parameters, we can observe important differences between the reference (local) values and those obtained after moving the local galaxies to the conditions of non-local surveys, indicating the influence that the observational bias from spatial resolution and depth can have on them. To quantify this bias, we measured the difference between the median values of the local and simulated distributions, normalized with respect to the median value of the local sample:  $\langle \text{local} \rangle - \langle \text{simulated} \rangle / \langle \text{local} \rangle$ . The obtained values are expressed in per cent in each plot, together with the corresponding range of interquartile range (IQR; between parentheses), measured as the difference between the third and first quartiles. In the following, we describe the obtained differences, considering as significant those within  $\geq 2$  IQR.

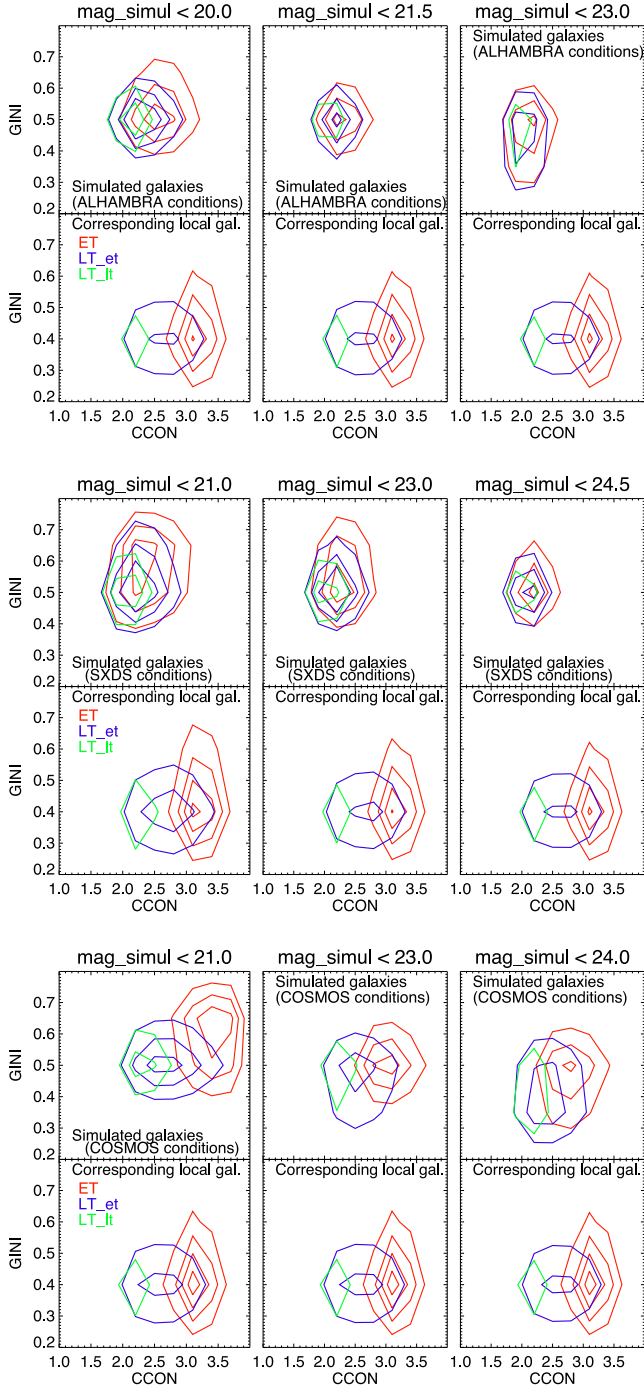
**CABR.** In ALHAMBRA conditions, CABR showed significant differences in the case of the faintest ETs and brightest LTs ( $\sim 30$  per cent within 2 IQR), with galaxies appearing less and more concentrated, respectively. In the conditions of SXDS, we did not find any differences (all within 1 IQR). In COSMOS, we obtained differences of  $\sim 20$ – $36$  per cent (within 2 IQR) at the brightest magnitude cuts for all three morphological types, with simulated galaxies appearing more concentrated. In ALHAMBRA, 2–3 per cent of sources showed invalid measurements (with values outside the typical range of this parameter, between 0 and 1), independently of the morphological type; in SXDS, this went between 3 and 5 per cent for ET and LT\_et, and 7–10 per cent for LT\_lt galaxies; while in COSMOS, 2–3 per cent for ET and LT\_et, and 8–10 per cent for LT\_lt galaxies.

**GINI.** In ALHAMBRA, this parameter showed differences of 10–20 per cent (2 IQR) at the first magnitude cut for all types, and



**Figure 11.** Relation between the Conselice–Bershady and CABR, in the three analysed non-local surveys: ALHAMBRA (top plots), SXDS (middle plots), and COSMOS (bottom plots). In each survey, top rows represent the morphological parameters of simulated sample obtained after moving the local galaxies to higher redshifts, considering three magnitude cuts (the first and the last column showing the lowest and the highest analysed magnitude cut, respectively), while bottom rows show the corresponding reference (local) values. The red, blue, and green contours represent ET, LT\_et, and LT\_lt galaxies, respectively.

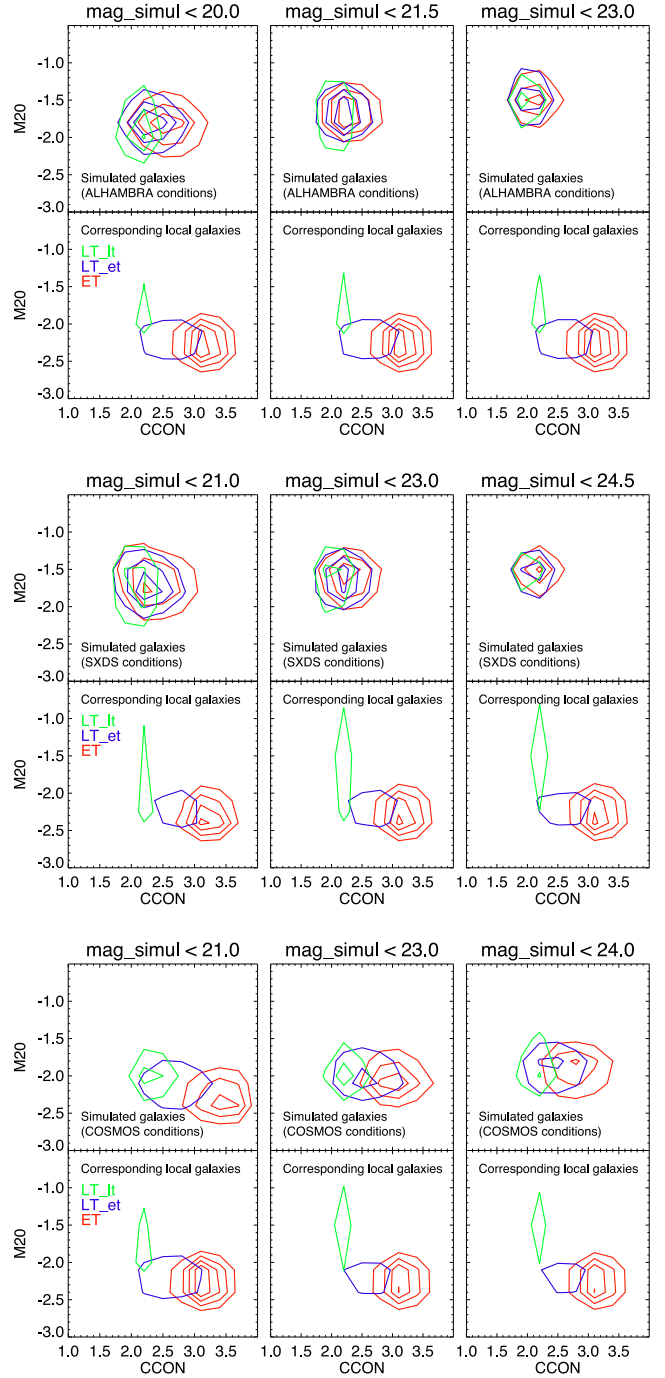
at the second magnitude cut for LT\_lt galaxies. In SXDS, the difference for LT\_et galaxies was observed at all three magnitude cuts (15–18 per cent within 2 IQR). The same was obtained for LT\_lt sources but with slightly higher differences. In the case of



**Figure 12.** Same as Fig. 11, but representing the relation between the CCON and GINI.

COSMOS, the only differences were observed at the first magnitude cut:  $\sim 20$  per cent (2 – 3 IQR) for all three types. In all cases where the differences were detected, we obtained negative values, showing that the simulated galaxies would appear more concentrated than they really are. With respect to the measured values of simulated galaxies that are outside of the valid ranges, the populations are in agreement with those for the CABR parameter.

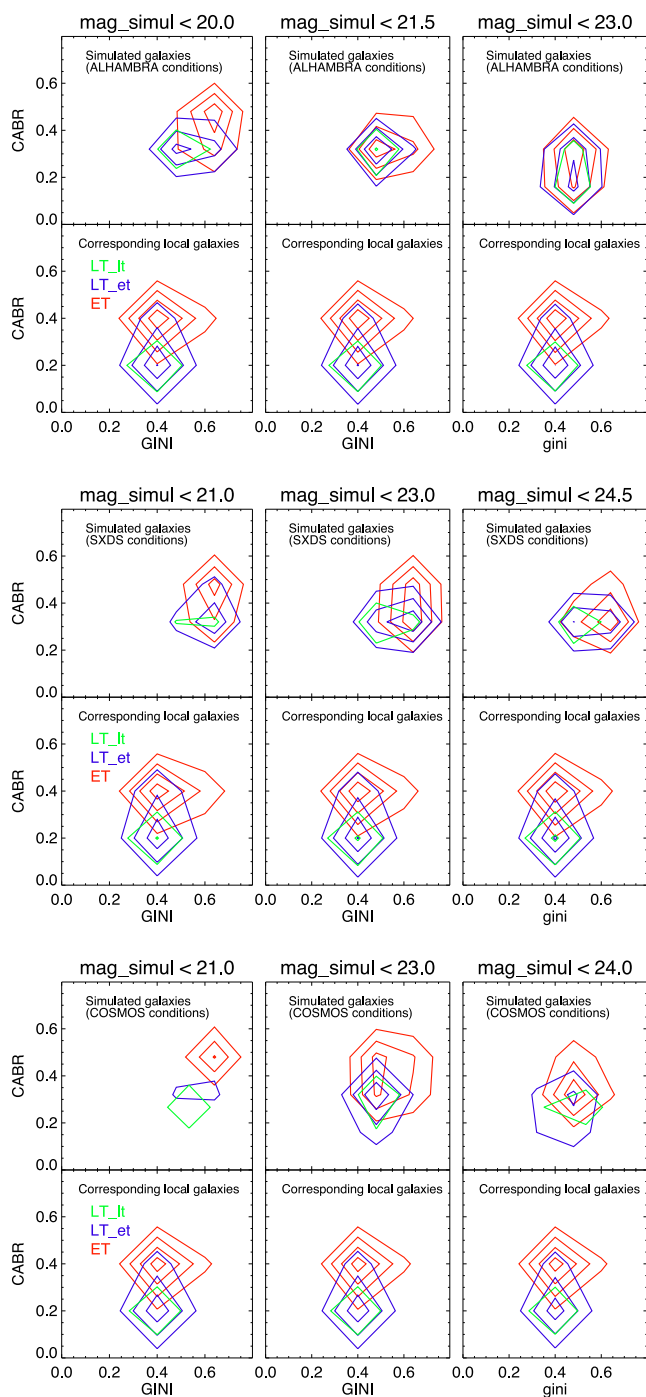
*CCON*. A different behaviour of *CCON* was detected for the conditions of the analysed ground- and space-based surveys. In the case of COSMOS, we did not detect any differences, independently of the magnitude cut and morphological type. This index showed



**Figure 13.** Same as Fig. 11, but representing the relation between the CCON and *M20*.

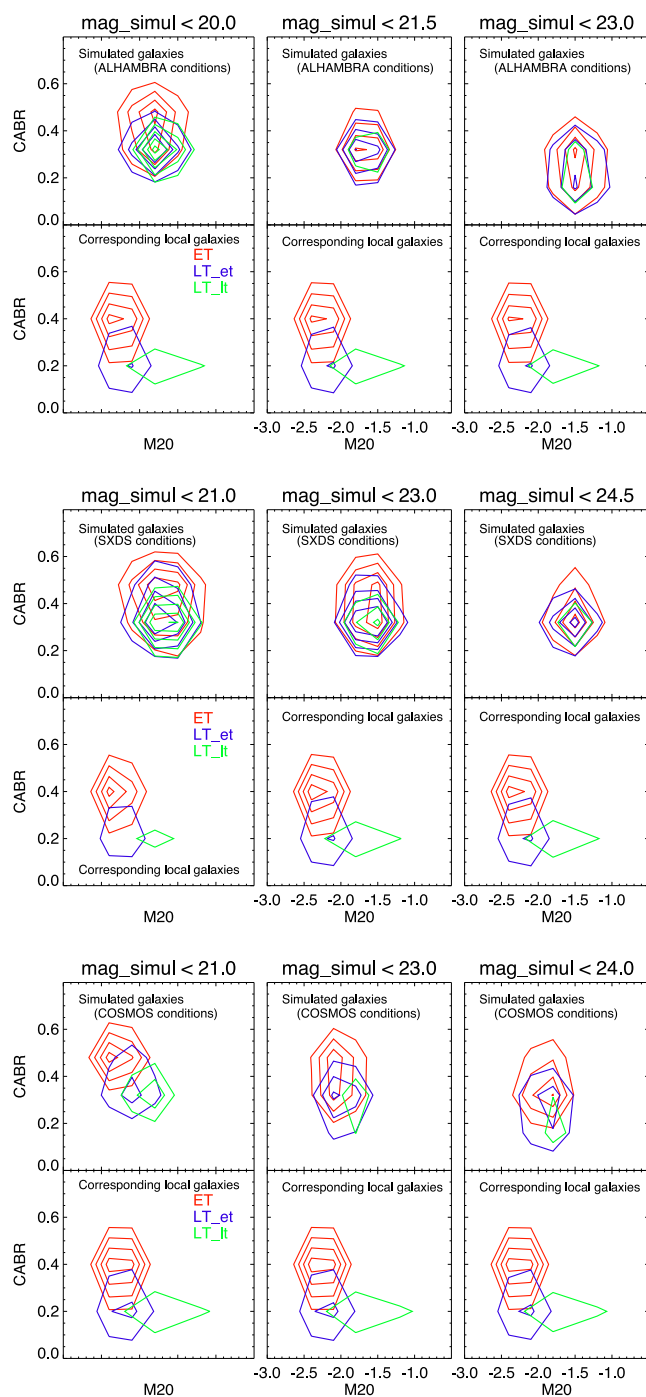
to be especially sensitive to the spatial resolution in comparison to CABR and GINI. In the case of ALHAMBRA and SXDS, the highest differences are detected for ET galaxies at the highest magnitude cuts,  $\sim 30$  per cent at 3 IQR and 4 IQR. In all cases where differences were observed in the ground-based surveys, lower concentrations appeared for simulated galaxies in comparison with the reference (local) values, independently on the magnitude cut and morphological type. Again, the population of the obtained invalid measurements is consistent with those presented for CABR.

*M20*. As described in Section 3.1, this parameter is sensitive to galaxy structures as the central light from the bulge, bars, spiral



**Figure 14.** Same as Fig. 11, but representing the relation between the GINI and CABR.

arms, strong star formation regions, etc. Therefore, for local ET galaxies it mainly maps the central light, having a much narrower distribution than in the case of LT\_et, or even more in the case of LT\_It galaxies, where the distribution is broader (filled-colour histograms in Figs 8–10). However, when moved to higher redshift surveys the distribution of this parameter changed significantly (both, shape and range; filled-line histograms), and showed in all cases the information coming mainly from the centre of the galaxy, and independently on the morphological type. In ALHAMBRA, the highest differences are of the order of  $\geq 30$  per cent within 2–3 IQR, for all three types. In SXDS conditions, for ET galaxies M20 suffered sim-

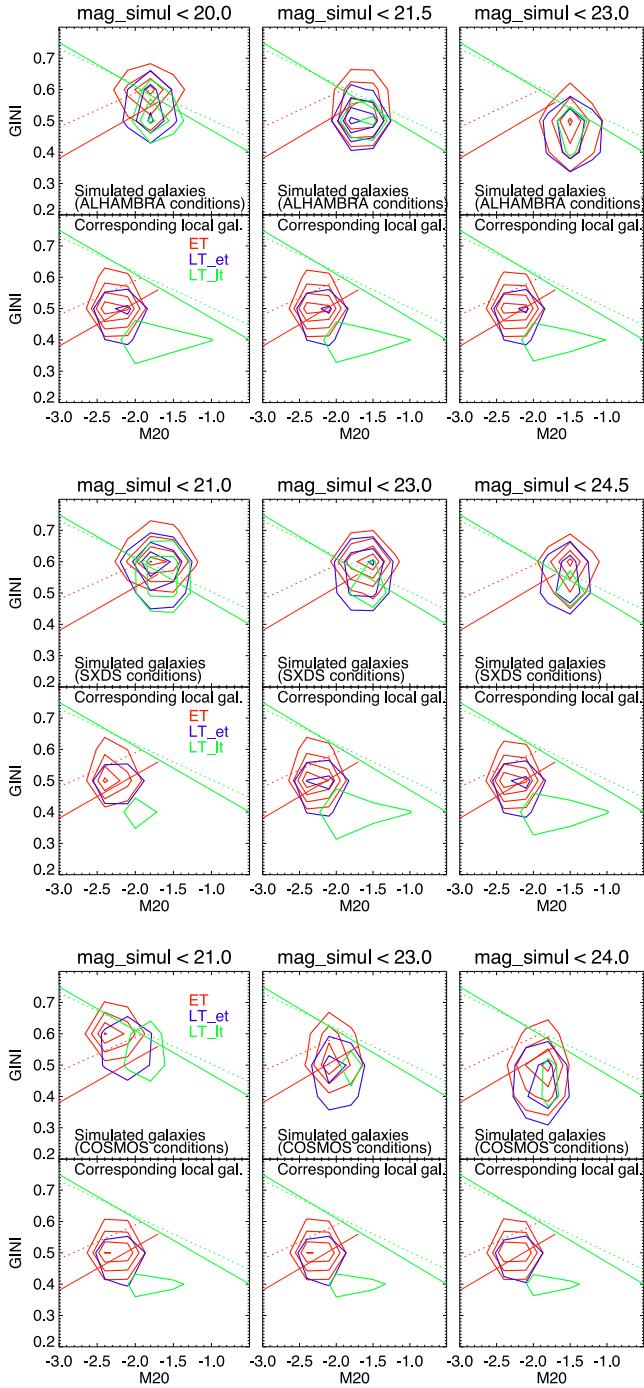


**Figure 15.** Same as Fig. 11, but representing the relation between the M20 and CABR.

ilar changes as in ALHAMBRA (at higher magnitude cuts), except for LT\_LT sources where no difference were found respect to local sample. In COSMOS conditions, M20 shows higher concentrations for simulated galaxies in comparison with the local ones, affecting only LT\_It galaxies, with differences  $\leq 31$  per cent (within 2 IQR). The population of sources with invalid measurements of M20 is consistent with that for CABR.

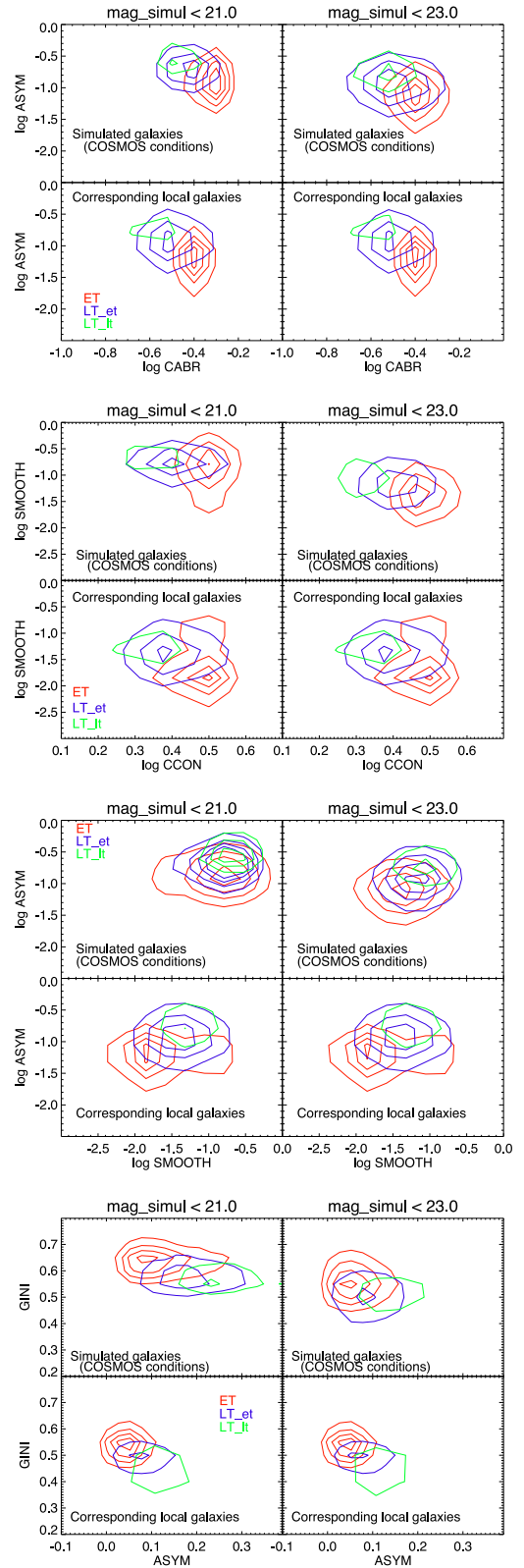
*ASYM.* This parameter showed the highest differences in the case of ALHAMBRA (even up to  $\sim 80$  per cent within 3 IQR at the faintest magnitude cut  $F613W \leq 23.0$ ), affecting LT galaxies in the sense of obtaining more symmetric galaxies in comparison





**Figure 16.** Same as Fig. 11, but representing the relation between the M20 and GINI. The green and red lines represent the limits of Lotz et al. (2008) to distinguish between the normal galaxies and mergers, and between ET and LT galaxies, respectively. In their classifications, mergers occupy the regions above the green lines, ETs (LTs) below the green lines and above (below) the red ones. Dotted lines (both green and red) provided this separation for a sample of local galaxies, while solid lines were used to classify sources at  $0.2 < z < 0.4$ .

with their reference ( $z \sim 0$ ) values. In COSMOS, all three morphological types suffered similar shifts ( $\sim 30$  per cent within 2 IQR), being the strongest one at the lowest magnitude cut  $F814W \leq 21.0$ , but showing positive values in comparison to ALHAMBRA, and so producing more asymmetric galaxies in



**Figure 17.** Relation between the CBR and ASYM (top), CCON and SMOOTH (middle, top), ASYM and SMOOTH (middle, bottom), and ASYM and GINI (bottom) in COSMOS, in the two analysed magnitude cuts, where the noise effect on ASYM and SMOOTH parameters is still insignificant (see Section 3.3). For more explications see the caption of Fig. 11.

simulated conditions. The population with invalid measurements was higher (<20 per cent) than in the case of CABR, GINI, and CCON. All these measurements were obtained without taking into account the noise introduced by the galaxy itself once being dropped into the background of non-local surveys. However, as showed in Section 3.3, in the case of ASYM and SMOOTH this noise could have a significant effect on these parameters, and the differences measured there should be added to those measured in this section. Therefore, we found these parameters unreliable for the use at higher redshifts in the case of ground-based surveys. In the space-based surveys similar to COSMOS, it becomes unreliable for the use in the case of galaxies fainter than  $F814W = 24.0$ .

**SMOOTH.** Without taking into account the noise studied in Section 3.3, the differences in SMOOTH in the case of ALHAMBRA were detected only for LT galaxies, being lower than 36 per cent (within 2 IQR); in the case of SXDS, no differences were observed between the simulated and local samples; nevertheless, in COSMOS conditions it showed 40–50 per cent differences (within 2–3 IQR) at the brightest magnitude cuts  $F814W \leq 21.0$ , resulting in galaxies to appear more clumpy. Again, to these differences those related to the inclusion of noise (see Section 3.3) should be added. Moreover, this parameter showed to be significantly more affected with both spatial resolution and depth, when considering the number of invalid measurements of the simulated sample (up to 50 and 20 per cent at the faintest magnitude cuts in the case of ALHAMBRA and COSMOS, respectively). Taking all this into account, we found this parameter especially sensitive to all, noise, spatial resolution, and data depth, and therefore unreliable for the use in the morphological classification of galaxies in the case of ground-based surveys, and for faintest sources in space-based surveys.

### 4.3 Behaviour of morphological diagnostic diagrams

In this section, we analyse how the observational bias, discussed in the previous section, affects some of the commonly used diagnostic diagrams, that allow us to separate between the ET and LT galaxies, to select disturbed galaxies, and/or interacting (merging) systems. To do this, we analysed the simulated sample in the conditions of the three non-local surveys at different magnitude cuts, and compared to the corresponding local sample. We can directly use the obtained diagrams to quantify how strong is the bias for each of these diagrams when taking into account the survey observational properties: basically the spatial resolution (comparing the results between the ground-based and space-based surveys) and survey depth (observing the changes at different magnitude cuts).

As already mentioned in Section 3.3, we analysed diagrams based on CABR, GINI, CCON, and M20 in all surveys, since these parameters are more stable to the noise effects. Figs 11–16 show the relation between CCON and CABR, CCON and GINI, CCON and M20, GINI and CABR, M20 and CABR, and finally M20 and GINI, respectively. In each figure are shown the results obtained for the conditions in the three surveys: ALHAMBRA (top), SXDS (middle), and COSMOS (bottom). For each survey, we plot the locus of the three analysed morphological groups, ET (red contours), LT\_et (blue contours), and LT\_lt (green contours), obtained once local galaxies (bottom plots) were moved to higher redshift conditions (top plots). Although initially we simulated the same sample of 3000 local galaxies to the conditions of the three non-local surveys, in some plots of the reference (local) values small differences can be observed, which is the reason why we represent them in each survey and for each magnitude cut. To understand these differences, it has to be considered that, first, the local measurements corre-

spond to the rest-frame band of the simulated sample, and hence slightly changes depending on the magnitude cut and survey (see Section 4.2). Secondly, the simulated galaxies with invalid measurements of morphological parameters (discussed in the previous section) are excluded in all plots, and therefore the corresponding local galaxies are also excluded from the bottom diagrams. ASYM and SMOOTH were only used in the case of COSMOS and its first two magnitude cuts. Fig. 17 shows some of the commonly used relations, representing the relation between CABR and ASYM (top plots), CCON and SMOOTH (middle plots), and finally ASYM and GINI (bottom plots). In all three plots, we represent the same comparisons as in Figs 11–16.

In all figures, when observing the reference parameters for local galaxies, we can clearly separate the regions typically occupied by ET, LT\_et, and LT\_lt galaxies, as expected. On the other side, we can observe how the position and the shape of the same regions change once we go to fainter magnitudes (higher redshifts). We measured the level of contamination for the highest density population (50 per cent) of each analysed morphological group (ET, LT\_et, and LT\_lt) with the other two types, for the morphological diagrams presented in Figs 11–16. We applied the Tukey’s Five Number Summary statistic (Tukey 1977), and measured for each parameter, and each morphological type: the sample minimum, first quartile, median value, third quartile, and sample maximum. The first and third quartiles, present the median values of the lower half and the upper half of the sample, respectively, with respect to the median value of the total sample. Therefore, measuring the first and the third quartile of each parameter on the particular diagnostic diagram, we are able to determine the regions that correspond to the highest population (50 per cent of sources) for ET, LT\_et, and LT\_lt around their median values. From this, for each morphological type, we can estimate the level of contamination by the other two types, within their densest regions: e.g. contamination of the densest ET region with the densest region of the LT\_et galaxies (as a number of LT\_et galaxies in the overlapping area, normalized with the number of ET galaxies), and/or with the densest region of the LT\_lt galaxies, and vice versa. Table 2 shows these measurements for simulated galaxies in ALHAMBRA (top table), SXDS (middle table), and COSMOS (bottom table), for their corresponding magnitude cuts. These measurements represent a quantification of the bias represented in the diagrams 11–16 in all the three surveys, and in Fig. 17 in the case of COSMOS. We can again observe that the contamination levels in the morphological types increase with the data depth in all three surveys, being however significantly lower in the case of COSMOS.

## 5 DISCUSSION

### 5.1 Individual parameters

As seen in Section 4.2, the morphological parameters show complex behaviours when going from one survey to another, from one magnitude cut to another, and when dealing with different morphological types. In principle, they suffer changes due to the impact of spatial resolution and survey depth, and depending on how strong these two effects are, they might cause

- (i) only a re-distribution of the light within the galaxy, without changing much the total flux, resulting in a weak impact, and/or
- (ii) the loss of the galaxy low surface brightness structures and the change of the total flux, hence producing a more significant impact.

However, how the distributions of morphological parameters behave in each survey when going from brighter to fainter magnitude cuts (from lower to higher redshifts), depends on their definitions. In general, for weak impacts of spatial resolution and data depth, the galaxies might appear as more concentrated, asymmetric and clumpy, while for stronger impacts, when the galaxy total flux can be affected significantly, the galaxies might appear as less concentrated, more symmetric, and less clumpy. To trace the real impact from spatial resolution and depth on the parameters in Figs 8–10, it is not only important to take into account the difference between the simulated and real values (and its standard deviation), but also the sign of that difference. In general, for the concentration parameters (CABR, GINI, CCON), the distributions will be the most affected when the final differences have positive values (showing that the simulated galaxies are less concentrated in comparison with the real ones); negative in case of M20, ASYM, and SMOOTH due to the use of the logarithm scale in our analysis.

## 5.2 Diagnostic diagrams: impact from spatial resolution and data depth

To evaluate the separate effect of resolution and data depth, we compare two cases: first, the results obtained for the two ground-based surveys, where both of them are affected with the spatial resolution, and we can study how important is the survey depth, and secondly, the results obtained for space-based data in comparison to the ground-based surveys, where we can see the role that play the spatial resolution in morphological classification.

### 5.2.1 The effect of survey depth in deep ground-based surveys: ALHAMBRA and SXDS

From the morphological diagrams and from the Table 2, for the deep ground-based surveys, like ALHAMBRA and SXDS, diagrams combined with M20 show to be less effective in separating galaxies. On the other side, as seen in all figures, survey depth is crucial when classifying galaxies. In the case of ALHAMBRA, at its lowest magnitude cut ( $F613W \leq 20.0$ ), although in all diagrams the position/shape of locus of the three morphological types change significantly in comparison with the same sample at  $z \sim 0$ , in most of diagrams we are still able to distinguish between the ET and LT galaxies, or at least the regions of the highest source densities do not overlap. Observing for each type their contours enclosing 50 per cent of galaxies, we found an  $\sim 5$ –25 per cent contamination of ET with LT\_et sources, and an  $\sim 3$ –35 per cent, of LT\_et with ET sources, where the contamination level depends on the used diagram, with GINI versus CABR and CCON versus CABR diagrams showing the lowest contamination ( $< 5$  per cent), and CCON versus M20, and M20 versus GINI the highest one. Similar conclusions are obtained for classifying LTs, but with higher contamination by ET sources. When including fainter galaxies (second magnitude cut  $F613W \leq 21.5$ ), the mixing between the three regions increases significantly, but still for diagrams that do not use M20 parameter the contamination levels of ET and LT\_et are between 20 and 30 per cent, depending on the diagram. In the case of the last magnitude cut ( $F613W \leq 23.0$ ) the mixing is much higher,  $\sim 60$ –100 per cent, making each of the diagrams useless if used separately. On the other side, in the case of SXDS we observe a similar situation at the different magnitude cuts as in the case of ALHAMBRA, but when including fainter galaxies at higher redshifts. The contamination levels obtained at the lowest mag-

nitude cut ( $i' \leq 21.0$ ) in SXDS (taking into account the central wavelength of the  $i'$  band, of  $\sim 760$  nm, this would corresponds to even higher magnitudes in ALHAMBRA), are higher than those obtained at the lowest analysed magnitude cut in ALHAMBRA ( $F613W \leq 20.0$ ), but lower than the second magnitude cut ( $F613W \leq 21.5$ ). Moreover, at the highest magnitude cut in the SXDS ( $i' \leq 24.5$ ), we obtained  $\sim 10$ –80 per cent lower contaminations, depending on the type, than in the case of the last magnitude cut in ALHAMBRA ( $F613W \leq 23.0$ ). The combination of other parameters with M20 (Figs 13, 15, and 16), which seems to be the most affected with the resolution, again shows to be the most ineffective. Finally, we should notice that either the effectiveness in separating between ET and LT\_et changes from one diagram to another and from one magnitude cut to another, LT\_lt galaxies in all cases, in both surveys, suffer very high levels of contamination ( $> 80$  per cent in most of cases) from other two morphological types.

### 5.2.2 The effect of resolution. COSMOS as a reference

Since the spatial resolution affects little the measurements of the morphological parameters in the case of COSMOS data, in comparison with the results obtained in ALHAMBRA and SXDS surveys, we can evaluate how much impact does it have when classifying galaxies. In all diagnostic diagrams (Figs 11–16), the loci of galaxies simulated to the conditions of COSMOS, irrespective of their morphological type, change significantly less at all magnitude cuts than in case of ALH and SXDS. Moreover, the diagrams that include ASYM and SMOOTH parameters provide the additional information, and can be used up to  $F814W = 24.0$  magnitudes. However, the data are still affected by the survey depth, as can be seen in all figures when comparing the diagrams at different magnitude cuts. When including only the brightest simulated galaxies in the sample ( $F814W \leq 21.0$ ) the locus of the three morphological groups basically does not change in comparison with that of the corresponding local values, and the contamination levels between ET and LT galaxies are  $\leq 10$  per cent. This picture starts to change when including fainter galaxies, down to  $F814W \leq 23.0$ , and even more at  $F814W \leq 24.0$ . Depending on the diagram, the contamination rate changes between 2 and 25 per cent at  $F814 \leq 23.0$ , and 6–35 per cent for ET and 20–70 per cent for LT galaxies at  $F814 \leq 24.0$ . However, even at the highest analysed magnitude cut, which corresponds to even fainter magnitudes in ALHAMBRA and SXDS selected bands, we still do not observe the complete overlapping of the ET and LT regions like in the case of ground-based surveys, and the contamination levels are still below those obtained for the brightest ALHAMBRA/SXDS magnitudes.

Although the impact from survey depth on the morphological parameters measured within COSMOS conditions is still present, it is significantly lower than in the case of deep ground-based surveys like ALHAMBRA and SXDS. We can assume that the spatial resolution is the main limitation that should be taken into account when dealing with the photometric data and morphological classification of galaxies.

## 5.3 Diagnostic diagrams: implications for morphological classification of galaxies

For intermediate- and high-redshift galaxy samples, and large data sets, typical for current (and future) deep photometric surveys,

**Table 2.** Expected level of contamination for densest regions (50 per cent of sources) of ET, LT\_et, and LT\_lt galaxies in some of the most used morphological diagrams in the ALHAMBRA (top), SXDS (middle), and COSMOS (bottom) surveys. Column description: Diagram – commonly used morphological diagrams (for the description of each parameter see Section 3.1); mag\_cut – for each survey, and in each diagram, the contamination levels are provided at three corresponding magnitude cuts (see Section 3.3); cont. ET, cont. LT\_et, and cont. LT\_lt – contamination level of ET, LT\_et, and LT\_lt simulated local galaxies, respectively, with the other two morphological types at the corresponding magnitude cuts.

	Diagram	mag1_cut ≤ 20.0			mag2_cut ≤ 21.5			mag3_cut ≤ 23.0		
		Cont. ET (per cent)	Cont. LT_et (per cent)	Cont. LT_lt (per cent)	Cont. ET (per cent)	Cont. LT_et (per cent)	Cont. LT_lt (per cent)	Cont. ET (per cent)	Cont. LT_et (per cent)	Cont. LT_lt (per cent)
A L H A M B R A	CCON versus CABR	– (LT_et) 3 (LT_lt)	4 (ET) 11 (LT_lt)	– (ET) 42 (LT_et)	24 (LT_et) – (LT_lt)	31 (ET) 20 (LT_lt)	2 (ET) 77 (LT_et)	58 (LT_et) 20 (LT_lt)	71 (ET) 35 (LT_lt)	92 (ET) 100 (LT_et)
	CCON versus GINI	11 (LT_et) – (LT_lt)	13 (ET) 16 (LT_lt)	– (ET) 58 (LT_et)	31 (LT_et) 1 (LT_lt)	38 (ET) 24 (LT_lt)	6 (ET) 86 (LT_et)	56 (LT_et) 22 (LT_lt)	78 (ET) 36 (LT_lt)	100 (ET) 100 (LT_et)
	CCON versus M20	– (LT_et) 34 (LT_lt)	46 (ET) 12 (LT_lt)	– (ET) 40 (LT_et)	50 (LT_et) 2 (LT_lt)	75 (ET) 19 (LT_lt)	13 (ET) 88 (LT_et)	65 (LT_et) 21 (LT_lt)	93 (ET) 34 (LT_lt)	100 (ET) 100 (LT_et)
	GINI versus CABR	3 (LT_et) – (LT_lt)	3 (ET) 22 (LT_lt)	16 (ET) 100 (LT_et)	29 (LT_et) 3 (LT_lt)	31 (ET) 36 (LT_lt)	49 (ET) 100 (LT_et)	58 (LT_et) 28 (LT_lt)	77 (ET) 42 (LT_lt)	100 (ET) 100 (LT_et)
	M20 versus CABR	6 (LT_et) – (LT_lt)	8 (ET) 20 (LT_lt)	– (ET) 64 (LT_et)	37 (LT_et) 5 (LT_lt)	48 (ET) 33 (LT_lt)	23 (ET) 100 (LT_et)	62 (LT_et) 27 (LT_lt)	80 (ET) 41 (LT_lt)	100 (ET) 100 (LT_et)
	M20 versus GINI	25 (LT_et) 2 (LT_lt)	36 (ET) 32 (LT_lt)	– (ET) 100 (LT_et)	54 (LT_et) 21 (LT_lt)	65 (ET) 40 (LT_lt)	85 (ET) 100 (LT_et)	60 (LT_et) 30 (LT_lt)	92 (ET) 42 (LT_lt)	100 (ET) 100 (LT_et)
		Diagram	cont. ET (per cent)	cont. LT_et (per cent)	cont. LT_lt (per cent)	cont. ET (per cent)	cont. LT_et (per cent)	cont. LT_lt (per cent)	cont. ET (per cent)	cont. LT_et (per cent)
S X D S	CCON versus CABR	31 (LT_et) – (LT_lt)	29 (ET) 15 (LT_lt)	– (ET) 81 (LT_et)	45 (LT_et) 4 (LT_lt)	41 (ET) 19 (LT_lt)	11 (ET) 86 (LT_et)	51 (LT_et) 18 (LT_lt)	42 (ET) 22 (LT_lt)	14 (ET) 98 (LT_et)
	CCON versus GINI	29 (LT_et) 0.5 (LT_lt)	40 (ET) 15 (LT_lt)	2 (ET) 83 (LT_et)	34 (LT_et) 2 (LT_lt)	41 (ET) 22 (LT_lt)	8 (ET) 93 (LT_et)	48 (LT_et) 6 (LT_lt)	55 (ET) 26 (LT_lt)	28 (ET) 94 (LT_et)
	CCON versus M20	52 (LT_et) 4 (LT_lt)	69 (ET) 20 (LT_lt)	27 (ET) 100 (LT_et)	56 (LT_et) 11 (LT_lt)	77 (ET) 23 (LT_lt)	64 (ET) 100 (LT_et)	71 (LT_et) 13 (LT_lt)	77 (ET) 26 (LT_lt)	73 (ET) 100 (LT_et)
	GINI versus CABR	24 (LT_et) – (LT_lt)	36 (ET) 18 (LT_lt)	– (ET) 88 (LT_et)	30 (LT_et) 2 (LT_lt)	33 (ET) 21 (LT_lt)	6 (ET) 94 (LT_et)	47 (LT_et) 11 (LT_lt)	42 (ET) 25 (LT_lt)	13 (ET) 100 (LT_et)
	M20 versus CABR	38 (LT_et) – (LT_lt)	46 (ET) 18 (LT_lt)	– (ET) 98 (LT_et)	48 (LT_et) 10 (LT_lt)	56 (ET) 25 (LT_lt)	34 (ET) 100 (LT_et)	56 (LT_et) 47 (LT_lt)	49 (ET) 24 (LT_lt)	28 (ET) 90 (LT_et)
	M20 versus GINI	35 (LT_et) 2 (LT_lt)	55 (ET) 27 (LT_lt)	10 (ET) 100 (LT_et)	38 (LT_et) 6 (LT_lt)	63 (ET) 24 (LT_lt)	31 (ET) 100 (LT_et)	60 (LT_et) 10 (LT_lt)	67 (ET) 30 (LT_lt)	53 (ET) 100 (LT_et)
		Diagram	cont. ET (per cent)	cont. LT_et (per cent)	cont. LT_lt (per cent)	cont. ET (per cent)	cont. LT_et (per cent)	cont. LT_lt (per cent)	cont. ET (per cent)	cont. LT_et (per cent)
C O S M O S	CCON versus CABR	– (LT_et) – (LT_lt)	– (ET) 7 (LT_lt)	– (ET) 36 (LT_et)	– (LT_et) – (LT_lt)	2 (ET) 9 (LT_lt)	– (ET) 31 (LT_et)	6 (LT_et) – (LT_lt)	17 (ET) 15 (LT_lt)	– (ET) 62 (LT_et)
	CCON versus GINI	– (LT_et) – (LT_lt)	– (ET) 11 (LT_lt)	– (ET) 48 (LT_et)	4 (LT_et) – (LT_lt)	11 (ET) 12 (LT_lt)	– (ET) 40 (LT_et)	11 (LT_et) – (LT_lt)	30 (ET) 19 (LT_lt)	– (ET) 75 (LT_et)
	CCON versus M20	– (LT_et) – (LT_lt)	– (ET) 3 (LT_lt)	– (ET) 17 (LT_et)	7 (LT_et) – (LT_lt)	15 (ET) 6 (LT_lt)	– (ET) 20 (LT_et)	22 (LT_et) – (LT_lt)	38 (ET) 12 (LT_lt)	– (ET) 47 (LT_et)
	GINI versus CABR	– (LT_et) – (LT_lt)	– (ET) 28 (LT_lt)	0 (ET) 100 (LT_et)	2 (LT_et) – (LT_lt)	4 (ET) 35 (LT_lt)	– (ET) 100 (LT_et)	13 (LT_et) – (LT_lt)	30 (ET) 34 (LT_lt)	– (ET) 100 (LT_et)
	M20 versus CABR	– (LT_et) – (LT_lt)	– (ET) 8 (LT_lt)	– (ET) 30 (LT_et)	2 (LT_et) – (LT_lt)	5 (ET) 12 (LT_lt)	– (ET) 47 (LT_et)	15 (LT_et) – (LT_lt)	32 (ET) 12 (LT_lt)	– (ET) 77 (LT_et)
	M20 versus GINI	5 (LT_et) – (LT_lt)	7 (ET) 10.5 (LT_lt)	– (ET) 43 (LT_et)	24 (LT_et) – (LT_lt)	55 (ET) 17 (LT_lt)	– (ET) 63 (LT_et)	35 (LT_et) 7 (LT_lt)	75 (ET) 26 (LT_lt)	– (ET) 100 (LT_et)
	logCABR versus logASYM	– (LT_et) – (LT_lt)	– (ET) 13 (LT_lt)	– (ET) 45 (LT_et)	1.3 (LT_et) – (LT_lt)	3 (ET) 18 (LT_lt)	– (ET) 55 (LT_et)	13 (LT_et) – (LT_lt)	26 (ET) 24 (LT_lt)	– (ET) 80 (LT_et)
	logCCON versus logSMOOTH	– (LT_et) – (LT_lt)	– (ET) 6.3 (LT_lt)	– (ET) 38 (LT_et)	5.5 (LT_et) – (LT_lt)	11.3 (ET) 10 (LT_lt)	– (ET) 34 (LT_et)	11 (LT_et) – (LT_lt)	17 (ET) 16 (LT_lt)	– (ET) 60 (LT_et)
	logSMOOTH versus logASYM	56 (LT_et) – (LT_lt)	28 (ET) 11 (LT_lt)	64 (ET) – (LT_et)	35 (LT_et) – (LT_lt)	39 (ET) 18 (LT_lt)	– (ET) 78 (LT_et)	42 (LT_et) 5 (LT_lt)	47 (ET) 22 (LT_lt)	46 (ET) 100 (LT_et)
	ASYM versus GINI	7 (LT_et) – (LT_lt)	6 (ET) 20 (LT_lt)	– (ET) 67 (LT_et)	16 (LT_et) – (LT_lt)	31 (ET) 25 (LT_lt)	– (ET) 84 (LT_et)	28 (LT_et) 4 (LT_lt)	60 (ET) 30 (LT_lt)	40 (ET) 100 (LT_et)

the application of automated non-parametric methods, that use the parameters analysed within this work, is one of the ways to obtain the information about the morphology. In some of the latest works, the classification based on non-parametric methods was done using a set of parameters simultaneously, therefore decreasing the impact of spatial resolution and depth (e.g. Huertas-Company et al. 2009, 2011; Pović et al. 2012, 2013; Carollo et al. 2013; Cibinel et al. 2013; Cotini et al. 2013; Holwerda et al. 2014). However, for classifications based on only one, two, or three parameters, special attention should be taken in relation to the magnitude and redshift distribution of the analysed sample, and the data quality.

The ASYM index was used in some of previous works to select mergers and disturbed galaxies (Lotz et al. 2010a,b, 2011; Villforth et al. 2014). ASYM is especially sensitive to the survey noise, as showed in Section 3.3, and to both, the spatial resolution and depth, as showed in Section 4.2. In surveys similar to COSMOS, for sources fainter than  $F814 = 23.0$ , we again do not recommend the use of a single criterion based on ASYM when classifying them morphologically, since the contamination levels of 50 per cent highest density regions of ET and LT rise to above 40 per cent. In the case of ground-based surveys, we do not recommend the use of this single criterion in any case (except for maybe the brightest galaxies), since the contamination levels are much higher, taking the noise effect especially into account; we suggest instead the simultaneous use of multiple diagnostic diagrams.

Over the past 10 yr, the M20 - GINI diagram (Fig. 16) found wide applications. It was mainly used in identifying galaxy mergers and peculiar (disturbed) galaxies (e.g. Lotz et al. 2004, 2008; Pierce et al. 2007; Conselice et al. 2008; Mendez et al. 2011; Chen et al. 2009; Stott et al. 2013; Petty et al. 2014; Hung et al. 2014); often it was used in the analysis where different types of galaxies, e.g. ULIRGs, dust obscured galaxies, star-forming galaxies, or quiescent galaxies, were compared with control samples (Bussmann et al. 2009; Pentericci et al. 2010; Lanyon-Foster et al. 2012; Wang et al. 2012; Böhm et al. 2013; Chung et al. 2013); or for selecting ET and LT galaxies (Lotz et al. 2004, 2008).

From our results, in ALHAMBRA and SXDS surveys even for the brightest analysed sources ( $F613 \leq 20.0$  and  $i' \leq 21.0$ , respectively) the contamination levels between the ET and LT galaxies in M20 - GINI diagram are about 30 per cent. In COSMOS however, at the lowest magnitude cut  $F814 \leq 21.0$  ( $z \lesssim 1$ ) the contamination is insignificant ( $\lesssim 10$  per cent, except for the youngest spiral/irregular galaxies), but already at the second magnitude cut  $F814 \leq 23.0$  ( $z \lesssim 1.5$ ) the contamination is  $>20$  per cent. We represented in Fig. 16 the limits established by Lotz et al. (2004, 2008) to separate between normal ET and LT galaxies and major mergers, in the local universe (dotted lines), and at  $0.2 < z < 0.4$  (solid lines), using the All-wavelength Extended Groth strip International Survey (AEGIS) *HST*/ACS data. The solid-lines criterion (non-local sample) was later applied by Lotz et al. (2008, 2010a,b) to  $z \leq 1.2$  sources to select major mergers, reporting a population of  $\sim 15$  per cent of misclassified sources, and was used in other works, as mentioned above. From the established limits, we can see again how sensitive the diagram is to spatial resolution and data depth, changing significantly between the ground- and space-based surveys. While in COSMOS the contamination by normal galaxies in the region assigned for major mergers, does not exceed the values measured by the authors (even at higher redshifts), in the analysed ground-based surveys the contamination can be  $>70$  per cent.

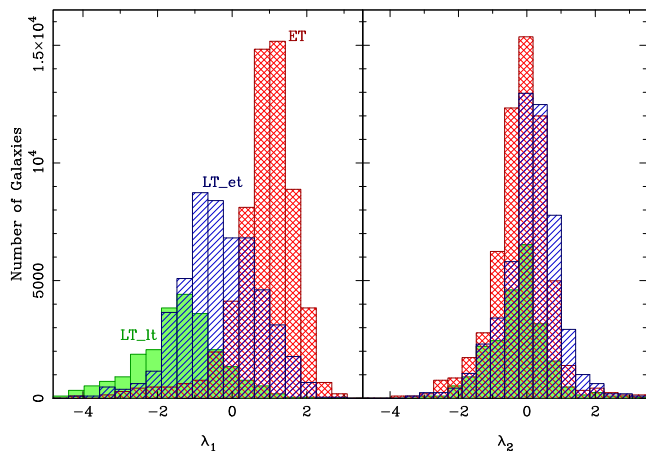
When applying the M20 - GINI diagram to classify galaxies and select mergers, we highly recommend to use it always with additional criteria when dealing with ground-based data, among which: a

combination with other morphological diagrams, a careful selection of the control sample and evaluation of the diagram effectiveness, and/or application of probabilistic approaches. In space-based surveys like COSMOS to distinguish between the ET and LT sources additional criteria should be taken into account at least for galaxies fainter than  $F814 = 23.0$ .

Finally, different combinations of two to three parameters analysed in the previous sections, were used in many studies to distinguish between the ET and LT galaxies, and also to study the structure of galaxies with other properties, like SFR, mass, environment, nuclear activity, etc. (e.g. Cassata et al. 2007; Pierce et al. 2007; Bussmann et al. 2009; Pović et al. 2009b,a; Pentericci et al. 2010; Böhm et al. 2013; Chung et al. 2013). The combination between the concentration parameters (CABR and/or CCON) and ASYM was also used for selecting ET and LT sources (Abraham et al. 1994, 1996; Conselice 2006; Huertas-Company et al. 2007; Pović et al. 2009b), while in some cases the combination between the GINI and CCON was used to study the host properties of AGN and quasars (e.g. Urrutia et al. 2008). The CCON - ASYM - SMOOTH diagram was used in classifying both galaxies and mergers (Conselice 2003; Cassata et al. 2007; Conselice et al. 2008). We found SMOOTH to be especially sensitive to both spatial resolution and depth, and the most unstable parameter of all analysed. Moreover, together with ASYM, we also found it to be more sensitive to noise, as showed in Section 3.3. As already mentioned, we do not recommend the use of these two parameters in the ground-based surveys if the noise effects are not taken into account carefully. In surveys similar to COSMOS, they can be used in combination with other parameters but for sources brighter than  $F814 = 24.0$ . The combination of only ASYM and SMOOTH gives high contamination levels even in the case of COSMOS (as showed in Fig. 17 and Table 2).

We suggest the consultation of Table 2 and Figs 11–17 when using diagnostic diagrams in morphological studies, especially if they are based on two to three parameters. We discussed in Section 4.3 the effectiveness of a particular diagnostic diagram in the ground-based surveys and noted that in cases similar to both ALHAMBRA and SXDS, special care has to be taken into account when classifying galaxies at all magnitude cuts (redshifts). Our suggestion when using ground-based data is to avoid the diagnostic diagrams based on two to three parameters, and to apply instead all morphological parameters simultaneously, and to use the statistical approaches providing a probability for a galaxy to belong to a given morphological type (e.g. approaches similar to that used in the *GALSVM* code and tested in Huertas-Company et al. 2008). Diagrams that include M20 showed to be more affected by spatial resolution, and less effective in separating galaxies. Moreover, the noise should be taken into account if using ASYM and SMOOTH parameters.

In space-based surveys similar to COSMOS, even at the highest magnitude cuts ( $F814 \leq 24.0$ ), in all diagnostic diagrams the contamination levels are significantly lower than at the brightest magnitudes in the ground-based surveys. The observational bias is insignificant for all parameters at magnitudes brighter than  $F814 = 21.0$  (except for the youngest spiral and irregular galaxies), as shown in Section 4.3; however when going to fainter magnitudes the contamination increases. Therefore, even when dealing with data sets similar to COSMOS, we remind that eventual contamination levels are present at fainter magnitudes; the diagnostic diagrams that show to be more effective in this case are CCON - CABR, CCON - GINI, CCON - SMOOTH, and/or CABR - ASYM. To apply again, when possible, different diagnostics will minimize the effect of the observational bias.

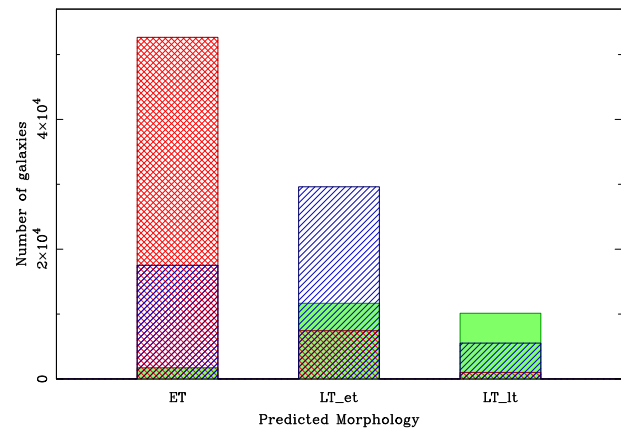


**Figure 18.** Histograms of the  $\lambda_1$  (left-hand panel) and  $\lambda_2$  (right-hand panel) eigenvectors resulting from the LDA of the local sample, without including the SMOOTH and ASYM parameters. The red cross-hatched, blue hatched, and green solid histograms correspond to the ET, LT\_et, and LT\_lt galaxies, respectively. As it can be seen,  $\lambda_1$  is the main morphology discriminator, whereas  $\lambda_2$  only slightly improves the classification for the LT\_et galaxies.

#### 5.4 Results from linear discriminant analysis

In order to merge all the analysed morphological parameters, and to provide an additional and statistical test of our previous results, a linear discriminant analysis (LDA; as implemented by Pedregosa et al. 2011) was performed. The LDA method allows us to find the linear combination of measured parameters that best separates two or more classes of objects or events. We tested the combination of the four morphological parameters (CABR, GINI, CCON, and M20) to distinguish between ET, LT\_et, and LT\_lt galaxies in the three analysed non-local surveys (ALHAMBRA, SXDS, and COSMOS). Knowing that ASYM and SMOOTH are found to be affected by noise, we performed the LDA without these two parameters. However, we also performed an LDA including all the six parameters for comparison and to check the effect of noise since both ASYM and SMOOTH parameters resulted to be the ones more significantly affected by noise, mainly in the two ground-based surveys, as we described in Section 3.3.

The training phase of the LDA was obtained by fitting the coefficients over the measured parameters of the galaxies of the local sample. We found that all the possible discrimination is based on two eigenvectors ( $\lambda_1$  and  $\lambda_2$ ) both being linear combinations of the four parameters entering in the classification. The first one,  $\lambda_1$ , is dominated by GINI, CCON, and CABR, whereas the second one,  $\lambda_2$ , is based on CABR, GINI, and M20. Almost all the power of the discrimination arise from the  $\lambda_1$  eigenvector whereas  $\lambda_2$  only slightly helps to the classification of LT\_et galaxies, as it can be seen in Fig. 18 where the distributions of both  $\lambda_1$  and  $\lambda_2$  for the three morphological types are shown. The three concentration indexes show to be more stable, discriminating better between different morphological types in comparison with M20. An attempt to improve the LDA was also carried out by doing a second LDA iteration including only those galaxies that show the same input and predicted output morphology. We did not find any significant change, and therefore did not proceed further with this additional step. As a result of the LDA of the local sample, we found that about 30 per cent of the objects show a different input morphology than the one predicted by the resulting eigenvectors of the LDA. This fact could be caused by errors in the measured parameters or even in the original classification. To illustrate the results obtained by the LDA, we present



**Figure 19.** Histogram of the predicted morphology obtained from the LDA. For each predicted morphology, the number of galaxies of each different input morphological type are displayed. Colours and representations are the same as in Fig. 18. ET galaxies are found to be well classified, however LT\_ET spread between adjacent input types, and LT\_LT galaxies are split between both late-types.

**Table 3.** LDA without SMOOTH and ASYM.

ALHAMBRA	Ngal	Score
20.0	139 470	0.583
21.5	139 046	0.510
23.0	137 463	0.356
SXDS	Ngal	score
21.0	13 872	0.478
23.0	14 579	0.533
24.5	14 427	0.525
COSMOS	Ngal	score
21.0	49 890	0.549
23.0	50 150	0.601
24.0	48 519	0.505

in Fig. 19, the histogram of the predicted morphologies after applying the eigenvector discrimination over the local sample, where for each output morphological type (ET, LT\_et, LT\_lt) we show the contribution of galaxies with different input morphology. As it clearly can be seen, whereas ET galaxies tend to be well classified, LT\_ET spread between adjacent input types, and LT\_LT are split between both input late-types.

After obtaining the eigenvectors that better discriminate the morphology in the local sample, LDA transformations to the ( $\lambda_1$  and  $\lambda_2$ ) space were performed for each non-local sample (ALHAMBRA, SXDX, COSMOS). The results so obtained at the three magnitude cuts and for each non-local sample are presented in Table 3. In this table, for each survey, the first column represents the three corresponding magnitude cuts, the second column corresponds to the total number of galaxies (Ngal) at each magnitude cut and sample (i.e. galaxies with valid measured parameters after applying the filtering criteria explained in Section 4.3) while the score column represents the ratio between the galaxies having the same LDA output morphology as the input one normalized to the total number of galaxies Ngal; a score of 1 would indicate a perfect LDA classification. We also tested the results when including ASYM and SMOOTH. In Table 4, we present the results obtained from the LDA taking into account all morphological parameters. In general, the scores obtained after removing these two parameters (see

**Table 4.** LDA using all parameters.

ALHAMBRA	Ngal	score
20.0	114 271	0.539
21.5	88 186	0.525
23.0	66 978	0.479
SXDS	Ngal	score
21.0	10 752	0.491
23.0	11 067	0.512
24.5	10 077	0.489
COSMOS	Ngal	score
21.0	49 292	0.653
23.0	47 988	0.655
24.0	43 061	0.600

**Table 5.** LDA results by morphological type.

ALHAMBRA	ET	LT_ET	LT_LT
20.0	0.86	0.51	0.02
21.5	0.47	0.74	0.10
23.0	0.17	0.55	0.41
SXDS	ET	LT_ET	LT_LT
21.0	0.73	0.33	0.15
23.0	0.72	0.51	0.10
24.5	0.55	0.67	0.13
COSMOS	ET	LT_ET	LT_LT
21.0	0.99	0.29	0.00
23.0	0.83	0.54	0.15
24.0	0.55	0.55	0.15

Table 3) are found to be similar to those that would be obtained including also ASYM and SMOOTH, however the total number of valid galaxies increase significantly and this is more notable in the case of the ground base surveys (ALHAMBRA and SXDS). If ASYM and SMOOTH are excluded, the total number of galaxies with good morphology is found to be larger, thus the final classification get worst when including these two parameters as they are indeed affected by noise. Only in surveys similar to COSMOS, the inclusion of ASYM and SMOOTH would have a slightly positive net effect.

Table 5 corresponds to the obtained scores (ratio of galaxies with same output/input morphology over total galaxies) for each morphological type and magnitude cut in each survey, as resulting from the LDA with neither SMOOTH nor ASYM included. As can be seen, in all surveys, ET galaxies are better recognized at the brighter magnitude cuts, whereas LTs tend to be better recognized at fainter magnitude cuts.

## 6 SUMMARY AND CONCLUSIONS

In this paper, we present for the first time a systematic study of the impact from spatial resolution and depth on the six morphological parameters commonly used for galaxy morphological classification: CABR, GINI, CCON, M20, ASYM, and SMOOTH. We studied how strong the impact is for three data sets that have different spatial resolutions and depth: ALHAMBRA and SXDS, as an example of ground-based data, and COSMOS, as an example of deep space-based data. Our results correspond to maximum analysed redshift values of: 1.0, 2.2, and 1.6 in the three surveys, respectively, where the provided values describe the distribution of 95 per cent of

each sample. We used a sample of 3000 early- and late-type local galaxies, with the available visual morphological classification, and we measured their parameters in two cases:

- (i) first, the reference values, that correspond to the real galaxy redshifts (magnitudes), and
- (ii) secondly, the simulated values, obtained after moving the local galaxies to the observational conditions and magnitude/redshift distribution of three selected surveys.

Comparing the results obtained in these two cases we showed how each parameter changes in each survey, at three particular magnitude cuts. Finally, we analysed and quantified how the impact from spatial resolution and depth affects some of the most used morphological diagrams, gave some suggestions for the galaxy/merger classification studies, and used the LDA statistical approach to analyse the most effective combination of parameters to distinguish between different types.

In the following, we summarize some of our main findings as follows.

- (i) All six analysed morphological parameters suffer from significant biases related to the spatial resolution and data depth.
- (ii) The impact of the spatial resolution on the morphology is much stronger in comparison with the data depth, being therefore the most responsible for changing the parameters in the ground-based surveys, making in general the galaxies to appear less concentrated and more symmetric.
- (iii) We stress that ASYM and SMOOTH are more sensitive to noise effects. Survey noise should be taken into account carefully when using these two parameters in morphological classification of galaxies, especially when dealing with ground-based data.
- (iv) M20 results to be also significantly affected in all surveys, changing both the shape and range of its distribution at all brightness levels. However, M20 - GINI is the most used diagram for selecting interacting (merging) systems. We highly recommend the use of this diagram simultaneously with other morphological parameters, when dealing with ground-based data sets, and in space-based surveys like COSMOS at least when studying sources fainter than  $F814 = 23.0$ .

(v) CCON shows to be more sensitive to the spatial resolution in comparison to GINI and CABR. It works therefore better for space-based data, whereas the other two concentration indexes behave better for ground-based surveys.

(vi) In surveys similar to ALHAMBRA, when analysing the highest density regions of early- and late-type galaxies in the main morphological diagnostic diagrams, the impact from spatial resolution and data depth introduces contamination levels of 5–25 per cent for ET and 3–35 per cent for LT galaxies (depending on the diagram), for the brightest galaxies with  $F613 \leq 20.0$ . The diagrams that seem to work the best in this case are those that do not include M20. At the faintest analysed magnitudes ( $F613 \leq 23.0$ ) the contamination levels increase significantly, being as high as 60–100 per cent, making each of the diagnostics useless if used separately. Similar results are obtained in the case of SXDS, but at higher magnitude cuts (of the order of 1–2). Moreover, in both surveys, at all magnitude cuts, classification of LT<sub>lt</sub> galaxies suffers very high levels of contamination. Taking all this into account, when dealing with ground-based data sets, we suggest to avoid the use of two to three parameter diagnostic diagrams in morphological classification, and to apply instead the use of all morphological parameters simultaneously, and to use statistical approaches based on probability distributions that the galaxy is ET or LT.

(vii) In space-based surveys similar to COSMOS, even at the highest magnitude cuts ( $F814 \leq 24.0$ ), the contamination levels are significantly lower than those at the brightest magnitudes in the ground-based surveys. The observational bias is insignificant for all parameters at magnitudes brighter than  $F814 = 21.0$  (except for the LT<sub>It</sub> galaxies). However, when going to fainter magnitudes the contamination increases, and depending on the diagram it changes from 2–25 per cent at  $F814 \leq 23.0$ , to 6–35 per cent for ET, and 20–70 per cent for LT galaxies at  $F814 \leq 24.0$ . In surveys similar to COSMOS, we again suggest to use several diagnostics to classify galaxies when going to magnitudes fainter than  $F814 = 23.0$ .

(viii) Through LDA analysis, we obtained that the combination of CABR, GINI, and CCON parameters, is the most effective to distinguish between different morphological types.

The results presented in this paper can be directly applied to any survey similar to ALHAMBRA, SXDS and COSMOS, and also can serve as an upper/lower limit to take into account when classifying galaxies using shallower/deeper data sets.

## ACKNOWLEDGEMENTS

We thank Chris Lintott for accepting to review this paper, giving us constructive comments that in our believe improved the paper significantly. We also appreciate that he signed his report. We thank Kaz Sekiguchi and the SXDS team for informing us about the stage of the photometric redshift measurements. MP acknowledge financial support from JAE-Doc programme of the Spanish National Research Council (CSIC), co-funded by the European Social Fund. This research was supported by the Junta de Andalucía through project TIC114, and the Spanish Ministry of Economy and Competitiveness (MINECO) through projects AYA2010-15169, AYA2013-42227-P, and AYA2013-43188-P. The work uses the observations collected at the Centro Astronómico Hispano Alemán (CAHA) at Calar Alto, operated jointly by the Max-Planck Institut für Astronomie and the Instituto de Astrofísica de Andalucía (CSIC). The CEFCA is funded by the Fondo de Inversiones de Teruel, supported by both the Government of Spain (50 per cent) and the regional Government of Aragón (50 per cent). In this work, we made use of Virtual Observatory Tool for OPERations on Catalogues And Tables (TOPCAT). IRAF is distributed by the National Optical Astronomy Observatories, which are operated by the Association of Universities for Research in Astronomy, Inc., under cooperative agreement with the National Science Foundation. Funding for SDSS-III has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, and the U.S. Department of Energy Office of Science. The SDSS-III web site is <http://www.sdss3.org/>. SDSS-III is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS-III Collaboration including the University of Arizona, the Brazilian Participation Group, Brookhaven National Laboratory, Carnegie Mellon University, University of Florida, the French Participation Group, the German Participation Group, Harvard University, the Instituto de Astrofísica de Canarias, the Michigan State/Notre Dame/JINA Participation Group, Johns Hopkins University, Lawrence Berkeley National Laboratory, Max Planck Institute for Astrophysics, Max Planck Institute for Extraterrestrial Physics, New Mexico State University, New York University, Ohio State University, Pennsylvania State University, University of Portsmouth, Princeton University, the Spanish Participation Group, University of Tokyo, University of Utah, Vanderbilt University, University of Virginia, University of Washington, and Yale University.

## REFERENCES

- Abraham R. G., Valdes F., Yee H. K. C., van den Bergh S., 1994, *ApJ*, 432, 75
- Abraham R. G., van den Bergh S., Glazebrook K., Ellis R. S., Santiago B. X., Surma P., Griffiths R. E., 1996, *ApJS*, 107, 1
- Abraham R. G., van den Bergh S., Nair P., 2003, *ApJ*, 588, 218
- Adams T. F., 1977, *ApJS*, 33, 19
- Andrae R., Jahnke K., Melchior P., 2011a, *MNRAS*, 411, 385
- Andrae R., Melchior P., Jahnke K., 2011b, *MNRAS*, 417, 2465
- Aparicio-Villegas T. et al., 2010, *AJ*, 139, 1242
- Baillard A. et al., 2011, *A&A*, 532, 74
- Baldry I. K., Glazebrook K., Brinkmann J., Ivezić Ž, Lupton R. H., Nichol R. C., Szalay A. S., 2004, *ApJ*, 600, 681
- Baldwin J. A., Phillips M. M., Terlevich R., 1981, *PASP*, 93, 5
- Barden M., Häußler B., Peng C. Y., McIntosh D. H., Guo Y., 2012, *MNRAS*, 422, 449
- Bell E. F., McIntosh D. H., Katz N., Weinberg M. D., 2003, *ApJS*, 149, 289
- Benítez N., 2000, *ApJ*, 536, 571
- Bershady M. A., Jangren A., Conselice C. J., 2000, *AJ*, 119, 2645
- Bertin E., Arnouts S., 1996, *A&AS*, 117, 393
- Blanton M. R. et al., 2003, *ApJ*, 594, 186
- Böhm A. et al., 2013, *A&A*, 549, 46
- Brammer G. B., van Dokkum P. G., Coppi P., 2008, *ApJ*, 686, 1503
- Bussmann R. S. et al., 2009, *ApJ*, 693, 750
- Carollo C. M. et al., 2013, *ApJ*, 773, 112
- Casata P. et al., 2007, *ApJS*, 172, 270
- Chen Y., Lowenthal J. D., Yun M. S., 2010, *ApJ*, 712, 1385
- Choi Y.-Y., Woo J.-H., Park C., 2009, *ApJ*, 699, 1679
- Chung J., Rey S.-C., Sung E.-C., Yeom B.-S., Humphrey A., Yi W., Kyeong J., 2013, *ApJ*, 767, 15
- Cibinel A. et al., 2013, *ApJ*, 776, 72
- Cirasuolo M., Magliocchetti M., Celotti A., 2005, *MNRAS*, 357, 1267
- Conselice C. J., 2006, *MNRAS*, 373, 1389
- Conselice C. J., 2003, *ApJ*, 147, 1
- Conselice C. J., Bershady M. A., Jangren A., 2000, *ApJ*, 529, 886
- Conselice C. J., Rajgor S., Myers R., 2008, *MNRAS*, 386, 909
- Cotini S., Ripamonti E., Caccianiga A., Colpi M., Della Ceca R., Mapelli M., Severgnini P., Segreto A., 2013, *MNRAS*, 431, 2661
- de Souza R. E., Gadotti D. A., dos Anjos S., 2004, *ApJS*, 153, 411
- de Vaucouleurs G., 1948, *Ann. Astrophys.*, 11, 247
- Deng X.-F., 2013, *Res. Astron. Astrophys.*, 13, 651
- Doi M., Fukugita M., Okamura S., 1993, *MNRAS*, 264, 832
- Folkes S. R., Lahav O., Maddox S. J., 1996, *MNRAS*, 283, 651
- Furusawa H., 2008, *ApJS*, 176, 1
- Gabor J. M. et al., 2009, *ApJ*, 691, 705
- Graham A. W., Erwin P., Caon N., Trujillo I., 2001b, *ApJ*, 563, L11
- Graham A. W., Trujillo I., Caon N., 2001a, *AJ*, 122, 1707
- Gunn J. E. et al., 1998, *AJ*, 116, 3040
- Hartley W. G. et al., 2013, *MNRAS*, 431, 3045
- Heckman T. M., 1978, *PASP*, 106, 113
- Hickox R. C. et al., 2009, *ApJ*, 696, 891
- Ho L. C., Filippenko A. V., Sargent W. L., 1995, *ApJS*, 98, 477
- Hogg D. W. et al., 2003, *ApJ*, 585, 5
- Holwerda B. W. et al., 2014, *ApJ*, 781, 12
- Huertas-Company M., Rouan D., Soucaïl G., Le Fèvre O., Tasca L., Contini T., 2007, *A&A*, 468, 937
- Huertas-Company M., Rouan D., Tasca L., Soucaïl G., Le Fèvre O., 2008, *A&A*, 478, 971
- Huertas-Company M., Foex G., Soucaïl G., Pelló R., 2009, *A&A*, 505, 83
- Huertas-Company M., Aguerri J. A. L., Bernardi M., Mei S., Sánchez Almeida J., 2011, *A&A*, 525, 157
- Huertas-Company M. et al., 2014, preprint ([arXiv:1406.1175](https://arxiv.org/abs/1406.1175))
- Humason M. L., 1931, *ApJ*, 74, 35
- Hung C.-L. et al., 2014, *ApJ*, 791, 63
- Ilbert O. et al., 2009, *ApJ*, 690, 1236
- Jiménez-Teja Y., Benítez N., 2012, *ApJ*, 745, 150
- Kauffmann G. et al., 2003, *MNRAS*, 346, 1055



- Kelly B. C., McKay T. A., 2004, *ApJ*, 127, 625  
 Kelly B. C., McKay T. A., 2005, *AJ*, 129, 1287  
 Kelm B., Focardi P., Sorrentino G., 2005, *A&A*, 442, 117  
 Koekemoer A. et al., 2007, *ApJS*, 172, 196  
 Kormendy J., Richstone D., 1995, *ARA&A*, 33, 581  
 Lanyon-Foster M. M., Conselice C. J., Merrifield M. R., 2012, *MNRAS*, 424, 1852  
 Leauthaud A. et al., 2007, *ApJS*, 172, 219  
 Lintott C. J. et al., 2008, *MNRAS*, 389, 1179  
 Lintott C. J. et al., 2011, *MNRAS*, 410, 166  
 Lisker T., 2008, *ApJS*, 179, 319  
 Lotz J. M., Primack J., Madau P., 2004, *AJ*, 128, 163  
 Lotz J. M. et al., 2008, *ApJ*, 672, 177  
 Lotz J. M., Jonsson P., Cox T. J., Primack J. R., 2010a, *MNRAS*, 404, 575  
 Lotz J. M., Jonsson P., Cox T. J., Primack J. R., 2010b, *MNRAS*, 404, 590  
 Lotz J. M., Jonsson P., Cox T. J., Croton D., Primack J. R., Somerville R. S., Stewart K., 2011, *ApJ*, 742, 103  
 McLure R. J., Dunlop J. S., Kukuła M. J., 2000, *MNRAS*, 318, 693  
 Marconi A., Hunt L. K., 2003, *ApJ*, 589, L21  
 Melbourne J., Phillips A. C., Harker J., Novak G., Koo D. C., Faber S. M., 2007, *ApJ*, 660, 81  
 Mendez A. J., Coil A. L., Lotz J., Salim S., Moustakas J., Simard L., 2011, *ApJ*, 736, 110  
 Moles M. et al., 2008, *AJ*, 136, 1325  
 Molino A. et al., 2014, *MNRAS*, 441, 2891  
 Morgan W. W., Mayall N. U., 1957, *PASP*, 69, 291  
 Nair P. B., Abraham R., 2010, *ApJS*, 186, 427 (N&A)  
 Ngan W., van Waerbeke L., Mahdavi A., Heymans C., Hoekstra H., 2009, *MNRAS*, 396, 1211  
 Oke J. B., Gunn J. E., 1983, *ApJ*, 266, 713  
 Pearson K., 1895, *Proc. R. Soc.*, 58, 240  
 Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., Dubourg V., 2011, *J. Mach. Learning Res.*, 12, 2825  
 Peng C. J., Ho L. C., Impey C. D., Rix H.-W., 2002, *AJ*, 124, 266  
 Peng C. J., Ho L. C., Impey C. D., Rix H.-W., 2010, *AJ*, 139, 2097  
 Pentericci L., Grazian A., Scarlata C., Fontana A., Castellano M., Giallongo E., Vanzella E., 2010, *A&A*, 514, 64  
 Petty S. M. et al., 2014, *ApJ*, 148, 111  
 Pierce C. M. et al., 2007, *ApJ*, 660, L19  
 Pović M. et al., 2009a, *ApJ*, 702, L51  
 Pović M. et al., 2009b, *ApJ*, 706, 810  
 Pović M. et al., 2012, *A&A*, 541, 118  
 Pović M. et al., 2013, *MNRAS*, 435, 3444  
 Sánchez-Almeida J., Aguerri J. A. L., Muñoz-Tuñón C., de Vicente A., 2010, *ApJ*, 714, 487  
 Scoville N. et al., 2007, *ApJS*, 172, 1  
 Sekiguchi K. et al., 2004, *Ann. Rep. Natl. Astron. Obs. Japan*, 5, 40  
 Sérsic J. L., 1963, *Bol. Asoc. Argentina Astron.*, 6, 41  
 Simard L. et al., 2002, *ApJS*, 142, 1  
 Simard L., Mendel J. T., Patton D. R., Ellison S. L., McConnachie A. W., 2011, *ApJS*, 196, 11  
 Simpson C. et al., 2012, *MNRAS*, 421, 3060  
 Simpson C., Westoby P., Arumugam V., Ivison R., Hartley W., Almaini O., 2013, *MNRAS*, 433, 2647  
 Stott J. P., Sobral D., Smail I., Bower R., Best P. N., Geach J. E., 2013, *MNRAS*, 430, 1158  
 Strateva I. et al., 2001, *AJ*, 122, 1861  
 Tasca L. A. M., White S. D. M., 2011, *A&A*, 530, 106  
 Tukey J. W., 1977, *Exploratory Data Analysis*. Addison-Wesley, Reading, MA  
 Urrutia T., Lacy M., Becker R. H., 2008, *ApJ*, 674, 80  
 Villforth C. et al., 2014, *MNRAS*, 439, 3342  
 Wang T. et al., 2012, *ApJ*, 752, 134  
 Weiner B. J. et al., 2005, *ApJ*, 620, 595  
 Willett K. W. et al., 2013, *MNRAS*, 435, 2835  
 Yamauchi C. et al., 2005, *AJ*, 130, 1545
- <sup>1</sup>*Instituto de Astrofísica de Andalucía (IAA-CSIC), E-18008 Granada, Spain*  
<sup>2</sup>*Astrophysics Research Institute, Liverpool John Moores University, Liverpool L35RF, UK*  
<sup>3</sup>*Instituto de Astrofísica de Canarias (IAC), E-38200 La Laguna, Tenerife, Spain*  
<sup>4</sup>*GEPI, Paris Observatory, F-75014 Paris, France*  
<sup>5</sup>*Observatório Nacional, 20921-400 Rio de Janeiro, Brazil*  
<sup>6</sup>*Centro de Estudios de Física del Cosmos de Aragón (CEFCA), E-44001 Teruel, Spain*  
<sup>7</sup>*Instituto de Astronomia, Geofísica e Ciências Atmosféricas, Universidade de São Paulo, 05508-090 São Paulo, Brazil*  
<sup>8</sup>*Departamento de Astrofísica, Facultad de Física, Universidad de la Laguna, E-38200 La Laguna, Spain*  
<sup>9</sup>*Department of Theoretical Physics, University of Basque Country, E-48080 Bilbao, Spain*  
<sup>10</sup>*IKERBASQUE, Basque Foundation for Science, E-48011 Bilbao, Spain*  
<sup>11</sup>*Facultad de Física. Departamento de Física Atómica, Molecular y Nuclear, Universidad de Sevilla, E-41012 Sevilla, Spain*  
<sup>12</sup>*Institit de Ciències de l'Espai, IEEC/CSIC, E-08193 Barcelona, Spain*  
<sup>13</sup>*Instituto de Física de Cantabria (CSIC-UC), E-39005 Santander, Spain*  
<sup>14</sup>*Unidad Asociada Observatori Astronòmic (IFCA - UV), E-46980 Valencia, Spain*  
<sup>15</sup>*Departamento de Astronomía, Pontificia Universidad Católica, 782-0436 Santiago de Chile, Chile*  
<sup>16</sup>*Observatori Astronòmic de la Universitat de València, E-46980 Valencia, Spain*

This paper has been typeset from a  $\text{\TeX}/\text{\LaTeX}$  file prepared by the author.