

## Shadows in the Dark: Low-Surface-Brightness Galaxies Discovered in the Dark Energy Survey

D. TANOGLIDIS,<sup>1</sup> AND MANY OTHERS,  
(DES COLLABORATION)

<sup>1</sup> University of Chicago

Submitted to ApJ

### ABSTRACT

We present a catalog of extended low-surface-brightness galaxies (LSBGs) identified in  $\sim 5000 \text{ deg}^2$  from the first three years of imaging data from the Dark Energy Survey (DES). Based on the DES coadd object catalog, we identify 21,420 LSBGs, defined as galaxies with  $r_{1/2} > 2''.5$  and  $\mu_{\text{eff}} > 24.3 \text{ mag arcsec}^{-2}$ . We perform a multi-band fit to the flux and morphology of each of these galaxies using galfitm. We find that the distribution of LSBGs is bimodal in  $(g - r)$  vs  $(g - i)$  color-color space. We find that redder LSBGs are more highly clustered than their blue counterparts, and are correlated with the distribution of nearby ( $z < 0.05$ ) bright galaxies. We find that  $\sim XX\%$  of the red galaxies are located within 0.5 degrees of low-redshift galaxy groups and clusters. For nine of the most prominent galaxy groups and clusters, we calculate the physical properties of associated LSBGs assuming a redshift derived from the host. In these systems, we identify XXX objects that can be classified as ultra-diffuse galaxies based on the criteria  $r_{\text{eff}} > 1.5 \text{ kpc}$  and  $\mu_{\text{eff}} > 24 \text{ mag arcsec}^{-2}$ . These measurements can be used to test the role of environment on models of LSBG formation and evolution.

*Keywords:* Low surface brightness galaxies, galaxies, catalogs — surveys

### 1. INTRODUCTION

The low surface brightness universe has been notoriously difficult to characterize due to the significant role of observational selection effects (Disney 1976; McGaugh et al. 1995). Low-surface-brightness galaxies (LSBGs) are conventionally defined as galaxies with central surface brightnesses fainter than the night sky (Bothun et al. 1997). These galaxies are expected to dominate the galaxy population by number (e.g.), and may account for a significant fraction of the dynamical mass budget ( $\sim 15\%$ ; e.g., Driver 1999; O’Neil et al. 2000; Minchin et al. 2004) in the present-day Universe. These faint galaxies are thought to contribute a minority (a few per cent) of the local luminosity and stellar mass density (e.g., Bernstein et al. 1995; Driver 1999; Hayward et al. 2005; Martin et al. 2019). Due to their difficulty to detect, low-surface brightness galaxies remain chal-

lenging to study as an unbiased population. LSBGs are known to exist across a wide range of physical sizes and environments, ranging from the ultra-faint satellites of the Milky Way (e.g., McConnachie 2012), to members of massive clusters like Virgo (e.g., Sabatini et al. 2005; Mihos et al. 2015, 2017), Perseus (e.g., Wittmann et al. 2017), Coma (e.g., Adami et al. 2006; van Dokkum et al. 2015; Koda et al. 2015) and Fornax (e.g., Muñoz et al. 2015; Venhola et al. 2017). LSBGs exist on the tails of the galaxy surface-brightness distribution, and understanding the formation and demographics of the LSBG population plays an important role in testing models of cosmology and galaxy formations.

The standard model of cosmology ( $\Lambda$ CDM) predicts that galaxies form hierarchically with smaller galaxies forming first and assembling to form larger galaxies, galaxy groups, and galaxy clusters (e.g., Peebles 1980; Davis et al. 1985; White & Frenk 1991). The formation and growth of galaxies over cosmic time is connected to the growth of the dark matter halos in which they reside (the so called “galaxy–halo connection”, e.g., Wechsler

Corresponding author: D. Tanoglidis  
[dtanoglidis@uchicago.edu](mailto:dtanoglidis@uchicago.edu)

& Tinker 2018). Many attempts have been made to use the properties of dark matter halos to predict the properties of the galaxies that inhabit them (e.g., Behroozi et al. 2013; Moster et al. 2013). As the extreme tails of the relationship between galaxy size and luminosity, LSBGs provide a litmus test for models of galaxy formation and cosmology (e.g., Ferrero et al. 2012; Papastergis et al. 2015). It has been suggested that LSBGs may form naturally within the  $\Lambda$ CDM framework, primarily in halos with high angular velocity (Dalcanton et al. 1997; Amorisco & Loeb 2016) or in dense environments (Martin et al. 2019). The role of environment in the formation of On the other hand, observations of low-surface brightness galaxies with anomalously low dark matter content (van Dokkum et al. 2018, 2019) have been suggested to necessitate modifications to the theories of galaxy formation (e.g., ) and/or dark matter physics (e.g., Carleton et al. 2019). Resolving the contributions of various mechanisms of LSBG formation have been historically challenging due to the small volume and highly biased observational data available.

The search for successively fainter populations of galaxies represents an active area of research, constantly plagued by observational systematics (Freeman 1970; Disney 1976; McGaugh et al. 1995) The wide-area, homogeneous, digital imaging of the Sloan Digital Sky Survey (SDSS) enabled statistical studies of large samples of LSBGs down to central surface brightnesses of  $\mu_0(B) \sim 24 \text{ mag arcsec}^{-2}$  (Zhong et al. 2008; Rosenbaum et al. 2009; Galaz et al. 2011). Small robotic telescopes, such as the Dragonfly Telephoto Array (Abraham & van Dokkum 2014), can make unique contributions by optimizing for extremely low levels of scattered light. These surveys have paid dividends in the discovery of populations of LSBGs in nearby groups (Merritt et al. 2016; Danieli et al. 2017) and clusters (van Dokkum et al. 2015). Recently, the Hyper Suprime-Cam Subaru Strategic Program (HSC SSP) revealed a large population of LSBGs with  $\mu_{\text{eff}}(g) > 23.5 \text{ mag arcsec}^{-2}$  in an untargeted search of the first  $\sim 200 \text{ deg}^2$  from the Wide layer of the HSC-SSP (Greco et al. 2018).

Untargeted searches for LSBGs are essential to understand the role that environment plays in their formation and evolution. However, such searches are challenging due to the depth that must be achieved over a wide area in order to provide a statistically significant population of LSBGs. The Dark Energy Survey (DES) covers  $\sim 5,000 \text{ deg}^2$  to a point-source depth of  $g \sim 24.5$  (two magnitudes fainter than SDSS). Here we use data from the first three years of DES to detect LSBGs with flux radii  $> 2''.5$  and central surface brightness of  $\bar{\mu}_{\text{eff}}(g) > 24.3 \text{ mag arcsec}^{-2}$ . Through a combination

of classical cut-based selections, machine learning techniques, and visual inspection, we produce a high-purity catalog of 21,420 LSBGs. We present the spatial, morphological, and photometric properties of this sample. We find significant evidence for bimodality in the color distribution of LSBGs, with redder galaxies ( $g-i > 0.6$ ) preferentially clustered around nearby galaxy groups and clusters. We identify a 96 nearby galaxy groups and clusters that are spatially coincident with  $\gtrsim 5\sigma$  overdensities in the LSBG population density. For these associations, we are able to use the measured distance of the associated object to derive the effective radii and luminosities of associated LSBGs, identifying objects with effective radii  $r_{\text{eff}} > 1.5 \text{ kpc}$  that can be sub-categorized as ultra-diffuse galaxies (e.g., van Dokkum et al. 2015).

This paper is organized as follows. In Section 2 we describe the DES data set used for our search. In Section 3 we describe our multi-step detection and measurement pipeline, resulting in a catalog of 21,420 LSBGs. In Section 4 we test the efficiency of our detection pipeline by comparing the number of LSBGs we detected around the nearby Fornax cluster with those detected in a deeper survey dedicated only to that cluster. In Section 5, we describe the observed properties of this sample, in particular emphasizing differences between the red and blue subsamples of the LSBG population. In Section 6 we present clustering statistics of the LSBG sample, showing in particular that the red subsample of LSBGs is highly clustered. In Section 7, we focus on LSBGs surrounding nearby galaxy groups and clusters, where it is possible to extract some physical properties of LSBGs through their statistical association to these well-studied systems. We summarize and conclude in Section 8.

## 2. DES DATA

DES is an optical–near infrared imaging survey of  $\sim 5000 \text{ deg}^2$  in the south Galactic cap using the Dark Energy Camera (DECam; Flaugher et al. 2015) on the 4-m Blanco Telescope at Cerro Tololo Inter-American Observatory (CTIO). The DECam focal plane comprises 62  $2\text{k}\times 4\text{k}$  CCDs dedicated to science imaging and 12  $2\text{k}\times 2\text{k}$  CCDs for guiding, focus, and alignment. The DECam field-of-view covers  $3 \text{ deg}^2$  with a central pixel scale of  $0''.263$ .

In this analysis, we use data collected from the first three years of DES observing (DES Y3). This data set uses the same image-level processing as the first DES data release (DR1; DES Collaboration et al. 2018). Catalogs were created with SExtractor (Bertin 2006) using the  $r+i+z$  coadded detection image. Photometric measurements were made in each band using “dual image” mode using the band of interest in combination with

the detection image. Background subtraction was performed in two phases: first we perform a background subtraction on the full focal plane using a PCA algorithm (Bernstein et al. 2018), **SExtractor** performs a local background modeling and subtraction on a scale of **XX** pixels (?).

### 3. LSBG CATALOG

Here we describe the pipeline used to identify and measure LSBGs in the DES Y3 data. Briefly, we start with a generic catalog of **SExtractor** detections and use the morphological and photometric properties to identify a subset of LSBG candidates. We train a machine learning algorithm to remove artifacts, and visually inspect the resulting candidate list to assemble a high-purity catalog of LSBGs. We then use **galfit** to fit a Sérsic profile to each identified LSBG in order to determine photometric properties in a manner that is consistent with previous work (e.g. Greco et al. 2018). The full catalog of DES LSBGs is available as supplementary material.

#### 3.1. Initial sample selection

We began with the **DES Y3 GOLD coadd object catalog** (v2.2) assembled from **SExtractor** detections. We first removed objects classified as point-like based on the multi-object fitting (MOF) size parameter and the *i*-band **SExtractor spread\_model** parameter (see , for more details). Following Greco et al. (2018), we defined our initial sample of candidate LSBGs based on angular size and surface brightness. We required that sources have flux radii in the range  $2.5'' < r_{1/2} < 20''$ <sup>1</sup> and mean surface brightness in *g*-band:

$$24.3 < \bar{\mu}_{\text{eff}}(g) < 28.8 \text{ mag arcsec}^{-2} \quad (1)$$

We also restricted our selection to objects with colors (based on **SExtractor**'s **MAG\_AUTO** magnitudes) in the range:

$$-0.1 < g - i < 1.4 \quad (2)$$

$$(g - r) > 0.7 \times (g - i) - 0.4 \quad (3)$$

$$(g - r) < 0.7 \times (g - i) + 0.4. \quad (4)$$

The complete selection criteria are presented in Appendix A. After performing the cuts described above, our sample consisted of 413,608 objects from an initial catalog of  $\sim 4 \times 10^8$  objects.

<sup>1</sup> After assembling our catalog, we got worried we inspected all  $\sim 1500$  objects with  $r_{1/2} > 20''$  and found only a few potentially interesting cases. Thus we decided to keep our initial selection cut.

### 3.2. Machine Learning Classification

The cuts described in the previous section resulted in a sample that was found to be highly contaminated by spurious detections. Initial visual inspection of a subset of a few thousand candidates revealed that  $\lesssim 8\%$  of the objects passing these selections were LSBGs. The most common sources of contamination were:

1. Faint, compact objects blended in the diffuse light from nearby bright stars or giant elliptical galaxies.
2. Bright regions of Galactic cirrus.
3. Knots and star-forming regions in the arms of large spiral galaxies.
4. Tidal ejecta that are clearly connected to a high-surface-brightness host galaxy.

The large size and low purity of our initial candidate list was well-suited to the application of conventional machine learning (ML) algorithms. Our goal with ML classification was to reject a large fraction of false positives while retaining high completeness for true LSBGs.

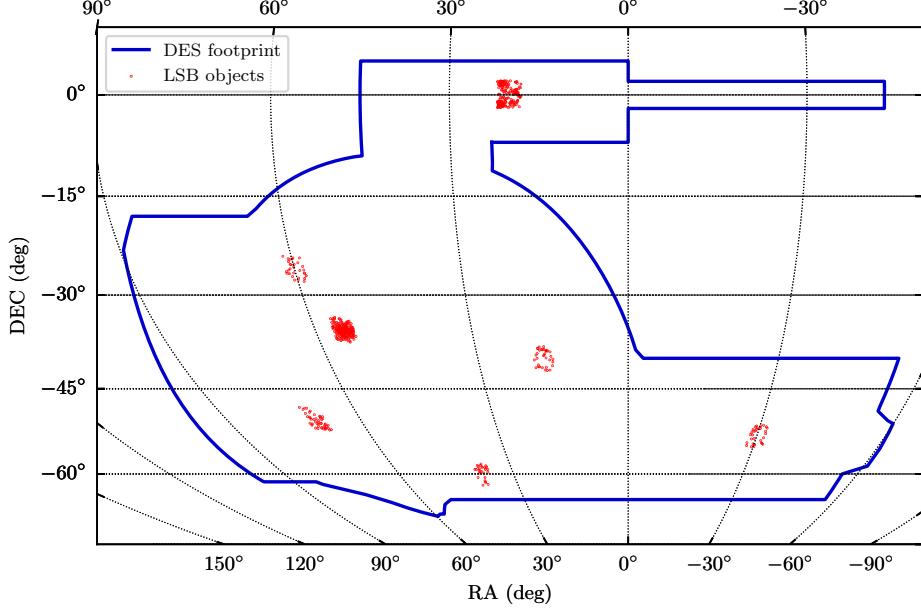
#### 3.2.1. Training Set

In order to train a supervised ML classification algorithm, we required a sample of objects where the true classification was known. To avoid biases when training the classifier, we seek to make this labeled training sample representative of full LSBG candidate sample.

We created a labeled sample by visually inspecting all objects that pass the cuts defined in section Section 5 in seven  $4^\circ \times 4^\circ$  patches spread over the DES footprint (Figure 1). One of these regions was centered on the Fornax galaxy cluster, which is known to contain a high concentration of LSBGs (e.g., Muñoz et al. 2015), while the other regions were selected at random. Our training set consists of 7760 visually inspected objects, of which 640 were classified as LSBGs.

#### 3.2.2. Features and Classifiers

We split the labeled set into two sets: a training set, consisted of the (randomly selected) 75% of the total population of labeled objects and a development set consisted of the remaining 25%. We used the development set to evaluate the performance of different classifiers and tune their hyperparameters. We did not create an independent (of the development set) testing set for two reasons: first, because we did not want to spare more of our limited training data. Second, because this was not absolutely necessary, since our plan is to visually inspect the positively classified (as LSB galaxies) objects, and reject the remaining false positives.



**Figure 1.** The distribution of the objects visually classified as LSBGs in the seven  $4^\circ \times 4^\circ$  regions used to create the labeled set for classification and validation.

In the classification we used 18 features for each object that are either **SExtractor** parameters, or combinations of them. Specifically, we use:

- The adaptive aperture magnitudes in  $g, r, i$  bands, **MAG\_AUTO**.
- The colors  $g - r$ ,  $g - i$ ,  $i - r$  derived from the aperture magnitudes.
- The angular size of the adaptive isophotal aperture (effective half-light radii) in the  $g, r, i$  bands, **FLUX\_RADIUS**.
- The effective surface brightness in the  $g, r, i$  bands, **MU\_EFF\_MODEL**.
- The maximum surface brightness measured by **SExtractor** in the  $g, r, i$  bands, **MU\_MAX**.
- The semi-major and semi-minor axes of the isophotal ellipse containing half the light, **A\_IMAGE** and **B\_IMAGE**.
- The isophotal ellipticity, defined as the ratio **ELL = 1 - B\_IMAGE/A\_IMAGE**.

We tested a number of popular classification algorithms, as implemented in the **scikit-learn** python library (Pedregosa et al. 2011).<sup>2</sup> Specifically, we tested the Naive Bayes, AdaBoost, Nearest Neighbor, Random

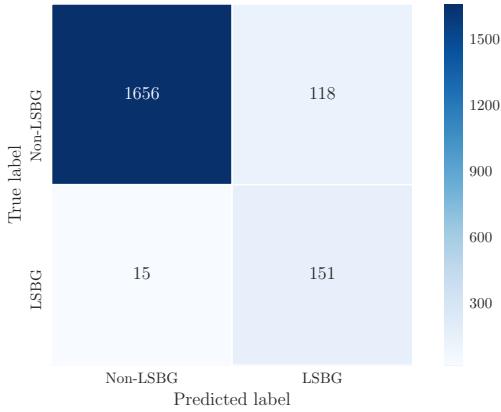
Forrest, Linear Support Vector Machines (SVM) and SVM with radial basis function (RBF) kernel classifiers.

Our goal was to find a classifier that minimized the false negative rate (FNR)—true LSBGs classified as false detections—while keeping the true positive rate (TPR) reasonably high. In other words, we favored completeness versus purity of the sample classified as LSBGs. This choice was motivated by our goal to reduce the candidate sample to a tractable size, that would be possible to visually inspect (thus reject remaining false positives), without losing many real LSBGs in the process.

Note that the samples in our training data are heavily imbalanced: from the 5820 objects ( $7760 \times 0.75$ ) only 480 ( $640 \times 0.75$ ) are true LSBGs. Class imbalance can lead to low accuracy in predicting the label of objects belonging to the less frequent class. We deal with this by weighting the classes using the **class\_weight** parameter, to make sure that we have a balanced mix of each class. Setting the above parameter equal to **balanced**, each class is weighted with a weight inversely proportional to its frequency,  $w_j = n/2n_j$ , where  $w_j$  the weight of the  $j$ -th class and  $n, n_j$  the total number of observations and observations of the  $j$ -th class, respectively.

We found that that the optimal classifier for our specified goal was a SVM classifier with a RBF kernel and parameters  $C = 10^4$  and  $\gamma = 0.012$ . In Figure 2 we

<sup>2</sup> <https://scikit-learn.org/stable/index.html>



**Figure 2.** The confusion matrix of our final SVM classifier evaluated on the development set. The false negative rate is  $\sim 9\%$ .

present the confusion matrix for this classifier, evaluated on the development set. We see that the false negative rate (FN), defined as the percentage of true LSBGs classified as non-LSBGs ( $FNR = FN/(FN + TP)$ ) is  $\sim 9\%$ . Furthermore, from the results of this matrix we expect that  $\sim 44\%$  of the objects classified as LSBGs are false positives. From subsequent visual inspection (Section 3.3) we find that the number of false positives is consistent with our expectations from the results presented here.

Using the optimized classifier, as described in the above section, we classified the 413,608 LSBG candidates that resulted after performing the cuts defined in Section 5. The classification returned 40,820 objects classified as LSBGs, thus reducing the sample by about an order of magnitude.

### 3.3. Visual Inspection

The final step in the generation of our LSB galaxy sample is to visually inspect the positively classified objects from the previous step. We split the candidates into groups of 500 each time. For each of these batches we generate cutouts  $30'' \times 30''$ , centered at the coordinates of each of the candidates, from the DECam Legacy Survey website<sup>3</sup>.

We show an example of 20 of these cutouts in Figure 3. To give you a sense of the procedure we used, in this sample of galaxies we immediately accept as true LSB galaxies the candidates with number 2, 3, 8, 12, 13, 14, 15, 18. As we can see some of them are ellipticals and some spirals. We see that candidates 10, 11 represent the same object. In a case like that (that was

frequent) we pick them most “centered” one, in the case presented here this is candidate 11. Candidates 0, 1, 5, 6, 7, 9, 16, 17, 19 are rejected as false positives. For some candidates (i.e., number 4), it is not immediately clear whether it is a LSBG or not. In these cases, we use the DES Sky Viewer<sup>4</sup> to inspect the region surrounding the candidate. The DES Sky Viewer allows more flexible zooming and scaling, and we ended up rejecting candidates like candidate number 4.

After visual inspection, our final sample contains 21,420 objects. Note that this sample may still contain a small fraction of false positives. The most common cases may include:

- Ejecta from large galaxies without a visible connection to them within the small angular size of the cutouts.
- Duplicates, like candidates 10, 11 in Figure 3 that were not projected at a close distance to each other, and thus their similarity was not identified. (ADW: This is a problem. These need to be identified and removed.)
- Small background galaxies in the halos of bright stars.

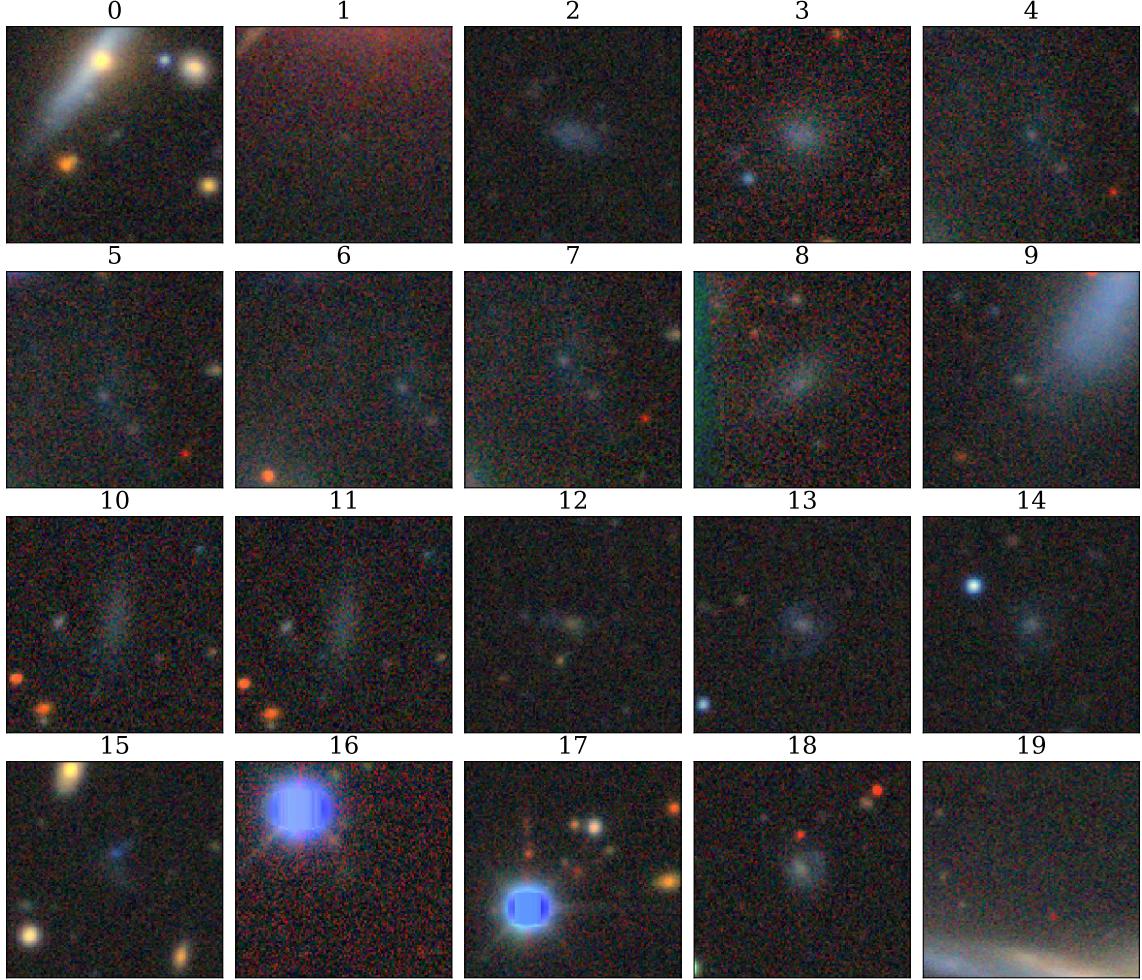
### 3.4. Sérsic Model Fitting

To compare the properties of our LSBG catalog against similar catalogs in the literature (e.g., Greco et al. 2018), we fit each galaxy with a Sérsic light profile. We use `galfitm`, a multi-band implementation of `galfit` developed in the context of the `MegaMorph` project (Peng et al. 2002; Barden et al. 2012; Häufbier et al. 2013), to perform a multi-band fit for each galaxy using the DES coadd images from the  $g, r, i$ -band. We started by creating square cutout images centered on each galaxy. The cutout size was set to be  $10 \times \text{FLUX\_RADIUS}$  of each galaxy (rounded up to the nearest 50 pixel step). A minimum cutout size of  $201 \times 201$  pix ( $\sim 50''$  on a side) was used for small galaxies. We assembled a mask in each band by combining the segmentation map from the DES detection coadd (a combination of the  $r, i, z$  images) with the bad pixel mask from each individual band. The `galfitm` “sigma image” was derived from the inverse variance weights plane associated produced by `SCAMP` for each of the DES coadded images.

Large LSBGs are often segmented into several catalog objects by `SExtractor`. Since we are using the segmentation map as a mask, regions of the image associated with other `SExtractor` sources are excluded from the `galfitm` analysis. However, we would like to use as much of the light of the galaxy as possible to provide an

<sup>3</sup> <http://legacysurvey.org/>

<sup>4</sup> <https://desportal2.cosmology.illinois.edu/dri/apps/sky/sky/23>



**Figure 3.**  $30'' \times 30''$  cutouts of twenty candidates, positively classified in the Machine Learning step (Section 3.2).

accurate fit. Thus, for each LSBG with `FLUX_RADIUS`  $> 15$  pix, we find the set of all `SExtractor` objects that share a parent with the LSBG. These “siblings” of the LSBG consist of foreground star, background galaxies, and various stellar overdensities associated with the LSBG itself (e.g., globular clusters, star forming regions, etc.). Since we are primarily interested in fitting the smooth light profile, we unmask regions associated with nearby siblings, except those that are compact, have high surface brightness, or are located at farther than three times the flux radius.

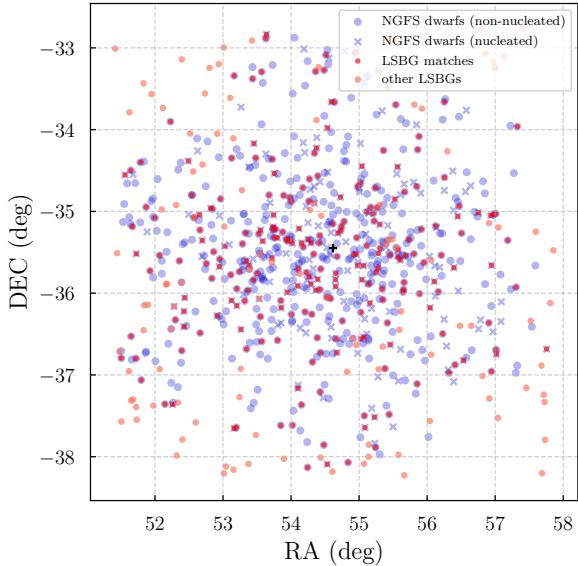
Sérsic model parameters were seeded based on the values of the `SExtractor` catalog. The centroid was initialized at the position derived by `SExtractor`, and was constrained within 10% of the of the `FLUX_RADIUS`. The Sérsic  $r_{\text{eff}}$  was similarly initialized based on the `FLUX_RADIUS` and was allowed to vary by a factor of 5 from this initial value. The Sérsic index was initialized at a value of  $n = 1.0$  and was constrained to lie within

the range  $0.2 < n < 5.0$ . When performing the fit with `galfitm`, we tied the morphological parameters (centroid position, effective radius, Sérsic index, ellipticity, and position angle) across the three bands. In contrast, the flux normalization of the model was allowed to vary independently in each band according to a quadratic function of wavelength. We visually inspect the residuals of each fit to identify and correct any catastrophic errors. The resulting best-fit values for each fit are provided in Table ??.

While the Sérsic model fit provides consistent properties across all objects in our sample and allows comparison to similar catalogs in the literature, it is not a sufficiently complex model to provide a good fit of all LSBGs. In particular, we note that a subset of our objects would be fit better through the inclusion of a nuclear point source, while others show clear indications of irregular or spiral structure. We provide a local estimate of the reduced  $\chi^2$  of our model calculated with

$r_{\text{eff}}$  of each galaxies in each band. This information can be used to identify objects that were poorly fit by the simple Sérsic model, and can be followed up with more detailed modeling.

#### 4. DETECTION EFFICIENCY AROUND FORNAX



**Figure 4.** The dwarf galaxies present in the NGFS catalog (in blue) and the matches from our LSBG catalog (red). The NGFS catalog is separated into nucleated (denoted by an x) and non-nucleated (circles) galaxies. We also plot LSBGs that we detected in that region but were not matched to any object in the NGFS catalog (light red). The black cross denotes the nominal center of the Fornax cluster.

To quantitatively estimate the efficiency of our multi-step procedure, we compare our LSBG catalog to a catalog assembled from deeper DECam data as part of the *Next Generation Fornax Survey* (NGFS Muñoz et al. 2015). The Fornax cluster (Abell S373) lies at a distance of  $\sim 20$  Mpc (Blakeslee et al. 2009) and is known to host a large population of faint galaxies (??). The NGFS covers a region of  $\sim 3^\circ$  radius around the Fornax cluster, extending to approximately half the virial radius of the cluster (Ordenes-Briceo et al. 2018). The NGFS has reported a total dwarf galaxy population of 643 galaxies, which is split into nucleated (181) and non-nucleated (462) galaxies (Eigenthaler et al. 2018; Ordenes-Briceo et al. 2018).

The NGFS catalog was assembled through visual inspection of the region surrounding the Fornax cluster. The NGFS process was specifically focused on identifying dwarf galaxies/LSBGs and it did not apply any cuts similar to those that we imposed on the photometric DES catalog. This makes the NGFS an interesting

independent data set to test how our procedure and cuts affect the number of LSBGs retrieved.

In this section we report the fraction of objects from the NGFS catalog that are also present in the DES Gold catalog before any cuts, and how this number changes as we apply the various selection criteria defined in the previous section. This allows us to estimate the completeness of our final LSBG sample relative to the NGFS.

After performing each cut, we match the resulting catalog with the NGFS catalog using a matching radius of  $3''$  (we find that using a larger matching radius does not increase the number of matches but not significantly). Given the different nature of nucleated and non-nucleated galaxies in the NGFS sample (non-nucleated galaxies were found to be fainter and smaller than their nucleated counterparts), we also match to these two sub-samples separately, since we expect the detection efficiency to differ between the two. The results are summarized in Table 1.

We see that, before any cuts, the DES **SExtractor** detection efficiency (number of NGBS galaxies that are also present in the DES catalog) is  $\sim 77\%$ . As expected, the recovery fraction is higher for the nucleated LSBGs where the DES detection efficiency reaches  $\sim 90\%$ .

Our surface brightness cut significantly reduces the number of detected objects; we also see that it affects the nucleated galaxies more strongly. This is reasonable, since they generally have higher central surface brightnesses. The radius cut is more significant in reducing the detection efficiency in the non-nucleated galaxy population. We expect the angular radius cut,  $r_{1/2} > 2''.5$  to be even more significant in reducing the number of LSBG detections around more distant galaxy clusters. Fornax is at a significantly smaller distance compared to other such systems (see also Sec. 7, Table 1); thus even galaxies with small physical sizes will have angular radii large enough to pass the cut. For more distant clusters, our selection limits our sample to only the largest of satellite LSBGs.

Applying all cuts the detection efficiency drops to 42.4% overall, with a detection efficiency of 48.6% and 39.9% for the nucleated and non-nucleated sub-samples, respectively. We further examine the decrease in efficiency from applying our machine learning classification and visual inspection. We find that the drop in efficiency (difference between the last two rows of Table 1) corresponds to an absolute drop of  $\sim 13\%$  in the number of LSBGs in the field that were not detected. That number is consistent with our expectation that the machine learning classification has FNR  $\sim 10\%$  (Figure 2). Furthermore, a visual inspection of misclassified galaxies showed that most of them were either ex-

**Table 1.** Summary of the detection efficiency of the NGFS dwarfs

Cuts applied	All galaxies	Nucleated	Non-nucleated
No cuts	76.6%	89.5%	71.6%
Surface brightness cut only	60.6%	54.1%	63.1%
Angular size cut only	56.1%	80.6%	46.4%
Both cuts	42.4%	48.6%	39.9%
Final result (After ML/Vis. inspection)	36.8%	43.6%	34.1%

NOTE— In this table we present the detection efficiency of the NGFS objects, denoted as the percentage of these objects presented in the DES catalog after performing different cuts. We present the detection efficiency for nucleated and non-nucleated sub-samples separately.

tremely faint/hard to distinguish from random background brightness fluctuations, or very compact and thus were not included in the final catalog. Figure 4 shows a scatter plot of the NGFS dwarfs, matched LSBGs from our catalog, and unmatched LSGBs in the region around the Fornax cluster. Some of them ( $\sim 5$ ) are close to an NGFS object and could have been matched with a slightly larger matching radius. This figure also shows  $\sim 40$  LSBGs that appear in our catalog but were not included in the NGFS catalog.

## 5. LSBG PROPERTIES

The large sky area of DES ( $\sim 5000 \text{ deg}^2$ ) gives us a unique opportunity to study the statistical properties of the LSBG population. Our search resulted into a sample of **21,420** LSBGs with isophotal half-light radius  $> 2.5''$  and  $\bar{\mu}_{\text{eff}}(g) > 24.3 \text{ mag arcsec}^{-2}$ . To our knowledge, this is the largest such catalog of LSBGs to date. In this section, we divide our catalog of LSBGs into red and blue subsamples and compare the properties of these samples to each other and to previous results (i.e., Greco et al. 2018).

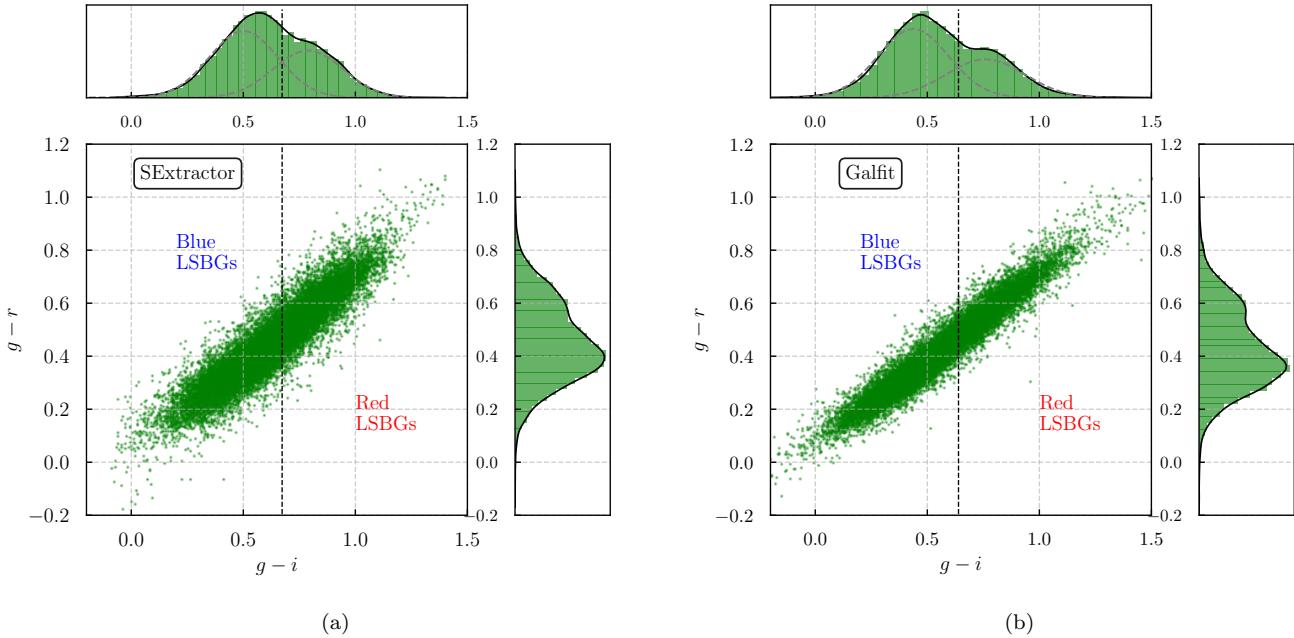
The optical colors of galaxies are indicative of their stellar populations. Colors are known to correlate strongly with galaxy morphology and environment. In particular, galaxies have been conventionally divided into two well-known sequences of red and blue galaxies (e.g., Strateva et al. 2001; Blanton & Moustakas 2009). Less is known about how the colors of LSBGs correlate with morphology, star formation history, and environment. For example, O’Neil et al. (1997) found that classical disk LSBGs span a range of blue and red colors. Similar to HSBGs, blue colors are generally associated with actively star forming spiral or irregular systems, while red colors tend to be indicative of spheroidal or elliptical morphology (Larson et al. 1980; Strateva et al. 2001; Baldry et al. 2004; Lintott et al. 2011, e.g.). It is expected that red galaxies will be found preferentially

in denser environments, where quenching from massive hosts will prevent the recent formation of stars (Bamford et al. 2009; Geha et al. 2017; Román & Trujillo 2017). (Greco et al. 2018) found that LSBGs detected in HSC showed a clear bimodality in color, with two apparently distinct populations separated at  $g - i = 0.64$ . They found that blue LSBGs had brighter mean surface brightness, while galaxies that are large ( $r_{\text{eff}} > 6''$ ) and faint ( $\bar{\mu}_{\text{eff}}(g) > 26 \text{ mag arcsec}^{-2}$ ) are almost exclusively red.

In Figure 5, we present the distribution of our LSBG sample in the  $g - i$  vs.  $g - r$  color space. We show the color-color diagrams derived from the **SExtractor** **MAG\_AUTO** quantities (left panel), and the magnitudes derived from the **galfit** Sérsic model fit (right panel). The color distributions are similar and present signs of bimodality that are slightly more prominent using colors from the Sérsic model fit. A similar bimodality in  $g - i$  and  $g - r$  colors was observed in Greco et al. using data from HSC.

Having established the similarity of the color distributions in the two cases, in the remainder of this paper we quote photometric (magnitudes, colors, surface brightness) and structural (Sérsic index, effective radius) parameters derived from the **galfit** model, in order to consistently compare existing literature results.

We separate the total LSBG sample into red and blue subsamples, according their  $g - i$  color. To do so, we use the following procedure: We fit a two-component Gaussian Mixture Model (GMM) to the 1-D  $g - i$  color distribution. The components can be seen in the top panels of Figure 5 (dashed gray lines). We find that the two Gaussians intersect at  $g - i = 0.639$  (**galfit** case; for comparison using the distribution coming from the **SExtractor** quantities the same point is at  $g - i = 0.673$ ). We then define as red those galaxies that lie right to that point,  $g - i \geq 0.639$ , (**7366** galaxies) and



**Figure 5.** Color-color diagram of our LSBG galaxy sample, using (a) the **SExtractor**’s `MAG_AUTO` parameters and (b) magnitudes returned from our **Galfit** fitting. In both cases bimodality is present in the  $g - i$  and  $g - r$  color distributions, which is more prominent in case (b). We separate the total sample into red and blue galaxies, based on their  $g - i$  color value: we fit a Gaussian Mixture Model with two Gaussians in the  $g - i$  distribution (gray dashed lines in the top panels) and we find the intersection point of them. This is at  $g - i = 0.673$  and  $0.639$  for the **SExtractor** and **galfit** cases, respectively (black vertical dashed lines). We define as red the galaxies that lie right to that value, while those left of it are defined as being blue.

as blue those left of that point,  $g - i < 0.639$  (14057 galaxies).

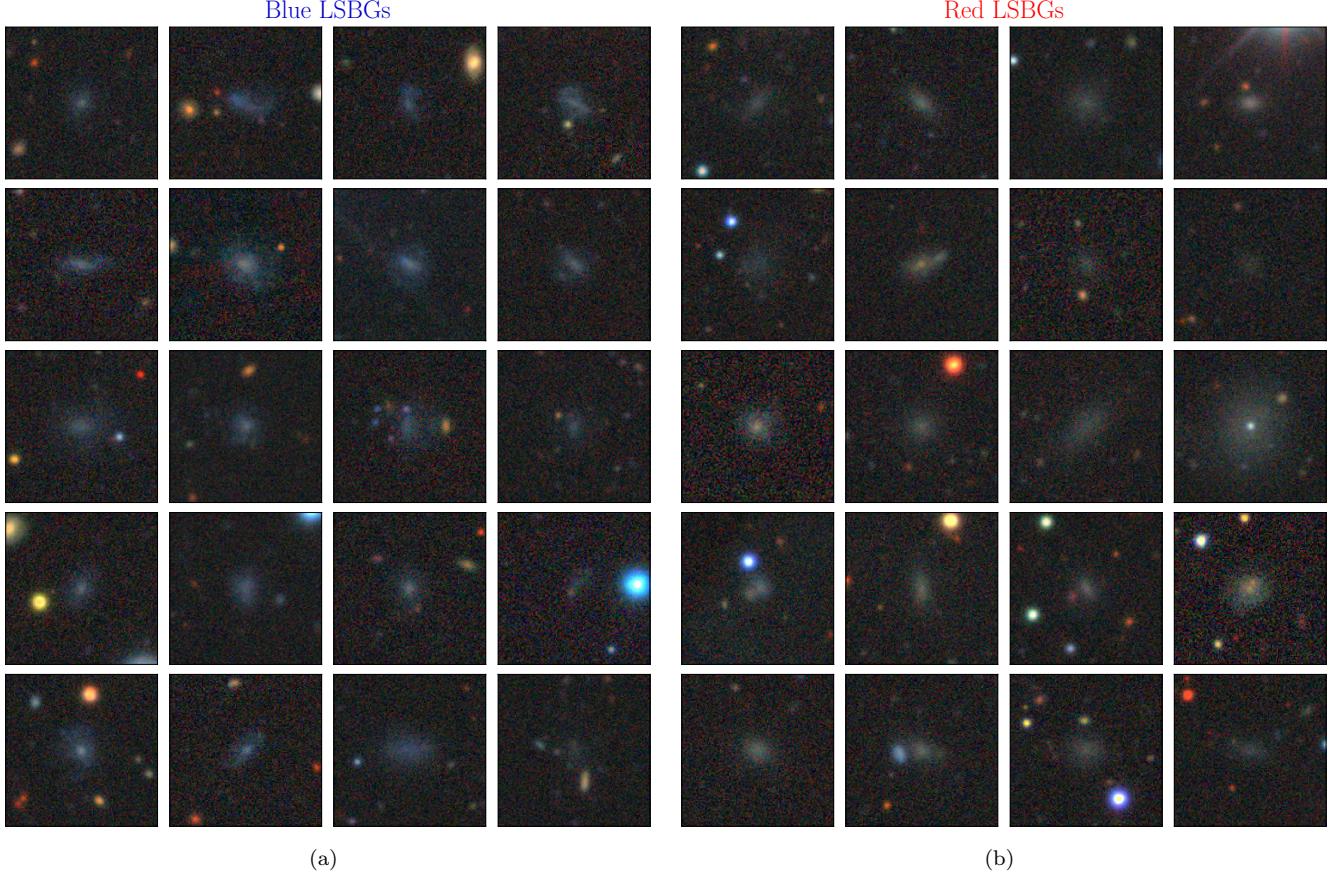
Our  $g - i$  separation threshold is in good agreement with the one used in Greco et al. (median HSC  $g - i = 0.64$ ). Note that in that work the authors used the median of the distribution to separate the two populations. However, there, the two modes of the distribution had similar amplitude and thus the median is  $\sim$  equivalent the way we define our threshold. If we transform from the DES bandpasses to the HSC bandpass, we find that this color corresponds to a DES  $g - i = XX$ . (ADW: The HSC bandpasses are different from DES. Conversions can be found... ) here. The median  $g - i$  of our red and blue LSBG subsamples are 0.79 and 0.44, respectively, again in good agreement with the results of Greco et al. (0.8 and 0.47).

In Figure 6 we show examples of randomly selected blue galaxies with  $g - i < 0.44$  (below the median of the blue population) and red galaxies with  $g - i > 0.79$  (above the median of the red population). As we can see, the two sub-samples show morphological differences as well. The bluer sample is composed of irregular galaxies and galaxies with signs of spiral structure. The redder

sample consists of spherical and elliptical galaxies, with some of them being nucleated.

In the left-hand side panel (a) of Figure 7 we present the joint distribution of our red and blue LSB galaxy samples in the space of effective radius,  $r_{\text{eff}}(g)$ , and mean surface brightness (within the effective radius),  $\bar{\mu}_{\text{eff}}(g)$ . Both populations populations present a similar size distribution. Their value lies in the range  $2''.5 - 16''$ . Despite the wide angular range, most of them (90%) have radius less than  $6''$ , with a median of  $\sim 4''$ . Moving to the mean surface brightness distribution, we see that the red galaxies present a fatter tail towards the faint end (larger values of  $\bar{\mu}_{\text{eff}}(g)$ ). In terms of the 50th, 80th and 90th percentiles, for the red sample these are  $\bar{\mu}_{\text{eff}}(g) = 25.1, 25.5, 25.8 \text{ mag arcsec}^{-2}$ , while for the blue sample the same percentiles are  $\bar{\mu}_{\text{eff}}(g) = 25.5, 26.1, 26.4 \text{ mag arcsec}^{-2}$ ; blue galaxies have generally higher mean surface brightnesses.

In the right-hand side panel (b) of Figure 7 we plot the Sérsic index,  $n$ , versus the central surface brightness,  $\mu_0(g)$ , for the red and blue LSBG samples, as before. The distribution in the Sérsic index is similar for two samples (although red LSBGs seem to avoid very low values), with a range  $n \sim 0.2 - 4$  and median  $n \sim 1.0$ . In



**Figure 6.** Examples of (a) blue and (b) red LSB galaxies in our sample. We randomly selected red galaxies with  $g - i$  above the median for the red population (that is,  $g - i > 0.79$ ) and blue galaxies below the median of the blue population ( $g - i < 0.44$ ) to make the color difference more prominent. Each of the cutouts is  $30'' \times 30''$  in size.

the central surface brightness, we find that, again, blue galaxies tend to have higher surface brightness, however the difference between the two distributions is not so striking as before: The median of the red population is at  $\mu_0(g) = 24.2$  mag arcsec $^{-2}$ , while of the blue population at  $\mu_0(g) = 23.9$  mag arcsec $^{-2}$ .

## 6. CLUSTERING OF LSBGS

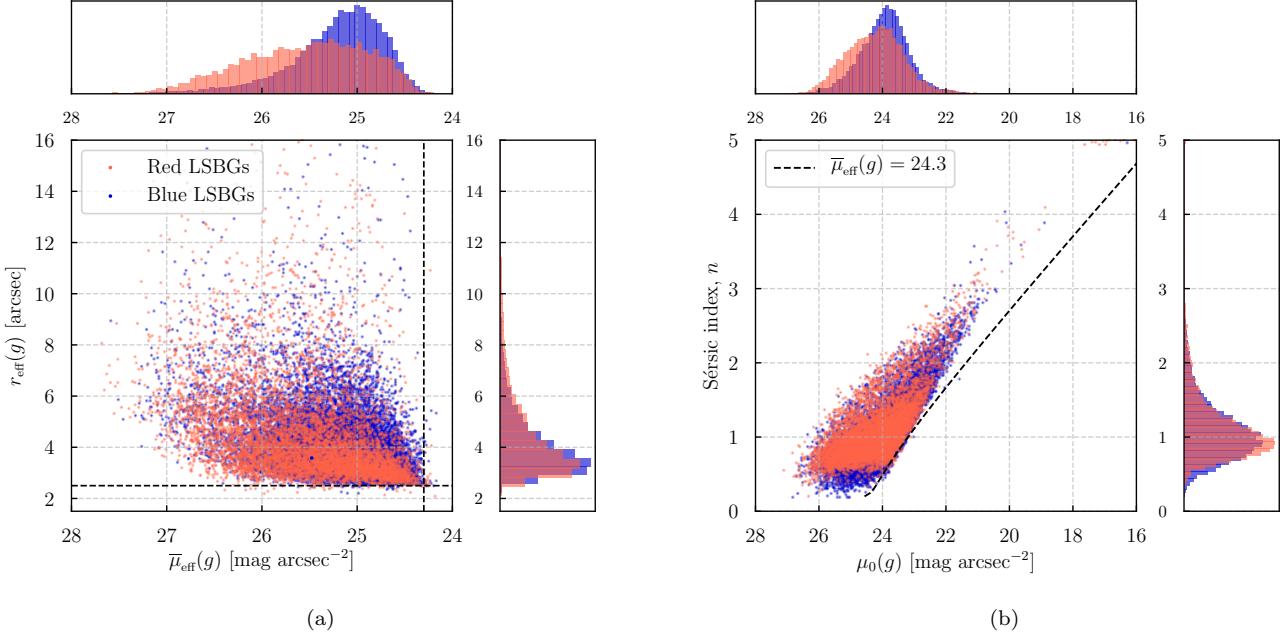
We expect that our selection on large size ( $r_{1/2} > 2.5''$ ) likely restricts our LSBG sample to low redshift. Greco et al. tentatively suggested that the spatial distribution of LSBGs in HSC SSP may be correlated with other low-redshift galaxies from the NASA-Sloan Atlas (?). However, due to the relatively small and non-contiguous coverage of their HSC SSP data set ( $\sim 200$  deg $^2$ ), they were unable to make any firm statistical statement about possible correlations. Our DES Y3 catalog covers a contiguous region  $\sim 25$  times larger than that of Greco et al., allowing us to perform a detailed exploration of the spatial distribution of LSBG populations. In particular, we are able to *separately* explore the clustering of red and blue LSBGs. We find that

a rather stark contrast exists between these two LSBG subpopulations: red LSBGs are highly clustered, while blue galaxies seems to be more uniformly distributed.

In Figure 8 we present the spatial distribution of our LSBG sample over the DES footprint. As before, we color-code the points to indicate red and blue galaxies based on the  $g - i$  color split described in the previous section. It is obvious that the spatial distribution of the two populations differs significantly, with the red LSBGs showing a higher level of clustering compared to the blue LSBGs, which appear to have a more homogeneous distribution on the sky. To see the difference more clearly, we plot the sky positions of the blue and red galaxies separate plots in Figure 9.

To quantify the level of clustering of our LSBG sample, we use `treecorr` (Jarvis 2015)<sup>5</sup> to calculate the angular two-point auto-correlation function for LSBGs. In Figure 10 we compare the two-point function of the full

<sup>5</sup> <https://github.com/rmjarvis/TreeCorr>



**Figure 7.** (a) Joint distribution of the red and blue LSBGs in the space of effective radius,  $r_{\text{eff}}$ , and mean surface brightness (within the effective radius),  $\bar{\mu}_{\text{eff}}$ , both in the  $g$  band. The two populations are defined according to the  $g - i$  color criterion described in the main text of Section 5 and in Figure 5. The dashed horizontal and vertical lines correspond to the limits of the selection criteria  $r_{1/2} > 2.5''$  and  $\bar{\mu}_{\text{eff}}(g) > 24.3 \text{ mag arcsec}^{-2}$ , respectively (However, notice that the selection was based on the **SExtractor** parameters, while here we plot those from our **galfit** fitting). (b) Sérsic index,  $n$ , versus the central surface brightness,  $\mu_0(g)$ , for the galaxies in our sample, again split into red and blue sub-samples, as before. The black dashed line corresponds to  $\bar{\mu}_{\text{eff}}(g) = 24.3 \text{ mag arcsec}^{-2}$ , our selection criterion.

LSBG sample (gray line) and the red and blue subsamples. As expected from the previous discussion, we find that the auto-correlation function of the red sample has a significantly (one to two orders of magnitude) higher amplitude than the blue sample (and the full sample) across all angular scales.

Next we compare the clustering properties of our LSBG catalog and two other catalogs: A catalog of high-surface brightness galaxies (HSBGs) extracted from the DES Y3 data and an external low-redshift galaxy catalog.

We construct the HSBG catalog by applying to the DES Y3 catalog the same selection criteria as those used to generate the LSBG sample, except, of course, those in the mean  $g$ -band surface brightness. Here we select the HSBG sample by applying the cut:

$$20.0 < \bar{\mu}_{\text{eff}}(g) < 22.0 \text{ mag arcsec}^{-2}. \quad (5)$$

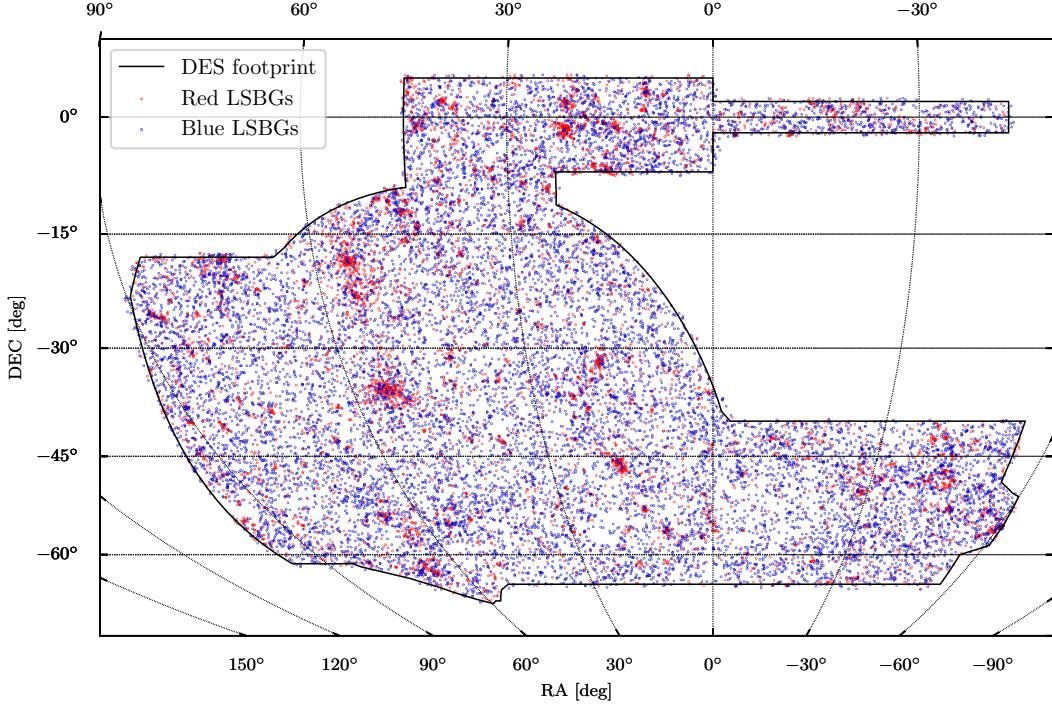
Motivated by the fact that our LSBGs most probably lie in low-redshifts, we performed a series of redshift cuts on the HSBGs with the intention to find the cut that gives a sample with an auto-correlation that is as close as possible to that of the LSBGs. For each cut

we select all HSBGs below that threshold, and then a random sub-sample selected in such a way that it has the same distribution in the  $g$ -band magnitude as the LSBG sample. Our best results come for a photometric redshift cut  $z < 0.07$ ; we plot the auto-correlation function of that sample in Figure 11 (blue line).

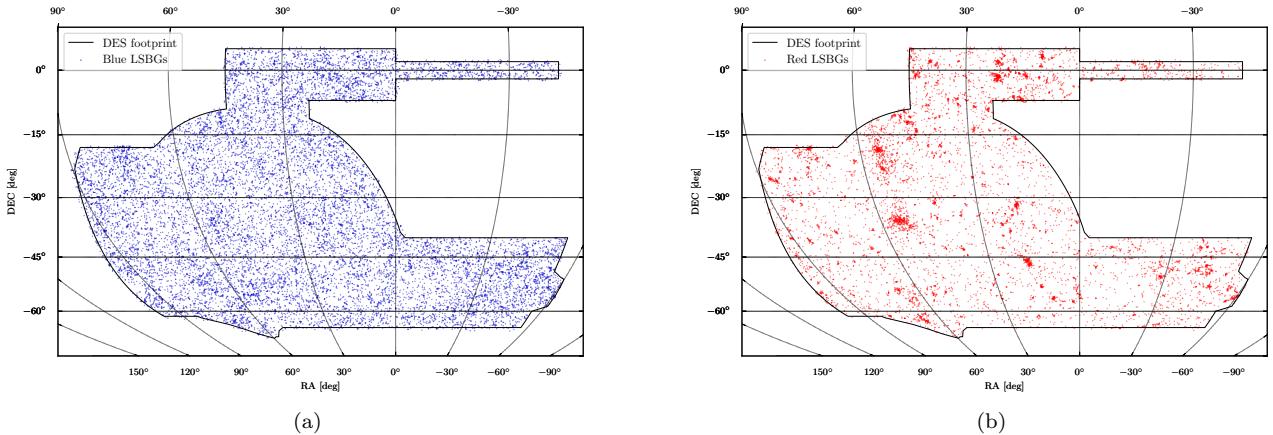
We also utilize the 2MASS Photometric Redshift (2MPZ) catalog (2MPZ Bilicki et al. 2014), an optical-IR all-sky photometric redshift catalog based on SuperCOSMOS, 2MASS, and WISE extending to  $z \sim 0.3$  (peaking at  $z \sim 0.07$ ) with a photometric redshift accuracy of  $\sigma_z = 0.015$  and very small percentage of outliers. We note that the 2MPZ galaxies have a very different selection function due to the required detections in the IR bands. By matching the 2MPZ catalog with galaxies from the full DES Gold Catalog (before any cuts) we were able to retrieve information about its magnitude and surface brightness distributions. We find that the mean surface brightness lies in the range:

$$19.0 < \bar{\mu}_{\text{eff}}(g) < 23.0 \text{ mag arcsec}^{-2}, \quad (6)$$

thus it is a sample with much higher mean surface brightness than the LSBG sample. The  $g$ -band magni-



**Figure 8.** Sky positions of the blue ( $g - i < 0.639$ ) and red ( $g - i \geq 0.639$ ) LSBGs over the DES footprint. It is clear that the distribution of the blue galaxies is relatively homogeneous, while that of the red galaxies is strongly clustered.



**Figure 9.** Similar to Figure 8, but here we have plotted separately the distribution of (a) the blue galaxies and (b) the red galaxies, to better see the difference in the level of the spatial clustering properties of the two populations.

tude (MAG\_AUTO) of the 2MPZ sample lies in the range:

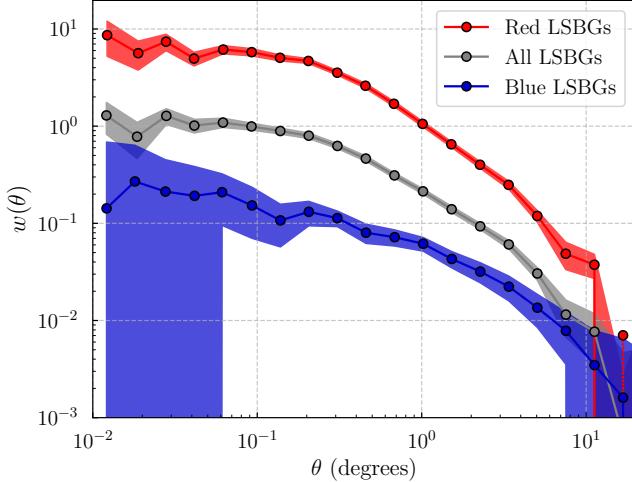
$$14.0 < g < 18.5, \quad (7)$$

while for the LSBG sample the range is, approximately,  $18 < g < 22$  (see also Appendix B). The 2MPZ sample thus consists of brighter, higher stellar mass galaxies compared to the LSBG sample.

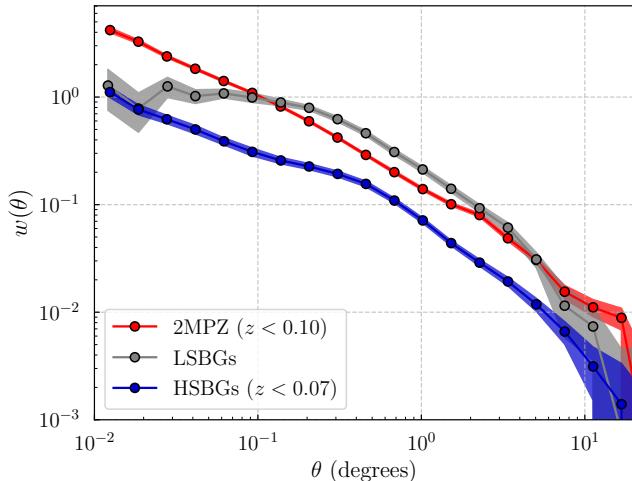
As before, we perform a series of redshift cuts to identify the one that gives an auto-correlation function that

is closest to that of the LSBGs. In this case that threshold is  $z < 0.10$ . We plot the auto-correlation function of the 2MPZ catalog (with  $z < 0.10$ ) in Figure 11 (red line). We also plot the auto-correlation function of the LSBG sample in Figure 11 (gray line).

We see that the HSB sample has a significantly lower clustering amplitude, both from the HSB and 2MPZ samples, especially in intermediate angular scales. LSBGs and the 2MPZ catalog give a more similar clustering amplitude; LSBGs are slightly more clustered than



**Figure 10.** The angular auto-correlation function of the total LSBG sample (dark gray line), and the red and blue LSBG subsamples (red and blue lines, accordingly). The errors were calculated using the Jackknife method. The correlation function of the red LSBGs has higher amplitude than that of the blue LSBGs across all angular scales.



**Figure 11.** The angular auto-correlation function of all LSBGs (gray line), the HSBG sample extracted from the DES data (blue line) and the 2MPZ sample (red line). We see that the LSBG exhibits a turnover at lower angular scales that is not observed either at the HSBG or 2MPZ samples.

2MPZ galaxies, with a turnover at  $\sim 0.1$  deg, which corresponds to a physical scale of 150 kpc at a distance of 60 Mpc ( $z = 0.02$ ). This turnover is observed in the LSBG sample, but neither in the HSBG or 2MPZ samples, at any redshift cut. Similar behaviour has been

noted before, specifically that LSBGs appear to avoid the centers of clusters (Wittmann et al. 2017).

The previous analysis probes the similarity in the clustering properties of the LSB, HSB and 2MPZ catalogs: which tend to have a similar (or stronger, or weaker) signal of presence of clusters in the the distribution of galaxies, and the dependence of that signal on the angular scale. However, this analysis does not inform us if the different galaxy samples probe the underlying matter overdensity field in a similar way; i.e. if, on a statistical basis, the peaks and troughs in their distribution coincide.

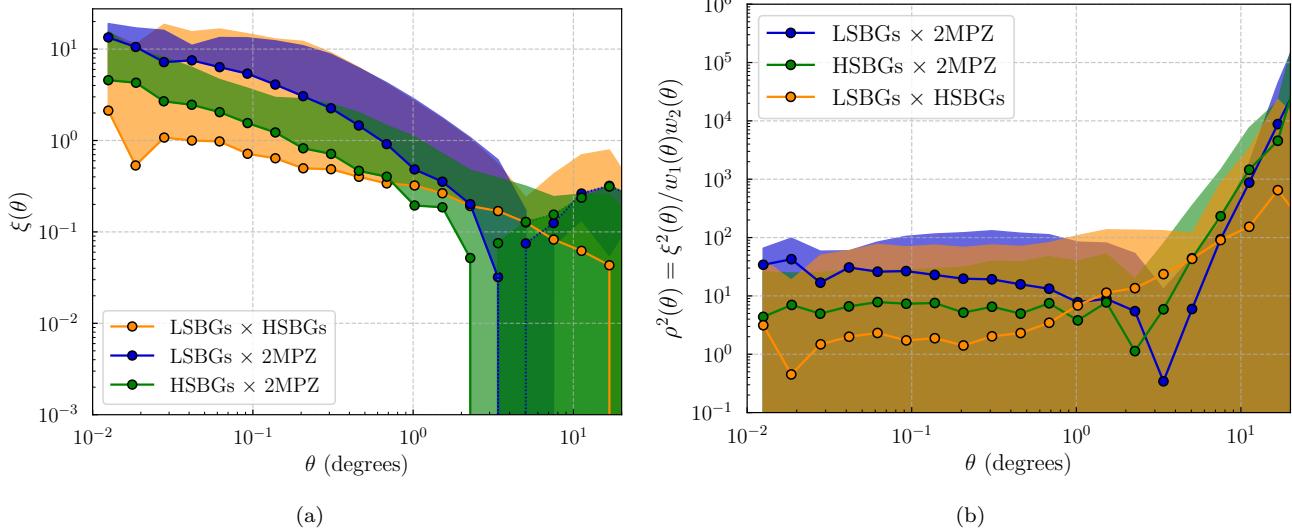
To answer this question we calculate and plot in the left-hand side panel of Figure 12 the cross correlation function,  $\xi(\theta)$ , between the LSB and HSB galaxy samples (orange line), between the LSB and the 2MPZ samples (blue line) and finally that between the HSB and 2MPZ samples (green line).

It is known that galaxies are biased traces of the underlying matter overdensity field. For large angular scales the two fields are connected by a (linear) galaxy bias factor,  $b_g$  as:  $\delta_g(z) = b_g(z)\delta_m(z)$ , where  $\delta$  refers to the overdenity field and the subscripts  $g, m$  to galaxies and matter, respectively. In general these are functions of redshift, as indicated, while the bias factor is different for different galaxy samples. The galaxy angular auto-correlation function is then defined as  $w(\theta) = \langle \delta_g(\hat{\mathbf{n}})\delta_g(\hat{\mathbf{n}} + \theta) \rangle = b_g^2 \langle \delta_m(\hat{\mathbf{n}})\delta_m(\hat{\mathbf{n}} + \theta) \rangle$ , where  $\hat{\mathbf{n}}$  the direction in the sky. For a detailed description of the modeling of the angular auto-correlation function see... . The cross correlation between two galaxy samples, say 1 and 2, is similarly defined as  $\xi_{12}(\theta) = \langle \delta_{g,1}(\hat{\mathbf{n}})\delta_{g,2}(\hat{\mathbf{n}} + \theta) \rangle = b_{g,1}b_{g,2}\langle \delta_m(\hat{\mathbf{n}})\delta_m(\hat{\mathbf{n}} + \theta) \rangle$ . If we define the cross-correlation coefficient between the two samples as:

$$\rho_{12}(\theta) = \frac{\xi_{12}(\theta)}{\sqrt{w_1(\theta)w_2(\theta)}}, \quad (8)$$

where  $w_{1,2}(\theta)$ , the auto-correlation functions of the individual samples, we can cancel out the corresponding bias factors present in the different samples. Thus, we can compare the correlations between the matter fields probed by the two samples. We plot the (square of the) cross-correlation coefficient between the same samples as those described above in the right-hand side panel of Figure 12.

Although the error bars are large, we see that the 2MPZ sample exhibits a significantly higher cross-correlation signal with the LSBG sample, compared to what the HSBG sample does. Thus, we conclude that 2MPZ sample, statistically, probes the same overdenities as those present in our LSBG sample (visual inspection



**Figure 12.** (a) The cross-correlation function,  $\xi(\theta)$ , between: (i) the LSB and HSB galaxy samples (orange line), (ii) the LSB and 2MPZ samples (blue line) and (iii) the HSB and 2MPZ galaxy samples (green line). We plot only the errors bands above the estimated  $\xi(\theta)$  values since they are large and because of the log-plot they would cover the whole area below the correlation function. (b) We also plot the (square of the) cross-correlation coefficient between the same samples as in (a), in order to cancel out the contribution of the different galaxy biases and compare the different cross-correlation levels.

confirmed that as well). We use this fact in the following section to construct radial profiles of the distribution of low- and high-surface brightness galaxies around the most prominent overdensity peaks.

## 7. ASSOCIATIONS WITH GALAXY CLUSTERS AND GROUPS

The analysis described in the previous section was of a statistical nature: we calculated the two point correlation function of the full LSBG sample and of the red and blue sub-samples and compared their values across different scales, between them, as well as with catalogs of HSBGs, either selected from the DES data or using external catalogs. The strong clustering of LSBGs demonstrated by the two-point auto-correlation function can be seen visually in the spatial distribution of our LSBG sample over the DES footprint (Figures 8 and 9). Here we turn our focus to the individual density peaks present in the LSBG distribution. We identify the most prominent spatial overdensities and we try to associate them with known galaxy clusters, galaxy groups and individual bright galaxies from external catalogs.

There are several reasons why it can be useful to associate peaks in the LSBG distribution to external catalogs:

1. Associating a peak in the LSBG distribution with a known object, for which we have distance information, can allow us to estimate the distance to the LSBGs themselves, assuming that all off them that lie within a specific radius from the peak are in the same distance. Distances allow us to estimate the intrinsic properties of

the LSBGs, such as physical sizes, absolute magnitudes etc.

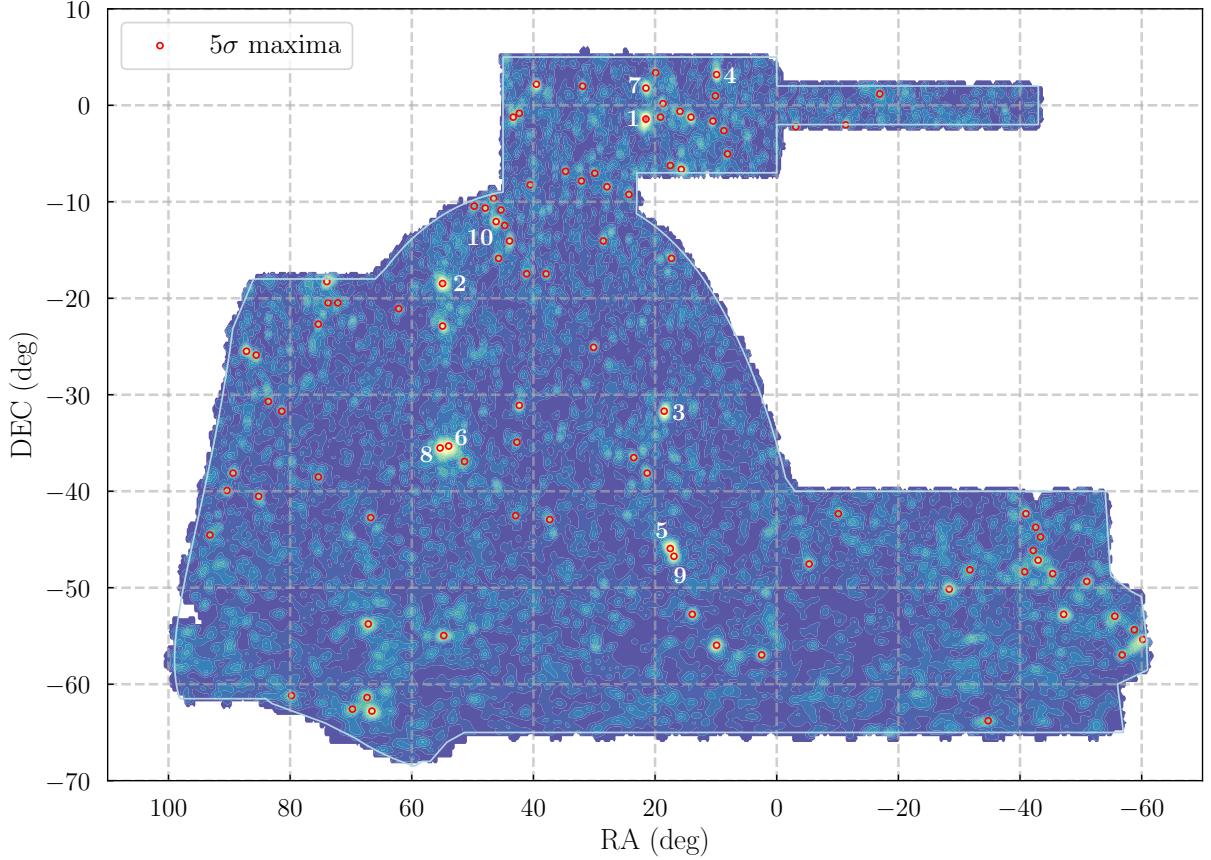
2. Defining a sample of likely LSB cluster members allows us to compare the properties of the LSBGs in cluster environments to those in the field. Such comparisons can be useful for testing models of LSBG formation and evolution.

3. We can compare the radial distributions of LSB and HSB cluster members to test for observable signatures of environmental effects that may be responsible for the creation of LSBGs.

4. Peaks in the LSBG density that are not associated to known clusters or groups can be potentially interesting, indicating different clustering patterns for LSBGs and HSBGs.

As a first step, we would like to identify the most prominent density peaks in a rigorous way. To start with, we have to go from the distribution map of the discrete data points, Figure 8, to a continuous density map defined over the DES footprint.

For that reason we perform a Kernel Density Estimation (KDE) using our full LSBG data sample. We use a Gaussian smoothing Kernel with a bandwidth of 0.3 degrees. The bandwidth was selected in such a way that does not erase the structure that can be seen in Figure 8; it is further justified by the radial profiles of LSBGs around peaks (see Figure 14). Note however that since clusters may lie in different radial distances and thus have different angular extends, the bandwidth may be



**Figure 13.** KDE map of the distribution of our LSBG sample. Blue regions denote areas of low density while yellow/red regions denote areas of higher density. With open red circles we denote the positions of the 96 prominent density peaks identified as described in Section 7. We have also numbered the ten most prominent (according to their KDE value) of them, with 1 denoting the most prominent peak.

small compared to the angular size of the nearest clusters and thus not fully smoothing cluster substructure. This is evident, for example, in the case of the Fornax cluster that lies at  $\text{RA} \sim 55^\circ$ ,  $\text{DEC} \sim -35^\circ$ , which is a nearby cluster at a distance of  $\sim 19$  Mpc, and as we can see it has an angular radius significantly larger than our bandwidth.

The resulting KDE map is presented in Figure 13, with bluer regions representing areas of lower density and more yellow/red regions representing areas of higher density. Next, we perform a  $5\sigma$ -clipping procedure on this KDE values map. Briefly, this procedure consists of an iteration over data, till convergence, where at each step values that exceed the median by  $5\sigma$  or more, are rejected. To identify the peaks, we find the local maxima in the regions of the KDE map that are above the threshold value returned from the  $5\sigma$ -clipping process. We locate 96 peaks passing the criteria of the above procedure; the red open circles in Figure 13 indicate the positions of these peaks. We furthermore number the ten most prominent of them (as defined by their KDE

value) and present their coordinates in Table 1. In the fifth column of that table we also present the number of LSBGs within 0.5 degrees from the center of each peak. The complete catalog can be found in a machine readable form in ....

Now, we would like to find how many of these peaks in the LSBG distribution can be associated with known galaxy clusters, galaxy groups and individual bright galaxies (since we expect some of our LSBGs to be dwarf satellites of bright galaxies). For that reason we employ the following external catalogs:

1. The Abell catalog of rich clusters (Southern survey, Abell et al. 1989).
2. The ROSAT-ESO Flux Limited X-ray (REFLEX) Galaxy cluster survey (Böhringer et al. 2004).
3. A catalog of galaxy groups built from the sample of the 2MASS Redshift Survey (Tully 2015). We keep only those groups that have more than five members.
4. The bright galaxies from the revised New General Catalogue (Sulentic & Tifft 1999).

**Table 1.** The ten most prominent density peaks and their associations

Peak	(RA,Dec) <sub>peak</sub>	Best	(RA,Dec) <sub>assoc</sub>	Redshift	Distance	$N(< 0.5^\circ)$
Number	(deg)	Association	(deg)	$z$	(Mpc)	
1	(21.5012, -1.4286)	Abell 194	(21.4200, -1.4072)	0.018	$75.07 \pm 5.26$	65
2	(54.9388, -18.4712)	RXC J0340.1-1835	(55.0475, -18.5875)	0.0057	$23.41 \pm 1.64$	45
3	(18.4983, -31.7043)	Abell S141	(18.4758, -31.7519)	0.020	$84.80 \pm 5.94$	41
4	(9.8887, 3.1829)	NGC 199	(9.8882, 3.1385)	0.0153	$62.81 \pm 4.41$	40
5	(17.4972, -45.9398)	Abell 2877	(17.6017, -45.9228)	0.0247	$106.61 \pm 7.45$	39
6	(53.9377, -35.3133)	Fornax (Abell S373)	(54.6162, -35.4483)	0.0046	$18.97 \pm 1.33$	32
7	(21.5017, 1.7794)	RXC J0125.5+0145	(21.3746, 1.7627)	0.01739	$72.32 \pm 5.10$	36
8	(55.3393, -35.5138)	Fornax (Abell S373)	(54.6162, -35.4483)	0.0046	$18.97 \pm 1.33$	27
9	(16.8965, -46.7418)	Abell 2870	(16.9299, -46.9165)	0.0237	$102.03 \pm 3.89$	30
10	(46.1290, -12.0551)	NGC 1200	(45.9770, -11.9918)	0.01305	$57.03 \pm 4.01$	32

NOTE— In this table we present the coordinates (RA, Dec)<sub>peak</sub> of the ten most prominent density peaks (second column) and their best association according to the criteria described in the main text, Section 7 (third column); in the fourth column we present the coordinates of these associations, (RA, Dec)<sub>assoc</sub>. In the fifth and sixth columns we present the redshift and the distance, respectively, of the associations, retrieved from the NASA Extragalactic Database <sup>a</sup>. Finally, in the seventh column we present the number of LSBGs that lie within  $0.5^\circ$  from the center of each peak.

<sup>a</sup> <https://ned.ipac.caltech.edu/>

For each one of the peaks we plotted the LSBGs and the objects in the mentioned catalogs in a region  $\pm 0.5$  degrees from the nominal center of the peak. To find the association (if any) we proceed as follows: We select the object from the external catalogs that is closest to the center, by giving priority to objects according to the above bullet point ordering. That means that, even if a NGC galaxy is closer to the peak than, say, an Abell cluster, we still associate the peak with Abell cluster, provided that it is close enough.

From the 96 peaks, 34 were associated with an Abell cluster, 9 with a REFLEX cluster, 7 with a 2MASS group, 31 with an NGC galaxy while 15 remained unassociated.

Due to their special interest, we visually inspected the regions around these 15 unassociated peaks, using the DES Sky Viewer tool. In  $\sim 8$  cases we were able to identify nearby bright galaxies/galaxy clusters, thus indicating that these are not special cases of clusters of LSB galaxies without luminous galaxy counterparts; these counterparts were simply not included in the catalogs of objects we used for the matching. Interestingly enough, though,  $\sim 7$  of them do not seem to lie to a HSB galaxy cluster and thus are genuine LSB clusters, that may potentially be an interesting target for further investigation.

A second question that naturally arises is the following: are there HSB galaxy clusters that do not have

high number of LSBGs, in other words do not present a peak in the LSBG distribution? Inside the DES footprint there are, for example, 1762 Abell catalog objects. Selecting those that have a redshift entry  $z > 0$  (sometimes an entry  $z = 0$  may indicate that the Abell object is not a real galaxy cluster), we remain with 241 clusters. Why only 34 of these show prominent LSB peaks? A first examination showed that all of the clusters associated with a peak have a low  $z$  value; 33 of them have  $z < 0.05$ . Out of the total 241 Abell cluster sample, 91 of them have  $z < 0.05$ . Thus  $\sim 1/3$  of the low- $z$  clusters present a peak. In terms of richness we didn't find significant differences between those presenting and not presenting an LSB peak. Although certainly requires further examination, proximity of a cluster seems to be the main factor for the identification of an associated peak in the LSB overdensity map.

In Table 1 we present the coordinates of the ten most prominent overdensity peaks, their best associations and the coordinates, redshifts and distances of these associations (retrieved from the NASA Extragalactic Database). Finally we also show the number of LSBGs within  $0.5^\circ$  from the center of each peak. Note that two peaks are both associated with the Fornax cluster (Abell S373). We note that the associations to these prominent overdensity peaks all seem to be, after visual inspection, very good. The full table can be found in supplemental material, where we add an extra column

characterizing the quality of association, ranging from I (very good), II (good) to III (not so good).

The association of the peaks with clusters of known distance allows us to estimate intrinsic physical properties for some of our LSBGs (under the assumption that LSBGs within a certain angular distance from the peak belong to that cluster and lie at the same distance). This information is otherwise absent due to the photometric nature of DES. In the remainder of this section we will use distance information from the nine most prominent associations to: (i) study the radial distribution of LSBGs around clusters, and (ii) get the size-luminosity relation for these galaxies.

### 7.1. Radial Profiles

The study of the distribution of the LSBGs inside clusters, and the comparison with the distribution of normal galaxies, can reveal environmental effects and probe the formation histories of LSGBGs. For that reason, in Figure 14 we plot the radial distribution (projected number density as a function of the projected physical distance from the center of the cluster) of LSBGs around the nine associations (clusters and NGC galaxies) described above, see Table 1. We also present, for comparison, the radial distribution of the 2MPZ galaxy sample (with redshift  $z < 0.10$ ) that is significantly correlated with the LSBG sample.

For each of the clusters, we select all LSBGs and 2MPZ galaxies that lie within an angular distance  $\theta$  from the center from the center of the cluster, given by  $\theta = 1.5\text{Mpc}/D$ , where  $D$  the distance to the cluster in Mpc. We thus assume that all these galaxies lie at the same radial distance.

We calculate the radial profiles as follows: We divide the total area around the center of each cluster (disk of radius 1.5 Mpc) into fifteen annuli of constant width 0.1 Mpc. We calculate the number density of galaxies in each annulus and we normalize to the mean number density of galaxies (of the particular sample) within the disk. With such a normalization, a flat line with unit amplitude indicates a homogeneous distribution of galaxies within the 1.5 Mpc region. The errors, presented as shaded regions in Figure 14, are estimated assuming Poisson noise in the measured number of galaxies per annulus and the total number of galaxies in the disk.

In all cases we see that the LSBG distribution is significantly peaked in the inner  $\sim 0.5$  Mpc region and flattens out at distances  $> 1.0$  Mpc. From the normalized number density, we find that most clusters present LSBG overdensity peaks of similar amplitude; from our sample, Fornax cluster (Abell S373) presents the most flat

(lower amplitude in the peak) radial distribution, while the distribution around NGC 199 is extremely peaked. This points towards the fact that most of the LSB galaxies around NGC 199 are dwarf satellite galaxies of a single central bright galaxy.

Comparing with the distribution of galaxies from the 2MPZ samples, we see that in all but three cases the normalized radial distributions of the LSB and the brighter galaxy samples are quite similar. For the three cases that the distributions do not match, it is easy to understand the origins of this discrepancy.

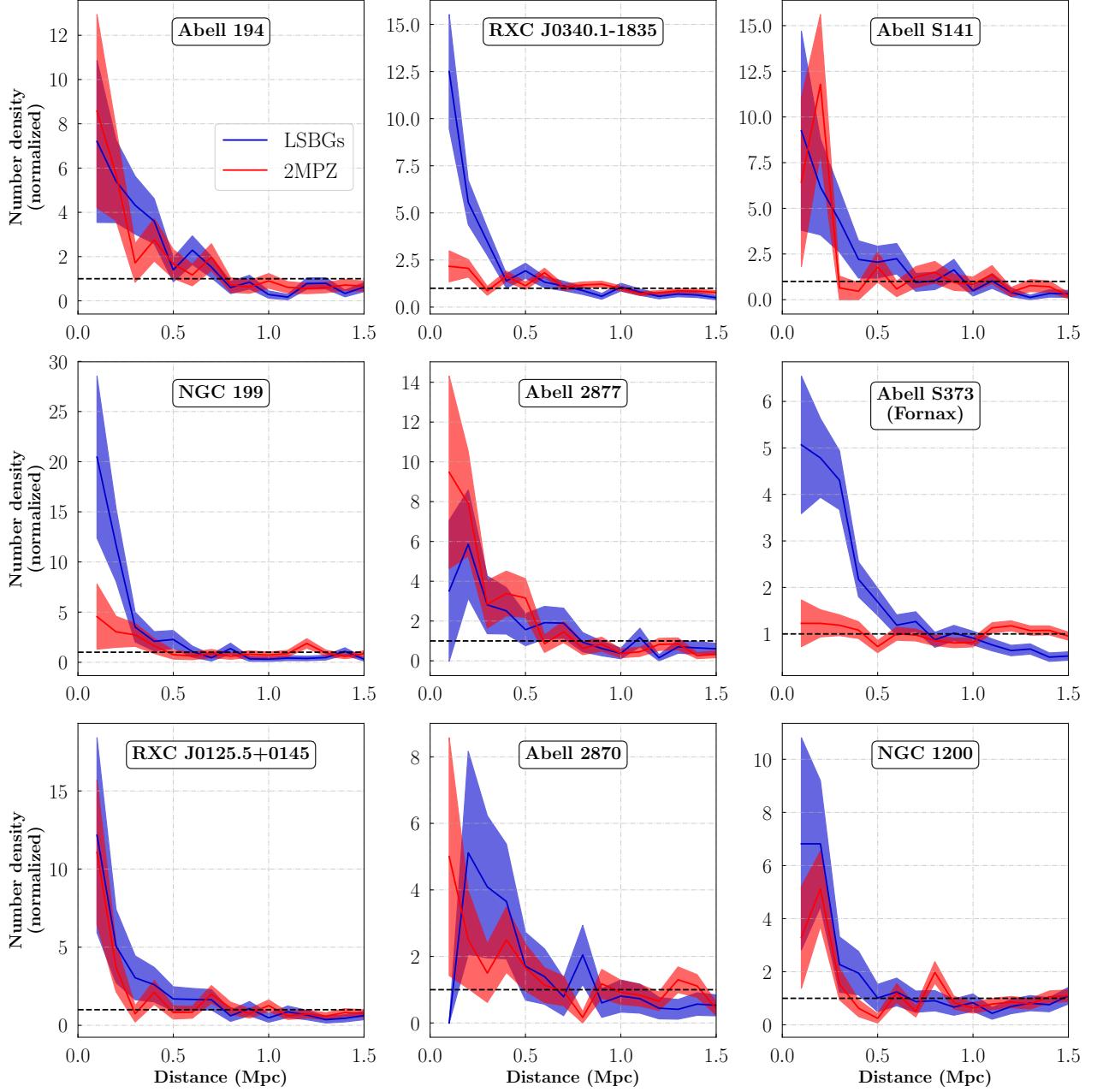
The two clusters, RXC J0340.1-1835 and Abell S373 (Fornax), panels 2 and 6, respectively, are significantly closer to us than the other systems(see Table 1). They lie at redshift  $z = 0.0057$  and  $z = 0.0046$ , respectively. The next closest association is NGC 1200 at  $z = 0.013$ . The 2MPZ catalog includes just a few objects with such low redshifts; there are only 24 objects with  $z < 0.005$  and 42 objects with  $z < 0.006$ . So, what we actually see in these two cases is the homogeneous background distribution of the 2MPZ sample. The third case where the distribution of 2MPZ and LSBG galaxies differ is around NGC 199. Again, the LSBGs are much more peaked than the 2MPZ sample, suggesting that the observed LSBG overdensity is caused by dwarf galaxies surrounding a single central host. Despite the small sample size, we can say qualitatively that we do not find significant differences in the distribution of LSBGs and HSBGs associated to other galaxy clusters.

### 7.2. Size-Luminosity Relation

The distance information stemming from the association of LSBG distribution overdensities with known clusters can also be used to get insight into the physical properties of the individual galaxies belonging to that cluster.

For the same nine associations of the most prominent peaks, as those described previously, we select all LSBGs that lie within 1 Mpc from the center of the association. By assuming that LSBGs within this radius are likely to belong to the corresponding cluster and are at the same distance, we can estimate their physical effective radius (in pc) and their absolute magnitude (note that this is more conservative compared to Section 7.1 in our selection). **(ADW: We should make an estimate of the contamination fraction based on the background in the outermost annulus.)**

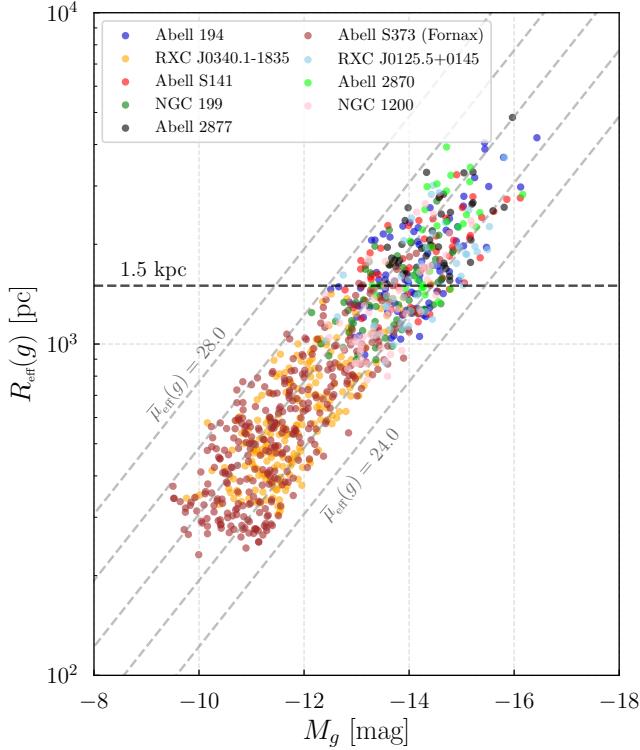
In Figure 15 we present the size-luminosity relationship for these galaxies based on the physical effective radius,  $R_{\text{eff}}(g)$ , and the absolute magnitude in the  $g$  band,  $M_g$ . We color-code the galaxies according to the cluster/association they belong to. The total sample



**Figure 14.** Normalized radial profiles of the distribution of LSB galaxies (blue) and galaxies from the 2MPZ catalog (red) around the associations of the most prominent LSBG overdensity peaks, presented in Table 1. We have assumed that all galaxies that are within a radius that corresponds to a physical scale of 1.5 Mpc at the distance of the association belong to that association. The normalization constant corresponds to the mean number density of galaxies within the 1.5 Mpc radius.

consists of 937 LSBGs. The largest number of objects (352) is associated to the Fornax cluster and the second largest number (210) belongs to RXC J0340.1-1835. These two clusters, as we have seen, are the closest ones, and thus more galaxies (galaxies with smaller physical sizes) pass the angular selection criterion,  $r_{1/2}(g) > 2''.5$ .

The LSBGs in the sample studied here span a wide range physical sizes, from  $R_{\text{eff}}(g) \sim 0.23 \text{ kpc}$ , up to  $R_{\text{eff}}(g) \sim 4.83 \text{ kpc}$ , with a median value of  $R_{\text{eff}}(g) \sim 0.90 \text{ kpc}$ . The lower limit is largely set by the angular size selection limit, translated to a physical size for the nearest cluster (Fornax). 228 galaxies have sizes  $> 1.5 \text{ kpc}$  and surface brightness  $\bar{\mu}_{\text{eff}}(g) > 24.0 \text{ mag arcsec}^{-2}$ , thus satisfying the conventionally definition of ultra-



**Figure 15.** Size-Luminosity relation for LSBGs around the associations of the most prominent overdensity peaks, presented in 1. We have assumed that all LSBGs within an angle corresponding to a physical radius of 1 Mpc at the distance of the association belong to that association. The sample consists of 937 galaxies, with 228 having effective radii exceeding 1.5 kpc (black dashed line), under our stated assumptions. We also show (dashed, diagonal gray lines) the lines of constant mean-surface brightness.

diffuse galaxies (UDGs; ?). The LSBGs belonging to more distant clusters are systematically biased to larger sizes, which is again a result of our angular selection criterion.

The sample covers a range of  $\sim 7$  orders of magnitude in absolute,  $g$ -band, magnitude, from  $M_g \sim -9$  to  $M_g \sim -17$ , with a median of  $M_g \sim -12.7$ . In Figure 15 we also show lines of constant mean surface brightness. The bright-end limit is largely set by the requirement  $\bar{\mu}_{\text{eff}}(g) > 24.3 \text{ mag arcsec}^{-2}$  used to produce our catalog (note however that that cut was imposed on the `SExtractor` mean surface brightness parameter, while here we use the `galfit` quantities). Only 5 galaxies lie in the extremely faint end,  $\bar{\mu}_{\text{eff}}(g) > 27.0 \text{ mag arcsec}^{-2}$ .

We see that the galaxies in the sample discussed here span the same range in mean surface brightness ( $\bar{\mu}_{\text{eff}}(g) \sim 24.3 - 27.0$ ), regardless their size: small or

large galaxies do not seem to prefer the bright or the faint end. Thus, UDGs seem to be a natural continuation of the general LSBG population in the regime of large size and low surface brightness, and not a distinct population that is well-separated in the size-luminosity parameter space from other LSBGs (?).

## 8. SUMMARY AND CONCLUSIONS

In this paper we have selected and analyzed 21,420 extended, low-surface-brightness galaxies (LSBGs) from the first three years of DES imaging data. Our detection pipeline consists of three main steps:

1. Cuts on the DES Y3 Gold catalog based on the `SExtractor` parameters. The selection criteria are summarized in Appendix A. The most important of them were on half-light radius ( $r_{1/2} > 2''.5$ ) and mean surface brightness  $\bar{\mu}_{\text{eff}}(g) > 24.3 \text{ mag arcsec}^{-2}$ .
2. A machine learning classification step, using a SVM classifier, tuned to reduce the incidents of false negatives (LSBGs classified as non-LSBGs).
3. A final visual inspection that eliminated the remaining false positives from the previous step, resulting in the final, highly pure sample of LSBGs.

We divide the total LSBG sample into two subsamples according to their  $g-i$  color. We study the photometric, structural and spatial clustering properties of the red ( $g-i \geq 0.639$ ) and blue ( $g-i < 0.639$ ) subsamples. Our main findings are the following:

- The distributions in angular size (effective radius) parameter are similar for the two subsamples with the red population having slightly higher median value ( $\sim 3.93 \text{ arcsec}$ ) compared to the blue population ( $\sim 3.74 \text{ arcsec}$ ).
- The median value of the Sérsic index parameter of the both samples is the same,  $n \sim 1.0$ .
- The mean surface brightness distribution present some noticeable difference between the two populations: blue galaxies have a distribution that is more peaked towards the high brightness end. The 80th percentile is at  $\bar{\mu}_{\text{eff}}(g) = 25.5 \text{ mag arcsec}^{-2}$  for the blue population and at  $\bar{\mu}_{\text{eff}}(g) = 26.1 \text{ mag arcsec}^{-2}$  for the red one. We note this behavior (blue galaxies having higher mean brightness surface) is not as prominent as the one previously observed in (Greco et al. 2018). The distribution in the central surface brightness,  $\mu_0(g)$ , does not present significant differences between the two subsamples.
- The spatial distribution of the red LSBGs exhibits a much higher level of clustering compared that of the blue LSBGs, which have an almost homogeneous distribution. Correlation analysis confirms that we can see

by visually inspecting the distribution of the two populations: the two-point angular correlation function of the red subsample has an amplitude  $\sim$  an order of magnitude higher than that of the blue subsample.

Furthermore, we compare the clustering level of the full LSBG sample with that of a high-surface brightness galaxies (HSBGs) selected from the DES year data (selecting galaxies with  $20.0 < \bar{\mu}_{\text{eff}}(g) < 22.0 \text{ mag arcsec}^{-2}$ ) and with the external 2MPZ catalog that contains nearby (peak at  $z \sim 0.07$ ) galaxies with accurately measured photometric redshifts. We find a similar clustering amplitude (and also a high cross-correlation signal) between the LSBG sample and the 2MPZ catalog, after imposing the redshift cut  $z < 0.1$  on the latter (which points to the fact that most of our LSBGs lie at low redshifts, as expected, because of the angular extension requirement). An interesting feature is the the lower clustering (turnover) of the LSBG sample for angular scales less than  $\sim 0.1$  deg.

The distribution of LSBGs presents some prominent overdensity peaks. We cross-match the 96 most prominent of them (returned from a  $5\sigma$  clipping procedure on the KDE map of the spatial distribution of LSBGs with a Gaussian smoothing kernel of bandwidth equal to 0.3 deg), with external catalogs of galaxy clusters, groups and individual bright galaxies. We find that out of the 96 peaks:

- 34 were associated with an Abell cluster.
- 9 were associated with a cluster from the REFLEX catalog.
- 7 were associated with a galaxy group from the 2MASS catalog.
- 31 were associated with a galaxy from the NGC catalog.
- 15 had no association in any of the above catalogs.

The association of peaks with objects (clusters, galaxies) of known distance allows us to gain some radial information about our sample, otherwise absent due to the poor quality of photo- $z$  estimation for nearby LSBG galaxies. The median distance of the associations for which such information is available is  $\sim 82$  Mpc, with a range starting at  $\sim 16.5$  Mpc (NGC 456) (closest cluster is Fornax at a distance of  $\sim 19$  Mpc) and going up to  $\sim 354$  Mpc (Abell 2911). The mean distance is  $\sim 106$  Mpc with a standard deviation of  $\sim 66$  Mpc.

This radial information allows further exploration for some of the LSBGs and their host systems. As two applications of that, we construct:

- Radial profiles of the distribution of the LSBG and 2MPZ samples around the nine most prominent associations, as a function of the physical distance from the center of the association. Overall normalization aside,

we find that for cases where meaningful results can be extracted (peaks that correspond to clusters and not individual bright galaxies, distances are not very low) the distributions from the two samples are very similar.

- A (physical) size - (absolute) magnitude relationship plot, for the LSBGs belonging the nine most prominent associations, as above (assuming that every LSBG within a an angular distance corresponding to 1 Mpc belongs to that association). We find that LSBGs in our sample, span a range in physical size (effective radius) from  $\sim 0.23$  kpc up to  $\sim 4.83$  kpc, with a median size of 0.90 kpc. Out of the 937 LSBGs studied, 228 are exceeding the conventional size definition for Ultra Diffuse Galaxies (UDGs), namely an effective radius  $R_{\text{eff}} > 1.5$  kpc. In terms of absolute magnitude (in the  $g$  band) they span  $\sim 7$  seven orders of magnitude ( $M_g \in [-9, -17]$ ) with a median of  $M_g \sim -12.7$ .

Our catalog is the largest catalog of LSBGs so far; it opens a new window for exploration and discovery in a regime of galaxy formation relatively unstudied, especially in terms of large, wide-field surveys (and not targeted observations around clusters). We have presented a preliminary, mainly descriptive, analysis of our catalog; further analyses can try to test it against galaxy formation predictions in the low-surface-brightness regime; further study the properties of LSBGs in different environments (clusters/field); getting constraints on the mean mass of LSBGs using weak lensing can also be attempted () .

Of equal importance is the fact that it can be used to better prepare for the next generation galaxy surveys, like the upcoming LSST project. Given its large size, it can be used, and we plan to use it, to try to improve LSBG detection pipelines: because of the detection challenges and large number of false positives, a visual inspection part was inevitable so far. Given the large amount of data they will collect, this may be infeasible in future surveys, unless crowdsourcing techniques are employed. An alternative to that can be to train machine and deep learning algorithms to fully automate the detection process. We will explore this possibility in an upcoming project.

## 9. ACKNOWLEDGMENTS

This research received support from the National Science Foundation (NSF) under grant no. NSF PHY17-48958 through the Kavli Institute for Theoretical Physics program “The Small-Scale Structure of Cold(?) Dark Matter,” and grant no. NSF DGE-1656518 through the NSF Graduate Research Fellowship received by E.O.N. This research made use of computational resources at SLAC National Accelerator Laboratory, a

U.S. Department of Energy Office; the authors thank the support of the SLAC computational team. This research has made use of the WEBDA database, operated at the Department of Theoretical Physics and Astrophysics of the Masaryk University.

Funding for the DES Projects has been provided by the U.S. Department of Energy, the U.S. National Science Foundation, the Ministry of Science and Education of Spain, the Science and Technology Facilities Council of the United Kingdom, the Higher Education Funding Council for England, the National Center for Supercomputing Applications at the University of Illinois at Urbana-Champaign, the Kavli Institute of Cosmological Physics at the University of Chicago, the Center for Cosmology and Astro-Particle Physics at the Ohio State University, the Mitchell Institute for Fundamental Physics and Astronomy at Texas A&M University, Financiadora de Estudos e Projetos, Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro, Conselho Nacional de Desenvolvimento Científico e Tecnológico and the Ministério da Ciência, Tecnologia e Inovação, the Deutsche Forschungsgemeinschaft and the Collaborating Institutions in the Dark Energy Survey.

The Collaborating Institutions are Argonne National Laboratory, the University of California at Santa Cruz, the University of Cambridge, Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas-Madrid, the University of Chicago, University College London, the DES-Brazil Consortium, the University of Edinburgh, the Eidgenössische Technische Hochschule (ETH) Zürich, Fermi National Accelerator Laboratory, the University of Illinois at Urbana-Champaign, the Institut de Ciències de l'Espai (IEEC/CSIC), the Institut de Física d'Altes Energies, Lawrence Berkeley National Laboratory, the Ludwig-Maximilians Universität München and the associated Excellence Cluster Universe, the University of Michigan, the National Optical Astronomy Observatory, the University of Nottingham, The Ohio State University, the University of Pennsylvania, the University of Portsmouth, SLAC National Accelerator Laboratory, Stanford University, the Univer-

sity of Sussex, Texas A&M University, and the OzDES Membership Consortium.

Based in part on observations at Cerro Tololo Inter-American Observatory, National Optical Astronomy Observatory, which is operated by the Association of Universities for Research in Astronomy (AURA) under a cooperative agreement with the National Science Foundation.

The DES data management system is supported by the National Science Foundation under Grant Numbers AST-1138766 and AST-1536171. The DES participants from Spanish institutions are partially supported by MINECO under grants AYA2015-71825, ESP2015-66861, FPA2015-68048, SEV-2016-0588, SEV-2016-0597, and MDM-2015-0509, some of which include ERDF funds from the European Union. IFAE is partially funded by the CERCA program of the Generalitat de Catalunya. Research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Program (FP7/2007-2013) including ERC grant agreements 240672, 291329, and 306478. We acknowledge support from the Brazilian Instituto Nacional de Ciência e Tecnologia (INCT) e-Universe (CNPq grant 465376/2014-2).

This manuscript has been authored by Fermi Research Alliance, LLC under Contract No. DE-AC02-07CH11359 with the U.S. Department of Energy, Office of Science, Office of High Energy Physics. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.

*Facilities:* Blanco, DECam

*Software:* astropy (Astropy Collaboration et al. 2013), Cloudy (Ferland et al. 2013), SExtractor (Bertin & Arnouts 1996), sci-kit learn (Pedregosa et al. 2011)

## REFERENCES

- Abell, G. O., Corwin, Jr., H. G., & Olowin, R. P. 1989, ApJS, 70, 1
- Abraham, R. G., & van Dokkum, P. G. 2014, PASP, 126, 55
- Adami, C., Scheidegger, R., Ulmer, M., et al. 2006, A&A, 459, 679
- Amorisco, N. C., & Loeb, A. 2016, MNRAS, 459, L51
- Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33
- Baldry, I. K., Glazebrook, K., Brinkmann, J., et al. 2004, ApJ, 600, 681
- Bamford, S. P., Nichol, R. C., Baldry, I. K., et al. 2009, MNRAS, 393, 1324

- Barden, M., Häußler, B., Peng, C. Y., McIntosh, D. H., & Guo, Y. 2012, MNRAS, 422, 449
- Behroozi, P. S., Wechsler, R. H., & Conroy, C. 2013, ApJ, 770, 57
- Bernstein, G. M., Nichol, R. C., Tyson, J. A., Ulmer, M. P., & Wittman, D. 1995, AJ, 110, 1507
- Bernstein, G. M., Abbott, T. M. C., Armstrong, R., et al. 2018, PASP, 130, 054501
- Bertin, E. 2006, in Astronomical Society of the Pacific Conference Series, Vol. 351, Astronomical Data Analysis Software and Systems XV, ed. C. Gabriel, C. Arviset, D. Ponz, & S. Enrique, 112
- Bertin, E., & Arnouts, S. 1996, A&AS, 117, 393
- Bilicki, M., Jarrett, T. H., Peacock, J. A., Cluver, M. E., & Steward, L. 2014, ApJS, 210, 9
- Blakeslee, J. P., Jordán, A., Mei, S., et al. 2009, ApJ, 694, 556
- Blanton, M. R., & Moustakas, J. 2009, ARA&A, 47, 159
- Böhringer, H., Schuecker, P., Guzzo, L., et al. 2004, A&A, 425, 367
- Bothun, G., Impey, C., & McGaugh, S. 1997, PASP, 109, 745
- Carleton, T., Errani, R., Cooper, M., et al. 2019, MNRAS, 485, 382
- Dalcanton, J. J., Spergel, D. N., Gunn, J. E., Schmidt, M., & Schneider, D. P. 1997, AJ, 114, 635
- Danieli, S., van Dokkum, P., Merritt, A., et al. 2017, ApJ, 837, 136
- Davis, M., Efstathiou, G., Frenk, C. S., & White, S. D. M. 1985, ApJ, 292, 371
- DES Collaboration, Abbott, T. M. C., Abdalla, F. B., et al. 2018, ApJS, 239, 18
- Disney, M. J. 1976, Nature, 263, 573
- Driver, S. P. 1999, ApJL, 526, L69
- Eigenthaler, P., Puzia, T. H., Taylor, M. A., et al. 2018, ApJ, 855, 142
- Ferland, G. J., Porter, R. L., van Hoof, P. A. M., et al. 2013, RMxAA, 49, 137
- Ferrero, I., Abadi, M. G., Navarro, J. F., Sales, L. V., & Gurovich, S. 2012, MNRAS, 425, 2817
- Flaugher, B., Diehl, H. T., Honscheid, K., et al. 2015, AJ, 150, 150
- Freeman, K. C. 1970, ApJ, 160, 811
- Galaz, G., Herrera-Camus, R., Garcia-Lambas, D., & Padilla, N. 2011, ApJ, 728, 74
- Geha, M., Wechsler, R. H., Mao, Y.-Y., et al. 2017, ApJ, 847, 4
- Greco, J. P., Greene, J. E., Strauss, M. A., et al. 2018, ApJ, 857, 104
- Häußler, B., Bamford, S. P., Vika, M., et al. 2013, MNRAS, 430, 330
- Hayward, C. C., Irwin, J. A., & Bregman, J. N. 2005, ApJ, 635, 827
- Jarvis, M. 2015, TreeCorr: Two-point correlation functions, ascl:1508.007
- Koda, J., Yagi, M., Yamanoi, H., & Komiyama, Y. 2015, ApJL, 807, L2
- Larson, R. B., Tinsley, B. M., & Caldwell, C. N. 1980, ApJ, 237, 692
- Lintott, C., Schawinski, K., Bamford, S., et al. 2011, MNRAS, 410, 166
- Martin, G., Kaviraj, S., Laigle, C., et al. 2019, MNRAS, 485, 796
- McConnachie, A. W. 2012, AJ, 144, 4
- McGaugh, S. S., Bothun, G. D., & Schombert, J. M. 1995, AJ, 110, 573
- Merritt, A., van Dokkum, P., Danieli, S., et al. 2016, ApJ, 833, 168
- Mihos, J. C., Harding, P., Feldmeier, J. J., et al. 2017, ApJ, 834, 16
- Mihos, J. C., Durrell, P. R., Ferrarese, L., et al. 2015, ApJL, 809, L21
- Minchin, R. F., Disney, M. J., Parker, Q. A., et al. 2004, MNRAS, 355, 1303
- Moster, B. P., Naab, T., & White, S. D. M. 2013, MNRAS, 428, 3121
- Muñoz, R. P., Eigenthaler, P., Puzia, T. H., et al. 2015, ApJL, 813, L15
- O’Neil, K., Bothun, G. D., & Cornell, M. E. 1997, AJ, 113, 1212
- O’Neil, K., Bothun, G. D., & Schombert, J. 2000, AJ, 119, 136
- Ordenes-Briceo, Y., Eigenthaler, P., Taylor, M. A., et al. 2018, The Astrophysical Journal, 859, 52
- Papastergis, E., Giovanelli, R., Haynes, M. P., & Shankar, F. 2015, A&A, 574, A113
- Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, Journal of Machine Learning Research
- Peebles, P. J. E. 1980, The large-scale structure of the universe
- Peng, C. Y., Ho, L. C., Impey, C. D., & Rix, H.-W. 2002, AJ, 124, 266
- Román, J., & Trujillo, I. 2017, MNRAS, 468, 4039
- Rosenbaum, S. D., Krusch, E., Bomans, D. J., & Dettmar, R. J. 2009, A&A, 504, 807
- Sabatini, S., Davies, J., van Driel, W., et al. 2005, MNRAS, 357, 819
- Strateva, I., Ivezić, Ž., Knapp, G. R., et al. 2001, AJ, 122, 1861

- Sulentic, J. W., & Tifft, W. G. 1999, VizieR Online Data Catalog, 7001
- Tully, R. B. 2015, AJ, 149, 171
- van Dokkum, P., Danieli, S., Abraham, R., Conroy, C., & Romanowsky, A. J. 2019, ApJL, 874, L5
- van Dokkum, P., Danieli, S., Cohen, Y., et al. 2018, Nature, 555, 629
- van Dokkum, P. G., Abraham, R., Merritt, A., et al. 2015, ApJL, 798, L45
- Venhola, A., Peletier, R., Laurikainen, E., et al. 2017, A&A, 608, A142
- Wechsler, R. H., & Tinker, J. L. 2018, ARA&A, 56, 435
- White, S. D. M., & Frenk, C. S. 1991, ApJ, 379, 52
- Wittmann, C., Lisker, T., Ambachew Tilahun, L., et al. 2017, MNRAS, 470, 1512
- Zhong, G. H., Liang, Y. C., Liu, F. S., et al. 2008, MNRAS, 391, 986

## APPENDIX

## A. SELECTION CRITERIA

Removal of point sources (star-galaxy separation):

```
(EXTENDED_CLASS_COADD != 0) &
(EXTENDED_CLASS_MOF != 0) &
(SPREAD_MODEL_I + 5/3*SPREADERR_MODEL_I > 0.007)
```

Selection of LSBG candidates:

- Surface brightness and radius cuts:

```
(FLUX_RADIUS_G > 2.5) & (FLUX_RADIUS_G < 20)
(MU_MEAN_MODEL_G > 24.3) & (MU_MEAN_MODEL_G < 28.8)
```

- Ellipticity cut:

```
(1 - A_IMAGE/B_IMAGE) < 0.7
```

- Color cuts:

```
-0.1 < (MAG_AUTO_G-MAG_AUTO_I) < 1.4
(MAG_AUTO_G - MAG_AUTO_R) > 0.7*(MAG_AUTO_G - MAG_AUTO_I) - 0.4
(MAG_AUTO_G - MAG_AUTO_R) < 0.7*(MAG_AUTO_G - MAG_AUTO_I) + 0.4
```

## B. MAGNITUDE DISTRIBUTIONS OF THE LSBG, HSB AND 2MPZ SAMPLES