

Observational properties of ultra-diffuse galaxies in low-density environments: field UDGs are predominantly blue and star forming

D. J. Prole,^{1,2★} R. F. J. van der Burg,² M. Hilker² and J. I. Davies¹

¹*School of Physics and Astronomy, Cardiff University, The Parade, Cardiff CF243AA, UK*

²*European Southern Observatory, Karl-Schwarzschild-Str 2, D-85748 Garching bei München, Germany*

Accepted 2019 July 2. Received 2019 July 1; in original form 2019 May 2

ABSTRACT

While we have learnt much about ultradiffuse galaxies (UDGs) in groups and clusters, relatively little is known about them in less dense environments. More isolated UDGs are important for our understanding of UDG formation scenarios because they form via secular mechanisms, allowing us to determine the relative importance of environmentally driven formation in groups and clusters. We have used the public Kilo-Degree Survey together with the Hyper Suprime-Cam Subaru Strategic Program to constrain the abundance and properties of UDGs in the field, targeting sources with low surface brightness ($24.0 \leq \bar{\mu}_{e,r} \leq 26.5$) and large apparent sizes ($3.0 \text{ arcsec} \leq \bar{r}_{e,r} \leq 8.0 \text{ arcsec}$). Accounting for several sources of interlopers in our selection based on canonical scaling relations, and using an empirical UDG model based on measurements from the literature, we show that a scenario in which cluster-like red-sequence UDGs occupy a significant number of field galaxies is unlikely, with most field UDGs being significantly bluer and showing signs of localized star formation. An immediate conclusion is that UDGs are much more efficiently quenched in high-density environments. We estimate an upper limit on the total field abundance of UDGs of $8 \pm 3 \times 10^{-3} \text{ cMpc}^{-3}$ within our selection range. We also compare the total field abundance of UDGs to a measurement of the abundance of H I-rich UDGs from the literature, suggesting that they occupy at least one-fifth of the overall UDG population. The mass formation efficiency of UDGs implied by this upper limit is similar to what is measured in groups and clusters.

Key words: galaxies: abundances – galaxies: dwarf – galaxies: evolution.

1 INTRODUCTION

According to the hierarchical model of galaxy formation (White & Rees 1978; Kauffmann, White & Guiderdoni 1993; Cole et al. 2000), dwarf galaxies ($M_* \leq \sim 10^9 M_\odot$) have constituted the most numerous population of galaxies over cosmic time. Despite this, their small physical sizes and relatively faint brightness can cause them to be proportionally overlooked in observational studies due to difficulties either in detecting them or in distinguishing them from background sources (e.g. Disney 1976; Davies, Davies & Keenan 2016; Williams et al. 2016).

Low surface brightness (LSB) galaxies are typically dwarf galaxies in terms of both their stellar and halo mass (e.g. Prole et al. 2019), but have a much lower density of stars. Traditionally, they have been defined by having surface brightnesses at least one magnitude fainter than the night sky, around $22.5 \text{ mag arcsec}^{-1}$ in the g band. However, galaxies with much lower surface brightness than this are known to exist (e.g. McConnachie 2012; Mihos et al.

2015). LSB galaxies were first proposed to exist (Disney 1976) and identified several decades ago (Bothun et al. 1987; Impey, Bothun & Malin 1988) and have been a subject of scientific discussion ever since (e.g. Davies, Phillipps & Disney 1989; McGaugh et al. 1996; Dalcanton et al. 1997; Conselice, Gallagher & Wyse 2003; Sabatini et al. 2003; Roberts et al. 2007; Lelli, Fraternali & Sancisi 2010).

Recently, there has been a resurgence of interest in LSB galaxies thanks to the deep imaging of the Dragonfly telephoto array (Abraham & van Dokkum 2014), with which van Dokkum et al. (2015) discovered a high abundance of large LSB galaxies in the Coma Cluster. The authors coined the term ‘ultradiffuse galaxy’ or UDG for such objects, a name that has been widely adopted throughout the literature. Specifically, UDGs are defined as objects comparable in effective (half-light) radii to the Milky Way ($r_{e,r} > 1.5 \text{ kpc}$) but of much lower surface brightness ($\bar{\mu}_{e,r} > 24$, where μ denotes a surface brightness in magnitude per square arcsecond and $\bar{\mu}_e$ is the average surface brightness within the effective radius).

There has been much debate over the significance of UDGs and as to whether they make up a different population (in terms of their formation mechanism and thus intrinsic properties) to other, smaller low surface brightness galaxies. There is a growing consensus

★ E-mail: ProleD@Cardiff.ac.uk

that UDGs share a continuum of properties with less extreme galaxies in terms of star formation rate (Leisman et al. 2017), size and luminosity (Conselice 2018; Danieli & van Dokkum 2018), metallicity (e.g. Fensch et al. 2019), and perhaps also mass-to-light ratios (e.g. Prole et al. 2019); however, some UDGs may be genuine outliers in the stellar mass–halo mass plane and are devoid of dark matter (e.g. van Dokkum et al. 2018, 2019, but see also Trujillo et al. 2019), perhaps suggesting separate formation mechanisms.

It is likely that the popularity of UDGs among the literature is thanks in part to their large sizes; this makes them easier to identify against background objects in groups and clusters. Indeed, this property has been exploited by several authors in their studies of UDGs in such environments (e.g. Koda et al. 2015; Muñoz et al. 2015; van der Burg, Muzzin & Hoekstra 2016; Yagi et al. 2016; Janssens et al. 2017; van der Burg et al. 2017; Venhola et al. 2017; Mancera Piña et al. 2019; Zaritsky et al. 2019). This bias towards studies in dense environments is exacerbated by the difficulty of obtaining spectroscopic redshifts (and therefore distances) for large samples of LSB galaxies. However, one recent development suggests that it may be possible to estimate distances to such galaxies by exploiting the ubiquity of the globular cluster luminosity function (Román et al. 2019).¹

UDGs in clusters are typically on the red sequence (Koda et al. 2015; van der Burg et al. 2016; RS et al. 2019) and show little evidence for tidal interaction even close to the cluster centres (Mowla et al. 2017), suggesting relatively high mass-to-light ratios. There is tentative evidence that UDGs tend to be bluer towards the outskirts of galaxy groups (Román & Trujillo 2017b; Alabi et al. 2018; Zaritsky et al. 2019) (and more generally in lower-density environments; cf. Greco et al. 2018a,b), suggesting that interactions with the environment during in-fall can diminish star formation in UDGs. While this is not surprising if UDGs are quenched during in-fall, a separate analysis by Román & Trujillo (2017a) did not show a significant trend between environmental density and colour.

One outstanding question regarding UDGs is whether they are able to form more efficiently in dense environments like groups and clusters, or whether density plays a detrimental role in UDG formation/survival efficiency. van der Burg et al. (2017) find that UDGs are relatively more common in higher-mass environments, but this is in tension with other studies (Román & Trujillo 2017b; Mancera Piña et al. 2018) that came to the opposite conclusion.

Whatever the role of the environment in UDG production, several authors (e.g. van der Burg et al. 2016; Wittmann et al. 2017; Mancera Piña et al. 2018) have observed a relative dearth in their number density towards the centres of massive clusters. This suggests that in very high density regions, UDGs are either destroyed quickly or do not form as efficiently. One scenario suggested by Janssens et al. (2017) is that UDGs dissolve in cluster cores, possibly depositing ultracompact dwarf galaxies in the process.

UDGs can also be understood from a theoretical point of view. UDG formation scenarios can be broadly classified as ‘*in situ*’ (i.e. secular formation in the absence of interactions with an exterior body or bodies), or environmentally driven. Di Cintio et al. (2017) showed through zoom-in cosmological simulations that gas outflows caused by internal feedback processes can produce UDGs within dwarf-sized haloes. Further, Amorisco & Loeb (2016) argue that UDGs can form *in situ* both in the field and in cluster environments, and should be expected to do so as the high angular

momentum tail of the dwarf galaxy distribution. The importance of high-spin haloes for UDG production was also noted by Rong et al. (2017) in their simulations. Indeed, there is much observational evidence suggesting UDGs reside in dwarf-sized haloes (e.g. Beasley & Trujillo 2016; Amorisco et al. 2018; Lim et al. 2018; Prole et al. 2019).

However, there are several other feasible formation mechanisms that involve the transformation of normal dwarf galaxies to UDGs through environmental effects. One example is tidal heating, whereby galaxy–galaxy interactions cause an expansion of the dwarfs (Collins et al. 2013; Carleton et al. 2018). There are several pieces of observational evidence showing that some UDGs are associated with tidal interactions (Wittmann et al. 2017; Bennet et al. 2018). Continually, ram-pressure stripping from the dense intracluster medium in groups and clusters is perhaps able to produce UDGs by quenching early in-fall galaxies (Yozin & Bekki 2015). Jiang et al. (2019) argue that ram-pressure stripping is the primary effect that causes UDGs to lose gas (and therefore shut down star formation) in dense environments. This can account for the red colours observed for UDGs in clusters.

While much is known about UDGs in dense environments, relatively little is known about the field population,² expected to form preferentially through secular mechanisms³ (see e.g. the work of Graham et al. 2017 and Janz et al. 2017 regarding the evolution of early-type dwarf galaxies). This is mainly because of the difficulties involved in measuring distances to large samples of LSB galaxies without prior information such as cluster association. Observational studies of groups and clusters alone have been unable to disentangle the relative importance of *in situ* versus environment-driven formation because of the need to perform a statistical background subtraction of interloping (i.e. non-group or cluster) UDG candidates. Some studies (Das 2013; Leisman et al. 2017; Papastergis, Adams & Romanowsky 2017; Greco et al. 2018a; Zaritsky et al. 2019) have shown that a field population of UDGs does indeed exist, yet the global properties of these galaxies are poorly understood.

Two particularly relevant pieces of work are those of Leisman et al. (2017) and Jones et al. (2018), who have shown not only that H I-bearing UDGs exist in the field (as theoretically predicted by Di Cintio et al. 2017), but also that their number density is too high to be explained by an extrapolation of the empirical relation between the number of UDGs and M_{halo} measured by van der Burg et al. (2017). Further, their samples appear systematically bluer than anticipated for UDGs in the field when compared with semi-analytic models (Rong et al. 2017; Jones et al. 2018). However, this analysis was limited to H I-rich field UDGs and it is unknown how this population relates to the overall field UDG population.

In this work, we use deep wide-field imaging combined with an empirical UDG model to statistically constrain the global properties of UDGs in the field without knowing the distances to any of our sources. This includes an analysis of their colours, number density, and mass formation efficiency. The paper is structured as follows. We describe our data in Section 2. We describe our sample of UDG candidates in Section 3 and quantify our recovery efficiency. In Section 4 we describe our empirical model of UDGs and potential

¹Although it is not currently clear whether the globular cluster luminosity function is ubiquitous for UDGs (e.g. van Dokkum et al. 2018, 2019).

²We note that our working definition of the field is a representative piece of the Universe in which galaxy groups and clusters are included, but massive haloes naturally make up a relatively small fraction by mass.

³Although for this study we cannot rule out all external processes such as accretion from gas clouds.

interlopers in our UDG candidate sample. Our results are presented and discussed in Sections 5 and 6, respectively. We conclude in Section 7. All magnitudes are quoted in the AB magnitude system. Cosmological calculations are performed assuming lambda cold dark matter (Λ CDM) cosmology with $\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$, $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$.

2 DATA

For source detection and structural parameter estimation, we use a 180 deg^2 subset of data from the Astrowise (McFarland et al. 2011) reduction of the Kilo-Degree Survey (KiDS; de Jong et al. 2013; Kuijken et al. 2019) that overlaps with the GAMA spectroscopic survey (Driver et al. 2011) equatorial fields. We use the r band for source detection because it is the deepest and has the best image quality. This is the same data⁴ used by van der Burg et al. (2017) in their study of the UDG populations of galaxy groups and so we can make direct comparisons with their findings. Despite the GAMA overlap, redshift measurements are not available for most of our sources because they are generally much fainter than the limiting depth of GAMA at $19.8m_r$.

The pixel size of KiDS is 0.2 arcsec, small enough to properly sample the point spread function (PSF) that has a typical full width at half-maximum (FWHM) < 1 arcsec. The sky background is estimated in meshes of 20 arcsec that are median filtered over in 3×3 meshes.

While the KiDS r band is sufficient to reach a limiting surface brightness of $\bar{\mu}_{e,r} \sim 26.5$, we additionally use the first data release of the overlapping Hyper-Suprime-Cam Subaru Strategic Program (HSC-SSP; Aihara et al. 2018) to measure colours. The HSC-SSP data is around $0.5m_r$ deeper than KiDS (more so in the g band), but has a smaller overlapping footprint by about a quarter compared to the KiDS area we consider. This leaves us with $\sim 39 \text{ deg}^2$ of unmasked data from which we can measure HSC-SSP colours.

Compared to the 180 deg^2 KiDS-GAMA overlap we use here, the remaining footprint that we have HSC-SSP data for is fairly limited. This may make us partially sensitive to cosmic variance. However, we note that our footprint is spread uniformly over three GAMA regions (G09, G12, and G15), each separated by at least 26 degrees. Additionally, we can account for local galaxy groups and clusters using the GAMA group catalogue (Robotham et al. 2011; see Section 6.3). In the future, our analysis can be easily upscaled to larger footprints.

We note that we do not use the HSC-SSP for detection because of its limited footprint and because its background subtraction is more aggressive compared to KiDS (mesh grid of ~ 20 arcsec but with no median filtering over meshes), meaning that it could restrict the maximum angular size of sources we could measure accurately. For the present analysis, we restrict ourselves to the g and r bands but note that this can be expanded in future studies.

3 MEASUREMENTS

Since we do not know the distance to any of our sources (apart from a small subset; cf. Appendix D), we must rely on selection criteria defined in observable parameter space (i.e. that which we measure as projected on the 2D surface of the sky). Specifically, we target the LSB regime $24.0 \leq \bar{\mu}_{e,r} < 26.5$, where the lower limit is chosen for

consistency with the literature and the upper limit is defined by the depth of KiDS. The upper selection limit on \bar{r}_e is chosen to be much smaller than the spatial scale of the KiDS background subtraction and we set it as 8.0 arcsec in line with van der Burg et al. (2017). The lower limit on \bar{r}_e is more difficult to set; while technically we are limited by the size of the PSF (FWHM ~ 0.6 arcsec), it is also worth considering that the number of contaminant sources (i.e. non-UDGs) that satisfy our selection criteria quickly increases as this limit is lowered because of massive galaxies in the background. This issue is compounded by the fact that we do not have the advantage of a directly measurable background surface density compared to similar studies on groups and clusters. Here, we use a lower limit of $\bar{r}_e \geq 3$ arcsec for our selection (e.g. Sabatini et al. 2003; Davies et al. 2016). At a redshift of $z = 0.2$, $\bar{r}_e = 3$ arcsec corresponds to ~ 10 kpc. Our upper limit of $\bar{r}_e = 8$ arcsec corresponds to 1.6 kpc at $z = 0.01$.

3.1 Source detection and measurement

We choose to improve upon the catalogue used in van der Burg et al. (2017), who used SExtractor for source extraction, by using software optimized for the detection of LSB sources; this enables us to probe slightly deeper than their catalogue. We have experimented with several different detection and segmentation algorithms, including MTOBJECTS (Teeninga et al. 2016), PROFOUND (Robotham et al. 2018),⁵ and DEEPCAN (Prole et al. 2018).⁶ After some consideration, we selected MTOBJECTS as the most suitable for our analysis because it seemed to produce less spurious detections around large, bright galaxies in our pipeline. We note that during this work PROFOUND has been updated with an alternative segmentation algorithm that improves its reliability around such objects, but we have not tested this. We used default parameters from MTOBJECTS: $\alpha = 10^{-6}$ and `move_factor` = 0.5, where α sets the statistical significance level for the deblending and `move_factor` determines the spread of large objects.

We used the KiDS weight images to mask out regions in the data that have less than three exposures contributing to the imaging prior to the MTOBJECTS run. This was done to ensure uniform sensitivity over the full data set, as MTOBJECTS relies on a global estimate of the background distribution.

3.1.1 Point spread function measurement

We took advantage of our decision to split the KiDS frames into 3×3 subframes by making one PSF model per subframe (i.e. nine PSF models per square degree). This was accomplished by targeting point sources in the `R_e` and `mag` plane (MTOBJECTS estimates of the effective radius and total magnitude, respectively) based on our MTOBJECTS catalogues from each subframe. Point sources were required to have an axis ratio as estimated by MTOBJECTS greater than 0.9. We then fitted Moffat profiles to the individual point source candidates using GALFIT (Peng et al. 2002). Our final PSF model for each subframe was taken as the model corresponding to the mean Moffat FWHM after a sigma-clipping algorithm was applied to remove outliers. We measure a mean FWHM of 0.6 arcsec with a standard deviation of 0.1 arcsec over the full KiDS area that we use. We measure a median value of the Moffat β parameter of 2.2, with a standard deviation of 0.1.

⁴Although they use the THELI (Erben et al. 2013) KiDS reduction, the depth is essentially equivalent.

⁵<https://github.com/asgr/ProFound>

⁶<https://github.com/danjampro/DeepScan>

3.1.2 Source extraction pipeline

Following measurement of the PSF, our overall detection and measurement consists of the following steps.

(i) Use MTOBJECTS to produce a segmentation image and preliminary source catalogue for each subframe.

(ii) Apply a pre-selection to the preliminary catalogue to identify candidates suitable for input to GALFIT. This is necessary to ensure a practically feasible number of fits. Specifically, we required $R90 > 1.5$ arcsec and $\mu_{\text{mean}} > 23.5$ mag arcsec⁻², where $R90$ and μ_{mean} are a proxy for the radius containing 90 per cent of the galaxy light and the average surface brightness, respectively.

(iii) Use GALFIT to fit a combined Sérsic plus inclined sky plane model to each pre-selected source, ignoring masked pixels and additionally masking other segments from the MTOBJECTS segmentation images. Parameter estimates from MTOBJECTS were used as the initial guesses for GALFIT. The sky RMS is estimated directly from pixels in the cut-out region that were unmarked in the segmentation image. The Sérsic profile is defined as

$$I(r) = I_e \exp \left\{ -b_n \left[\left(\frac{r}{r_e} \right)^{\frac{1}{n}} - 1 \right] \right\} \quad (1)$$

where $I(r)$ is the galaxy's intensity as a function of radius, I_e is the intensity at the effective (half-light) radius r_e , and b_n is a constant determined only by the index n , which in turn governs the profile slope. We note that all conversions between Sérsic parameters are performed using the prescriptions of Graham & Driver (2005).

(iv) Apply the selection criteria to the resulting GALFIT models in order to produce a final catalogue of UDG candidates.

The PSF models were used as an input to GALFIT such that the measurements correspond to deconvolved parameters. The GALFIT cut-out size was 400 pixels, large enough to recover the intrinsic Sérsic parameters properly over the full range of parameter space we explore here, and was tested with the synthetic source injections (see Section 3.2). We note that it was important to include the inclined sky plane component in the GALFIT modelling in order to retrieve unbiased measurements in the cases of high- \bar{r}_e or high- n profiles.

Following the full set of pre-selections, we are left with $\sim 2 \times 10^6$ sources that are input to GALFIT over the full 180 degree squared KiDS area. Following the selection using GALFIT parameters ($24.0 \leq \bar{\mu}_{e,r} \leq 26.5$, 3.0 arcsec $\leq \bar{r}_{e,r} \leq 8.0$ arcsec, $n \leq 2.5$), we are left with 829 UDG candidates. After selecting sources also within the HSC-SSP footprint, our final catalogue of UDG candidates consists of 212 sources; some examples are shown in Appendix A. We note that contrary to UDGs in clusters and groups, our sample comprises sources that appear far more irregular, with features suggestive of active star formation.

3.2 Recovery efficiency

We define the recovery efficiency ϵ as the fraction of sources with intrinsic observable parameters (i.e. without the effects of measurement uncertainty) that have measurements that meet our selection criteria. As such, sources that do not meet our selection criteria in terms of their intrinsic observable parameters may be selected ($\epsilon > 0$) because of measurement uncertainty. Anticorrelated with ϵ is therefore the selection purity, defined as the fraction of detections with intrinsic observable parameters that do not meet the selection criteria but have measured properties that do and thus make it into the final catalogue of UDG candidates. This is different compared to the purity of the UDG candidates, which is defined

Table 1. Parameter ranges for the artificial galaxy injections. The parameter realizations are drawn uniformly within the ranges, which are much wider than our selection criteria for UDG candidates.

| Parameter | Lower limit | Upper limit |
|--|-------------|-------------|
| $\bar{\mu}_e$ (mag arcsec ²) | 22 | 30 |
| \bar{r}_e (arcsec) | 1 | 25 |
| q | 0.1 | 1 |
| n | 0.2 | 2.5 |

as the fraction of sources in the UDG candidate catalogue that are intrinsically UDGs as defined by their physical properties.

We have used synthetic source injections to quantify ϵ as a function of intrinsic Sérsic parameters, $\epsilon(\bar{\mu}_{e,\text{int}}, \bar{r}_{e,\text{int}})$. To do this, we create mock images by inserting artificial galaxy profiles (PSF-convolved, one-component Sérsic) into each frame of the real KiDS data and run them through the full detection and measurement pipeline described in Section 3.1. The Sérsic parameters were drawn uniformly from the ranges presented in Table 1, where q is the observed axis ratio. As noted by van der Burg et al. (2016), \bar{r}_e (the circularized half-light radius) is robust against the intrinsic distribution of axis ratios and $\bar{\mu}_e$ is a better indicator of a source's detectability than other parameters such as the central surface brightness. Note that we do not include the Sérsic index n as a free parameter in $\epsilon(\bar{\mu}_{e,\text{int}}, \bar{r}_{e,\text{int}})$ in order to simplify the analysis. This does not severely impact our results because the intrinsic range in n for UDGs is narrow (e.g. Koda et al. 2015 find a mean of $n \sim 1$ with a standard deviation of 0.34).

The sources were injected at a surface density low enough to ensure that injected profiles were separated on average by 6.5 times the maximum value of \bar{r}_e given in Table 1. We repeated the process several times in order to increase the number statistics for the $\epsilon(\bar{\mu}_{e,\text{int}}, \bar{r}_{e,\text{int}})$ measurement, simulating ~ 735 000 sources overall.

We only considered unmasked sources for the estimate of $\epsilon(\bar{\mu}_{e,\text{int}}, \bar{r}_{e,\text{int}})$, which was measured with ~ 550 000 artificial galaxy injections spread evenly over our full KiDS subset. We show our ability to precisely measure the intrinsic parameters of our injected sources in Fig. 1. Our fiducial measurement of $\epsilon(\bar{\mu}_{e,\text{int}}, \bar{r}_{e,\text{int}})$ is shown in Fig. 2.

We note that we additionally tested IMFIT (Erwin 2015) in place of GALFIT in the pipeline described in Section 3.1 and found that it made no significant difference to our measurement of $\epsilon(\bar{\mu}_{e,\text{int}}, \bar{r}_{e,\text{int}})$.

A criticism of the above method is that it relies on field galaxies being well fitted by Sérsic profiles. While this is certainly justified in dense environments where UDGs show little evidence of tidal features (Mowla et al. 2017), it is not clear whether this is justified for the field population. Since field galaxies are expected to be relatively isolated, they are likely to show little evidence of tidal disruption. This does not however rule out irregular morphologies caused by bright star-forming regions and secular processes like stellar feedback from e.g. supernovae.

In addition to the injections described above and in equal measure, we inject nucleated profiles (Sérsic + Moffat PSF model) to the data assuming that approximately 1 per cent of the galaxy light is contained within the nucleus. This allows us to quantify any systematic differences in our recovery efficiency that might be caused by the presence of nuclei and adapt our selection criteria accordingly. We note that we do not attempt to fit nucleated profiles for our measurements because experiments with artificial galaxy profiles showed that the fits were not reliable. We find that the presence of a nucleus is sufficient to positively bias recovered values of the

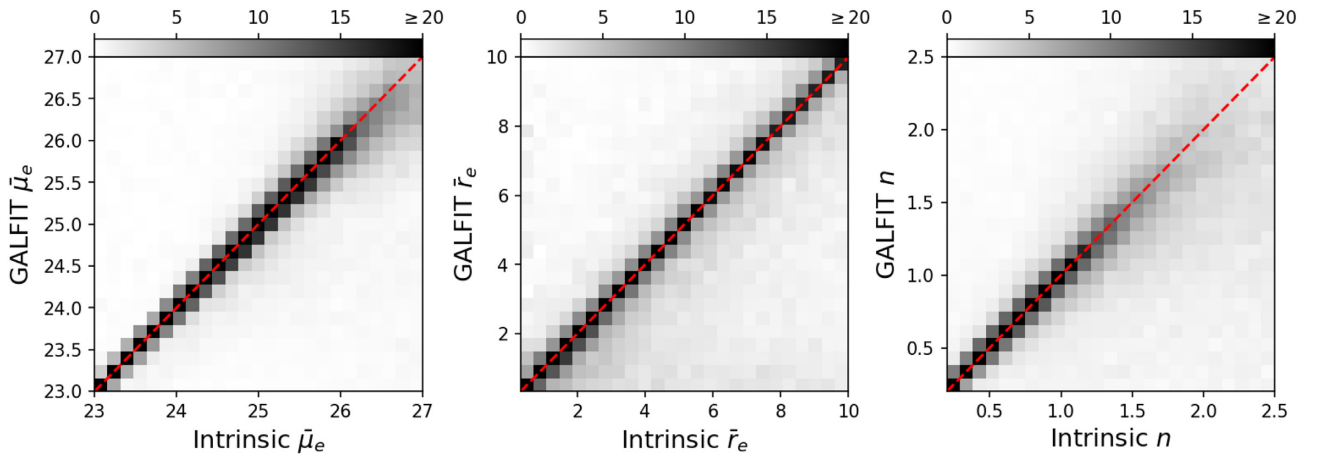


Figure 1. This figure shows how the measured GALFIT parameters compare with the intrinsic Sérsic parameters of the synthetic galaxies that we inject into the data. The black 2D histograms show this data for our sample of recovered injections. The red dashed line is the one-to-one relation and is not a fit. Clearly we are able to recover the intrinsic parameters with good accuracy and precision over the range of parameter space that we are interested in. However, our precision diminishes slightly when recovering high- n (Sérsic index) sources, but we retain accuracy such that there is little bias in the recovered parameter estimates. The units of the histogram are the percentage of sources with intrinsic parameters (given by the column) that occupy a particular bin in measured parameter space. The colour map is capped at 20 per cent in order to increase the contrast for bins of lower completeness.

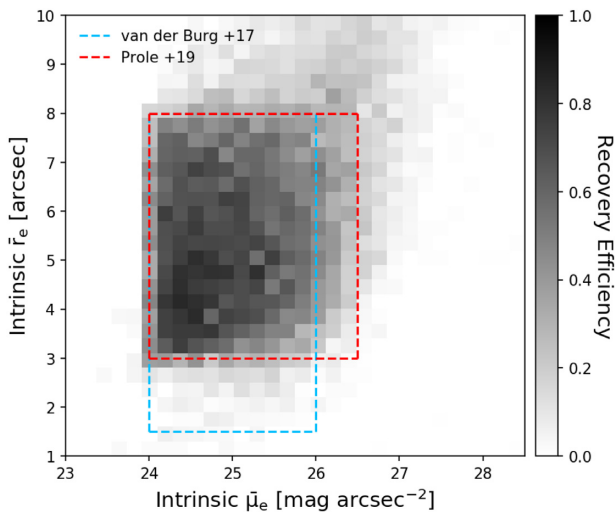


Figure 2. The recovery efficiency of synthetic sources injected into the data as a function of circularized effective radius and mean surface brightness within the effective radius given our selection criteria. The red box indicates our selection criteria, while the blue box is that used in van der Burg et al. (2017).

Sérsic index by approximately 30 per cent at $n = 1$. Since almost all recorded UDGs have Sérsic indices less than around 1.5 (there is both observational and theoretical evidence for this; e.g. Koda et al. 2015; Di Cintio et al. 2017; Román & Trujillo 2017b; Venhola et al. 2017), our upper selection limit of $n = 2.5$ is justified. The effect of the selection in n on our results is discussed further in Section 5.2.

In comparison to van der Burg et al. (2017), our selection criteria allow us to probe ~ 0.5 mag deeper in surface brightness. The increased depth comes in part from the proficiency of MTOBJECTS over SEXTRACTOR for detecting LSB sources. We also use a higher cut in \bar{r}_e ; since we do not have the advantage of a measurable background level, imposing a higher minimum cut in \bar{r}_e allows us to mitigate against an excessively contaminated sample of UDG candidates. In Fig. 2, it can be seen that we expect some

contamination from (apparently) large, faint sources that do not intrinsically meet our selection criteria (top right of the figure). However, since the number of apparently large sources is very small compared to the number of smaller ones, the decrease in purity from such sources is negligible.

3.3 Colours

We exploit the overlap of the KiDS survey with the HSC-SSP footprint in order to measure $(g - r)$ colours for our UDG candidates. We remind the reader that while HSC-SPP is ~ 0.5 mag deeper than KiDS, the background subtraction in HSC-SPP is slightly more aggressive.

We use an aperture-based strategy to measure $(g - r)$ colours. Specifically, all colours are measured within the $1\bar{r}_{e,r}$ ellipses from our GALFIT measurements. We estimate the sky level along with its uncertainty using a random aperture approach, whereby we place 100 equally shaped apertures in close vicinity to (but not touching) the source. Before measuring the median background level and its uncertainty, fluxes are sigma-clipped at $2\sigma^7$ in order to lower the potential for overestimating the background level because of nearby sources. We do not perform additional aperture corrections because our sources are much more extended than the PSF.

Due to the increased depth offered by HSC-SSP, we are able to measure positive fluxes in the g and r bands for close to 100 per cent of our sources. The typical measurement error in $(g - r)$ due to the background fluctuations is approximately 0.04 mag; this is discussed further in Section 4.3. We show comparisons between KiDS and HSC-SSP imaging in Fig. A1.

4 THE EMPIRICAL MODEL

Without knowing the distances to any of our sources, it is difficult to tell how many are intrinsically UDGs and how many are

⁷The bias in the recovered standard deviation when sigma-clipping normally distributed data at 2σ is approximately 25 per cent; we therefore correct our estimates by this factor.

cosmologically dimmed background galaxies. In this section, we describe an empirical model that can be used to generate a synthetic population of UDGs in order to compare with observation. This is supported by an additional model for massive galaxies that allows us to estimate the number of non-UDG contaminants in our observational sample.

4.1 Empirical UDG model

4.1.1 Empirical properties of UDGs

One of the simplest models can be created by assuming field UDGs share similar empirical properties with UDGs in clusters. Of importance for our analysis are prescriptions for $\bar{\mu}_{e,\text{int}}$ (the surface brightness corrected for cosmological projection effects), $\bar{r}_{e,\text{phys}}$ (physical size), and $(g - r)_{\text{int}}$, the rest-frame colour. As discussed in Section 3.2, we assume that all UDGs occupy the range of 0.2 to 2.5 in Sérsic index.

van der Burg et al. (2016) recorded that the distribution of average surface brightness $\bar{\mu}_e$ is approximately uniform in group environments for UDGs. This has been complimented by the findings of Danieli & van Dokkum (2018), who found that the distribution of absolute magnitude at a fixed size is approximately uniform for large, red galaxies in the Coma Cluster after accounting for the newly discovered UDGs. These two observations are equivalent, since at a given size the mean surface brightness is uniquely defined by its magnitude (i.e. there is no dependence on Sérsic index). However, the work of Danieli & van Dokkum (2018) showed that this relation extends from the low surface brightness regime to much brighter galaxies. We therefore adopt a uniform distribution $U_{x,\text{min}}^{x,\text{max}}(x)$ for $\bar{\mu}_{e,\text{int}}$:

$$\bar{\mu}_{e,\text{int}} \sim U_{24.0}^{26.5}(\bar{\mu}_{e,\text{int}}). \quad (2)$$

The subsequent observational study of van der Burg et al. (2017) (supported theoretically by Carleton et al. 2018) has shown that the size distribution of UDGs in groups and clusters is well described by a power law of slope -2.71 ± 0.33 in logarithmic size bins, such that smaller UDGs are much more common than larger ones. The intrinsic distribution of physical sizes in kpc is therefore taken as

$$\bar{r}_{e,\text{phys}}[\text{dex}] \sim \bar{r}_{e,\text{phys}}^{-2.71}, \quad (3)$$

where we assume the range of $\bar{r}_{e,\text{phys}}$ lies between 1.5 and 7.0 kpc, consistent with van der Burg et al. (2017). We probe the effect of varying the power-law slope on our result in Appendix B.

It has been noticed by several authors that UDGs in clusters tend to lie on the red sequence (e.g. Koda et al. 2015; van der Burg et al. 2016) and this is also expected theoretically. There have been hints that UDGs may tend to be much bluer in less dense environments (Román & Trujillo 2017b; Jiang et al. 2019; Jones et al. 2018), although this is not always clear from an observational perspective (Román & Trujillo 2017a). As such, colours of field UDGs remain relatively poorly understood. We therefore leave the distribution of $(g - r)_{\text{int}}$ as a variable of our model, and discuss it further in Section 4.1.4.

4.1.2 Estimated number density of UDGs

We use the (almost linear) empirical relation between the number of UDGs and the mass of their parent halo measured by van der Burg et al. (2017) to estimate the formation efficiency of UDGs per unit mass in clusters and groups. From this, we can calculate the

total number of UDGs that should exist out to redshift z_{max} , using our cosmological model to estimate the total enclosed mass. We calculate the total mass M_{tot} contained within the volume V probed by solid angle ω out to z_{max} using equation (4),

$$M_{\text{tot}} = \int_0^{z_{\text{max}}} \Omega_m(z) \rho_{\text{crit}}(z) \frac{dV}{dz d\omega} dz d\omega \quad (4)$$

where Ω_m is the fractional contribution of matter to ρ_{crit} , the critical density of the Universe. We additionally assume that the UDGs are spatially distributed smoothly according to the integrand of equation (4), such that the redshift distribution of field UDGs follows the mass.

Note that the adopted value of z_{max} does not impact the result, provided that the number of sources we predict to observe out to z (given our recovery efficiency) has converged, i.e. has stopped increasing, before z_{max} . For our modelling we use $z_{\text{max}} = 1$, which meets this criterion (see Section 5.1).

If we assume that UDGs form with an average efficiency equivalent to a $10^{15} M_{\odot}$ cluster according to equation (1) of van der Burg et al. (2017), we derive a volume density of $\sim 9 \times 10^{-3} \text{ cMpc}^{-3}$. This is a factor of 6 higher than the total number density of H I-bearing UDGs measured by Jones et al. (2018), who measured a value of $1.5 \pm 0.6 \times 10^{-3} \text{ cMpc}^{-3}$. Using a value different from $10^{15} M_{\odot}$ for the halo mass would not strongly modify the initial number density estimate since the slope of the relation between M_{halo} and the number of UDGs hosted by the halo is approximately 1 (at least down to $M_{\text{halo}} \sim 10^{12} M_{\odot}$; Román & Trujillo 2017b; van der Burg et al. 2017; Mancera Piña et al. 2018). However, by selecting a halo mass of $10^{15} M_{\odot}$, we are essentially comparing the field abundance with that in a $10^{15} M_{\odot}$ cluster in our later analysis. We estimate the impact that the uncertainty in the van der Burg et al. (2017) relation has on our result in Appendix B.

4.1.3 Accounting for cosmological effects

We account for the cosmological distance modulus, angular diameter distance d_a (describing how physical sizes map to angular sizes as a function of the redshift, z), and k -corrections [the filter and spectral energy distribution (SED) dependent effect that modifies a source's apparent brightness with z , independently from the distance modulus; Hogg et al. 2002]. In combination, these quantities allow us to project the surface brightnesses and angular sizes of our mock sources out to a certain redshift.

While d_a is simple to account for, the exact k -correction depends on the assumed SED for the UDGs. Quiescent UDGs are thought to be old, metal-poor galaxies (e.g. Fensch et al. 2019; Ferré-Mateu et al. 2018; Ruiz-Lara et al. 2018). We adopt the average UDG properties from Ferré-Mateu et al. (2018) to estimate the k -corrections for such galaxies, namely an age of 6.7 Gyr and $[Z/H] = -0.66$. In the case of star-forming UDGs, we assume the same age and metallicity as in the quiescent model, but introduce star formation at a uniform rate until the time of observation. While this is an idealized scenario, we probe the significance on the assumed model for k -corrections in Appendix B. All stellar population models and k -correction estimates are calculated using the Flexible Stellar Population Synthesis (FSPS; Conroy, Gunn & White 2009; Conroy & Gunn 2010) code. For the KiDS r band, we assume SDSS-like filters for the k -correction estimates. For the HSC-SSP colours, we use Subaru Suprime Cam filters.

We construct mock catalogues by sampling intrinsic parameters from the appropriate distributions. Following this, we convert the

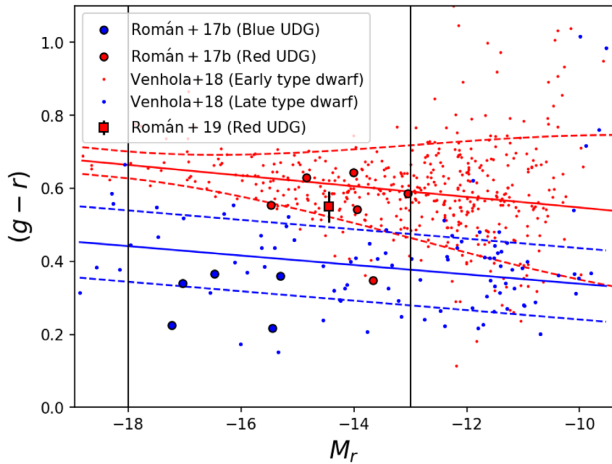


Figure 3. Our model colour distributions for the mock UDG catalogues. The data we fit are from Venhola et al. (2018) and correspond to late-type (small blue points) and early-type (small red points) dwarf galaxies in the Fornax cluster. We fit simple linear models (uninterrupted coloured lines) and their 1σ uncertainties (dashed lines) after clipping outliers at 3σ . Under the assumption that UDGs have similar colours to dwarf galaxies, we use the late/early-type fits for our star-forming/quiescent mock UDG catalogues. The vertical black lines span the approximate range of absolute magnitudes occupied by UDGs (van der Burg et al. 2016). The bold points show the red/blue UDGs of Román & Trujillo (2017b) and Román et al. (2019).

units into apparent observed quantities through the cosmological distance modulus, angular diameter distance (for the angular sizes), and band-specific k -corrections.

4.1.4 UDG colour models

The inclusion of colour into our analysis is critical because the colour of a galaxy contains some information about its distance thanks to the cosmological redshifting of spectroscopic features. While we have assumed a stellar population model for the UDGs in order to estimate the k -corrections, we cannot use these models to assign colours to our mock catalogue because of the need to include some intrinsic scatter. One alternative approach is to model the colours using measurements from the literature.

Several authors have shown that UDGs occupy the red sequence in clusters (e.g. Koda et al. 2015; van der Burg et al. 2016) and this is also supported theoretically (Rong et al. 2017). However, modelling the colour distribution of star-forming UDGs is slightly harder because there is not as much available data for them. Since UDGs have stellar populations similar to dwarf galaxies (e.g. Fensch et al. 2019), one viable method is to assume that star-forming UDGs have colours similar to late-type dwarf galaxies.

Venhola et al. (2018) have measured the $(g-r)$ colours as a function of absolute magnitude for such galaxies in the Fornax cluster. Using these measurements, it is possible to fit the relationship between colour and absolute magnitude with a simple linear model separately to each of their early- and late-type samples. For the late-type galaxies, we use a constant scatter term, while we interpolate the standard deviation of the colours in bins of absolute magnitude for the early-type galaxies (ETGs). We show the corresponding fits in Fig. 3, where we have clipped outliers at 3σ . We note that our fit to the early-type dwarf galaxies is consistent with the approximate fit to the red-sequence UDGs in clusters from van der Burg et al. (2016).

In Fig. 3 we also compare with the UDGs discovered by Román & Trujillo (2017b),⁸ which have been decomposed into red and blue populations based on their $(g-i)$ colour. While their red population is fairly consistent with our red-sequence model, the blue UDGs seem to be systematically bluer than our colour model for blue galaxies. This may be explained by the fact that our model is based on measurements from the Fornax galaxy cluster where environmental processes, for example ram-pressure stripping, may cause reddening of the galaxies. In comparison, the UDGs of Román & Trujillo (2017b) are found in isolated galaxy groups where such effects are less prolific.

4.2 Empirical model for massive galaxies

Not all of the UDG candidates in our observational sample are intrinsically UDGs. As large, bright galaxies are shifted towards higher redshift, they become both fainter and smaller in terms of their angular size and may eventually satisfy our selection criteria. Equally, small foreground dwarf galaxies not meeting the UDG criteria have the potential to contaminate the sample. Since we are not able to directly measure the number of these interlopers (as can easily be done when considering a group or cluster environment), we are forced to use empirical relations from the literature to estimate the level of contamination.

It is standard practice to broadly categorize galaxies as either late type or early type based on their morphology and/or colour (e.g. Bell et al. 2003; Baldry et al. 2004; Driver et al. 2006; Taylor et al. 2015). Massive ETGs are typically quiescent and therefore redder than late types. Additionally, massive ETGs generally have higher Sérsic indices compared to late types. For ETGs, the Sérsic index increases with total stellar mass (e.g. Caon, Capaccioli & D’Onofrio 1993; Graham et al. 1996; Danieli & van Dokkum 2018). While early-type dwarf galaxies exist with Sérsic indices around 1 (e.g. Prole et al. 2018), we are probing the field population and therefore expect that the main contribution from ETG interlopers will be from higher-mass galaxies with correspondingly higher Sérsic indices. We discriminate against recovering such objects in our UDG candidate sample through the upper-limit cut in Sérsic index at $n = 2.5$. As such, we expect the dominant source of contamination in terms of massive galaxies ($M_* > 10^9 M_\odot$) to be mainly constituted of massive late-type galaxies. By contrast to massive ETGs, late-type galaxies are systematically bluer, with Sérsic indices $n < 2.5$ (Vulcani et al. 2014).

4.2.1 Canonical empirical distributions

While relatively little is known about the population of field galaxies with stellar masses lower than around $10^9 M_\odot$, much is known about objects at higher stellar masses. We can therefore produce mock observational catalogues of high-stellar-mass galaxies and use them to estimate the contamination level in our actual observed catalogue. Such an estimate would be naturally conservative owing to the truncation at $10^9 M_\odot$ that essentially excludes all dwarf galaxies, including UDGs (e.g. Prole et al. 2019). The ingredients of our model catalogues are:

(i) The stellar mass function (SMF). We have used measurements from GAMA (Baldry et al. 2012) and COSMOS/UltraVISTA (Muzzin et al. 2013) to model the galaxy SMF of field galaxies,

⁸Where the $(g-r)$ colours have been kindly provided by Javier Román.

including its redshift dependence between $z = 0$ and $z = 1$. These measurements are additionally decomposed into red and blue galaxy populations and we have incorporated this into our mock catalogues.

(ii) The stellar mass to size relation. We used the measurements of van der Wel et al. (2014) to assign effective circularized radii to each of our random samples of stellar masses, as a function of redshift. Again, we make use of their early/late-type colour decomposition. We also include scatter in the sampling based on their measurements.

(iii) We assign $(g - r)$ colours to our sampled galaxies as a function of their stellar mass by using empirical data gathered by the GAMA survey⁹ (Taylor et al. 2011). Specifically, we bin their $(g - r)$ measurements in stellar mass and assign intrinsic (i.e. observed at $z = 0$) colours to our mock observations in each bin by randomly sampling from the corresponding GAMA $(g - r)$ measurements.

(iv) We calculate k -corrections in the same way as described in Section 4.1.3, this time assuming an onset of star formation 8.9 Gyr ago and $[Z/H] = -0.38$, with a uniform star formation rate. These values are based on the high surface brightness, late-type sample of Ferré-Mateu et al. (2018). The effects of modifying this are discussed in Appendix B.

(v) As before, we assume the redshift distribution of our catalogue to be smoothly distributed with the mass in the Universe, according to equation (4).

Clearly such an exercise is approximate in nature, and is designed only to get a first-order estimate of the number of contaminants in our UDG sample. A discussion regarding the contribution to our observations from LSB galaxies smaller than the fiducial UDG limit of $\bar{r}_e = 1.5$ kpc can be found in Appendix C.

Many of the mock massive galaxies are brighter than $m_r = 19.8$. This means that it is possible to compare the redshift distribution of our mock catalogue with that of the GAMA spectroscopic survey. We show in Appendix D that our mock catalogues are consistent with that observed by GAMA.

4.3 Measurement errors

Before the catalogues can be directly compared with our observations, it is important to consider the effect of measurement errors on the predicted distributions of observed parameters. Of particular importance is the uncertainty in \bar{r}_e , which increases for larger and fainter galaxies. This is significant because there is typically a steep gradient in the distribution of \bar{r}_e , whereby there are far less large objects than small ones, both in terms of physical and angular size. Thus, including the measurement error in the mock catalogues causes an increase in the predicted number of galaxies observed with large angular sizes.

The measurement uncertainty on the Sérsic parameters is estimated directly from the synthetic source injections described in Section 3.2. Measurements of the injected sources are used to estimate the recovery efficiency, defined in intrinsic observable parameter space. Consequently, the effects of measurement error (including any bias) as a function of intrinsic size and surface brightness are already contained in the recovery efficiency estimate. We can therefore account for the effect of the measurement uncertainty in our mock catalogues by using the recovery efficiency to assign probabilities of detection (see Section 5.1). The limitation of this

approach is that we cannot directly compare structural parameters in our mock catalogues with the observations.

Also of importance is the measurement error in $(g - r)$ colour. Starting from our estimates of fluxes and their errors described in Section 3.3, we perform Monte Carlo realizations of flux ratios in order to estimate the distribution of uncertainties in the magnitude. We fit a lognormal distribution to the result, and use it to randomly sample uncertainties in colour; we then ‘jiggle’ (randomly perturb within error) the colours in the mock catalogues according to the result. The mean uncertainty in colour is ~ 0.04 mag.

5 RESULTS

5.1 Observations versus model predictions

We are now in a position to compare our mock catalogues with the observations. We note that for this analysis, all absolute numbers are normalized to an area of 180 square degrees. Each source in our mock catalogues is assigned a probability of recovery using the recovery efficiency discussed in Section 3.2, which are used as weights in the analysis. We note that after using such weights, the number of UDGs we predict to observe converges (i.e. does not increase further) by $z \sim 0.2$ (see Fig. D1). Similarly, the mock massive galaxy catalogue converges by $z \sim 0.5$. This is mainly an effect of the lower limit angular size cut at $\bar{r}_e \geq 3$ arcsec. We probe the accuracy of our modelling with reference to the redshift distribution of GAMA spectroscopic sources in Appendix D.

We compare the $(g - r)$ histogram of our observed UDG candidate catalogue with each of our mock catalogues (red UDGs, blue UDGs, massive blue galaxies) in Fig. 4, the results of which are fairly striking. Clearly either the assumption that all UDGs are on the red sequence as they are in clusters is not correct (as made clear by the significant offset between the peaks of the observed and predicted distributions) or UDGs in general do not form in the field with a mass efficiency anywhere near what they do in clusters. However, since we already know that blue UDGs do exist in abundance in the field (e.g. Leisman et al. 2017; Jones et al. 2018), it is clear that we can rule the latter hypothesis out completely. From this result we would expect isolated red UDGs, like the ones found by Martínez-Delgado et al. (2016) and Román et al. (2019), to be relatively rare.

The discrepancy is further compounded if one considers the estimates for the massive blue galaxy interlopers. We argue that since the massive blue galaxies are the dominant source of contamination in our UDG catalogue (see Section 4.2 and Appendix C), we can obtain an observational sample that is representative of the UDG population by statistically subtracting the massive blue galaxy population from the observed catalogue of UDG candidates. The result is displayed in the left-hand panel of Fig. 5, along with our mock UDG catalogues.

It is clear that the mock blue UDG catalogue is in much better agreement with the observed colour distribution than the red UDG catalogue. However, the observations are ~ 0.05 mag bluer than our empirical models predict. This means that the colours may be more consistent with the blue UDGs of Román & Trujillo (2017b) (see Fig. 3). This is not particularly surprising; late-type galaxies in clusters are typically redder than those in the field because of environmental quenching from e.g. ram-pressure stripping.

By comparing our mock catalogues to the interloper-corrected observations, it is possible to estimate the total field density of UDGs, along with a corresponding mass formation efficiency. This is accomplished by comparing the predicted number of UDGs from

⁹Specifically, we use the public `StellarMasses_v19` catalogue available from <http://www.gama-survey.org/dr3/>.

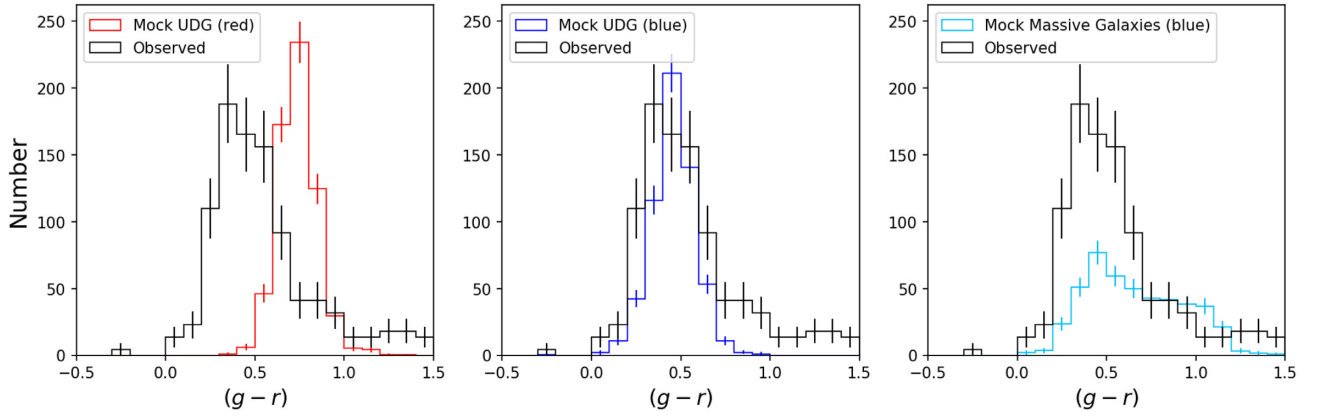


Figure 4. Synthetic distributions of $(g-r)$ colour for our mock red UDG, blue UDG, and massive blue galaxy catalogues, weighted by the probability of observation, compared to the actual observed histogram. The absolute numbers are normalized to an area of 180 square degrees. The error bars show the Poisson uncertainties in each bin. Colours are in the observed reference frame. We note that we include the effect of measurement error in our mock colours. It is clear that the red UDG model is not consistent with the observations, being much more consistent with the blue model.

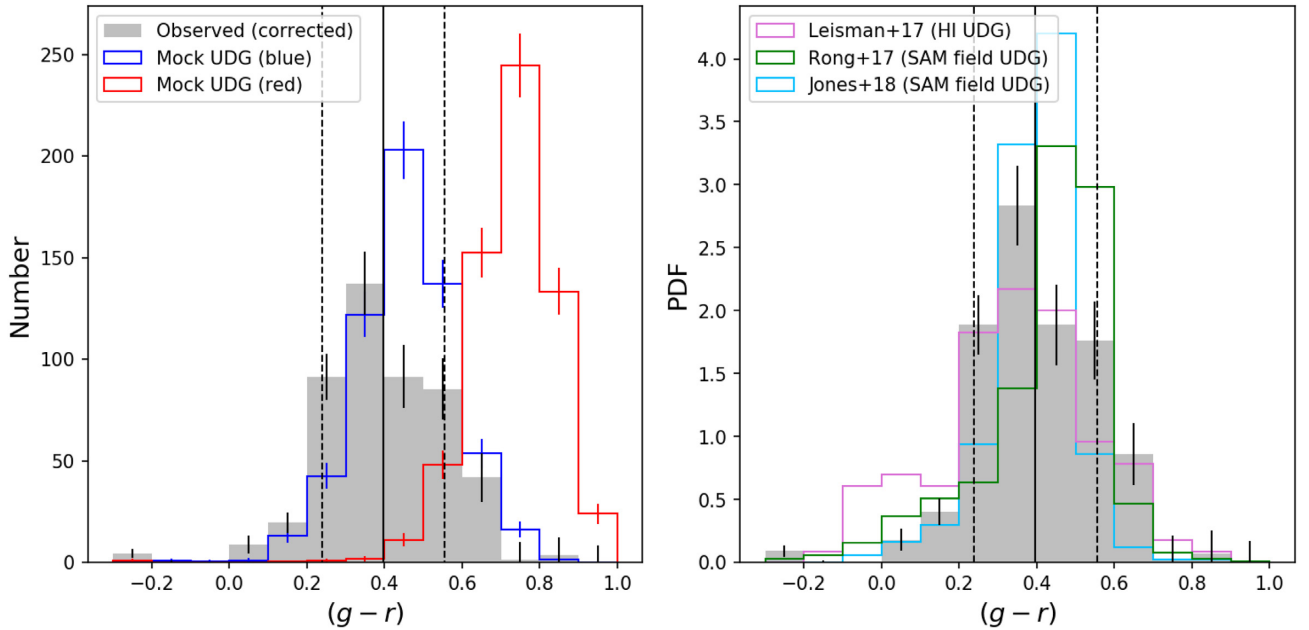


Figure 5. The observed distribution of colour after subtracting the estimated contribution from massive blue galaxies (grey histogram). *Left:* comparison with the empirical red and blue UDG models from this study. We show the mean and 1σ dispersion of the observations with the vertical lines. *Right:* normalized comparison with observations of HI-bearing field UDGs in the literature (Leisman et al. 2017) and predictions from the semi-analytical models (SAM) of Rong et al. (2017) and Jones et al. (2018). Poisson error-bars are shown. See text for discussion.

our empirical model with the estimated number of observed UDGs. From the appearance of Fig. 5 (left-hand panel), it is clear that we have overestimated the number density (and therefore mass formation efficiency) of UDGs in our model.

For the empirical model to predict the correct number of UDGs, we would require a mass formation efficiency $\sim 0.8 \pm 0.2$ times what it is in clusters, taking into account uncertainties discussed in Appendix B. This translates into a field density of $8 \pm 3 \times 10^{-3} \text{ cMpc}^{-3}$. This is an upper limit on the true field abundance of UDGs because the estimated number of observed UDGs is likely an overestimate; we have only considered massive blue galaxies as contaminant sources. We note that these estimates apply only to the range of physical parameters that we have probed here, i.e. sizes in the range $1.5 \leq \bar{r}_{e,r} [\text{kpc}] \leq 7.0$ and intrinsic (i.e. not cosmologically dimmed) surface brightnesses spanning $24.0 \leq$

$\bar{\mu}_{e,r} \leq 26.5$. If we were to consider even fainter sources, this number density would likely increase.

5.2 Dependence on Sérsic index

UDGs typically have Sérsic indices $n < 1.5$. The justification for our cut at $n = 2.5$ is as follows.

- (i) Our measurements are conflated by measurement error, which gets worse as a function of surface brightness. In Fig. 1 we show that we are never the less able to recover essentially all non-nucleated profiles with $n < 1.5$ by imposing a cut at $n \sim 2$.
- (ii) Some UDGs are nucleated (e.g. Venhola et al. 2017). In this analysis, we have only fitted single Sérsic profiles. The presence of

a nucleus can bias our recovery of n by +30 per cent, so a higher cut than $n = 2$ is justified to preserve completeness.

(iii) We do not explicitly include the Sérsic index distribution in our empirical model for background interlopers,¹⁰ which is statistically subtracted from the observational sample in the analysis. It is typical in the literature to take $n = 2.5$ as the dividing line between ‘early’ and ‘late’ type samples (e.g. van der Wel 2008; Vulcani et al. 2014; Vika et al. 2015). Since we do not want to oversubtract the interloper population, it is important to use a consistent cut for the sample selection.

(iv) In this analysis, we are striving to place upper limits on quantities like the UDG field number density; this is motivated by the fact that our interloper subtraction is likely incomplete. Lowering our Sérsic index selection cut would reduce the size of the observational sample and thus lower the inferred number density. In the interests of upper limits, it is therefore prudent to keep a relatively high cut in n .

(v) Our results are closely compared with the work of van der Burg et al. (2016), who used an even higher cut at $n = 4$.

In summary, while the cut at $n = 2.5$ might be relatively high compared to the observed values of n for UDGs, we account for the resulting contamination in our observational sample using the empirical model. However, it is important to discuss the effects of varying the index cut on our results. Recall that with the cut at $n = 2.5$, the upper-limit mass formation efficiency is estimated to be $\sim 0.8 \pm 0.2$ times that in clusters. If we instead take the cut at $n = 2.0$, this drops to ~ 50 per cent of the value for clusters. Here we have likely increased the purity of UDGs in our sample, but for the reasons given above it is difficult to quantify the effect on the completeness. If instead we drop to $n = 1.5$, the formation efficiency estimate drops to ~ 30 per cent of its value in clusters. However, it is likely that this value suffers from significant completeness effects and is an underestimate.

6 DISCUSSION

6.1 Comparison with H I-bearing UDGs

In this section, we compare our observed, contaminant-corrected ($g - r$) histogram to other measurements from the literature. We do not consider values of $(g - r) > 1$ because they are almost certainly not part of the UDG population. One catalogue that we can directly compare with is that of Leisman et al. (2017),¹¹ who measured the colours of isolated H I-bearing UDGs using SDSS data. While these measurements are conflated with measurement error because of the limited depth of SDSS, we can perform a qualitative comparison between the reported results (Fig. 5).

In Fig. 5, we show how the colour distribution of our corrected observational sample of UDG candidates compares with that of Leisman et al. (2017). The two PDFs are very similar, providing an indication that UDGs in the field are predominantly blue independently of the colour models we assumed in Section 4.1.4. There are some differences between the two distributions: We observe slightly more UDG candidates on the red side of the peak. There are several possible explanations: They could be sources that Leisman et al. (2017) were not sensitive to thanks to low

H I content, or they are contaminant objects that we have not properly accounted for in our UDG sample such as massive early-type galaxies. Alternatively, since we are sensitive to UDGs out to $z \sim 0.2$ comparing to their maximum distance of 120 Mpc ($z \sim 0.03$), it may be that k -corrections play a role. The Leisman et al. (2017) catalogue also seems to have an excess of blue UDGs compared to what we observe. This could be explained either by the measurement error arising from the limited SDSS depth or perhaps because blue UDGs are intrinsically brighter and we miss them in our selection (Leisman et al. 2017 use a slightly brighter bright-end selection cut).

Jones et al. (2018) used the Leisman et al. (2017) catalogue to estimate the field density of H I-bearing UDGs, obtaining a value of $1.5 \pm 0.6 \times 10^{-3} \text{ cMpc}^{-3}$. This is approximately one-fifth of our upper-limit estimate of the overall UDG field density. However, comparing such field densities is difficult because the limited depth of the SDSS imaging used by Leisman et al. (2017) to identify UDGs creates significant measurement uncertainty, blurring their selection boundaries and leading to an uncertainty of ~ 25 per cent in their sample size. However, using the fact that our estimate of the UDG density is an upper limit, we can estimate that H I-bearing UDGs comprise at least one-fifth of the overall population.

6.2 Comparison with semi-analytic models

We also compare our results with the work of Jones et al. (2018), who used the Santa Cruz semi-analytic model (SAM; Somerville, Popping & Trager 2015) to generate a UDG sample in order to compare with the observations of Leisman et al. (2017). Their results are also displayed in Fig. 5. We note that we jiggle their ($g - r$) colours to match our measurement error for the comparison. The peak of their ($g - r$) distribution is in reasonable agreement with our observations, yet it is slightly shifted towards the red and narrower. This may be because our observed catalogue is not entirely made from UDGs but also contains some contaminant sources, or perhaps because the SAM does not reproduce the correct amount of scatter for UDG colours. Alternatively, it may be a projection effect; our observed colours are in the observed frame and therefore are subject to k -corrections.

A similar comparison can be made with the work of Rong et al. (2017), who used the Guo et al. (2013) SAM to obtain a catalogue of simulated UDGs. We again jiggle their colours using our measurement error for the comparison. The colour distribution of their field UDGs is shifted towards the red compared to our observations, as also noted by Jones et al. (2018). If we were to include additional reddening of their colours because of k -corrections (i.e. to make a fair comparison with our observations), then this discrepancy would be exaggerated.

We calculate the total field density of UDGs in the Santa-Cruz SAM by integrating the stellar mass function for UDGs (fig. 4 of Jones et al. 2018). We obtain a value of $2 \times 10^{-2} \text{ cMpc}^{-3}$, approximately twice the upper-limit estimate from our measurements.

6.3 Impact of nearby galaxy groups

While our observed catalogue of UDG candidates is dominated by field sources, it is important to consider the effects of nearby galaxy groups on our result. After all, if such sources are predominantly quiescent and exist in similar number to our field sample, we should expect to find a population of red UDG candidates.

Fortunately, our KiDS/HSC-SSP footprint overlaps with the GAMA spectroscopic survey (Driver et al. 2011) and thus the

¹⁰Although this could potentially be implemented in future studies.

¹¹We note that we use the HUD-B sample, which contains 115 sources and was selected using selection criteria consistent with those of van der Burg et al. (2016).

GAMA group catalogue (Robotham et al. 2011). We can therefore make an estimate of the contribution of group/cluster UDGs to our field sample using a similar method to van der Burg et al. (2017). Working in our favour is the fact that we have applied a relatively high cut in angular size ($\bar{r}_e \geq 3$ arcsec) and there are no massive groups that are close enough to dominate our selection.

We select all GAMA groups from the group catalogue that satisfy $N_{\text{fof}} \geq 5$ and $0.01 < z_{\text{fof}} < 0.2$ (where N_{fof} and z_{fof} are respectively the number of friends-of-friends sources and an estimate of the group redshift). For each group, we count the number of sources that are compatible with being UDGs at that redshift, using our selection criteria and a physical radius range of $1.5 \leq \bar{r}_e$ [kpc] ≤ 7.0 . We also subtract a statistical background correction based on the total number of sources across the entire footprint that meet the same criteria. We estimate that up to 8 per cent of our UDG candidate catalogue is associated with relatively massive groups (i.e. the ones that have at least five friends-of-friends members), with the uncertainty coming from the background count estimate. The colour histogram of these sources is statistically indistinguishable from that of the whole catalogue, and we conclude that their inclusion does not significantly impact our results.

Using the empirical scaling relation between the group mass and total r -band luminosity from Viola et al. (2015), we estimate that ~ 6 per cent of the available mass out to $z = 0.2$ is taken up by the groups we consider here. This is very similar to the fraction of observed UDG candidates associated with groups. Taken with the fact that we expect all observed UDGs to be at $z < 0.2$, this provides an independent indication that the mass formation efficiency of UDGs in the field is comparable to that in groups and clusters. It also shows that the presence of the massive groups does not severely impact our result.

7 CONCLUSIONS

In this paper we have used deep wide-area optical imaging from the KiDS survey to detect sources with low surface brightness ($24.0 \leq \bar{\mu}_{e,r} \leq 26.5$) and large angular sizes ($3.0 \text{ arcsec} \leq \bar{r}_{e,r} \leq 8.0 \text{ arcsec}$). Following the detection and measurement of these sources with MTOBJECTS and GALFIT, we measured colours using the HSC-SSP survey data. Our catalogue of UDG candidates consists of 212 sources over ~ 39 square degrees. Compared to UDGs in groups and clusters, our sample consists of sources that appear to have much more irregular morphologies and show hints of active star formation.

These observations were compared to mock observations of UDGs created by sampling empirical distributions of UDG properties from the literature. Our key assumptions were intrinsic size, surface brightness, and colour distributions for the UDGs. All the assumptions we made are justified based on the current understanding of UDGs.

By comparing our mock catalogues with the observations, we have shown that it is very unlikely for a significant population of UDGs that are as red in colour as they are in clusters to exist in the field. It is much more likely that almost all UDGs in the field are instead much bluer, with colours similar to late-type dwarf galaxies in clusters. An immediate conclusion based on the predominantly blue colours is that secular evolutionary processes are not producing large numbers of cluster-like quenched (red) UDGs.

This finding means that isolated red UDGs, like the ones found by Martínez-Delgado et al. (2016) and Román et al. (2019), should be quite rare. At first glance this contrasts with the work of RS et al. (2019), who find a population of UDGs with low specific

star formation rates (compared to the star-forming main sequence) in the field. Taking the appearance of our detections in Fig. A1 into account, it is likely that UDGs in the field are forming stars only at a few localized sites; this locality may result in relatively low specific star formation rates that never the less make their integrated colours bluer. As noted by RS et al. (2019) and Zaritsky et al. (2019), their sample of field UDGs is still systematically bluer than that observed for cluster UDGs. Assuming that UDGs across different environments share similar metallicities, this is good evidence that star formation in field UDGs can be quite tentative and easily quenched in cluster environments.

We also created mock observations of massive blue galaxies, thought to be the primary source of contamination in our UDG candidate sample, using canonical empirical relations. We statistically subtracted these from our observations to acquire a contaminant-corrected catalogue of UDGs. The normalized distribution of $(g - r)$ colour is very similar to that estimated for H I-bearing field UDGs measured by Leisman et al. (2017). The observed distribution is also similar to that predicted for UDGs in SAMs (Rong et al. 2017; Jones et al. 2018), but slightly bluer. While our colour distribution appears to have greater dispersion, this is likely due to systematic shortcomings in comparing simulations with observations.

Using our mock catalogues as a reference, we estimate an upper limit on the field density of UDGs as $8 \pm 3 \times 10^{-3} \text{ cMpc}^{-3}$, equivalent to a mass formation efficiency $\sim 0.8 \pm 0.2$ times that in clusters. Perhaps surprisingly, this density actually implies that UDGs form with a mass efficiency in the field that is quite close to that in cluster environments. The field density applies for UDGs with physical sizes $1.5 \leq \bar{r}_{e,r}$ [kpc] ≤ 7.0 and intrinsic (i.e. not cosmologically dimmed) surface brightnesses $24.0 \leq \bar{\mu}_{e,r} \leq 26.5$. This number density also suggests that current SAMs overpredict the number of UDGs by at least a factor of 2. However, we note that if UDGs exist in abundance at lower surface brightnesses than we have probed here, the total number density of large LSB objects could be much higher. Based on the field density measured by Jones et al. (2018), H I-bearing UDGs comprise at least one-fifth of the overall UDG population in the field. This is consistent with what is predicted from the Santa-Cruz SAM.

We note that the analysis we have performed in this work has been approximate in nature because of the absence of any distance measurements. Acquiring large samples of spectroscopic redshifts for LSB galaxies in the field is not currently feasible. In the near term the second data release of the HSC-SSP will provide an opportunity to follow up this work thanks to its expanded footprint; this analysis can easily be expanded to larger areas. In the longer term, deep all-sky imaging (perhaps combined with photometric redshifts) from LSST may provide the ultimate data set for providing statistical constraints on LSB galaxies in the field.

ACKNOWLEDGEMENTS

We are thankful to Caroline Haigh and the MTOBJECTS team for providing their software.

We are also grateful to Javier Román, Yu Rong, and Michael G. Jones (and collaborators) for kindly sharing their data.

We are additionally thankful to Alister Graham, Ivan Baldry, and Benne Holwerda for providing insightful comments and suggestions.

We would also like to acknowledge the contribution that the reviewer, Javier Román, has made to the quality of the paper.

Based on data products from observations made with ESO Telescopes at the La Silla Paranal Observatory under programme IDs

177.A-3016, 177.A-3017, and 177.A-3018, and on data products produced by Target/OmegaCEN, INAF-OACN, INAF-OAPD, and the KiDS production team, on behalf of the KiDS consortium. OmegaCEN and the KiDS production team acknowledge support by NOVA and NWO-M grants. Members of INAF-OAPD and INAF-OACN also acknowledge the support from the Department of Physics & Astronomy of the University of Padova, and of the Department of Physics of Univ. Federico II (Naples).

GAMA is a joint European–Australasian project based around a spectroscopic campaign using the Anglo-Australian Telescope. The GAMA input catalogue is based on data taken from the Sloan Digital Sky Survey and the UKIRT Infrared Deep Sky Survey. Complementary imaging of the GAMA regions is being obtained by a number of independent survey programmes including *GALEX* MIS, VST KiDS, VISTA VIKING, *WISE*, *Herschel*-ATLAS, GMRT, and ASKAP providing UV to radio coverage. GAMA is funded by the STFC (UK), the ARC (Australia), the AAO, and the participating institutions. The GAMA website is <http://www.gama-survey.org/>.

This research made use of ASTROPY,¹² a community-developed core PYTHON package for Astronomy (Astropy Collaboration 2013, 2018).

This research has made use of the VizieR catalogue access tool, CDS, Strasbourg, France (DOI: 10.26093/cds/vizier). The original description of the VizieR service was published in *A&AS* 143, 23.

This research has made use of NASA’s Astrophysics Data System.

REFERENCES

- Abraham R. G., van Dokkum P. G., 2014, *PASP*, 126, 55
- Aihara H. et al., 2018, *PASJ*, 70, S8
- Alabi A. et al., 2018, *MNRAS*, 479, 3308
- Amorisco N. C., Loeb A., 2016, *MNRAS*, 459, L51
- Amorisco N. C., Monachesi A., Agnello A., White S. D. M., 2018, *MNRAS*, 475, 4235
- Astropy Collaboration, 2013, *A&A*, 558, A33
- Astropy Collaboration, 2018, *AJ*, 156, 123
- Baldry I. K., Glazebrook K., Brinkmann J., Ivezić Ž., Lupton R. H., Nichol R. C., Szalay A. S., 2004, *ApJ*, 600, 681
- Baldry I. K. et al., 2012, *MNRAS*, 421, 621
- Beasley M. A., Trujillo I., 2016, *ApJ*, 830, 23
- Bell E. F., McIntosh D. H., Katz N., Weinberg M. D., 2003, *ApJS*, 149, 289
- Bennet P., Sand D. J., Zaritsky D., Crnojević D., Spekkens K., Karunakaran A., 2018, *ApJ*, 866, L11
- Bothun G. D., Impey C. D., Malin D. F., Mould J. R., 1987, *AJ*, 94, 23
- Caon N., Capaccioli M., D’Onofrio M., 1993, *MNRAS*, 265, 1013
- Carleton T., Cooper M., Kaplinghat M., Errani R., Penarrubia J., 2018, *AAS*, 231, 412.05
- Cole S., Lacey C. G., Baugh C. M., Frenk C. S., 2000, *MNRAS*, 319, 168
- Collins M. L. M. et al., 2013, *ApJ*, 768, 172
- Conroy C., Gunn J. E., 2010, *ApJ*, 712, 833
- Conroy C., Gunn J. E., White M., 2009, *ApJ*, 699, 486
- Conselice C. J., 2018, *RNAAS*, 2, 43
- Conselice C. J., Gallagher J. S. III, Wyse R. F. G., 2003, *AJ*, 125, 66
- Dalcanton J. J., Spergel D. N., Gunn J. E., Schmidt M., Schneider D. P., 1997, *AJ*, 114, 635
- Danieli S., van Dokkum P., Conroy C., 2018, *ApJ*, 856, 69
- Das M., 2013, *JA&A*, 34, 19
- Davies J. I., Phillipps S., Disney M. J., 1989, *MNRAS*, 239, 703
- Davies J. I., Davies L. J. M., Keenan O. C., 2016, *MNRAS*, 456, 1607
- de Jong J. T. A., Verdoes Kleijn G. A., Kuijken K. H., Valentijn E. A., 2013, *Exp. Astron.*, 35, 25
- Di Cintio A., Brook C. B., Dutton A. A., Macciò A. V., Obreja A., Dekel A., 2017, *MNRAS*, 466, L1
- Disney M. J., 1976, *Nature*, 263, 573
- Driver S. P. et al., 2006, *MNRAS*, 368, 414
- Driver S. P. et al., 2011, *MNRAS*, 413, 971
- Erben T. et al., 2013, *MNRAS*, 433, 2545
- Erwin P., 2015, *ApJ*, 799, 226
- Fensch J. et al., 2019, *A&A*, 625, A77
- Ferré-Mateu A. et al., 2018, *MNRAS*, 479, 4891
- Graham A., Lauer T. R., Colless M., Postman M., 1996, *ApJ*, 465, 534
- Graham A. W., Driver S. P., 2005, *Publ. Astron. Soc. Aust.*, 22, 118
- Graham A. W., Janz J., Penny S. J., Chilingarian I. V., Ciambur B. C., Forbes D. A., Davies R. L., 2017, *ApJ*, 840, 68
- Greco J. P. et al., 2018a, *ApJ*, 857, 104
- Greco J. P., Goulding A. D., Greene J. E., Strauss M. A., Huang S., Kim J. H., Komiyama Y., 2018b, *ApJ*, 866, 112
- Guo Q. et al., 2013, *MNRAS*, 435, 897
- Hogg D. W., Baldry I. K., Blanton M. R., Eisenstein D. J., 2002, preprint (arXiv:astro-ph/0210394)
- Impey C., Bothun G., Malin D., 1988, *ApJ*, 330, 634
- Janssens S., Abraham R., Brodie J., Forbes D., Romanowsky A. J., van Dokkum P., 2017, *ApJ*, 839, L17
- Janz J., Penny S. J., Graham A. W., Forbes D. A., Davies R. L., 2017, *MNRAS*, 468, 2850
- Jiang F., Dekel A., Freundlich J., Romanowsky A. J., Dutton A., Maccio A., Di Cintio A., 2019, *MNRAS*, 487, 5272
- Jones M. G., Papastergis E., Pandya V., Leisman L., Romanowsky A. J., Yung L. Y. A., Somerville R. S., Adams E. A. K., 2018, *A&A*, 614, A21
- Kauffmann G., White S. D. M., Guiderdoni B., 1993, *MNRAS*, 264, 201
- Kelvin L. S. et al., 2012, *MNRAS*, 421, 1007
- Koda J., Yagi M., Yamanoi H., Komiyama Y., 2015, *ApJ*, 807, L2
- Kuijken K. et al., 2019, *A&A*, 625, A2
- Leisman L. et al., 2017, *ApJ*, 842, 133
- Lelli F., Fraternali F., Sancisi R., 2010, *A&A*, 516, A11
- Lim S., Peng E. W., Côté P., Sales L. V., den Brok M., Blakeslee J. P., Guhathakurta P., 2018, *ApJ*, 862, 82
- Lupton R., Blanton M. R., Fekete G., Hogg D. W., O’Mullane W., Szalay A., Wherry N., 2004, *PASP*, 116, 133
- Mancera Piña P. E., Peletier R. F., Aguerri J. A. L., Venhola A., Trager S., Choque Challapa N., 2018, *MNRAS*, 481, 4381
- Mancera Piña P. E., Aguerri J. A. L., Peletier R. F., Venhola A., Trager S., Choque Challapa N., 2019, *MNRAS*, 485, 1036
- Martínez-Delgado D. et al., 2016, *AJ*, 151, 96
- McConnachie A. W., 2012, *AJ*, 144, 4
- McFarland J. P., Verdoes-Kleijn G., Sikkema G., Helmich E. M., Boxhoorn D. R., Valentijn E. A., 2011, preprint (arXiv:1110.2509)
- McGaugh S. S., 1996, *MNRAS*, 280, 337
- Mihos J. C. et al., 2015, *ApJ*, 809, L21
- Mowla L., van Dokkum P., Merritt A., Abraham R., Yagi M., Koda J., 2017, *ApJ*, 851, 27
- Muñoz R. P. et al., 2015, *ApJ*, 813, L15
- Muzzin A. et al., 2013, *ApJ*, 777, 18
- Papastergis E., Adams E. A. K., Romanowsky A. J., 2017, *A&A*, 601, L10
- Peng C. Y., Ho L. C., Impey C. D., Rix H.-W., 2002, *AJ*, 124, 266
- Prole D. J., Davies J. I., Keenan O. C., Davies L. J. M., 2018, *MNRAS*, 478, 667
- Prole D. J. et al., 2019, *MNRAS*, 484, 4865
- Roberts S., Davies J., Sabatini S., Auld R., Smith R., 2007, *MNRAS*, 379, 1053
- Robotham A. S. G. et al., 2011, *MNRAS*, 416, 2640
- Robotham A. S. G., Davies L. J. M., Driver S. P., Koushan S., Taranu D. S., Casura S., Liske J., 2018, *MNRAS*, 476, 3137
- Román J., Trujillo I., 2017a, *MNRAS*, 468, 703
- Román J., Trujillo I., 2017b, *MNRAS*, 468, 4039
- Román J., Beasley M. A., Ruiz-Lara T., Valls-Gabaud D., 2019, *MNRAS*, 486, 823
- Rong Y., Guo Q., Gao L., Liao S., Xie L., Puzia T. H., Sun S., Pan J., 2017, *MNRAS*, 470, 4231

¹²<http://www.astropy.org>

RS P., Zaritsky D., Donnerstein R., Spekkens K., 2019, preprint ([arXiv:1904.06262](https://arxiv.org/abs/1904.06262))
 Ruiz-Lara T. et al., 2018, *MNRAS*, 478, 2034
 Sabatini S., Davies J., Scaramella R., Smith R., Baes M., Linder S. M., Roberts S., Testa V., 2003, *MNRAS*, 341, 981
 Somerville R. S., Popping G., Trager S. C., 2015, *MNRAS*, 453, 4337
 Taylor E. N. et al., 2011, *MNRAS*, 418, 1587
 Taylor E. N. et al., 2015, *MNRAS*, 446, 2144
 Teeninga P., Moschini U., Trager S. C., Wilkinson M. H., 2016, International Symposium on Mathematical Morphology and Its Applications to Signal and Image Processing. p. 157
 Trujillo I. et al., 2019, *MNRAS*, 486, 1192
 van der Burg R. F. J., Muzzin A., Hoekstra H., 2016, *A&A*, 590, A20
 van der Burg R. F. J. et al., 2017, *A&A*, 607, A79
 van der Wel A., 2008, *ApJ*, 675, L13
 van der Wel A. et al., 2014, *ApJ*, 788, 28
 van Dokkum P. et al., 2018, *Nature*, 555, 629
 van Dokkum P., Danieli S., Abraham R., Conroy C., Romanowsky A. J., 2019, *ApJ*, 874, L5

van Dokkum P. G., Abraham R., Merritt A., Zhang J., Geha M., Conroy C., 2015, *ApJ*, 798, L45
 Venhola A. et al., 2017, *A&A*, 608, A142
 Venhola A. et al., 2018, *A&A*, 620, A165
 Vika M., Vulcani B., Bamford S. P., Häußler B., Rojas A. L., 2015, *A&A*, 577, A97
 Viola M. et al., 2015, *MNRAS*, 452, 3529
 Vulcani B. et al., 2014, *MNRAS*, 441, 1340
 White S. D. M., Rees M. J., 1978, *MNRAS*, 183, 341
 Williams R. P. et al., 2016, *MNRAS*, 463, 2746
 Wittmann C. et al., 2017, *MNRAS*, 470, 1512
 Wright A. H. et al., 2017, *MNRAS*, 470, 283
 Yagi M., Koda J., Komiyama Y., Yamanoi H., 2016, *ApJS*, 225, 11
 Yozin C., Bekki K., 2015, *MNRAS*, 452, 937
 Zaritsky D. et al., 2019, *ApJS*, 240, 1

APPENDIX A: EXAMPLES

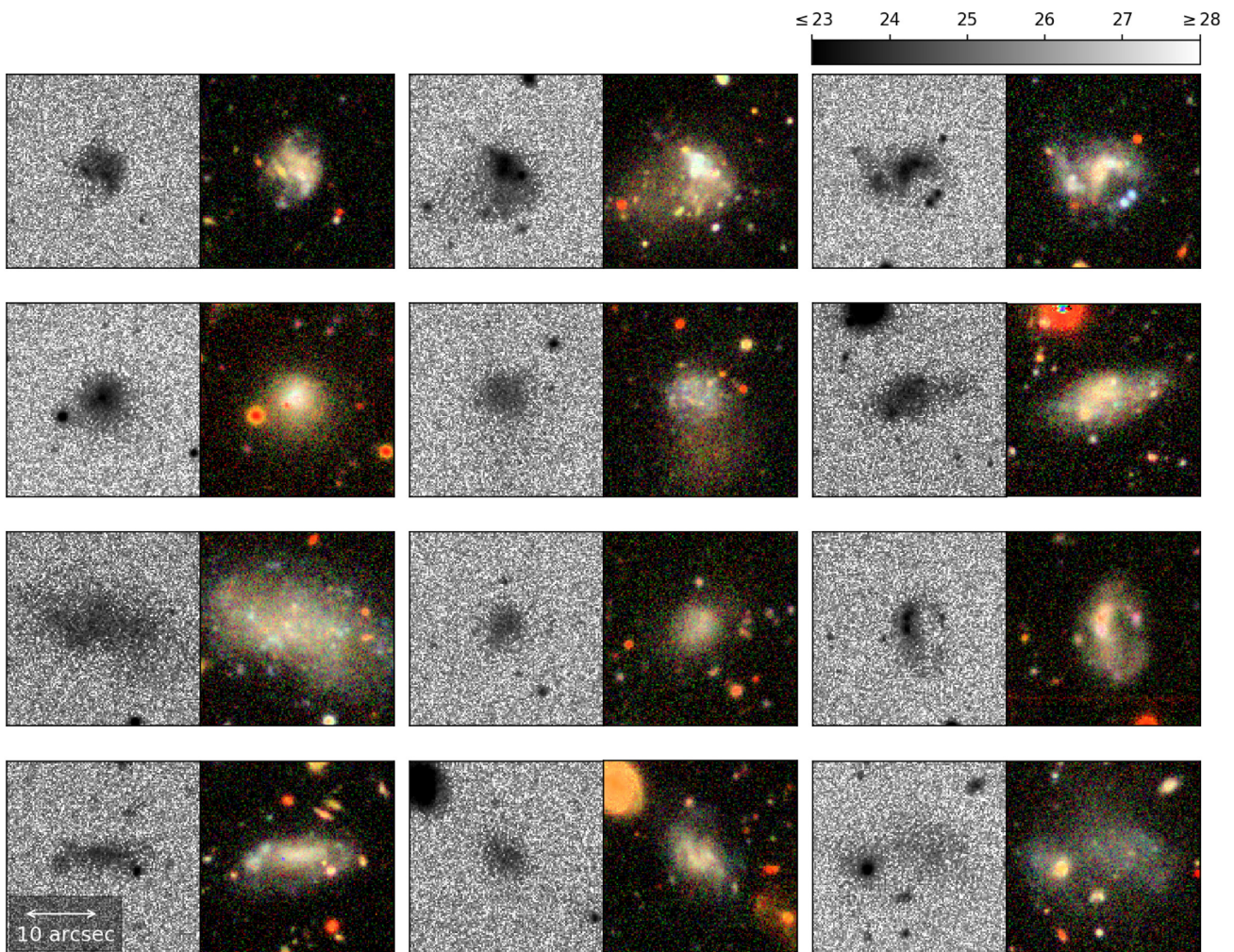


Figure A1. Examples of large LSB sources in our UDG candidate sample. Each source is shown in the KiDS r band (left-hand panels) versus a colour image made from the g , r , and i HSC-SSP bands according to Lupton et al. (2004) (right-hand panels). The cut-out size is 25 arcsec, much less than the 80 arcsec regions that we use to fit the sources. The colour bar shows the surface brightness in units of magnitude per square arcsecond for the KiDS data. In comparison to the regular morphologies of UDGs in groups and clusters, many of our sources are rather amorphous, with possible signs of discrete sites of active star formation.

APPENDIX B: MODEL UNCERTAINTIES

While our analysis here is first order in nature, it is still important to quantify how uncertainties in the model ingredients may impact our result. In particular, we have not discussed how the uncertainties in the assumed UDG size distribution propagate. van der Burg et al. (2017) have measured a power-law index of -2.71 ± 0.33 for the distribution of circularized radii (equation 3). The results of varying the slope by 1σ are shown in Fig. B1. It is clear that lowering the index (more small UDGs) lowers our estimate of the number of UDGs we expect to observe by around 15 per cent. Conversely, increasing the index (more large sources) causes the predicted number to increase by around 25 per cent.

An additional source of uncertainty is that which arises from our estimate of the mass formation efficiency. From the empirical relation of van der Burg et al. (2017), we estimate a ~ 20 per cent error. This uncertainty also propagates to our estimate of the field density of UDGs.

A separate issue is how the assumed stellar population (i.e. that which defines the k -corrections) affects our analysis. For the UDGs, we have explored a red and blue colour model, using quiescent and star-forming populations for the k -corrections, respectively. One of the uncertainties for the star-forming population model is the star formation history to assume; for this analysis we have assumed a uniform star formation rate. As a means to test whether this assumption impacts our result, we can also model the blue UDG population using the quiescent model for k -corrections. The results of this are shown in Fig. B2. From this figure, it appears that the change is small, with a slight shift towards redder colours. The impact on our analysis is negligible; this is not surprising as most of our observed UDGs are expected to be at low redshift where k -corrections are small.

We have repeated a similar process for our late-type interloper model, replacing the stellar population model with that used for star-forming UDGs. We also find that this makes no significant difference to our results.

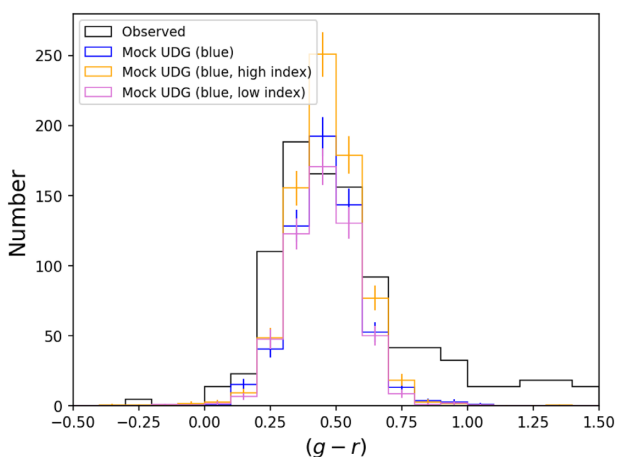


Figure B1. Impact of changing model parameters within their errors. The observed colour histogram is shown in back. The blue UDG model is shown in dark blue. The mock catalogue with the size power-law index lowered/raised by 1σ is shown in purple/orange. All error bars are Poisson uncertainties.

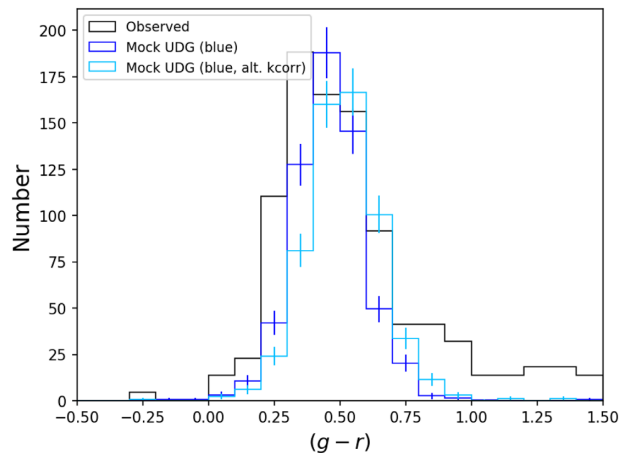


Figure B2. Comparison of the model colour distributions for blue UDGs assuming a star-forming stellar population for the k -corrections (dark blue) and a quiescent one (light blue). Overall, we find the difference to be negligible for our analysis.

APPENDIX C: EXTENDING THE MODEL

One shortcoming of our analysis is that we do not account for galaxies smaller than the fiducial UDG size limit of 1.5 kpc, a fairly arbitrary cut-off. One way to probe how the inclusion of such galaxies may alter the results presented in Fig. 4 is to extrapolate the empirical size distribution that we use in our UDG model (van der Burg et al. 2017) to lower size limits. This exercise is approximate in nature because it is not clear whether an extrapolation of this relation is valid for smaller galaxies. There are two competing effects: While smaller galaxies are more numerous because of the steep power law (equation 3), their smaller size means that they are much less likely to be observed given our selection criteria and the corresponding recovery efficiency.

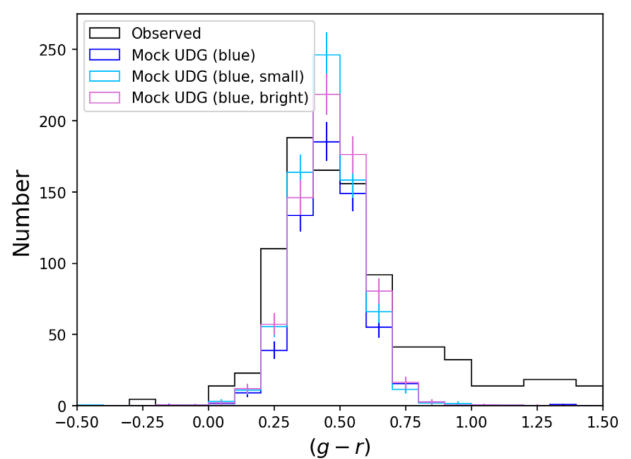


Figure C1. Results of extending our empirical model. The observed colour histogram is shown in back. The blue UDG model is shown in dark blue. The mock catalogue with the reduced lower limit of $\bar{r}_e \geq 0.5$ kpc is shown in light blue. The mock catalogue that includes bright galaxies ($\bar{\mu}_e \geq 22$) is shown in purple. All error-bars are Poisson uncertainties.

We display the result of reducing the lower physical size limit from 1.5 to 0.5 kpc in Fig. C1. The total number of sources we generate is increased proportionally to equation (3). Despite the increase in the number of sources (by a factor of ~ 20), there is only a minor difference between the number of sources we would expect to observe.

As a further extension to the model, we can consider what happens when we decrease the lower limit in surface brightness at $\bar{\mu}_e = 24$ to allow brighter sources into the selection. Since Danieli & van Dokkum (2018) have shown that the distribution of intrinsic size of large red galaxies is approximately uniform with absolute magnitude (and therefore surface brightness), this extension can be interpreted as including large red galaxies with Sérsic indices meeting our selection criteria. As an example, we show in the figure the effect of using a bright-end surface brightness cut of $\bar{\mu}_e = 22$, increasing the number of sources by 80 per cent according to equation (2). As in the previous test, the difference with the result in Fig. 4 is fairly insignificant. We are left to conclude that our observational sample is indeed likely made up of large low surface brightness galaxies.

APPENDIX D: COMPARISON WITH GAMA REDSHIFTS

We can compare our measurements and mock catalogues against measurements from the GAMA spectroscopic survey in order to test how well our mock catalogues represent reality. For this test, we use our best model: the combination of blue UDGs with massive blue galaxy interlopers. We assume that UDGs form with a mass efficiency as calculated in Section 5.1.

Using the public data release 3 data obtained from the GAMA website,¹³ we crossmatched the `SpecObj v27` catalogue (containing spectroscopic redshifts) with the `SersicCatSDSS v09` table (containing Sérsic profile fits to GAMA targets in SDSS data from Kelvin et al. 2012). We imposed our selection criteria on the Sérsic parameters and additionally required `SURVEY_CLASS` ≥ 4 in order to select legitimate sources with $m_r < 19.8$, leaving us with 209 GAMA sources. We also applied the $m_r < 19.8$ criterion to our mock catalogue, retrieving 379 sources. The results of the comparison are shown in the top panel of Fig. D1.

Despite the surface brightness limits of GAMA (e.g. Wright et al. 2017), we find that 45 of our UDG candidates (over the full unmasked KiDS area) have matches in the GAMA `SpecObj` catalogue within 3 arcsec. This allows us to make the same comparison as above, using our measurements in place of those of Kelvin et al. (2012); the result is also shown in the top panel of Fig. D1. Clearly the overall distribution of our mock catalogues is in good agreement with the observations.

¹³<http://www.gama-survey.org/>

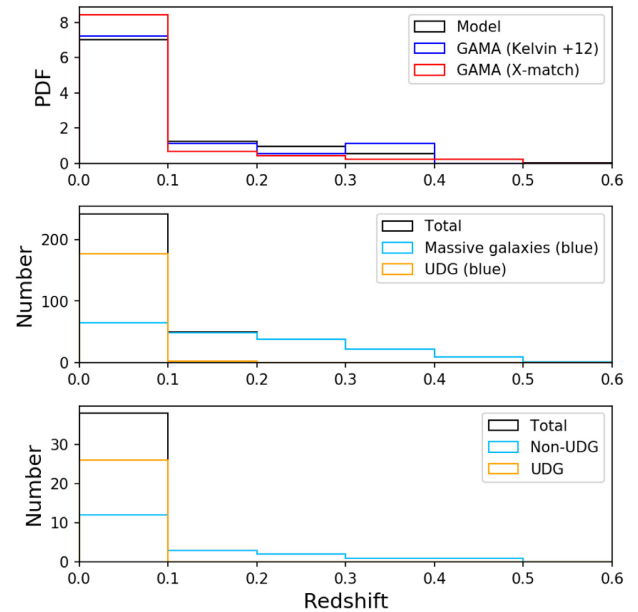


Figure D1. *Top:* comparison of the overall redshift distribution from our mock catalogue (black histogram) with that of Kelvin et al. (2012) (blue histogram) and a crossmatch between our observed UDG candidate catalogue and the GAMA spectroscopic survey (red histogram). *Middle:* decomposition of our mock catalogue into UDGs (orange) and interlopers (blue) as a function of redshift. *Bottom:* the same as the middle panel, but for our crossmatch with GAMA.

However, we can go one step further. We can now estimate the intrinsic parameters of the matching sources because we know their redshifts and apparent structural parameters. In doing so, we can test whether the individual distributions for UDGs and interlopers are approximately correct. For this test, we define a UDG as having $\bar{r}_e \geq 1.5$ kpc and $M_* \leq 10^9 M_\odot$. The stellar mass is estimated assuming our blue UDG stellar population model from Section 4.1.3 together with the GALFIT m_r measurements. We can then decompose the catalogues into UDG and non-UDG populations.

The decomposition of the mock catalogue is shown in the middle panel of Fig. D1. This compares with the decomposed observed catalogue in the lower panel. Clearly the distributions are similar; at low redshifts UDGs dominate our sample, while at higher redshifts, massive interlopers dominate.

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