The Astrophysical Journal, 979:118 (15pp), 2025 February 1 https://doi.org/10.3847/1538-4357/ad7119

The Prevalence of Star-forming Clumps as a Function of Environmental Overdensity in Local Galaxies

Dominic Adams[1], Hugh Dickinson[2], Lucy Fortson[1], Kameswara Mantha[1], Vihang Mehta[3], Jürgen Popp[2], Claudia Scarlata[1], Chris Lintott[4], Brooke Simmons[5], and Mike Walmsley[6]
[1] School of Physics and Astronomy, University of Minnesota, 116 Church St. SE, Minneapolis, MN 55455, USA
[2] School of Physical Sciences, The Open University, Walton Hall, MK7 6AA Milton Keynes, UK
[3] IPAC, Mail Code 314-6, California Institute of Technology, 1200 E California Blvd., Pasadena CA 91125, USA
[4] Department of Physics, University of Oxford, Denys Wilkinson Building, Keble Rd., Oxford, OX1 3RH, UK
[5] Department of Physics, Lancaster University, Lancaster LA1 4YB, UK
[6] Jodrell Bank Centre for Astrophysics, Department of Physics & Astronomy, University of Manchester, Oxford Rd., Manchester M13 9PL, UK
Received 2023 October 2; revised 2024 July 7; accepted 2024 August 6; published 2025 January 21

Abstract

At the peak of cosmic star formation ($1 \lesssim z \lesssim 2$), the majority of star-forming galaxies hosted compact, star-forming clumps, which were responsible for a large fraction of cosmic star formation. By comparison, $\lesssim$5% of local star-forming galaxies host comparable clumps. In this work, we investigate the link between the environmental conditions surrounding local ($z < 0.04$) galaxies and the prevalence of clumps in these galaxies. To obtain our clump sample, we use a Faster R-CNN object detection network trained on the catalog of clump labels provided by the Galaxy Zoo: Clump Scout project, then apply this network to detect clumps in approximately 240,000 Sloan Digital Sky Survey galaxies (originally selected for Galaxy Zoo 2). The resulting sample of 41,445 *u*-band bright clumps in 34,246 galaxies is the largest sample of clumps yet assembled. We then select a volume-limited sample of 9964 galaxies and estimate the density of their local environment using the distance to their projected fifth nearest neighbor. We find a robust correlation between environment and the clumpy fraction ($f_{\rm clumpy}$) for star-forming galaxies (specific star formation rate, sSFR $> 10^{-2}$ Gyr^{-1}) but find little to no relationship when controlling for galaxies' sSFR or color. Further, $f_{\rm clumpy}$ increases significantly with sSFR in local galaxies, particularly above sSFR $> 10^{-1}$ Gyr^{-1}. We posit that a galaxy's gas fraction primarily controls the formation and lifetime of its clumps, and that environmental interactions play a smaller role.

Unified Astronomy Thesaurus concepts: Galaxies (573); Star forming regions (1565); Starburst galaxies (1570); Galaxy evolution (594); Galaxy formation (595); Galaxy structure (622); Cosmological evolution (336); Convolutional neural networks (1938)

1. Introduction

It is well established that galaxies at high redshift ($z \gtrsim 1$) are more irregular and more highly star forming than their low-redshift counterparts. In particular, observations over the last 30 yr have revealed that a high fraction of high-redshift galaxies host giant, star-forming, off-center clumps (L. L. Cowie et al. 1995; D. M. Elmegreen et al. 2004a, 2004b). These clumps are considerably larger and brighter than typical star-forming regions in most low-redshift galaxies. It is common to define a clump as a star-forming region that emits at least 8% of its host galaxy's flux; under this definition, approximately 50% of star-forming galaxies at the peak of cosmic star formation ($1 \lesssim z \lesssim 2$) host at least one clump (Y. Guo et al. 2015; T. Shibuya et al. 2016), while fewer than 5% of star-forming galaxies at $z \sim 0$ host one (D. Adams et al. 2022). In order to fully understand how early galaxies evolved into modern ones, we must understand how these clumps formed and why they became so rare over time.

It is generally agreed that the majority of high-redshift clumps form "in situ," i.e., from gas already present within the host galaxy (e.g., F. Bournaud & B. G. Elmegreen 2009; F. Bournaud et al. 2009; A. Dekel et al. 2009; R. Genzel et al. 2011; N. Mandelker et al. 2014; S. Inoue et al. 2016; A. Zanella et al. 2019). The in situ mode of clump formation requires a source of available gas, as well as turbulent and unstable gas dynamics within the host galaxy's disk; under these conditions, gas is expected to collapse and form clumps (A. Dekel et al. 2009, 2022; F. Bournaud et al. 2014; N. Mandelker et al. 2014; D. B. Fisher et al. 2017). Alternatively, some clumps may form in the "ex situ" mode, in which the clumps begin their lives as small independent galaxies before joining the host in a minor merger. The remnant of this minor merger becomes a clump. Simulations (e.g., N. Mandelker et al. 2014, 2017) suggest that ex situ clumps are older, more massive, and larger in size than their in situ counterparts, as well as maintaining their own dark matter component.

At least one observational study (A. Zanella et al. 2019) has found that clump properties are bimodal, with one mode consisting of larger and redder clumps and another consisting of smaller and bluer clumps; this difference in properties may indicate a difference in their origins. It also is possible that the dominant mode of clump formation (in situ or ex situ) may be a function of galaxy mass or cosmic time (Y. Guo et al. 2015; D. Adams et al. 2022). In particular, the rate of cosmic star formation (P. Madau et al. 1996) and galaxy turbulence (S. A. Kassin et al. 2012) have been observed to decline faster than the minor merger rate (J. M. Lotz et al. 2011), which may suggest that ex situ clumps comprise a larger fraction of clumps in modern galaxies. However, it is very challenging to untangle

the origin and properties of clumps from direct observation alone. This is mainly due to their compact size: While early estimates suggested that clumps are kiloparsec scale objects (e.g., D. M. Elmegreen 2007), high-resolution observations of low-redshift clumpy galaxies (R. A. Overzier et al. 2009; D. B. Fisher et al. 2017; M. Messa et al. 2019) and lensed high-redshift galaxies (R. C. Livermore et al. 2012; A. Adamo et al. 2013; E. Wuyts et al. 2014; A. Cava et al. 2018; A. Claeyssens et al. 2023) reveal that many clumps are complexes of small star-forming regions with scales of 10–100 pc each, which cannot be resolved at high redshift with current instruments.

It should be noted that both the in situ and ex situ modes of clump formation require an external environmental trigger to occur. Theory (A. Dekel et al. 2009; R. C. Livermore et al. 2012), simulations (F. Bournaud et al. 2014; N. Mandelker et al. 2014; J. Fensch & F. Bournaud 2021), and observations (R. Genzel et al. 2011) all indicate that the in situ mode of clump formation is most likely to occur in regions of high gas concentration and turbulent gas dynamics. In particular, A. Dekel et al. (2009) put forward a model in which cold gas accretion along dark matter filaments can both supply a disk with pristine gas and add the necessary energy for this gas to become turbulent and unstable, creating the conditions for clump formation. On the other hand, ex situ clumps do not require a supply of gas; it is instead more reasonable to expect them in regions where the minor merger rate is higher, i.e., in cluster environments. Because of this, the environmental conditions surrounding clumpy galaxies can provide critical insights into their clumps' origins.

In this paper, we search for the presence of these clump formation trends in local galaxies from the Sloan Digital Sky Survey (SDSS) main survey. To do so we use a sample of clumps identified automatically by a deep learning model designed for generalized object detection, which was trained on the visually identified clump sample from the citizen science project Galaxy Zoo: Clump Scout (D. Adams et al. 2022). As in several past papers (V. Mehta et al. 2021; D. Adams et al. 2022), we choose to focus on local galaxies rather than high-redshift galaxies for two reasons: (a) Local galaxies are much easier to observe, making it easier to assemble a large sample of clumpy galaxies with uniform selection; and (b) By comparing the statistical properties of modern- and early-Universe clumpy galaxies, we can clarify the cosmic trends in the formation and properties of clumps.

This paper is organized as follows. In Section 2, we describe our process for identifying clumps and measuring their photometry. In Section 3, we describe the methods we use for computing galaxies' environmental densities as well as our selection of target galaxies and environment-tracing galaxies. In Section 4, we present our results on the environmental trends in SDSS clumpy galaxies. In Section 5, we discuss the implications of these results on the primary mode of clump formation in the local Universe. Section 6 contains our summary and conclusions. Throughout this paper, we assume a flat ΛCDM cosmology with $\Omega_m = 0.3$ and $\Omega_\Lambda = 0.7$, and take the Hubble constant to be $h = H_0/100\,\mathrm{km\,s^{-1}\,Mpc^{-1}} = 0.70$.

2. Clump Identification and Photometry

The challenge of clump identification has presented a major issue for the field of clump research over the last few decades. While clumps have historically been identified by visual classification alone (e.g., L. L. Cowie et al. 1995; D. M. Elmegreen et al. 2004a, 2004b; D. M. Elmegreen 2007), the need for larger and more consistent samples has led to the development of automated methods. For observational studies, these include the `clumpfind` algorithm (J. P. Williams et al. 1994), methods relying on source extraction using the `SExtractor` software package (E. Bertin & S. Arnouts 1996; e.g., Y. Guo et al. 2015; V. Mehta et al. 2021), machine learning methods (M. Huertas-Company et al. 2020), and others. To ensure consistency, one common strategy is to select for clumps whose near-UV emission (∼3500 Å) exceeds a specific fraction of its galaxy's total near-UV luminosity (as used by Y. Guo et al. 2015, 2018; T. Shibuya et al. 2016). In cases where near-UV fluxes have been unavailable, analogous selection limits have been used with different photometric bands (D. B. Fisher et al. 2017; D. Adams et al. 2022).

The largest catalog of clumps in SDSS galaxies was identified through the Galaxy Zoo: Clump Scout citizen science project (D. Adams et al. 2022; H. Dickinson et al. 2022), which recruited volunteers to visually identify and mark clumps in postage stamps of over 58,000 SDSS galaxies. These volunteers' marks were then aggregated to identify consensus clump locations. The Clump Scout catalog consists of 10,738 clumps in 7050 galaxies (with no flux-based selection limits applied), as well as 3861 objects identified by volunteers as probable contaminants. While this catalog's size makes it effective for statistical analyses, it leaves room for improvement in a few ways; namely (a) the catalog has limited clump identification completeness (reaching $\gtrsim$60% for only the brightest clumps), and (b) it covers only a small fraction of available SDSS galaxies. For these reasons, we expand the sample for this paper using machine learning methods. Clump detection is at its core an object identification task, and the Clump Scout sample—consisting of tens of thousands of human-annotated images—lends itself very naturally to use as training data for an object detection network.

2.1. Creating the Galaxy and Clump Parent Samples

The full training process for our clump detection model is provided in J. J. Popp et al. (2024), though we provide a brief description here. Using the clump labels from the Clump Scout project, we trained a Faster R-CNN (FRCNN, S. Ren et al. 2017) object detection framework to identify clumps. The FRCNN framework employs a convolutional neural network (CNN) to extract features from galaxies, as well as a region proposal network (RPN), which returns bounding boxes around candidate detections in the image. Each bounding box comes with a detection score weighing the machine's certainty about the detection. J. J. Popp et al. (2024) found that the best performance was achieved using an FRCNN whose CNN component had been pretrained to perform galaxy classification. Therefore, as our CNN, we used a version of the Zoobot model (M. Walmsley et al. 2023), which was trained to classify the morphologies of galaxies from SDSS as well as in images from the Hubble Space Telescope, the Hyper Suprime-Cam, and the Dark Energy Camera Legacy Survey. To further train and evaluate the FRCNN, we used the sample of 18,772 Clump Scout galaxies for which at least one clump was detected by the Clump Scout aggregation method.[7] This sample was randomly

[7] Note that this is larger than the published Clump Scout catalog, which only contains clumps whose estimated false positive probabilities fall below 0.6; no such cut is applied to the FRCNN training sample.

divided into training, validation, and test sets comprising 70%, 20%, and 10% of galaxies, respectively. The model was also trained to identify two classes of clumps, "normal" and "unusual," based on the labels provided for Clump Scout clumps. "Unusual" clumps typically correspond to foreground stars, cosmic rays, or other bright point sources that are not clumps; these objects can be filtered out to improve the final sample.

Our trained model was then applied to RGB galaxy images from the Galaxy Zoo 2 project, which were prepared by the same method as those in the training sample. From 243,500 Galaxy Zoo 2 galaxies, we selected the subset of 239,748 included in the SDSS DR8 spectroscopic and photometric catalogs (H. Aihara et al. 2011).[8] In total, 238,923 RGB cutouts were created. Additionally, we obtained estimates of stellar mass and specific star formation rates (sSFRs) from the MPA–JHU value-added catalog (G. Kauffmann et al. 2003; J. Brinchmann et al. 2004). A small number of galaxies (4219) were missing either mass or sSFR and were therefore excluded from the analysis in this paper. RGB cutouts were then produced from the *i*, *r*, and *g* image bands of these galaxies, following the same method as was used by Clump Scout (described in D. Adams et al. 2022); we produced cutouts this way to ensure that the FRCNN's clump identification was consistent with human identification, i.e., that it had access to the same information and was most likely to detect similar objects. Each RGB cutout is 200 by 200 pixels in size and scaled to 3 times the size of the galaxy's 90% Petrosian radius in the *r* band. (Image sizing and rescaling were performed by the Montage Library, J. C. Jacob et al. 2010.)

In each galaxy image, the FRCNN network predicted bounding boxes at the locations of candidate clumps. These bounding boxes are then filtered down to a list of candidate clump locations in several steps. (A more detailed description of our training and processing steps is included in J. J. Popp et al. 2024.)

1. Nonmaximum suppression: For each pair of overlapping bounding boxes (those whose areas had an intersection over the union of $\geqslant 0.2$), the bounding box with the lower confidence score was discarded. We then removed all bounding boxes with confidence scores <0.3; this was found to result in a fairly complete sample without sacrificing purity (see Section 2.3).
2. Centroids: Each bounding box was replaced with its centroid, which was taken to be the location of the clump candidate.
3. Deduplication: Each pair of locations separated by less than one full width at half-maximum of the image's *r*-band point-spread function (PSF-FWHM) was merged, and replaced with a single location at the midpoint of the pair. If either location in the pair was labeled as "unusual" by the machine, the merged location was marked as "unusual" as well.
4. Masking: Each galaxy image is segmented using the Python package *photutils* (L. Bradley 2023). The central segment's mask is taken to belong to the target galaxy, and locations that do not overlap with this segment are removed.
5. Removal of unusuals: Any location labeled as "unusual" is removed, as it is likely a contaminant.

[8] Galaxy Zoo 2 galaxies were selected from the DR7 photometric catalog. Due to the reassignment of photometric IDs between DR7 and DR8, not all of these galaxies have associated photometric IDs in the DR8 catalog.

After following these filtering steps, we are left with 75,468 candidate locations for clumps distributed across 57,851 galaxies.

2.2. Estimating Clump Fluxes

To determine the photometric properties of clumps, we used the same method as in D. Adams et al. (2022). Briefly, at the location of each clump, we place a "target aperture" with a diameter of 2.25 times the PSF-FWHM, as well as a "background annulus" spanning 3–5 times the PSF-FWHM. We then estimate the diffuse flux per pixel from the galaxy within the background annulus and subtract this flux from the target aperture. Finally, we multiply by an aperture correction (of 1.191) to account for clump flux falling outside of the target aperture. These values were adjusted for Milky Way dust extinction following D. J. Schlegel et al. (1998).

Given the fluxes of each clump, we made a cut on each clump's relative *u*-band flux compared to its host galaxy. To do so, we measured the background-subtracted fluxes of all clumps and computed the *u*-band luminosity fraction

$$f_{\rm Lu} = L_{u,\rm clump}/L_{u,\rm galaxy}. \quad (1)$$

Throughout this paper, we restrict our analysis to the sample of clumps with either $f_{\rm Lu} > 3\%$ (a moderately bright sample) or $f_{\rm Lu} > 8\%$ (a very bright sample).

2.3. Evaluating the Clump Detection Model

We evaluated the quality of the FRCNN-detected clumps in several ways. First, over the FRCNN's test sample (comprising 10% of galaxies for which the Clump Scout aggregator detected at least one clump), the locations of FRCNN-detected clumps were compared to the aggregated locations of Clump Scout clumps, which were treated as the "ground truth" clump sample. An FRCNN-detected clump was considered a "true positive" (TP) if it was separated from a "ground truth" clump by less than 0.75 of the image-specific *r*-band PSF-FWHM. The purity and completeness of this sample were then computed against the "ground truth" volunteer sample:

$$\begin{aligned} \text{Purity} &= \frac{N_{\rm TP}}{N_{\rm TP} + N_{\rm FP}} \\ \text{Completeness} &= \frac{N_{\rm TP}}{N_{\rm TP} + N_{FN}}, \end{aligned} \quad (2)$$

where $N_{\rm TP}$ is the number of true positive clumps, $N_{\rm FP}$ the number of false positives (detected clumps without corresponding ground truth clumps), and $N_{\rm FN}$ the number of false negatives (ground truth clumps without corresponding detections). For sufficiently bright clumps ($f_{\rm Lu} > 3\%$), the FRCNN detector achieved a completeness and purity of $\gtrsim$75% compared to the volunteer-provided sample (see J. J. Popp et al. 2024, Figures 10(b) and 11(b)). As an additional test, the completeness of the detected sample was evaluated using the simulated clump sample (comprising 84,565 simulated clumps inserted into 26,736 real galaxy images) produced for Galaxy Zoo: Clump Scout as the "ground truth" sample; a similar figure of 75% completeness was achieved for the same selection criteria (see J. J. Popp et al. 2024, Figure F2). A more complete description of the way purity and completeness were evaluated can be found in J. J. Popp et al. (2024).

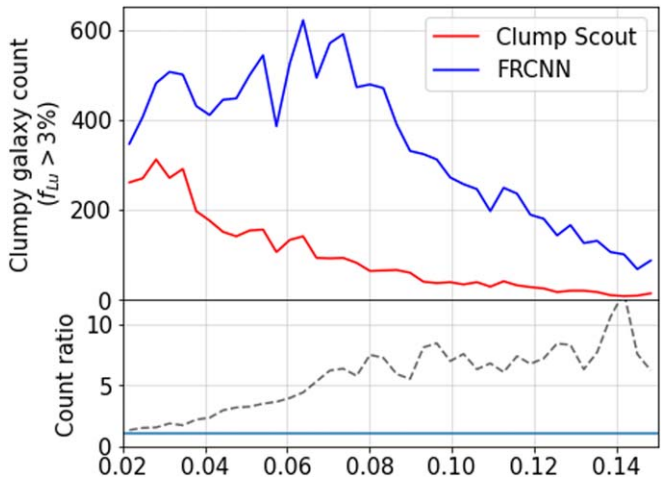

Figure 1. Comparison of the redshift distribution of clumpy galaxies from the Clump Scout catalog vs. those from the FRCNN. The top plot displays galaxy number counts binned by redshift, while the bottom plots the ratio between them, with a solid line marking unity. Compared with volunteers from Clump Scout, the FRCNN detects more clumpy galaxies at all redshifts, but it is far more efficient at detecting clumps in higher-redshift galaxies for which image resolution is lower.

When applied to the same sample of 58,550 galaxies examined by Clump Scout, the FRCNN detector detects a much larger sample, as it is much better at locating clumps at higher redshift and with smaller apparent size. The FRCNN finds 11,002 galaxies with at least one clump with $f_{\rm Lu} > 3\%$ compared with only 3002 such galaxies in the Clump Scout sample; the vast majority of this galaxy excess ($\gtrsim$80%) can be found at redshifts above 0.05 (see Figure 1). In spite of this difference, we find no significant evolution in the distribution of sSFR or clump $f_{\rm Lu}$ in galaxies with respect to redshift, as shown in Figure 2, and clump and galaxy properties appear constant out to a redshift of 0.08. Further, the distributions of mass and sSFR are very similar between the two samples. We therefore attribute the higher number of FRCNN clumpy galaxies at higher redshift to incompleteness in the Clump Scout sample. It is possible that human classifiers had greater difficulty identifying clumps in fainter, lower-resolution images, while the FRCNN was able to extend the sample to these images with greater completeness. For this paper, we limit our science sample to $z < 0.0325$, so evolution in the FRCNN sample at increasing redshift is unlikely to affect our results. Figure 3 contains a random selection of clumpy galaxies identified within our science sample.

2.4. Notes on Our Clump Selection Criteria

In order to be added to the science sample, a clump must be detected in an *irg* composite image and then meet a fractional luminosity criterion in the *u* band. We opted not to use the *u* band for the initial detection step, as it is noisy in SDSS and prone to false detections that must be confirmed in other bands. However, *u*-band photometry is the best available way to select clumps, because it is the closest available SDSS band to the near-UV (∼2500 Å), which has been used to select clumps in a number of high-redshift studies, such as T. Shibuya et al. (2016) and Y. Guo et al. (2018). It is therefore used to select the brightest clumps for which noise presents the least problem. Analysis by D. Adams et al. (2022) demonstrated that the clump sample selected by this *u*-band criterion is comparable to that selected by the near-UV, though values for $f_{\rm Lu}$ tend to be slightly smaller than their $f_{\rm LUV}$ counterparts. Our multistep filtering process therefore selects clumps that are not only visually similar to those identified by volunteers in the *i*, *r*, and *g* bands but also meet similar near-UV photometric criteria to those identified in high-redshift galaxy surveys. We plot the *u*- and *g*-band photometric properties of a sample of clumps in Figure 4 for reference.

Most past studies have selected for clumps with $f_{\rm Lu} > 8\%$; these clumps are extremely rare in local galaxies, and tend to dominate the morphologies of their hosts. However, in this paper (as in D. Adams et al. 2022), we also use a $f_{\rm Lu} > 3\%$ criterion to select clumps. This value was selected as it is just above the completeness limit for clumps in SDSS. While the clumps selected by the 3% criterion are less extreme, the larger size of the resulting sample enables more precise statistical analysis and comparison with other studies.

In Figure 5, we plot the distribution of galaxy mass and sSFR for clumpy galaxies selected using the 3% and 8% criteria. These distributions match quite closely and differ from those of the overall galaxy population: Clumpy galaxies tend to have lower masses and higher sSFRs than the general population, regardless of whether the 8% or 3% selection criterion is used; this bolsters the case for using the less stringent 3% criterion, since it selects for a similar clumpy galaxy population. Additionally, the galactocentric distances of clumps selected under the 3% or 8% criteria are very similar. It is notable that the clumps exceeding $f_{\rm Lu} > 8\%$ have a slightly (∼0.1 mag) redder $g - r$ color distribution than clumps for which $f_{\rm Lu} > 3\%$. One possibility is that highly luminous clumps tend to be longer lived: it has been predicted that ex situ clumps, as well as long-lived in situ clumps, are older, redder, and more massive than young in situ clumps (e.g., N. Mandelker et al. 2017; A. Zanella et al. 2019; A. Dekel et al. 2022), which could explain the color difference we observe.

3. Selecting the Environmental Analysis Sample

A major difficulty when estimating the environmental densities of galaxies in large surveys is that, while the 2D projected location of a galaxy can be calculated very precisely, the galaxy's line-of-sight distance is prone to scatter. In spectroscopic surveys, redshifts can be estimated with great precision; SDSS's spectroscopic redshifts typically have an error on the order of $\Delta z \sim 0.001$, corresponding to $\pm 300\,{\rm km\,s^{-1}}$ in recession velocity. However, the peculiar velocities of galaxies (particularly those in large clusters) can be up to $\pm 1000\,{\rm km\,s^{-1}}$. As a result, using redshift to estimate the line-of-sight distance to a galaxy is inherently prone to error (an effect known as "redshift space distortion"). In order to mitigate this effect, measures of galaxy environments include all galaxies with line-of-sight velocities within $\pm 1000\,{\rm km\,s^{-1}}$ of a target galaxy, and use the sky-projected 2D distances between galaxies to approximate their separation. M. C. Cooper et al. (2005) found that the projected Nth nearest neighbor distance is the most reliable estimator of local environmental density, and this metric has become standard in other works (e.g., I. K. Baldry et al. 2006; K. Kovač et al. 2010; M. R. Haas et al. 2012; L. Kawinwanichakij et al. 2017).

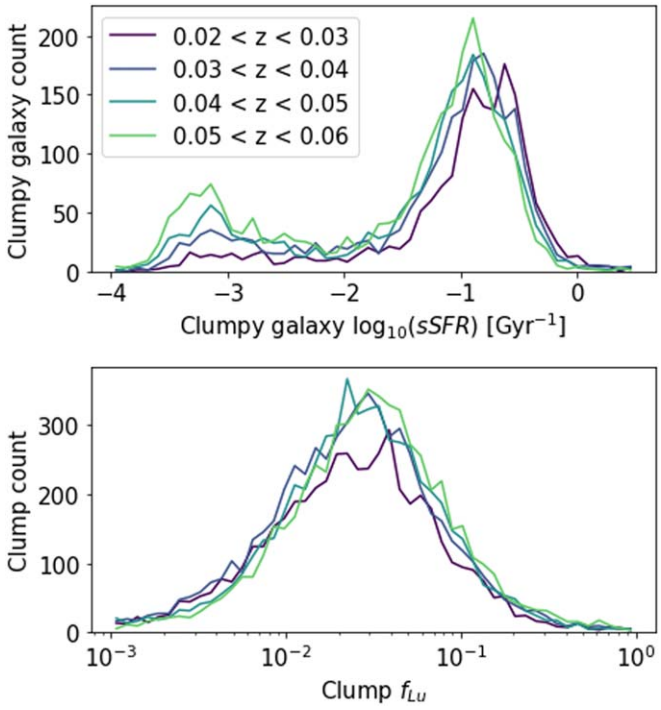

Figure 2. Comparison of clump and galaxy properties in different redshift bins for clumpy galaxies selected by the FRCNN. Clumps' brightness as well as their host galaxies' sSFRs do not significantly evolve between $0.02 < z < 0.06$, suggesting that the FRCNN is finding the same types of objects over this redshift range (at higher redshifts, the incompleteness of the SDSS sample makes this comparison difficult).

3.1. Selection of the Target and Environment Samples

From our parent sample, we select a smaller, volume-limited subsample of galaxies as our "target sample" for the environmental analysis. Target galaxies are limited by redshift and stellar mass to $0.02 < z < 0.0325$ and $M > 9.4\ M_{\odot}$ to ensure mass completeness. These limits are in addition to the cuts applied to the Galaxy Zoo 2 parent sample, i.e., a brightness limit (Petrosian half-light magnitude brighter than 17 in the r band) and a size limit (Petrosian 90% flux radius larger than 3″). The narrow redshift selection ensures that there should be minimal variation in identified clumps in the target sample, as the SDSS resolution and magnitude limits do not vary widely between the redshift bounds. In total, there are 9964 galaxies in the target sample.

To verify that our clump sample is complete within our target sample's bounds, we compute the fraction of galaxies with at least one detected clump exceeding 3% of its host's u-band luminosity, $f_{\mathrm{clumpy,3\%}}$ (see Figure 6). Since there is no significant evolution in the fraction of clumpy galaxies within our redshift limits, we find that the target sample is properly redshift and mass limited with respect to clumps. In order to constrain the detection incompleteness with respect to galaxy size, we additionally measure the trend in the number of clumps per galaxy as a function of galaxy Petrosian radius for star-forming galaxies (Figure 7). We do find that, in our faintest clump sample ($f_{\mathrm{Lu}} > 3\%$), there is a significant drop in the number of clumps detected per galaxy for galaxies with r-band Petrosian radii below 6″ or above 20. However, it is unclear whether this is an observational effect or an intrinsic one, as compact or extended galaxies may be less prone to producing or maintaining clumps. In particular, it is unlikely that detection incompleteness is responsible for the reduction in clumpiness for extended galaxies, which tend to be nearer, brighter, and easier to resolve. For compact galaxies, we found that doubling our size selection limit from 3″ to 6″ had a negligible effect on the environmental trends in Section 4. As a result, we note that there are fewer detected clumps in highly compact or extended galaxies but elect to leave these galaxies in the target sample for analysis.

We then select a wider galaxy sample from SDSS to trace the environment, known as the "environment sample." This sample consisted of all galaxies with spectroscopic redshifts between $0.017 < z < 0.355$ (i.e., a broader redshift range than the target sample by 0.003 on either side), as well as stellar masses $M > 9.5\ M_{\odot}$ to ensure mass completeness within this window. These selections yielded 20,024 galaxies for the environment sample. Figure 6 visually demonstrates our choice of sample and our tests for its completeness, while Figure 8 displays a map of the environment sample with clumpy galaxies highlighted within it.

3.2. Estimation of Environmental Density

In order to compute the density of the local environment around each galaxy, we use a projected fifth nearest neighbor technique. We first select all environment galaxies whose line-of-sight velocity is within $\pm$ 1000 km s^{-1} of the target galaxy to exclude foreground and background galaxies. Then, we compute the sky-projected distances from each target to neighboring environment galaxies (assuming all environment galaxies are at the same redshift as the target). Taking d_5 to be the distance to the fifth nearest neighbor, the surface overdensity of a target galaxy Σ_5 is computed as

$$\Sigma_5 = \frac{5}{\pi d_5^2}. \tag{3}$$

Using Σ_5, we compute the fractional overdensity for each target galaxy $(1 + \delta_5)$, where

$$\delta_5 = \frac{\Sigma_5 - \langle \Sigma_5 \rangle}{\langle \Sigma_5 \rangle}. \tag{4}$$

Here, $\langle \Sigma_5 \rangle$ is the mean projected surface density of all galaxies in the target sample.

Throughout this paper, we use $\log_{10}(1 + \delta_5)$ as our main density probe. (Physically, values of $\log_{10}(1 + \delta_5) = -2$, 0, and 2 correspond approximately to projected fifth nearest neighbor distances of 10 Mpc, 1 Mpc, and 100 kpc respectively.)

We ensure that our estimates of environmental density are reliable by recomputing them using the method from I. K. Baldry et al. (2006, herein called the Baldry+2006 method). Baldry+2006 also uses a projected Nth nearest neighbor method to compute environmental density, but defines their environment sample using an absolute magnitude cut of $M_r < -20$ to $1.6(z - z_0)$ rather than a mass cut, and computed the projected surface density by averaging Σ_4 and Σ_5. The absolute magnitude cut in practice selects for a more massive sample of galaxies than our environment selection cut of $M^* > 10^{9.5}\ M_{\odot}$, with a resulting median mass of $10^{10.4}\ M_{\odot}$ rather than $10^{10.1}\ M_{\odot}$ for our environment sample. Despite this, the fractional overdensity values computed by these methods are highly correlated, with a Pearson correlation coefficient of 0.92. We tested the consistency of these methods by computing the mean physical quantities of our target galaxies, including

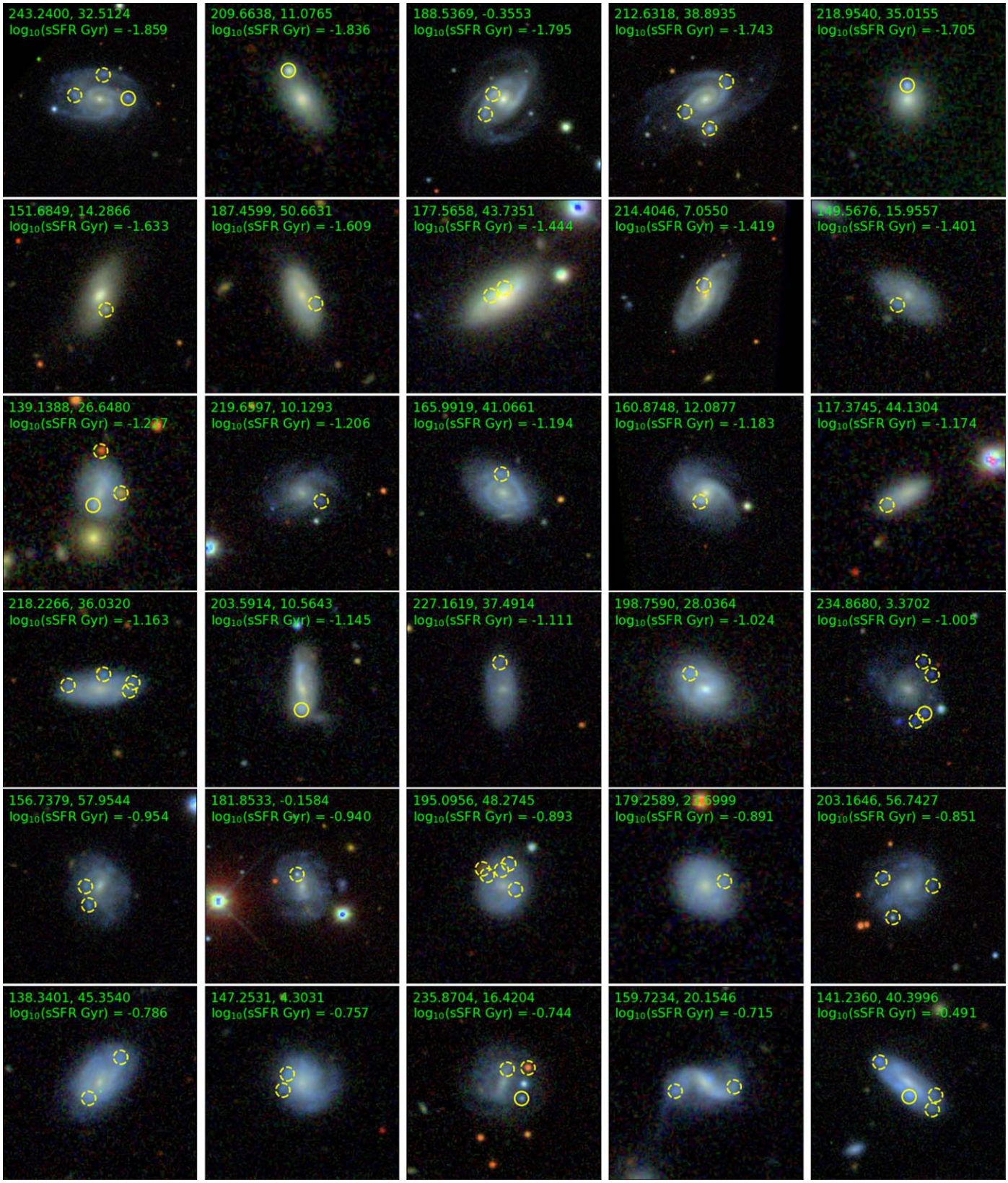

Figure 3. A stamp sheet of star-forming clumpy galaxies (galaxies with at least one clump with $f_{\rm Lu} > 3\%$) drawn at random from our target sample. Overlaid text lists each galaxy's R.A., decl., and sSFR. In rows 1 and 2, galaxies were selected such that sSFRs (in $\rm Gyr^{-1}$) ranged from 10^{-2} to $10^{-1.4}$; in rows 3 and 4, from $10^{-1.4}$ to 10^{-1}; and in rows 5 and 6, above 10^{-1}. Clumps are circled with a solid line if $f_{\rm Lu} > 8\%$ and a dashed line if $3\% < f_{\rm Lu} < 8\%$.

color, mass, and sSFR, in bins of fractional overdensity, using both the Baldry+2006 fractional overdensity estimates and our own. The resulting trends of galaxy properties with the environment were consistent (see Figure 9); from this, we conclude that the results drawn from our environmental density estimates are comparable to results from the Baldry+2006

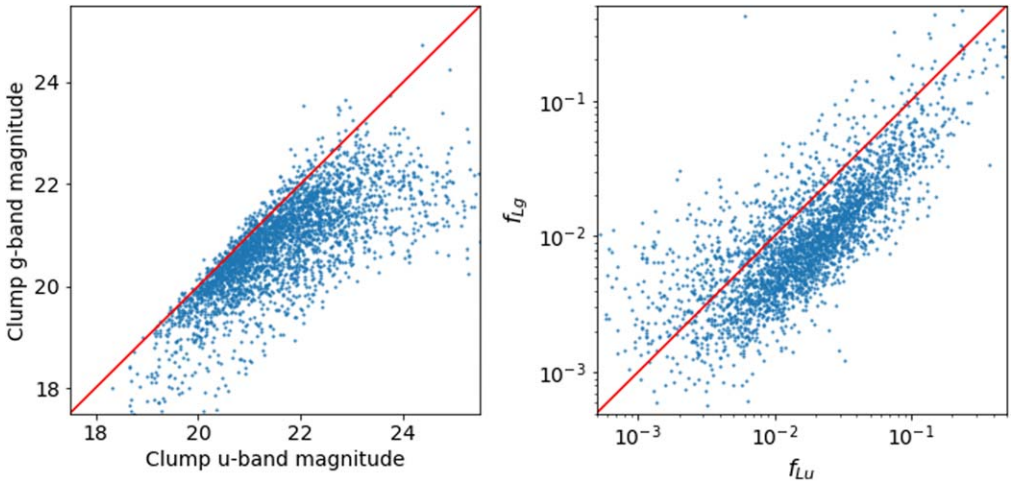

Figure 4. The *u*- and *g*-band distributions of clump magnitudes and luminosity fractions ($f_{\rm Lu}$ and $f_{\rm Lg}$), for clumps in a volume-limited galaxy sample (galaxy mass greater than $10^{9.4}$ $M_\odot$ and redshift between $0.02 < z < 0.0325$). Many clumps identified by the FRCNN detector in *irg* composite images are very faint in the *u* band; these clumps are removed from the science sample by the $f_{\rm Lu}$ selection criterion. $f_{\rm Lg}$ is significantly smaller than $f_{\rm Lu}$ for most clumps, and $f_{\rm Lg}$ is not as well correlated with clumps' near-UV fractional luminosity $f_{\rm LUV}$. For this and other reasons, we use a $f_{\rm Lu}$ selection for our science sample.

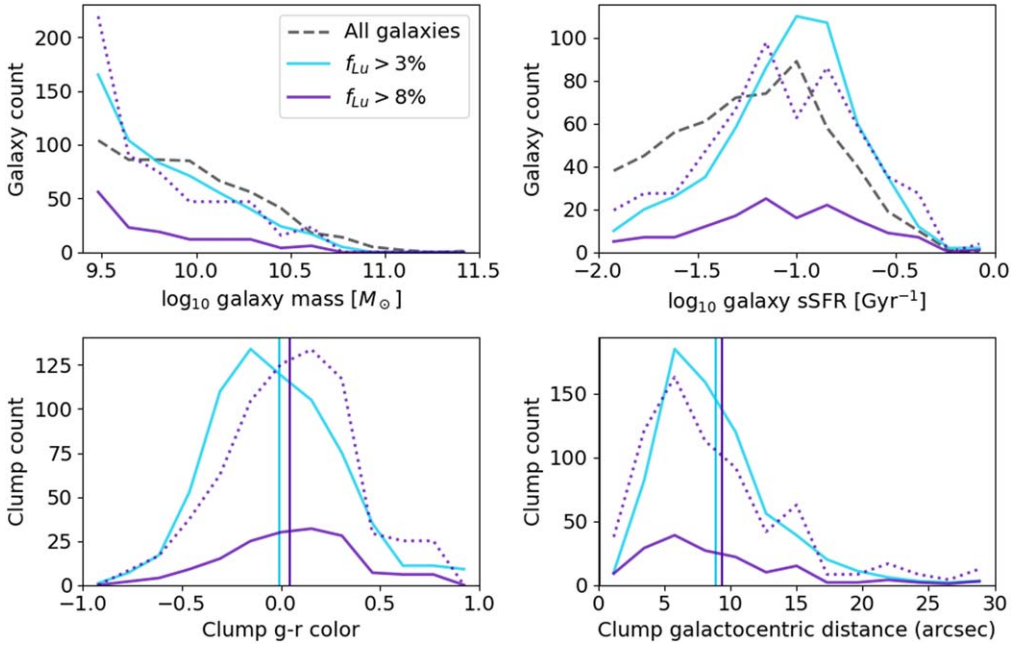

Figure 5. Top: the distributions of galaxy mass, galaxy sSFR, and clump color under three different selection criteria. The cyan and purple curves represent clumpy galaxies selected using $f_{\rm Lu} > 3\%$ and $f_{\rm Lu} > 8\%$, respectively, while the gray dashed curve represents a sample of all galaxies (clumpy and nonclumpy) whose number count matches that of the 3% sample. In addition, a purple dotted line represents the $f_{\rm Lu} > 8\%$ result adjusted upwards so that the total number count matches that of the 3% sample. A volume-limited galaxy sample is used (galaxy mass greater than $10^{9.4}$ $M_\odot$ and redshift between $0.02 < z < 0.0325$). Regardless of which criterion is used, clumpy galaxies tend to be lower mass and have a higher sSFR than the overall galaxy population. Bottom: the distributions of clump colors and galactocentric distances for clumps selected using $f_{\rm Lu} > 3\%$ and $f_{\rm Lu} > 8\%$. (Colors and styles match the top row.) Clumps occur at similar galactocentric distances irrespective of how they are selected. However, the most luminous clumps ($f_{\rm Lu} > 8\%$) are about 0.1 mag redder than the $f_{\rm Lu} > 3\%$ population.

method, while taking advantage of the lower mass limit of our environment sample.

4. Results: $f_{\rm clumpy}$ as a Function of Environment

In order to trace the prevalence of clumps across different environments, we compute the fraction of clumpy galaxies $f_{\rm clumpy}$ across many galaxy selection conditions, where

$$f_{\rm clumpy} = \frac{\rm N(star\text{–}forming\ galaxies\ with\ \geqslant\ 1\ clump)}{\rm N(star\text{–}forming\ galaxies)}, \quad (5)$$

$f_{\rm clumpy}$ has been used in many past papers to trace the rate of clumpiness across different galaxy masses and redshift (e.g., Y. Guo et al. 2015; T. Shibuya et al. 2016; D. Adams et al. 2022). We compute both $f_{\rm clumpy,3\%}$ (the clumpy fraction using only clumps for which $f_{\rm Lu} > 3\%$) and $f_{\rm clumpy,8\%}$ ($f_{\rm Lu} > 8\%$) for our sample. While some of these papers apply an "incompleteness correction" to $f_{\rm clumpy}$ to correct for clumpy galaxies that were not detected, this work uses the observed fraction. This is because we do not use the absolute values of $f_{\rm clumpy}$ in our analysis, but only compare $f_{\rm clumpy}$ between different selection conditions, and our selection limits make it unlikely that the

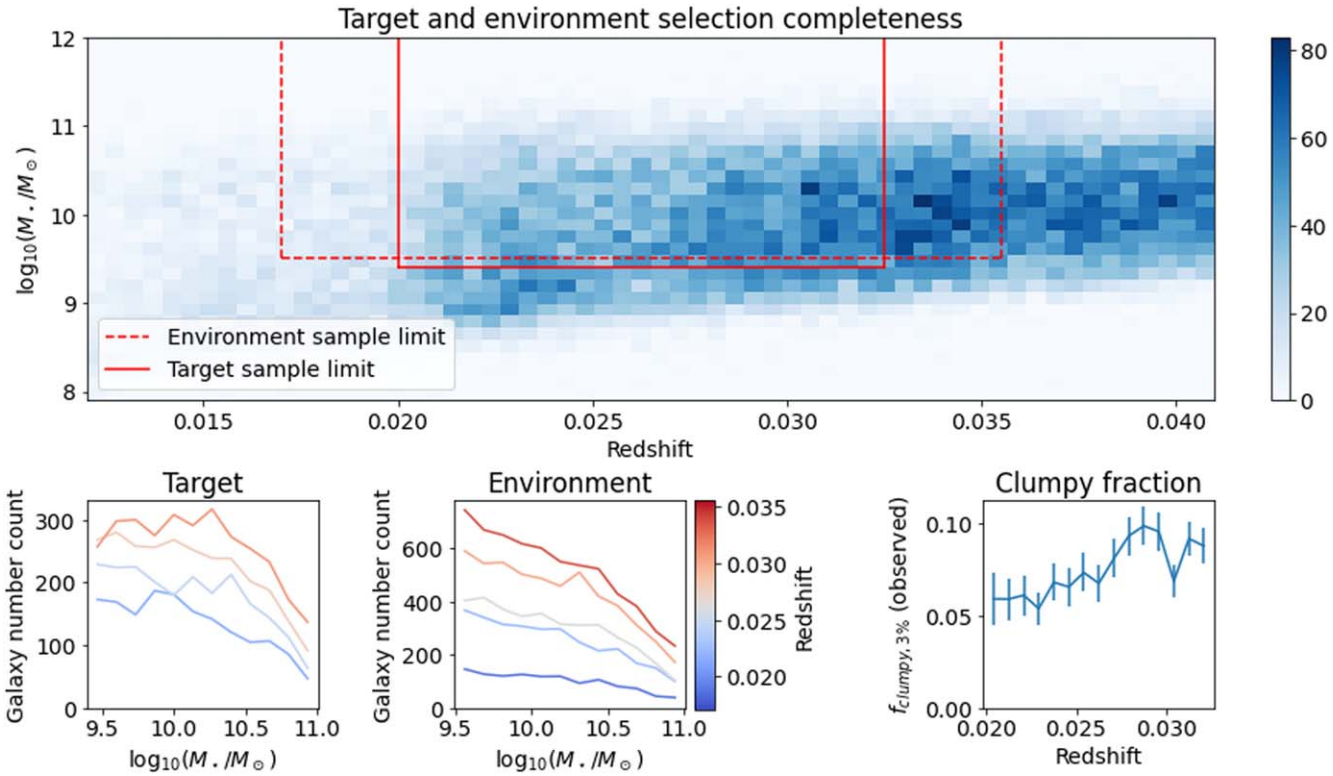

Figure 6. Selection of target and environment samples. Top: The 2D histogram plots the distribution in mass and redshift of all SDSS galaxies with spectroscopic redshifts between $0.012 < z < 0.041$. Solid lines mark the selection limits of the target sample, while dashed lines mark the selection limits of the environment sample. Bottom: To ensure that our sample is properly volume limited, we show the number counts of target and environment galaxies vs. mass, binned by redshift. In none of the redshift bins do we see a significant drop-off in the number count of galaxies as mass decreases. We also plot the clumpy fraction of target galaxies against redshift, using the $f_{\rm Lu} > 3\%$ rule to select clumps. The clumpy fraction also does not evolve significantly with redshift, confirming that our sample is complete with respect to clumps selected under this rule.

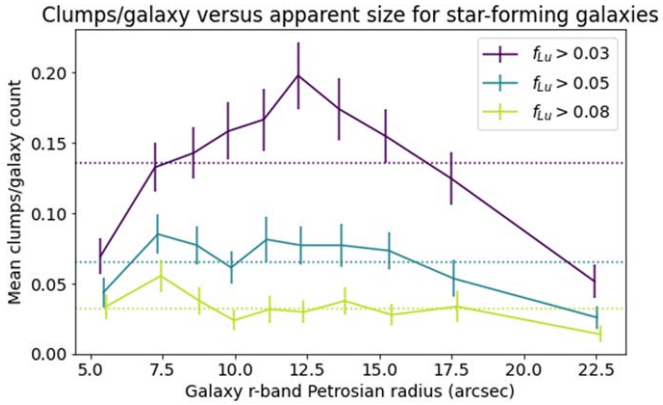

Figure 7. The mean number of clumps per galaxy as a function of galaxy apparent size (using the r-band 90% Petrosian radius) for star-forming galaxies ($\mathrm{sSFR} > 10^{-2}\,\mathrm{Gyr}^{-1}$). There is a peak in this function for galaxies with moderate radii, particularly for less luminous clumps. It is unclear whether galaxies at larger and smaller radii are prone to detection incompleteness, or if there is a physical cause for this trend; because of this, and the fact that their inclusion/exclusion had a negligible effect on our results, we opted to include these galaxies in the target sample.

completeness of our sample is significantly different under our different selection conditions.

In order to narrow down the cause of any environmental trends in $f_{\rm clumpy}$, we analyze the trends not only for all star-forming galaxies, but also for galaxy samples binned by several different tracers of star formation rate (SFR). This is motivated by the fact that clumps are thought to be far more common in star-forming galaxies, and SFR is highly correlated with environment. We therefore wish to know how much of the trend in $f_{\rm clumpy}$ is due to the correlation between SFR and environment, and how much is independent of it. In order to control for the star formation properties of galaxies, we divide them into bins by three different tracers of star formation: sSFR, SFR, and $g-r$ color. The values of sSFR and SFR were both selected from the MPA–JHU catalog (G. Kauffmann et al. 2003; J. Brinchmann et al. 2004), while $g-r$ color was computed as the difference of the g-band and r-band Petrosian magnitudes computed by SDSS. We use all three quantities because each probes the state of star formation in our target galaxies in a different way. The SFR is a direct estimate of the stellar mass forming per unit time while selecting star-forming galaxies using sSFR emphasizes lower mass galaxies with particularly high gas fractions. $g-r$ color was included as well because the g- and r-band magnitudes are computed over the entire spatial extent of each target galaxy, whereas both SFR and sSFR rely on the galaxy's spectrum, as measured with a 3″ diameter fiber around each galaxy's center. For brevity, we refer to these quantities (sSFR, SFR, and color) as "star formation tracers" or "SF tracers."

We selected bin boundaries for each SF tracer such that approximately one-third of the star-forming galaxy sample falls into each bin. For each tracer, star-forming galaxies were

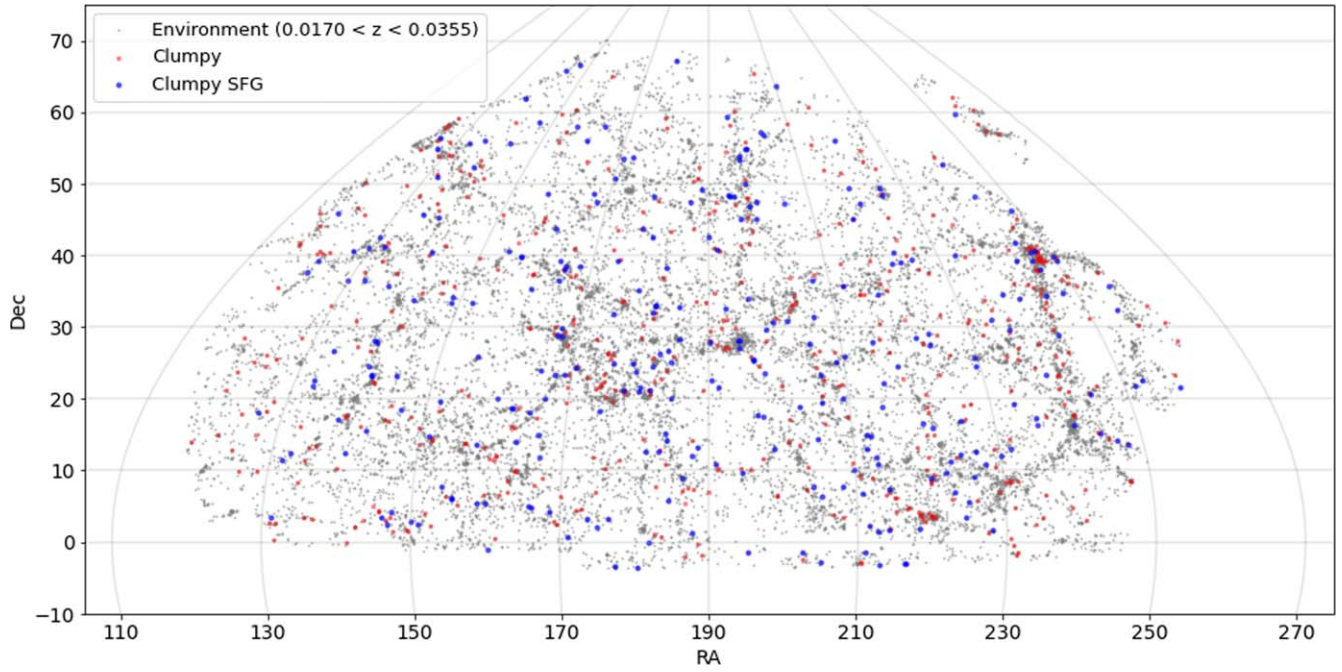

Figure 8. Projected spatial distribution of SDSS galaxies in the target and environment samples. Coordinates are plotted in an equal area projection, with vertical grid lines tracing lines of equal R.A. and horizontal grid lines tracing lines of equal decl. There are many visible large-scale structures present in the environment sample, including clusters, filaments, and voids. The target galaxies shown are selected to be clumpy (with at least $f_{\rm Lu} > 8\%$); blue points denote star-forming galaxies with sSFR > 0.1 Gyr^{-1}, while red points denote galaxies with sSFR < 0.1 Gyr^{-1}.

selected by the following cuts:

$$\begin{aligned} \log_{10}(\mathrm{sSFR}) &> -2 \\ \log_{10}(\mathrm{SFR}) &> -1 \\ M_g - M_r &> 0.7 \end{aligned} \tag{6}$$

(using units of Gyr^{-1} for sSFR and $M_\odot$ Gyr^{-1} for SFR). The particular bin boundaries chosen were

$$\begin{aligned} \log_{10}(\mathrm{sSFR}) &: (-1.37,\ -1.02) \\ \log_{10}(\mathrm{SFR}) &: (-1,\ -0.25) \\ M_g - M_r &: \ (0.75,\ 0.55). \end{aligned} \tag{7}$$

We refer to the resulting bins as the "low," "moderate," and "high star formation" bins, as the low and high boundaries fall just below and just above the star-forming main sequence respectively (see Figure 10).

For the purposes of plotting, we separate our target galaxies into five bins of $\log_{10}(1 + \delta_5)$ such that 20% of galaxies fall into each bin. When all galaxies in our target sample are divided this way, the most isolated galaxies are on average bluer by $\sim$0.1 mag, $\sim$0.1 dex less massive, and a factor of $\sim$5 higher in sSFR than galaxies in the densest environments (as shown in Figure 9). These changes are in line with past studies of galaxy properties with respect to the environment, such as G. Kauffmann et al. (2004) and I. K. Baldry et al. (2006).

In Figure 11, we plot the trend in $f_{\rm clumpy}$ versus environment across our full set of selection conditions. Figure 11(A) plots the trend for all star-forming galaxies in the target sample (sSFR $> 10^{-2}$ Gyr^{-1}), using both a $f_{\rm Lu}$ cut of 3% and 8% to select clumps. We refer to this sample as the "all galaxies" sample. In both cases, we find a factor of $\gtrsim$2 more clumpy galaxies in the most isolated bin compared to the most clustered bin. Under the more stringent $f_{\rm Lu} > 8\%$ clump selection, 4.0% of the galaxies in the most isolated bin were clumpy, compared to only 1.4% of those in the most clustered bin (for $f_{\rm Lu} > 3\%$, the numbers are 12.5% and 6.3% respectively).

To quantitatively test the significance of the $f_{\rm clumpy}$ trend, we perform a two-sample Anderson–Darling (AD) test on the environment $\log_{10}(1 + \delta_5)$ distributions for clumpy and non-clumpy galaxies. The AD test is used as it has been found to be more sensitive than similar tests (such as the Kolmogorov–Smirnov test) and because it emphasizes differences in the distribution tails (i.e., in the densest and most isolated environments) where the difference in galaxy properties is expected to be largest (S. Engmann & D. Cousineau 2011). For both the $f_{\rm Lu} > 8\%$ and 3% cases, we measure a difference between these distributions with significance less than 0.5%. To ensure that this test is sensitive enough to work on the smaller, binned samples of galaxies, we recompute this trend on a random subsample of one-third of the "all galaxies" galaxy sample (1722 of 5148). While the results are less significant, a statistically significant trend is still detected for each of the $f_{\rm Lu} > 8\%$ and 3% cases. (The full set of AD test statistics and significance values is presented in Table 1) We conclude that this test not only detects a consistent environmental trend in $f_{\rm clumpy}$, but that it is also sensitive enough to work on the binned galaxy samples with sizes on the order of $N \sim 1700$ galaxies.

In Figure 11(B), we plot the trend in $f_{\rm clumpy}$ versus environment for galaxies binned by each of our SF tracers. For two of the tracers, sSFR, and $g - r$ color, the majority of tests find no significant trend. Mildly significant trends are found for galaxies in the moderate sSFR bin ($p < 3.4\%$ for $f_{\rm Lu} > 8\%$), as well as in the reddest color bin ($p < 2.9\%$ for $f_{\rm Lu} > 3\%$). However, no similarly significant trends are recovered in other bins or in alternate clump selections, which suggests that these trends may be statistical anomalies rather than valid results. Further, none of the trends achieved the significance of the "all galaxies" sample, even when similarly

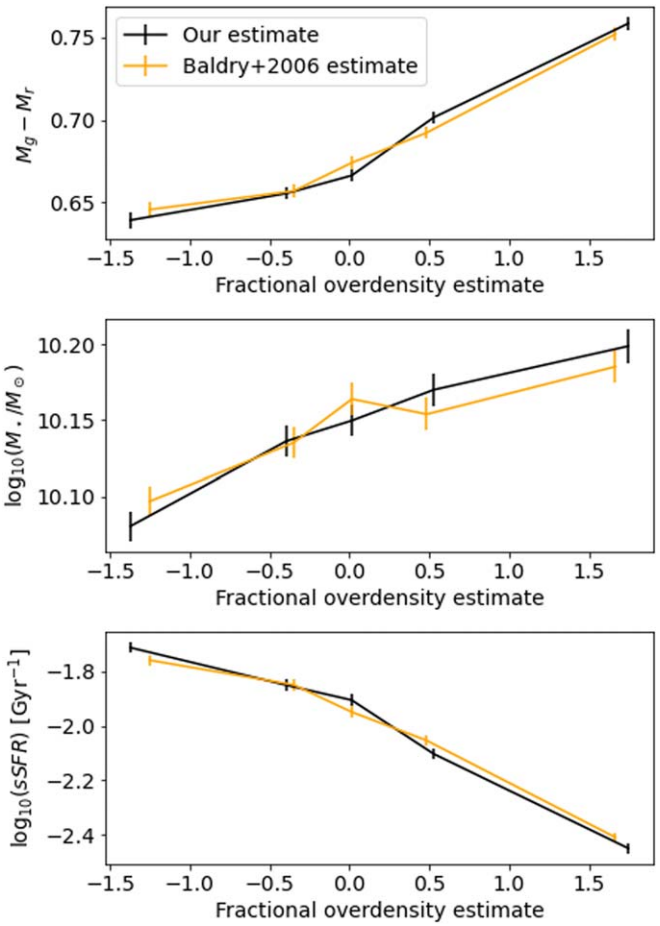

Figure 9. Comparison of average galaxy color, mass, and sSFR, binned by two different estimates of environmental density (our own method and the method in I. K. Baldry et al. 2006). Galaxy properties are statistically consistent between the two estimates, which suggests that our environment estimate is meaningfully similar to that of I. K. Baldry et al. (2006) despite the differences in selection criteria.

sized galaxy samples were used. Overall, the fact that much weaker $f_{\rm clumpy}$ trends are recovered when controlling for sSFR and color suggests that much of the environmental trend in $f_{\rm clumpy}$ can be attributed to the correlation between sSFR/color and environment, and is independent of other environmental influences.

Interestingly, we find that a few significant environmental trends in $f_{\rm clumpy}$ did exist when galaxies were binned by SFR. Of the six samples analyzed (three selection bins with two clump selection criteria each), AD tests found that three have significant differences in the environment distribution of their clumpy and nonclumpy populations (with p-values below 0.01). Notably, each galaxy's SFR is equal to its sSFR times its mass; the only difference between the sSFR and SFR binning conditions is therefore in the way that these bins control for mass. Given that there is a correlation between galaxy mass and environment, we posit that it is the correlation with mass, rather than with a galaxy's star formation history, that yields the trend we observed.

We conclude that there is an environmental trend in $f_{\rm clumpy}$ among all star-forming galaxies, and that this trend is detectable even in our smaller, binned galaxy samples. However, there does not appear to be a significant trend in $f_{\rm clumpy}$ when controlling for tracers of star formation. While we do detect some significant environmental trends when controlling for SFR in particular, we suggest that it is the correlation between environment and galaxies' mass, rather than their star formation history, that is responsible for these trends, since they are not recovered under other conditions.

5. Discussion

A major goal of this study is to investigate whether we can use the local environment around galaxies to distinguish between different possible clump formation modes, and in particular the in situ and ex situ modes. This is motivated by the fact that in situ and ex situ formation are both thought to be modulated by interactions between the host galaxy and its surroundings. Here, we briefly review the in situ and ex situ modes of clump formation and how they correlate with the host galaxy's environment.

In several past works, in situ clump formation has been linked to the availability of gas, particularly gas accreted along smooth, cold filaments. The infalling gas adds kinetic energy to disks, pushing them toward turbulence (A. Dekel et al. 2009); further, clumps that form within turbulent, gas-rich disks are able to accrete enough gas as they travel to withstand external perturbations (J. Fensch & F. Bournaud 2021). There is a broad body of evidence to suggest that gas availability should be much higher in low-mass halos. Several papers (e.g., D. Kereš et al. 2005, 2009; A. Dekel & Y. Birnboim 2006) have found that gas accretion broadly follows along two modes: the "hot mode," in which gas is shock heated as it accretes onto the halo, and the "cold mode," in which the infalling gas skips this shock-heating phase and is able to coalesce and form stars without requiring an intermediate cooling phase. Because higher-mass halos are more likely to shock-heat infalling gas, it is far more likely for cold-mode accretion to occur in low-mass halos—i.e., among isolated galaxies. As a result, isolated galaxies in low-mass halos are more likely to experience rapid gas accretion, star formation, and clump formation compared to their crowded, high-mass halo counterparts.

In contrast to in situ clump formation, ex situ clumps do not require gas availability in order to form. Instead, ex situ clumps are thought to be the remnants of minor mergers, whose size and density cause them to present as bright point sources in observations. It is expected that in denser environments, the merger rate, particularly the minor merger rate, is elevated compared to galaxies in the field (e.g., J. M. Lotz et al. 2013; N. K. Hine et al. 2016; C. Watson et al. 2019, though low-redshift results are somewhat limited). Additionally, because the formation of ex situ clumps does not directly require gas availability, ex situ clumps should be significantly more prevalent than in situ clumps in galaxies with lower SFRs. The strength of SFR enhancement diminishes as environments become denser; in clusters, interactions tend to quench star formation (S. L. Ellison et al. 2010; A. Das et al. 2021). Together, this suggests that the fraction of ex situ clumps should be elevated in dense environments and among galaxies with low SFRs.

Among the galaxies in our sample, we do recover a trend in $f_{\rm clumpy}$ with respect to the environment when not controlling for SFR. However, the most striking result in this work is that, when controlling for sSFR or color, this trend is significantly weaker or entirely absent. For galaxies in these groups, no specific environmental condition is preferential for clump formation. Notably, our trend in $f_{\rm clumpy}$ for local star-forming

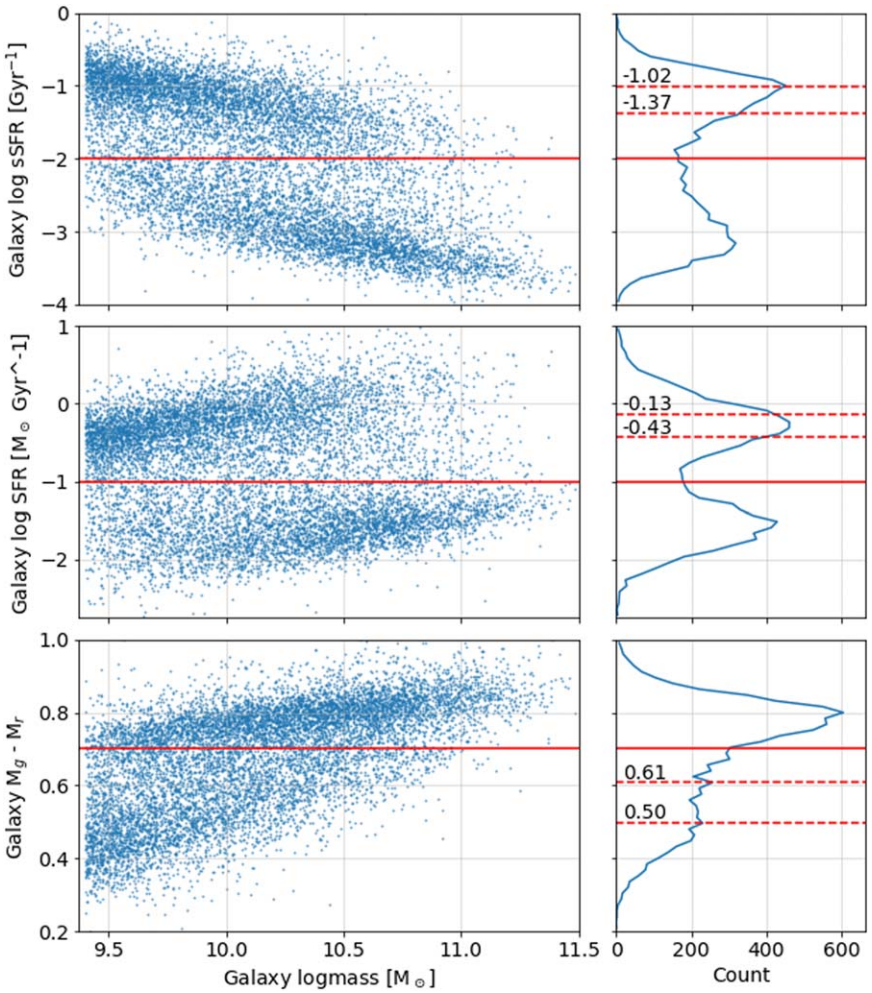

Figure 10. Choice of quantity bins to control for SFR, sSFR, and $g - r$ color. In the left column, we plot the sSFR, SFR and $g - r$ color of our target galaxy sample against their masses. Red sequence galaxies (bounded by a solid red line in each plot) are excluded from the environment analysis. The remaining galaxies are divided into three bins of equal number count. A histogram of each galaxy property is provided in the right column, with bin divisions marked by dashed red lines.

galaxies is very similar to the trend from Z. Sattari et al. (2023) for high-redshift star-forming galaxies, ranging from $0.5 < z < 3$. Across all of these redshifts, no correlation has been found between the environment of star-forming galaxies and the prevalence of clumps.

We broadly interpret the observed trends in $f_{\rm clumpy}$ when controlling for SFR as evidence that in situ clump formation, rather than ex situ, dominates in the local Universe. In situ clump formation is thought to be driven by many of the same basic secular processes as star formation (i.e., turbulent disk dynamics leading to gas collapse), while ex situ formation requires specific environmental interactions to occur. If ex situ formation was dominant, we would expect to find evidence that $f_{\rm clumpy}$ is more greatly impacted by the environment, particularly for galaxies with lower SFRs which are less likely to form clumps in situ. We would also expect to see a smaller gap between the clumpiness of highly star-forming galaxies, which have highly elevated $f_{\rm clumpy}$ values in our results, compared to their low star-formation counterparts.

Special attention should be given to star-forming clumpy galaxies in cluster environments. According to theory, cold smooth accretion should be far less likely in these galaxies; more generally, cluster galaxies are likely to have a different history of interactions with neighboring galaxies as well as the circumgalactic medium when compared to isolated galaxies. However, the presence of clumps in star-forming galaxies at approximately equal rates in both cluster and field environments broadly suggests that there is no preferred interaction history that produces clumps. Instead, a high rate of star formation seems to be a sufficient condition to explain most clump formation in local galaxies. Given that there is a strong correlation between SFR and gas surface density through the Kennicutt–Schmidt relation (R. C. Kennicutt 1998, 1989), this result also suggests that gas availability may be the primary driver of clumpiness in local galaxies, and the method by which that gas was accreted may play a smaller role.

Recent models and simulations, such as those in J. Fensch & F. Bournaud (2021) and A. Dekel et al. (2022), suggest that a galaxy's gas fraction plays a critical role in the process of forming and maintaining in situ clumps. Briefly, clumps form as local overdensities in a galaxy's gas; as they travel through the disk, they continue to accrete incident gas, replenishing the outflowing gas that they expel via feedback. Thus, both the formation and survival of clumps require a high gas fraction over the entirety of the host galaxy's disk. In galaxies with lower gas fractions (25%, as compared to a higher fraction of 50%, as examined in J. Fensch & F. Bournaud 2021), clump formation is mostly limited to the spiral arms of a galaxy.

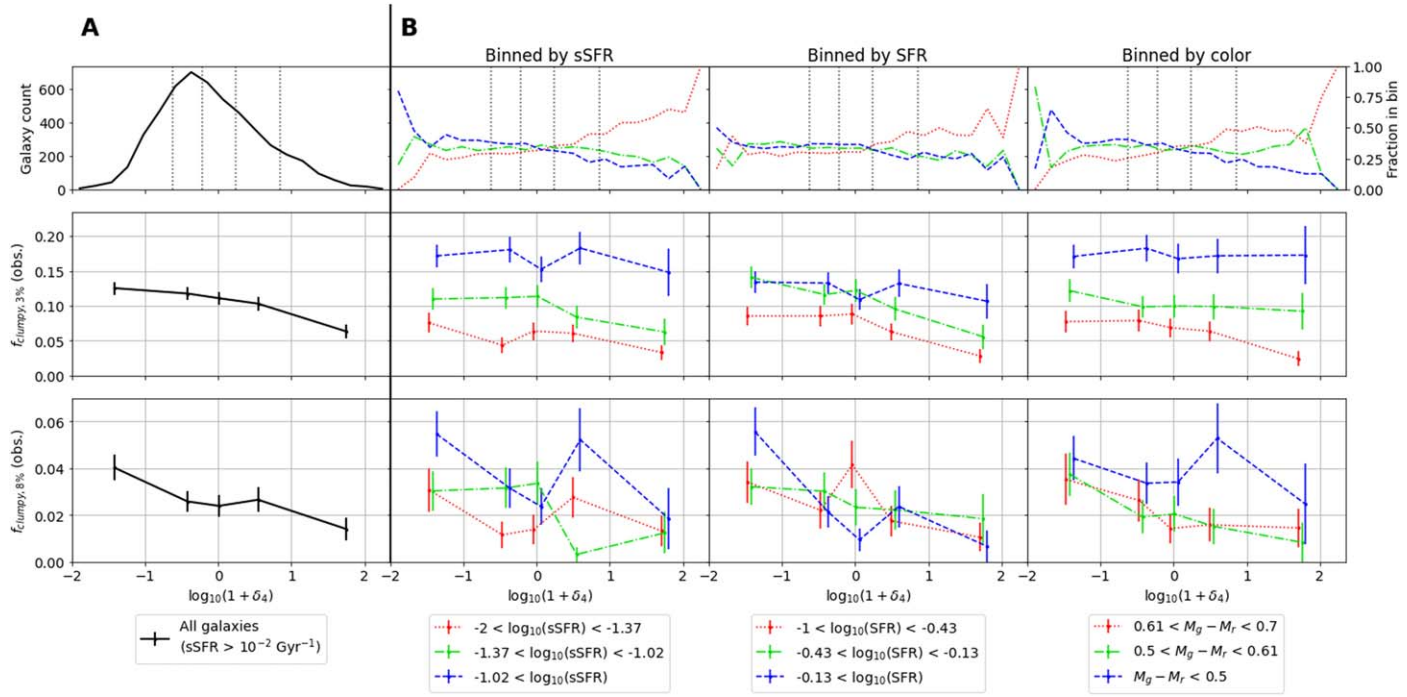

Figure 11. The trend in $f_{\rm clumpy}$ vs. environmental density under a number of conditions. In the left column (A), the trends for all target galaxies are shown together, while subsequent columns (B) show the trend for galaxies divided into bins by sSFR, SFR, and color, respectively. The top row of plots shows a histogram of target galaxies (A) as well as the fraction of target galaxies in each property bin with respect to the environment (B), with gray dashed lines demarcating the boundaries between density bins. Rows 2 and 3 plot the observed value of $f_{\rm clumpy}$ vs. environmental density for clumps that exceed 3% (row 2) and 8% (row 3) of the host galaxy's u-band flux. (Units: sSFR quantities are provided in Gyr^{-1}, and SFR in $M_\odot$ Gyr^{-1}.)

Table 1
Two-sample Anderson–Darling Test Statistics, Clumpy vs. Nonclumpy Galaxy Environments

		$f_{\rm Lu} > 3\%$		$f_{\rm Lu} > 8\%$	
	$N_{\rm galaxies}$	AD Stat.	$p_{\rm max}$	AD Stat.	$p_{\rm max}$
All galaxies ($\log_{10}$(sSFR) > -2)	5148	7.6	1.0×10^{-3}	4.8	1.6×10^{-3}
(Downsampled)	1722	4.6	4.9×10^{-3}	3.3	1.4×10^{-2}
$-2 < \log_{10}$(sSFR) < -1.43	1724	0.87	0.14	−0.30	0.25
$-1.37 < \log_{10}$(sSFR) < -1.02	1696	0.89	0.14	2.4	3.4×10^{-2}
$-1.02 < \log_{10}$(sSFR)	1729	−0.48	0.25	1.0	0.12
$-1 < \log_{10}$(SFR) < -0.43	1838	4.8	4.1×10^{-3}	1.0	0.12
$-0.43 < \log_{10}$(SFR) < -0.13	1788	4.0	7.7×10^{-3}	0.27	0.25
$-0.13 < \log_{10}$(SFR)	1819	−0.55	0.25	11	1.0×10^{-3}
$0.61 < M_g - M_r < 0.7$	1457	2.5	2.9×10^{-2}	1.5	7.6×10^{-2}
$0.5 < M_g - M_r < 0.61$	1514	−0.56	0.25	1.096	0.12
$M_g - M_r < 0.5$	1520	−1.1	0.25	−0.749	0.25

Note. Results from two-sample Anderson–Darling tests on clumpy and nonclumpy galaxy environment distributions under all selection conditions. The provided $p_{\rm max}$ values are approximate ceilings on the significance level (as precise significance values are not straightforward to compute). In columns 3 and 4, we select clumps using the $f_{\rm Lu} > 3\%$ criterion; in columns 5 and 6, we select them with the more stringent $f_{\rm Lu} > 8\%$. (Units: sSFR quantities are provided in Gyr^{-1}, and SFR in $M_\odot$ Gyr^{-1}.)

Between the shearing forces of the spiral arms and the lack of accretable gas in the rest of the disk, clumps typically survive for a much shorter time in disks with a low gas fraction. O. Ginzburg et al. (2021) similarly found that high-redshift clumps tend to be long lived only if they are sufficiently massive, requiring a large gas mass to collapse or be accreted during the clump's lifetime; smaller clumps tended to be disrupted much sooner by feedback and shocks.

Any theory of clump formation must explain the significant reduction in $f_{\rm clumpy}$ in star–forming galaxies between $z \gtrsim 0.5$ and $z \sim 0$. Even for star-forming galaxies selected by the same criteria (i.e., sSFR > 0.1 Gyr^{-1}), there is as much as an order-

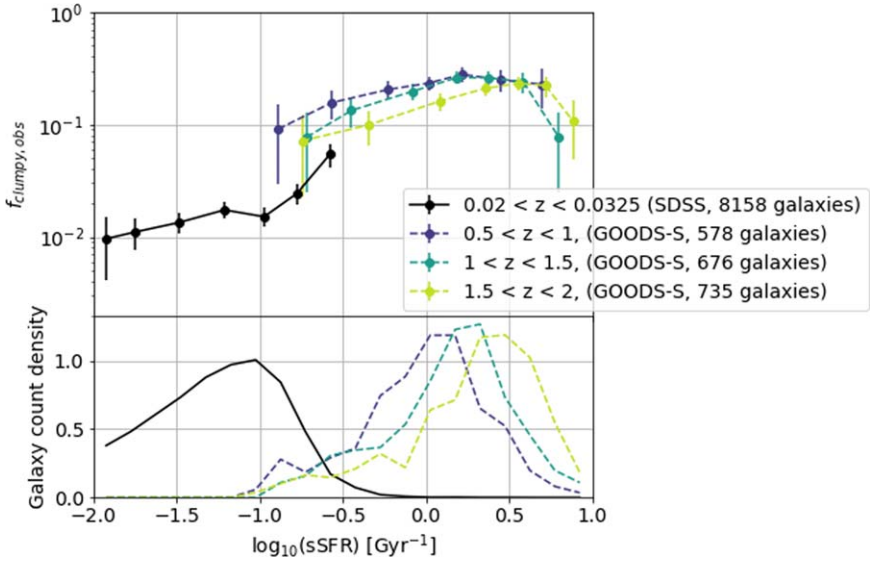

Figure 12. The observed (uncorrected for completeness) clumpy fraction of galaxies as a function of galaxy sSFR at several different redshifts. The SDSS result ($z \sim 0$) comes from this work, while the GOODS-S results ($0.5 < z < 2$) are computed from the Y. Guo et al. (2018) catalog for galaxies with sSFR > 0.1 Gyr^{-1}. Galaxies are mass limited to $M > 10^{9.4}$ $M_{\odot}$, and clumps are selected such that $f_{\rm Lu}$ (in SDSS) or $f_{\rm LUV}$ (in GOODS-S) is above 8%. Galaxies are binned by their sSFR, and error bars are computed as the standard error on a proportion (which underestimates the true error). The clumpy fraction of local galaxies rise significantly at sSFR > 0.1 Gyr^{-1}, and is comparable to, though still smaller than, $f_{\rm clumpy}$ for high-redshift galaxies with similar sSFR. The bottom subplot displays the distribution of galaxies in each sample.

of-magnitude drop in the clumpy fraction over this time period (Y. Guo et al. 2015; T. Shibuya et al. 2016; D. Adams et al. 2022, among others). This redshift trend can be best explained by more closely examining the galaxies selected for by the sSFR > 0.1 Gyr^{-1} cut. At low redshift, this cut selects for the high tail end of the sSFR distribution; at high redshift, galaxies tend to exceed this cut by a far more significant margin. If only galaxies with sSFR > 0.1 Gyr^{-1} are considered, the median sSFR of the high-redshift clumpy galaxy sample from Y. Guo et al. (2018) ($0.5 < z < 3$) is $\sim$41 Gyr^{-1}, compared to only 6 Gyr^{-1} for the $z \sim 0$ clumpy galaxy sample in this work. Galaxies at high redshift have a significantly higher gas fraction than local galaxies (L. J. Tacconi et al. 2010), and commensurately higher sSFRs. Therefore, the reduction in the gas fraction with cosmic time between $0 < z < 3$ could also partially explain the drop in the clumpy fraction over this time period.

In Figure 12, we examine the relationship between sSFR and clump formation across a broad range of redshifts by plotting the observed fraction of clumpy galaxies in bins of sSFR for both our own clumpy galaxy sample ($z \sim 0$) and that of Y. Guo et al. (2018) over $0.5 < z < 2$ (herein called the Guo+18 sample; galaxies at $z > 2$ are not used here, as the clump detection completeness for higher-redshift galaxies is significantly lower in Guo+18.). For each sample, $f_{\rm clumpy}$ is computed for clumps exceeding 8% fractional luminosity (in the u band for our sample and the near-UV, $\sim$3500 Å, for Guo+18), among galaxies exceeding the mass limit $10^{9.4}$ $M_{\odot}$. These samples should not be compared one to one, as the selection criteria and detection completeness are different between these samples, and this difference is not controlled for here. However, the trend in $f_{\rm clumpy}$ with respect to sSFR can help to determine how much of the change in $f_{\rm clumpy}$ between $z > 0.5$ and $z \sim 0$ can be explained by the change in sSFR alone and how much must be due to other environmental factors. Within the local sample, we observe a sharp increase in $f_{\rm clumpy}$ with increasing sSFR for highly star-forming galaxies, rising from 1.5% at sSFR ~ 0.11 Gyr^{-1} to 5.4% at sSFR ~ 0.3 Gyr^{-1}. For comparison, we plot clumpy fractions observed in the Guo +18 sample at three different redshift intervals ($0.5 < z < 1$, $1 < z < 1.5$, and $1.5 < z < 2$) as a function of sSFR. Each of the high-redshift $f_{\rm clumpy}$ values exceeds the local value by a factor of $\sim$2 at sSFR ~ 0.3 Gyr^{-1}; however, this is significantly less than the order-of-magnitude difference in $f_{\rm clumpy}$ when not controlling for sSFR.

It is clear that accounting for sSFR explains a significant amount of the difference in $f_{\rm clumpy}$ between $z \sim 0$ and $z \gtrsim 0.5$. This raises the question of what causes the remaining difference. Observationally, the clump detection completeness at high and low redshift may be different; the $f_{\rm Lu}$ criterion also selects for slightly fewer clumps than the $f_{\rm LUV}$ criterion. However, there are also a number of systematic physical differences between star-forming galaxies at high and low redshift, as high-redshift galaxies are on average more compact, more turbulent, and have a higher rate of interactions and mergers (J. M. Lotz et al. 2011; S. A. Kassin et al. 2012; C. L. Carilli & F. Walter 2013); it is possible that one or several of these differences results in increased clumpiness, independent of sSFR. Regardless, the fact that controlling for sSFR appears to remove the majority of the difference between the clumpy fraction at high and low redshift, coupled with the fact that the clumpiness of star-forming galaxies is not significantly affected by environment, suggests that the gas density within a galaxy is most responsible for its clumpiness irrespective of other environmental triggers.

6. Summary

In this paper, we examine the relationship between the environments of low-redshift galaxies from the SDSS and the presence of giant star-forming clumps within these galaxies. To do so, we train and deploy a machine learning model to identify clumps in Galaxy Zoo 2 galaxies at $z > 0.02$, using human-provided labels from the citizen science project Galaxy Zoo:

Clump Scout as training data. The resulting catalog is the largest and most complete local Universe catalog of clumps to date and allows us to compute statistics on the prevalence of clumps in local galaxies with much greater accuracy.

We then isolate a target sample of 9964 galaxies in a narrow redshift range ($0.02 < z < 0.035$) and broad mass range ($M > 10^{9.4}\ M_{\odot}$) to search for correlations between clump formation and galaxy environment. We trace environmental density by measuring the distance to the projected fifth nearest neighbor and computing the fractional overdensity ($1 + \delta_5$), in line with previous studies of the environment.

By selecting clumps whose fractional luminosity in the SDSS u band, $f_{\rm Lu}$, exceeds a threshold value (either 3% or 8% of its host galaxy's total flux), we measure the clumpy fraction of galaxies $f_{\rm clumpy}$ as a function of environment. For all star-forming galaxies (the "all galaxies" sample), we find a significant trend in the clumpy fraction: the most isolated 20% of galaxies are on the order of 2 times more likely to host clumps than the most clustered 20%. To investigate the cause of this trend, we control for the SFR of galaxies by dividing galaxies into bins across multiple star formation tracers (sSFR, SFR, and color). When galaxies are binned by sSFR and color, we do not find a significant correlation between $f_{\rm clumpy}$ and environment, even when testing a comparably sized sample of galaxies to the "all galaxies" sample. Our result is very similar to high-redshift trends in $f_{\rm clumpy}$, which also show no correlation with environment Z. Sattari et al. (2023). We do find several significant correlations between $f_{\rm clumpy}$ and environment when galaxies are divided by SFR; however, we attribute this to the correlation between environment and mass, as none of these correlations are found in the sSFR-binned galaxy samples.

Broadly, we interpret our $f_{\rm clumpy}$ result as evidence that in situ clump formation is the dominant mechanism by which clumps form in the local Universe. The signatures of ex situ clump formation—a higher fraction of clumps in low star formation galaxies, as well as a stronger environmental correlation—are not found. We additionally propose that most clumpiness can be explained by galaxies' sSFRs, as opposed to by environmental interactions. By comparing our low-redshift clumpy fraction results with the high-redshift results of Y. Guo et al. (2018) in bins of sSFR, we find that controlling for sSFR explains the majority of the order-of-magnitude difference in $f_{\rm clumpy}$ between $z \sim 0$ and $z \gtrsim 0.5$. We propose that it is the higher gas fraction within star-forming galaxies, rather than the mode of gas accretion or the influence of other external forces, that primarily enables clumps to form and survive over long timescales.

Acknowledgments

We extend an enormous thank you to the volunteers who participated in the Galaxy Zoo: Clump Scout project. It is the efforts of these volunteers that made all of this work possible.

This research is partially supported by the National Science Foundation under grants IIS 2006894 and AST 1716602, and is based upon work partially supported by the National Aeronautics and Space Administration (NASA) under grant No. HST-AR-15792.002-A and Award No. 80NSSC20M0057. Brooke Simmons acknowledges support through a UK Research and Innovation Future Leaders Fellowship [grant No. MR/T044136/1].

This publication uses data generated via the Zooniverse.org platform, the development of which is funded by generous support, including a Global Impact Award from Google, and a grant from the Alfred P. Sloan Foundation.

This research made use of Montage. It is funded by the National Science Foundation under grant No. ACI-1440620, and was previously funded by the National Aeronautics and Space Administration's Earth Science Technology Office, Computation Technologies Project, under Cooperative Agreement Number NCC5-626 between NASA and the California Institute of Technology.

This research made use of Photutils, an Astropy package for detection and photometry of astronomical sources (Bradley 2023).

ORCID iDs

Dominic Adams https://orcid.org/0000-0003-1939-5180
Lucy Fortson https://orcid.org/0000-0002-1067-8558
Vihang Mehta https://orcid.org/0000-0001-7166-6035
Claudia Scarlata https://orcid.org/0000-0002-9136-8876
Chris Lintott https://orcid.org/0000-0001-5578-359X
Brooke Simmons https://orcid.org/0000-0001-5882-3323
Mike Walmsley https://orcid.org/0000-0002-6408-4181

References

Adamo, A., Östlin, G., Bastian, N., et al. 2013, ApJ, 766, 105
Adams, D., Mehta, V., Dickinson, H., et al. 2022, ApJ, 931, 16
Aihara, H., Allende Prieto, C., An, D., et al. 2011, ApJS, 193, 29
Baldry, I. K., Balogh, M. L., Bower, R. G., et al. 2006, MNRAS, 373, 469
Bertin, E., & Arnouts, S. 1996, A&AS, 117, 393
Bournaud, F., & Elmegreen, B. G. 2009, ApJL, 694, L158
Bournaud, F., Elmegreen, B. G., & Martig, M. 2009, ApJL, 707, L1
Bournaud, F., Perret, V., Renaud, F., et al. 2014, ApJ, 780, 57
Bradley, L. 2023, astropy/photutils: 1.8.0, Zenodo, v1.8.0, Zenodo, doi:10.5281/zenodo.7946442
Brinchmann, J., Charlot, S., White, S. D. M., et al. 2004, MNRAS, 351, 1151
Carilli, C. L., & Walter, F. 2013, ARA&A, 51, 105
Cava, A., Schaerer, D., Richard, J., et al. 2018, NatAs, 2, 76
Claeyssens, A., Adamo, A., Richard, J., et al. 2023, MNRAS, 520, 2180
Cooper, M. C., Newman, J. A., Madgwick, D. S., et al. 2005, ApJ, 634, 833
Cowie, L. L., Hu, E. M., & Songaila, A. 1995, AJ, 110, 1576
Das, A., Pandey, B., Sarkar, S., & Dutta, A. 2021, arXiv:2108.05874
Dekel, A., & Birnboim, Y. 2006, MNRAS, 368, 2
Dekel, A., Mandelker, N., Bournaud, F., et al. 2022, MNRAS, 511, 316
Dekel, A., Sari, R., & Ceverino, D. 2009, ApJ, 703, 785
Dickinson, H., Adams, D., Mehta, V., et al. 2022, MNRAS, 517, 5882
Ellison, S. L., Patton, D. R., Simard, L., et al. 2010, MNRAS, 407, 1514
Elmegreen, D. M. 2007, in IAU Symp. 235, Galaxy Evolution Across the Hubble Time, ed. F. Combes & J. Palous (Cambridge: Cambridge Univ. Press), 376
Elmegreen, D. M., Elmegreen, B. G., & Hirst, A. C. 2004a, ApJL, 604, L21
Elmegreen, D. M., Elmegreen, B. G., & Sheets, C. M. 2004b, ApJ, 603, 74
Engmann, S., & Cousineau, D. 2011, JAQM, 6, 1, https://jaqm.ro/issues/volume-6,issue-3/pdfs/jaqm_vol6_issue3.pdf#page=5
Fensch, J., & Bournaud, F. 2021, MNRAS, 505, 3579
Fisher, D. B., Glazebrook, K., Damjanov, I., et al. 2017, MNRAS, 464, 491
Genzel, R., Newman, S., Jones, T., et al. 2011, ApJ, 733, 101
Ginzburg, O., Huertas-Company, M., Dekel, A., et al. 2021, MNRAS, 501, 730
Guo, Y., Ferguson, H. C., Bell, E. F., et al. 2015, ApJ, 800, 39
Guo, Y., Rafelski, M., Bell, E. F., et al. 2018, ApJ, 853, 108
Haas, M. R., Schaye, J., & Jeeson-Daniel, A. 2012, MNRAS, 419, 2133
Hine, N. K., Geach, J. E., Alexander, D. M., et al. 2016, MNRAS, 455, 2363
Huertas-Company, M., Guo, Y., Ginzburg, O., et al. 2020, MNRAS, 499, 814
Inoue, S., Dekel, A., Mandelker, N., et al. 2016, MNRAS, 456, 2052
Jacob, J. C., Katz, D. S., Berriman, G. B., et al. 2010, Montage: An Astronomical Image Mosaicking Toolkit, Astrophysics Source Code Library, ascl:1010.036
Kassin, S. A., Weiner, B. J., Faber, S. M., et al. 2012, ApJ, 758, 106

Kauffmann, G., Heckman, T. M., White, S. D. M., et al. 2003, MNRAS, 341, 33
Kauffmann, G., White, S. D. M., Heckman, T. M., et al. 2004, MNRAS, 353, 713
Kawinwanichakij, L., Papovich, C., Quadri, R. F., et al. 2017, ApJ, 847, 134
Kennicutt, R. C. 1989, ApJ, 344, 685
Kennicutt, R. C. 1998, ApJ, 498, 541
Kereš, D., Katz, N., Fardal, M., Davé, R., & Weinberg, D. H. 2009, MNRAS, 395, 160
Kereš, D., Katz, N., Weinberg, D. H., & Davé, R. 2005, MNRAS, 363, 2
Kovač, K., Lilly, S. J., Cucciati, O., et al. 2010, ApJ, 708, 505
Livermore, R. C., Jones, T., Richard, J., et al. 2012, MNRAS, 427, 688
Lotz, J. M., Jonsson, P., Cox, T. J., et al. 2011, ApJ, 742, 103
Lotz, J. M., Papovich, C., Faber, S. M., et al. 2013, ApJ, 773, 154
Madau, P., Ferguson, H. C., Dickinson, M. E., et al. 1996, MNRAS, 283, 1388
Mandelker, N., Dekel, A., Ceverino, D., et al. 2014, MNRAS, 443, 3675
Mandelker, N., Dekel, A., Ceverino, D., et al. 2017, MNRAS, 464, 635
Mehta, V., Scarlata, C., Fortson, L., et al. 2021, ApJ, 912, 49
Messa, M., Adamo, A., Östlin, G., et al. 2019, MNRAS, 487, 4238
Overzier, R. A., Heckman, T. M., Tremonti, C., et al. 2009, ApJ, 706, 203
Popp, J. J., Dickinson, H., Serjeant, S., et al. 2024, RASTI, 3, 174
Ren, S., He, K., Girshick, R., & Sun, J. 2017, IEEE PAMI, 39, 1137
Sattari, Z., Mobasher, B., Chartab, N., et al. 2023, ApJ, 951, 147
Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, ApJ, 500, 525
Shibuya, T., Ouchi, M., Kubo, M., & Harikane, Y. 2016, ApJ, 821, 72
Tacconi, L. J., Genzel, R., Neri, R., et al. 2010, Natur, 463, 781
Walmsley, M., Allen, C., Aussel, B., et al. 2023, JOSS, 8, 5312
Watson, C., Tran, K.-V., Tomczak, A., et al. 2019, ApJ, 874, 63
Williams, J. P., de Geus, E. J., & Blitz, L. 1994, ApJ, 428, 693
Wuyts, E., Rigby, J. R., Gladders, M. D., & Sharon, K. 2014, ApJ, 781, 61
Zanella, A., Le Floc'h, E., Harrison, C. M., et al. 2019, MNRAS, 489, 2792