Nearby dwarf galaxies with extreme star formation rates: a window into dwarf–galaxy evolution in the early Universe

S. Kaviraj[®],¹* B. Bichang'a,¹ I. Lazar[®],¹ A. E. Watkins,¹ G. Martin^{®2} and R. A. Jackson³

¹Centre for Astrophysics Research, Dept of Physics, Astronomy and Mathematics, University of Hertfordshire, Hatfield AL10 9AB, UK ²School of Physics and Astronomy, University of Nottingham, University Park, Nottingham NG7 2RD, UK ³Department of Physics and Astronomy, University of Victoria, Victoria, BC V8P 5C2, Canada

Accepted 2025 April 29. Received 2025 March 31; in original form 2024 July 27

ABSTRACT

We study a sample of nearby ($z \sim 0.2$) low-luminosity dwarf ($10^7 M_{\odot} < M_{\star} < 10^8 M_{\odot}$) galaxies that have extreme (0.1–3 M_{\odot} yr⁻¹) star formation rates (SFRs) for this mass regime, making them plausible analogues of dwarfs at $z \sim 5.5$. We compare the properties of these analogues to control samples of 'normal' dwarfs, which reside on the star formation main sequence (SFMS) at $z \sim 0.2$ and are matched in their stellar mass and redshift distributions to the analogue population. The analogue and normal populations do not show differences, either in their half-light radii or the projected distances to nodes, filaments, and massive galaxies. This suggests that the comparatively extreme SFRs in the analogues are not driven by them being anomalously compact or because they reside in specific environments that might provide a larger gas supply. However, the fractions of interacting galaxies and those that have early-type morphology are significantly elevated (by factors of ~5.6 and ~9, respectively) in the analogues compared to the normal population. Extrapolation of the redshift evolution of the SFMS into our mass range of interest appears to underestimate the SFRs of observed dwarfs at $z \sim 5.5$. Since current SFMS measurements remain dominated by low-and intermediate-redshift data (especially at low stellar masses), our study suggests that this underestimation may be driven by interactions (which are more frequent at earlier epochs) boosting the SFRs in the high-redshift dwarf population. Our results are consistent with a picture where higher gas availability, augmented by interactions, drives much of the stellar mass assembly of dwarf galaxies in the early Universe.

Key words: galaxies: dwarf – galaxies: evolution – galaxies: formation – galaxies: interactions.

1 INTRODUCTION

Dwarf galaxies are fundamental to our understanding of galaxy evolution. Not only do dwarfs dominate the galaxy number density at all epochs (e.g. Kaviraj et al. 2017; Wright et al. 2017; Weaver et al. 2023), their shallow potential wells make them sensitive laboratories for studying many of the processes that shape the evolution of galaxies over cosmic time, such as baryonic feedback (e.g. Manzano-King, Canalizo & Sales 2019; Martin et al. 2019; Koudmani, Sijacki & Smith 2022; Romano et al. 2023, 2024), stellar mass buildup (e.g. Flores-Freitas et al. 2024), mergers and interactions with neighbours (e.g. Martin et al. 2021; Jackson et al. 2021a, b; Watkins et al. 2023), and the formation of small-scale structure such as planes of satellites around massive galaxies (e.g. Revaz & Jablonka 2018; Uzeirbegovic et al. 2024; Madhani et al. 2025). Notwithstanding their significance, much of what is known about the dwarf regime is underpinned by studies in our local neighbourhood, e.g. in the Local Group (e.g. Tolstoy, Hill & Tosi 2009) or from surveys of dwarf satellites around nearby massive galaxies (e.g. Hunter et al. 2012; Madden et al. 2013; Duc et al. 2015; Geha et al. 2017; Venhola et al. 2018; Mao et al. 2021; Poulain et al. 2021; Trujillo et al. 2021). This is driven by the relative faintness of the dwarf population, which makes it difficult to assemble complete unbiased statistical samples outside the nearby Universe using past surveys like the SDSS (e.g. Alam et al. 2015), which are wide but relatively shallow (Kaviraj et al. 2025).

https://doi.org/10.1093/mnras/staf716

However, new surveys are poised to revolutionize our understanding of dwarf-galaxy evolution, both in the nearby Universe and at high redshift. Surveys such as the Dark Energy Survey (DES; Dey et al. 2019) and the Hyper Suprime-Cam Subaru Strategic Program (HSC-SSP; Aihara et al. 2018b) have recently enabled the detection and detailed study of dwarfs in low-density environments outside the local neighbourhood (e.g. Tanoglidis et al. 2021; Lazar et al. 2024a, b; Thuruthipilly et al. 2024; Kaviraj et al. 2025). The forthcoming Legacy Survey of Space and Time (Ivezić et al. 2019) will drive another leap in the scope of such work, by increasing the footprints where deep optical imaging is available to roughly half the sky. In a similar vein, the advent of the JWST (Gardner 2023) is offering novel insights into galaxy evolution in the early Universe, with recent work successfully probing star formation activity and metallicity in high-redshift dwarf galaxies (e.g. Nakajima et al. 2023; Boyett et al. 2024; Clarke et al. 2024; Curti et al. 2024; Duan et al. 2024; Harshan et al. 2024), down to stellar masses of $M_{\star} \sim 10^7 \text{ M}_{\odot}$ at $z \sim 5$ –6. Nevertheless, studying large statistical samples of dwarfs down to these stellar masses at high redshift remains challenging, even with the exquisite abilities of the JWST.

Published by Oxford University Press on behalf of Royal Astronomical Society. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

^{*} E-mail: s.kaviraj@herts.ac.uk

Several studies in the past literature have explored the evolution of high-redshift dwarf galaxies by studying nearby 'analogues' i.e. galaxies at low redshift that have similar properties [stellar masses, star formation rates (SFRs), etc.] to their counterparts in the early Universe (e.g. Cardamone et al. 2009; Lofthouse, Houghton & Kaviraj 2017; Yang et al. 2017; Bian, Kewley & Dopita 2018; Izotov et al. 2021; Nakajima et al. 2022; Shivaei et al. 2022). While lowredshift galaxies cannot fully mimic the conditions in which their high-redshift counterparts reside, studying nearby analogues offers an opportunity to study aspects of these galaxies in greater detail than might be possible at high redshift.

For example, given that they reside in the nearby Universe, higher signal-to-noise observations are possible of analogues, which may enable more accurate estimations of physical parameters like stellar mass, SFR, and local environment. Since they reside in the nearby Universe, analogues have larger angular sizes, which aids in the study of structure and morphology. This is particularly true of low-surface-brightness features like tidal tails that signpost recent interactions (e.g. Kaviraj 2014) but which, given the prohibitively large cosmological dimming,¹ are challenging to study outside the nearby Universe (e.g. Martin et al. 2022). Finally, the low redshift of analogues also makes it feasible to study these systems in the rest-frame optical using existing data sets.

Past dwarf analogue studies have offered interesting insights into the processes that may be important in driving stellar assembly in dwarfs at high redshift. For example, Cardamone et al. (2009) have studied a population of 'green pea' galaxies, which are compact low-metallicity systems with extreme SFRs, making them plausible analogues of relatively luminous dwarfs ($10^{8.5}$ M $_{\odot} < M_{\star} < 10^{10}$ M_o) at high redshift. A conclusion of this work, using images from the Hubble Space Telescope (HST), is that the green peas show evidence for star-forming clumps and low-surface-brightness features that suggest the presence of recent mergers and interactions. Other studies, which have probed green peas using integral field spectroscopy (e.g. Lofthouse et al. 2017) and HI imaging (e.g. Purkayastha et al. 2022), have shown that at least some of these systems indeed show strong evidence of interactions (see also Le Reste et al. 2024). The importance of interactions in driving star formation in such systems appears to extend to dwarf analogues that have lower stellar masses (Dutta et al. 2024), often called 'blueberry' galaxies (Yang et al. 2017). Taken together, these studies hint at galaxy interactions playing a key role in the stellar mass build-up of dwarfs in the early Universe.

The purpose of this work is to perform a homogeneous statistical study of nearby ($z \sim 0.2$) dwarfs in the stellar mass range $10^7 \text{ M}_{\odot} < M_{\star} < 10^8 \text{ M}_{\odot}$, which have extreme SFRs ($0.1-3 \text{ M}_{\odot} \text{ yr}^{-1}$) for this mass regime, making them plausible analogues of dwarfs at high redshift ($z \sim 5.5$). Our choice of $z \sim 5.5$ is driven by the fact that our analogue selection is based on extrapolations of the local star formation main sequence (SFMS) out to high redshift, complemented by *JWST* data of real dwarfs in the early Universe. Currently available SFMS extrapolations extend out to $z \sim 5.5$ at which *JWST* data are also available, making this epoch a natural one at which to base our study. We combine (1) size measurements, (2) estimates of the projected distances to nodes, filaments, and massive galaxies, and (3) morphological information derived from visual inspection of deep

HSC and *HST* images to explore the processes that drive the extreme SFRs in the analogue population.

The plan for this paper is as follows. In Section 2, we describe the COSMOS2020 catalogue (Weaver et al. 2022) that underpins our study, the selection of our analogue population, and a sample of 'normal' dwarfs that have typical SFRs at $z \sim 0.2$. In the analysis that follows, we compare these two populations to gain insights into the drivers of the star formation activity in the analogues. In Section 3 we compare the environments (Section 3.1), sizes (Section 3.2), and morphologies (Section 3.3) of the two populations and speculate on the factors that drive the extreme SFRs in our analogues. We summarize our findings in Section 4.

2 DATA

2.1 Physical parameters

Our study is based on physical parameters (photometric redshifts, stellar masses, and SFRs) available from the Classic version of the publicly available COSMOS2020 catalogue. COSMOS2020 offers a value-added catalogue of sources in the 2 deg² COSMOS field, calculated using deep multiwavelength UV to mid-infrared photometry, in 40 broad and medium-band filters from the following instruments: GALEX (Zamojski et al. 2007), MegaCam/CFHT (Sawicki et al. 2019), ACS/HST (Leauthaud et al. 2007), Subaru/Hyper Suprime-Cam (Aihara et al. 2019), Subaru/Suprime-Cam (Taniguchi et al. 2007, 2015), VIRCAM/VISTA (McCracken et al. 2012), and IRAC/Spitzer (Ashby et al. 2013, 2015, 2018; Steinhardt et al. 2014). A particular strength of this data set is the use of optical (grizy) data for object detection, from the 'Ultra-deep' layer of the HSC-SSP (Aihara et al. 2019), which has a point source depth of $r \sim 28$ mag. As a comparison, the detection limit of HSC-SSP Ultra-deep is ~ 5 mag fainter than standard-depth imaging from the SDSS and ~ 10 mag fainter than the magnitude limit ($r \sim 18$ mag) of the SDSS spectroscopic main galaxy sample.

Fluxes are extracted within circular apertures after the survey images are homogenized to a common point spread function. The SEXTRACTOR and IRACLEAN codes are used to process the UV/optical and infrared photometry, respectively. Physical parameters are then calculated by applying the LEPHARE spectral energy distribution (SED)-fitting algorithm (Arnouts et al. 2002; Ilbert et al. 2006) to this UV-mid-infrared photometry. The template library, described in Ilbert et al. (2009), includes templates from all galaxy morphological types (Polletta et al. 2007), blue star-forming galaxy models from Bruzual & Charlot (2003), exponentially declining star formation histories to improve the photometric redshifts of quiescent galaxies (Onodera et al. 2012) and templates that account for active galactic nucleus (e.g. Salvato et al. 2009, 2011). Extinction is included as a free parameter, with a flat prior in the reddening $(E_{\rm B-V} < 0.5)$. The attenuation curves considered in the SED fitting are taken from Calzetti et al. (2000), Prevot et al. (1984), and two variants of the Calzetti law, which include the bump at 2175 Å (Fitzpatrick & Massa 1986) with two different amplitudes. Emission lines are added using the relation between the UV luminosity and [O II] emission-line flux, following Ilbert et al. (2009). The accuracies of the resultant photometric redshifts are better than 1 and 4 per cent for bright (i < 22.5 mag) and faint (25 mag < i < 27 mag) galaxies, respectively. We direct readers to Weaver et al. (2022) for more details of the construction of the catalogue.

To construct an initial sample of dwarfs, we select objects that are (1) classified as galaxies by LEPHARE (type = 0 in the COSMOS2020 catalogue), (2) classified as 'extended' (i.e. galaxies)

¹Cosmological dimming in terms of magnitudes scales with redshift as $10 \times \log(1 + z)$. The cosmological dimming out to e.g. $z \sim 6$ is therefore around 8.5 mag arcsec⁻².

by the HSC pipeline in the HSC g, r, i, and z filters, (3) have stellar masses in the range $10^7 \text{ M}_{\odot} < M_{\star} < 10^8 \text{ M}_{\odot}$ and redshifts in the range z < 0.3, (4) have u band as well as near- and mid-infrared photometry (since a wide wavelength baseline aids the accuracy of the parameter estimation, see e.g. Ilbert et al. 2006), and (5) lie outside masked regions such as bright star masks and image edges (flag_combined = 0 in the COSMOS2020 catalogue). This produces a sample of ~6400 dwarf galaxies with a median redshift of $z \sim 0.2$. As described below, we use this sample to select both a population of high-redshift dwarf analogues and a sample of 'normal' dwarfs in the nearby Universe to which the analogues can be compared, in order to investigate the processes that drive their extreme SFRs.

2.2 Selection of candidate samples of high-redshift analogues and normal dwarfs

We use the dwarf galaxy sample described above to select two populations. The first is a sample of dwarfs that have, compared to dwarf galaxies at low redshift, extreme SFRs that make them plausible analogues of the dwarf population at high redshift ($z \sim$ 5.5). The second is a sample of 'normal' dwarfs that have typical SFR properties in the nearby Universe. These are defined as systems that lie between the 45th and 55th percentile values of the SFR distribution at a given stellar mass i.e. they straddle the evolving median value of the SFR distribution as a function of stellar mass in the nearby Universe (e.g. at $M_{\star} \sim 10^7$ M_{\odot} the log of the median SFR is around -2). In Section 3, we compare the analogues to control samples, constructed from the normal population, which are matched to the analogues in stellar mass and redshift, in order to explore the source of the extreme SFRs in our analogues.

The identification of analogues ideally requires statistical samples of SFR measurements of dwarf galaxies at high redshift in our mass range of interest ($10^7 \text{ M}_{\odot} < M_{\star} < 10^8 \text{ M}_{\odot}$). With the advent of *JWST*, small samples of galaxies in this mass range have become available at high redshift (e.g. Curti et al. 2024). To select our analogues, we use both measurements of the SFMS at $z \sim 5.5$ and measured dwarf SFRs at high redshift. As noted before, our choice of using $z \sim 5.5$ as our definition of the high-redshift Universe results from the fact that currently available measurements of the SFMS extend out to this epoch and *JWST* data are also available at these redshifts.

Figure. 1 shows SFR versus stellar mass for galaxies at z < 0.3in the COSMOS2020 catalogue. Recall that the median redshift of this population is $z \sim 0.2$. The heatmap shows all COSMOS2020 galaxies that reside at z < 0.3, while the thick dotted red line shows a running median that we use to define the ridgeline of the SFMS at $z \sim 0.2$. Other SFMS measurements at $z \sim 0.2$ from the literature (see legend), which agree well with the COSMOS2020 ridgeline, are shown using the solid lines. To select our analogue population we proceed as follows. Popesso et al. (2023, P23 hereafter) have measured the evolution of the SFMS out to high redshift and demonstrate that the offset between the local SFMS and those at progressively higher redshifts does not depend strongly on stellar mass at $M_{\star} < 10^{9.5}$ M_{\odot}. Therefore, we first calculate the P23 SFMS offset, measured at $10^{8.8}$ M_{\odot}, between the local SFMS and that at $z \sim 5.5$ (the highest redshift out to which P23 offsets are available). We then add this offset (which is around 1 dex) to the COSMOS2020 ridgeline at $z \sim 0.2$ to estimate the extrapolated position of the SFMS at $z \sim 5.5$ in our mass range of interest. This is shown using the thick dotted yellow line in Fig. 1. The purple triangles indicate SFR measurements of observed galaxies that have stellar masses in the



Figure 1. The SFMS for galaxies at z < 0.3 in the COSMOS2020 catalogue. The median redshift of this population is $z \sim 0.2$. The heatmap shows all COSMOS2020 galaxies that reside at z < 0.3, the thick dotted red line shows the ridgeline of the SFMS at $z \sim 0.2$, and the thick dotted yellow line shows the ridgeline of the SFMS at $z \sim 0.2$ plus the $z \sim 5.5$ offset calculated using P23 (see the text in Section 2.2 for details). Other SFMS measurements at $z \sim 0.2$ from the literature (Whitaker et al. 2012; Behroozi, Wechsler & Conroy 2013; Salim et al. 2016), which agree well with the COSMOS2020 $z \sim 0.2$ ridgeline, are shown using the thin solid lines. The average scatter in these SFMS measurements is shown using the red error bar. The blue and green circles show $z \sim 5.5$ analogues and the population of normal dwarfs, respectively (see Section 2.2 for how these populations are defined). SFR uncertainties are shown for individual galaxies. The purple triangles indicate SFR measurements of galaxies that have stellar masses in the range $10^7 M_{\odot}$ $< M_{\star} < 10^8 M_{\odot}$ and redshifts in the range 4.5 < z < 6 from Curti et al. (2024).

range $10^7 M_{\odot} < M_{\star} < 10^8 M_{\odot}$ and redshifts in the range 4.5 < z < 6 from Curti et al. (2024).

The measured SFRs of high-redshift dwarfs lie almost exclusively above the thick dotted yellow line, as opposed to straddling its position in the SFR versus stellar mass plane. This suggests that the extrapolation of the P23 SFMS - which is underpinned by relatively massive galaxies at low and intermediate redshift - into our mass range of interest may underestimate the SFRs of dwarfs at high redshift (as has been noted already in previous work, e.g. Curti et al. 2024).² We return to this point later during our study of the factors that drive the extreme SFRs in the analogue population. Given this comparison, we select an initial sample of analogue candidates as galaxies that, at a given stellar mass, have SFRs equal to or greater than the locus traced by the thick dotted yellow line. The final sample of analogues is selected after visually inspecting the HSC and HST images of each galaxy to remove cases where more than one source may be contributing to the HSC photometry, as described in the next section.

²We note that a potential reason for this discrepancy could be that the JADES survey (Eisenstein et al. 2023), which underpins the Curti et al. (2024), may preferentially detect dwarfs with very high SFRs. This is not possible to confirm without a deeper survey that is capable of observing more quiescent galaxies. However, it is worth noting that the spread (~1 dex) in the observed SFRs of high-redshift dwarfs does not decrease in our mass range of interest compared to slightly more massive galaxies. Furthermore, the fraction of galaxies for which spectroscopic redshifts can be successfully measured remains stable out to $z \sim 6$ (e.g. Bunker et al. 2024). Together this suggests that JADES is unlikely to be significantly skewed towards the objects with highest SFRs at $z \sim 6$



Figure 2. *HST F*814*W* images of a random sample of our dwarf galaxies. The first row shows examples of galaxies classified as ETGs, while the bottom row shows LTGs. The panels with a green border show examples of systems that are flagged as interacting. The ETG in column 3 shows an internal asymmetry, while the ETG in column 4 shows a tidal feature to the north east of the galaxy. The LTG in column 4 similarly shows a tidal feature to the east of the galaxy. Each image has a size of 3 arcsec on a side. At the median redshift of our dwarf sample ($z \sim 0.2$), this corresponds to a physical size of around 10 kpc.

Finally, our population of normal dwarfs, which is selected to lie around the ridgeline of the $z \sim 0.2$ SFMS, as described above, is shown using the green circles along with their error bars. As with the analogues, we select an initial sample of normal candidates, from which we then select a final sample after the visual inspection described in the next section.

2.3 Visual inspection of HSC and *HST* images – removal of blended sources and morphological classification

We visually inspect both the HSC gri colour composite and HST F814W(I-band) images (Koekemoer et al. 2007; Massey et al. 2010) of each analogue and normal galaxy. The HSC and HST images have 5σ point-source depths of 27.8 and 28 mag and angular resolutions of ~ 0.6 (Aihara et al. 2018a) and ~ 0.05 arcsec (Koekemoer et al. 2007), respectively. The visual inspection, performed by one expert classifier (SK), has two purposes. First, as noted above, we use the higher resolution HST images to identify cases where the HSC photometry may have contributions from two separate sources - we find eight objects matching this description.³ In such cases, the HSC centroid lies between the two HST objects that have been blended in the HSC photometry. Appendix A shows an example of such systems, which we remove from our analysis. Our final samples of analogue and normal dwarfs contain 116 and 590 galaxies, respectively. The blue and green circles in Fig. 1 show the final analogue and normal populations that underpin our study, with individual SFR error bars also shown for each galaxy.

The second purpose of the visual inspection is to perform morphological classification of our dwarf galaxies using their *HST* images. Visual inspection offers the most accurate route to the classification of galaxy morphologies and the identification of interacting systems

 3 The fraction of such blended sources is similar (~6 per cent) in the analogue and normal populations.

(e.g. Kaviraj 2014), against which automated methods that use quantitative parameters are often calibrated (e.g. Lazar et al. 2024a). We classify our galaxies into two broad morphological classes: earlytype galaxies (ETGs) and late-type galaxies (LTGs), following the traditional definitions of these morphological types. ETGs exhibit central light concentrations that are surrounded by smooth light distributions, while galaxies that do not exhibit these characteristics are classified as LTGs. We also flag interacting systems that show evidence of an ongoing or recent interaction e.g. tidal features, internal asymmetries, or tidal bridges due to an ongoing merger with another galaxy.

Note that none of the eight objects that were removed from our analysis because the HSC photometry may have contributions from two separate sources show any evidence of tidal features or tidal bridges between the two objects, suggesting that they are interacting. The removal of these objects therefore has no impact on the fraction of interacting galaxies in our sample. Finally, we note that our interacting galaxies are, by construction, those that are already in the process of interacting i.e. they are not 'close pairs', which are not possible to identify without spectroscopic redshifts.

Figure. 2 shows *HST* images of a random sample of our dwarf galaxies. The first row shows examples of galaxies classified as ETGs, while the bottom row shows LTGs. The panels with a green border show examples of systems that are flagged as interacting. The ETG in column 3 shows an internal asymmetry, while the ETG in column 4 shows a tidal feature to the north east of the galaxy. The LTG in column 4 similarly shows a tidal feature to the east of the galaxy. Each image has a size of 3 arcsec on a side. At the median redshift of our dwarf sample ($z \sim 0.2$), this corresponds to a physical size of around 10 kpc.

2.4 Estimation of environmental parameters

We measure the distances of our dwarfs from nodes, filaments, and massive $(M_{\star} > 10^{10} \text{ M}_{\odot})$ galaxies using the DisPerSE (Sousbie

2011) structure-finding algorithm. Our methodology follows Lazar et al. (2023) and Bichang'a et al. (2024), who have performed a similar density analysis using the COSMOS2020 catalogue. Dis-PerSE creates a density map using Delaunay tessellations, calculated using the positions of galaxies (Schaap & van de Weygaert 2000). Stationary i.e. critical points in the density map (minima, maxima, and saddles) correspond to voids, nodes, and the centres of filaments in the cosmic web, respectively. Segments are then used to connect the saddle points with the local maxima, forming a set of ridges that describes the network of filaments that define the cosmic web. The properties of the filament network thus created are determined by a 'persistence' parameter, which sets a threshold value for defining pairs of critical points within the density map. Only critical pairs with Poisson probabilities above $N\sigma$ from the mean are retained. We use a persistence value of 2 in our study, which removes ridges close to the noise level, where structures could be spurious.

We use massive $(M_{\star} > 10^{10} \text{ M}_{\odot})$ galaxies to construct our density maps, as they have the smallest redshift errors ($\delta z \sim 0.008$ at $z \sim 0.2$) and dominate the local gravitational potential wells. The accuracy of the COSMOS2020 redshifts enables us to employ welldefined and relatively narrow ($\delta z \sim 0.02$) redshift slices to build our density maps. When constructing each density map, individual galaxies are weighted by the area under their redshift probability density function that is contained within the slice in question. This ensures that the photometric redshifts of massive galaxies, albeit very accurate in COSMOS2020, are propagated through the construction of the density maps.

3 COMPARISON OF THE PROPERTIES OF ANALOGUE AND NORMAL DWARFS

We proceed by comparing the properties of the analogue and normal populations – local environment, galaxy sizes, the morphological mix, and the presence of interactions. In all cases, we compare the analogue population to a control sample drawn from the normal dwarfs, which has an identical number of galaxies and is matched to the analogues in its distributions of stellar mass and redshift.

3.1 Distances from nodes, filaments, and massive galaxies

We begin, in Fig. 3, by comparing the distributions of projected distances from the nearest filaments, nodes, and massive $(M_{\star} >$ $10^{10} M_{\odot}$) galaxies for the analogue and normal populations. Median values and their uncertainties are shown using the dashed and dotted lines, respectively. Compared to the normal population, the analogues reside slightly further away from filaments and nodes but, given the errors on the medians, show no difference in the median distance to the nearest massive galaxies. KS tests for the two populations yield p-values of 0.32, 0.06, and 0.52 for the projected distances to filaments, nodes, and massive galaxies, respectively. Recall that the null hypothesis in the KS test, i.e. that the two samples are drawn from the same distribution, is typically rejected if the p-value is lower than 0.05. While the median values indicate that the environments of the analogue and normal dwarfs are slightly different, the KS tests suggest that these differences are relatively small (except possibly for the distances to nodes, where the *p*-value is close to 5 per cent). These minor differences suggest that the large discrepancy in the SFRs between the analogue and normal populations is not driven by differences in local environment.



Figure 3. Distributions of projected distances from the nearest filaments (top), nodes (middle), and massive $(M_{\star} > 10^{10} \text{ M}_{\odot})$ galaxies (bottom) for the analogue (red) and normal (black) populations. NMG = nearest massive galaxy. Median values and their uncertainties are shown using the dashed and dotted lines, respectively. These are calculated using bootstrapping, where the distribution is resampled with replacement 1000 times. The median for each resampled distribution is calculated and the standard deviation of the resultant distribution of 1000 median values is used as the error on the median of the original distribution.

3.2 Galaxy size

Next, we consider the sizes of the analogue and normal populations. Figure. 4 compares the distributions of half-light radii, calculated in the I band (Leauthaud et al. 2007), of these two samples. The half-light radius is calculated using the curve of growth of the flux and



Figure 4. Distributions of half-light radii of the analogue (red) and normal (black) populations. Median values and their uncertainties (calculated using bootstrapping) are shown using the dashed and dotted lines, respectively. The inset shows the half-light radius plotted against the SFR. The heatmap shows all dwarfs that have stellar masses in the range $10^7 \text{ M}_{\odot} < M_{\star} < 10^8 \text{ M}_{\odot}$ and redshifts in the range z < 0.3. A random 10 per cent of galaxies are shown overplotted using the black dots.

Table 1. The fractions of galaxies that are interacting (second row) and those that have early-type morphology (third row) in the analogue (second column) and normal (third column) populations. Note that any galaxy that does not have early-type morphology is classed as having late-type morphology. The ratio between the analogue and normal fractions is shown in the last column. The uncertainties are calculated following Cameron (2011).

	Analogue	Normal	Ratio (Ana./Norm.)
Interacting fraction	0.36 ± 0.05	0.04 ± 0.02	9
ETG fraction	0.28 ± 0.03	0.05 ± 0.02	5.6

defined as the radius that contains half of the light from the galaxy. The median values are indistinguishable within the uncertainties and a KS test yields a *p*-value of 0.1, suggesting that the distributions are similar. The inset, which shows the half-light radii plotted against the SFRs of all dwarf galaxies in our mass range of interest, indicates that galaxy size generally does not show a correlation with SFR, in agreement with results in the literature at higher stellar masses (e.g. Whitaker et al. 2017; Lin et al. 2020). The lack of a correlation between SFR and size across the general dwarf population explains why the sizes of the analogues and normal samples are similar even though they have very different SFRs. In summary, the differences in the SFRs between the analogue and normal populations are unlikely to be driven by the analogues having different sizes (e.g. due to being more compact).

3.3 Morphology and the presence of interactions

Table 1 summarizes the fractions of galaxies classified as ETG and interacting in the analogue and normal populations. The uncertainties in the fractions are estimated following Cameron (2011) who calculates accurate Bayesian binomial confidence intervals using the quantiles of the beta distribution. These are more accurate than simpler techniques, like using the normal approximation, which tends to misrepresent the statistical uncertainty under the sampling conditions typically encountered in astronomical surveys. Both the interacting and ETG fractions are significantly elevated (by several factors) in the analogues, compared to their normal counterparts.

Given that interactions create and consolidate dispersiondominated components in galaxies (e.g. Martin et al. 2018), one expects a galaxy population with a higher incidence of interactions to also show an elevated ETG fraction, as appears to be the case here. While the larger interaction fractions in the analogues may be somewhat surprising given the similarities in galaxy environments, it is worth noting that dwarfs have shallow potential wells, which make them more susceptible to being morphologically affected by tidal perturbations from nearby objects, so that higher density environments may not be needed for generating the interaction signatures probed here (e.g. Jackson et al. 2021b). It is interesting to note that the median SFR of interacting galaxies is elevated by \sim 29 per cent compared to those of non-interacting galaxies. In a similar vein the median SFR of galaxies flagged as ETGs is elevated by \sim 55 per cent compared to those that are LTGs. Taken together, the higher interacting and ETG fractions and the elevations in the SFR in the interacting and ETG populations suggest that interactions play a significant role in driving the extreme SFRs seen in the analogue population.

3.4 What drives the extreme SFRs in the analogue population?

The analysis above indicates that the vastly different SFRs of the analogue and normal populations is unlikely to be driven either by a difference in local environment or galaxy size. However, the analogue and normal dwarfs do show significantly different interacting and ETG fractions, suggesting that interactions may play a role in the SFR divergence observed between these two populations.

In this context, recall that an extrapolation of the redshift evolution of the P23 SFMS measurements (which are largely based on relatively massive galaxies at low and intermediate redshift) into our mass range of interest appears to underestimate the measured SFRs of dwarfs at $z \sim 5.5$. In a similar vein, Liu et al. (2019) have estimated molecular gas fractions ($\mu_{molgas} \equiv M_{molgas}/M_{\star}$) and gas depletion time-scales ($\tau_{molgas} \equiv M_{molgas}/SFR$) in relatively massive galaxies ($M_{\star} > 10^{9.2} M_{\odot}$) out to $z \sim 6$. A caveat here is that their fits to the redshift evolution of μ_{molgas} and τ_{molgas} , particularly at lower stellar masses, rely mainly on data in the nearby Universe. Since $\mu_{molgas}/\tau_{molgas}$ yields the specific SFR (sSFR), these results offer a way of gauging the expected evolution of star formation activity due to the changes in the gas conditions between the local Universe and high redshift.

Liu et al. (2019) show that this redshift evolution in the star formation activity is stellar-mass-dependent. Figure. 5 shows the ratio of the sSFR (i.e. $\mu_{\text{molgas}}/\tau_{\text{molgas}}$) for galaxies on the SFMS between $z \sim 5.5$ and $z \sim 0.2$, as a function of stellar mass from Liu et al. (2019). While this study does not extend past $M_{\star} \sim 10^{9.2} \,\mathrm{M_{\odot}}$, the log of the sSFR ratio decreases almost linearly with the log of the stellar mass. Extrapolation of this behaviour into our mass range of interest is shown using the dotted line. At $M_{\star} \sim 10^7 \,\mathrm{M_{\odot}}$ and $M_{\star} \sim 10^8 \,\mathrm{M_{\odot}}$ the sSFR ratios are ~4 and ~8, respectively. Given that the observed SFRs of dwarfs in our mass range of interest are at least a factor of 14 (1.16 dex) higher at $z \sim 5.5$ than at $z \sim 0.2$, a simple extrapolation of the gas conditions out to high redshift and down to low stellar mass also appears to underpredict the evolution in the star formation activity.

Given the significantly elevated interacting and ETG fractions found in the analogue population, we speculate that the underestimates described above may be driven, at least partly, by interactiondriven enhancement of star formation in high-redshift dwarf galaxies. The star-formation and gas-mass calibrations in the current literature, particularly at low stellar masses, remain dominated by galaxies at low redshift, where interactions are less frequent (e.g. Conselice et al.



Figure 5. The ratio of the sSFR between $z \sim 5.5$ and $z \sim 0.2$ as a function of stellar mass, from Liu et al. (2019). While this study does not extend past $M_{\star} \sim 10^{9.2} \,\mathrm{M_{\odot}}$, the log of the sSFR ratio decreases almost linearly with the log of the stellar mass. Extrapolation of this behaviour into our mass range of interest is shown using the dotted line.

2014). However, it is well established that interactions enhance star formation in dwarf galaxies (e.g. Martin et al. 2022; Lazar et al. 2024a) in which interactions typically boost star formation across the entire body of the galaxy (Lazar et al. 2024b). For this reason, extrapolating the star formation and gas properties of nearby dwarfs out to high redshift may underestimate the star formation activity in the high-redshift dwarf population. We suggest therefore that while gas availability is higher in the early Universe, the stellar assembly in high-redshift dwarfs is catalysed by interactions. Indeed the shallow potential wells of dwarfs may make the impact of interactions greater in these objects than in the massive-galaxy regime, where interactions may not play as significant a role in galaxy assembly at high redshift (e.g. Duncan et al. 2019; Romano et al. 2021; Dalmasso et al. 2024).

Finally, our conclusions about the role of interactions is driven by the fact that this is where the analogue and normal populations differ the most, while other properties (e.g. galaxy size and environment) seem to be similar. However, a potential caveat is that our study necessarily uses extrapolations of low-redshift data that may not accurately represent the conditions in the high-redshift Universe, which is worth bearing in mind when considering our conclusions.

4 SUMMARY

We have studied a sample of nearby ($z \sim 0.2$) low-luminosity dwarf ($10^7 \text{ M}_{\odot} < M_{\star} < 10^8 \text{ M}_{\odot}$) galaxies with extreme SFRs, which make them plausible analogues of dwarfs at $z \sim 5.5$. We have first constructed control samples, which are drawn from a population of 'normal' dwarfs with typical SFRs at $z \sim 0.2$, and are matched in their stellar mass and redshift distributions to the analogue population. We have then compared the properties of the analogues to the normal dwarfs (via the control samples) to explore the processes and factors that are likely to drive the extreme SFRs in the analogues. Our main conclusions are as follows:

(i) The analogue and normal populations do not show strong differences in their distances to nodes, filaments, and massive galaxies, suggesting that the analogue SFRs are not the result of residing in a specific environment that may provide a larger gas supply.

(ii) The analogue and normal populations also do not show differences in their distributions of half-light radii, indicating that the extreme SFRs in the analogues are not driven by a difference in size (e.g. due to the analogues being anomalously compact). (iii) However, the fraction of interacting galaxies and those that are ETGs is significantly elevated (by several factors) in the analogues compared to the normal galaxies, suggesting that interactions play an important role in driving the extreme SFRs.

(iv) Extrapolation of the redshift evolution of the SFMS into our mass range of interest underestimates the SFRs of observed dwarfs at $z \sim 5.5$. Since current SFMS measurements remain dominated by data at low and intermediate redshift, we suggest that this underestimation could be driven by interactions, which are more frequent at earlier epochs, boosting the SFRs in the high-redshift dwarf population.

(v) Our results suggest that, while gas availability is higher in the early Universe, the stellar mass assembly of high-redshift dwarfs is catalysed by interactions.

ACKNOWLEDGEMENTS

We thank the referee for many constructive comments that helped us to significantly improve the quality of the original manuscript. SK, IL, and AEW acknowledge support from the STFC (grant numbers ST/Y001257/1, ST/S00615X/1, and ST/X001318/1). SK also acknowledges a Senior Research Fellowship from Worcester College Oxford. BB and IL acknowledge PhD studentships funded by the Centre for Astrophysics Research at the University of Hertfordshire.

The Hyper Suprime-Cam (HSC) collaboration includes the astronomical communities of Japan and Taiwan, and Princeton University. The HSC instrumentation and software were developed by the National Astronomical Observatory of Japan (NAOJ), the Kavli Institute for the Physics and Mathematics of the Universe (Kavli IPMU), the University of Tokyo, the High Energy Accelerator Research Organization (KEK), the Academia Sinica Institute for Astronomy and Astrophysics in Taiwan (ASIAA), and Princeton University. Funding was contributed by the FIRST program from the Japanese Cabinet Office, the Ministry of Education, Culture, Sports, Science and Technology (MEXT), the Japan Society for the Promotion of Science (JSPS), Japan Science and Technology Agency (JST), the Toray Science Foundation, NAOJ, Kavli IPMU, KEK, ASIAA, and Princeton University. This paper makes use of software developed for Vera C. Rubin Observatory.

This paper is based on data collected at the Subaru Telescope and retrieved from the HSC data archive system, which is operated by the Subaru Telescope and Astronomy Data Center (ADC) at NAOJ. Data analysis was in part carried out with the cooperation of Center for Computational Astrophysics (CfCA), NAOJ. We are honored and grateful for the opportunity of observing the Universe from Maunakea, which has the cultural, historical and natural significance in Hawaii.

For the purpose of open access, the authors have applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising from this submission.

DATA AVAILABILITY

The physical parameters used in this study are available in Weaver et al. (2022). The density parameters were produced using the DisPerSE algorithm (Sousbie 2011).

REFERENCES

Aihara H. et al., 2018a, PASJ, 70, S4 Aihara H. et al., 2018b, PASJ, 70, S8 Aihara H. et al., 2019, PASJ, 71, 114 Alam S. et al., 2015, ApJS, 219, 12 Arnouts S. et al., 2002, MNRAS, 329, 355 Ashby M. L. N. et al., 2013, ApJ, 769, 80 Ashby M. L. N. et al., 2015, ApJS, 218, 33 Ashby M. L. N. et al., 2018, ApJS, 237, 39 Behroozi P. S., Wechsler R. H., Conroy C., 2013, ApJ, 770, 57 Bian F., Kewley L. J., Dopita M. A., 2018, ApJ, 859, 175 Bichang'a B., Kaviraj S., Lazar I., Jackson R. A., Das S., Smith D. J. B., Watkins A. E., Martin G., 2024, MNRAS, 532, 613 Boyett K. et al., 2024, MNRAS, 535, 1796 Bruzual G., Charlot S., 2003, MNRAS, 344, 1000 Bunker A. J. et al., 2024, A&A, 690, A288 Calzetti D., Armus L., Bohlin R. C., Kinney A. L., Koornneef J., Storchi-Bergmann T., 2000, ApJ, 533, 682 Cameron E., 2011, Publ. Astron. Soc. Aust., 28, 128 Cardamone C. et al., 2009, MNRAS, 399, 1191 Clarke L., Shapley A. E., Sanders R. L., Topping M. W., Brammer G. B., Bento T., Reddy N. A., Kehoe E., 2024, ApJ, 977, 133 Conselice C. J., Bluck A. F. L., Mortlock A., Palamara D., Benson A. J., 2014, MNRAS, 444, 1125 Curti M. et al., 2024, A&A, 684, A75 Dalmasso N. et al., 2024, MNRAS, 533, 4472 Dey A. et al., 2019, AJ, 157, 168 Duan Q. et al., 2024, preprint (arXiv:2411.04944) Duc P.-A. et al., 2015, MNRAS, 446, 120 Duncan K. et al., 2019, ApJ, 876, 110 Dutta S., Bera A., Bait O., Narayan C. A., Sebastian B., Vaddi S., 2024, MNRAS, 531, 5140 Eisenstein D. J. et al., 2023, preprint (arXiv:2306.02465) Fitzpatrick E. L., Massa D., 1986, ApJ, 307, 286 Flores-Freitas R., Trevisan M., Mückler M., Mamon G. A., Schnorr-Müller A., Bootz V., 2024, MNRAS, 528, 5804 Gardner J. P., et al., 2023, PASP, 135, 068001 Geha M. et al., 2017, ApJ, 847, 4 Harshan A. et al., 2024, MNRAS, 532, 1112 Hunter D. A. et al., 2012, AJ, 144, 134 Ilbert O. et al., 2006, A&A, 457, 841 Ilbert O. et al., 2009, ApJ, 690, 1236 Ivezić Ž. et al., 2019, ApJ, 873, 111 Izotov Y. I., Guseva N. G., Fricke K. J., Henkel C., Schaerer D., Thuan T. X., 2021, A&A, 646, A138 Jackson R. A. et al., 2021a, MNRAS, 502, 1785 Jackson R. A. et al., 2021b, MNRAS, 502, 4262 Kaviraj S., 2014, MNRAS, 440, 2944 Kaviraj S. et al., 2017, MNRAS, 467, 4739 Kaviraj S., Lazar I., Watkins A. E., Laigle C., Martin G., Jackson R. A., 2025, MNRAS, 538, 153 Koekemoer A. M. et al., 2007, ApJS, 172, 196

- Koudmani S., Sijacki D., Smith M. C., 2022, MNRAS, 516, 2112
- Lazar I., Kaviraj S., Martin G., Laigle C., Watkins A., Jackson R. A., 2023, MNRAS, 520, 2109
- Lazar I., Kaviraj S., Watkins A. E., Martin G., Bichang'a B., Jackson R. A., 2024a, MNRAS, 529, 499
- Lazar I., Kaviraj S., Watkins A. E., Martin G., Bichang'a B., Jackson R. A., 2024b, MNRAS, 533, 3771
- Le Reste A. et al., 2024, MNRAS, 528, 757
- Leauthaud A. et al., 2007, ApJS, 172, 219
- Lin L. et al., 2020, ApJ, 899, 93
- Liu D. et al., 2019, ApJ, 887, 235
- Lofthouse E. K., Houghton R. C. W., Kaviraj S., 2017, MNRAS, 471, 2311
- Madden S. C. et al., 2013, PASP, 125, 600
- Madhani J. P. et al., 2025, preprint (arXiv:2504.18515)
- Manzano-King C. M., Canalizo G., Sales L. V., 2019, ApJ, 884, 54
- Mao Y.-Y., Geha M., Wechsler R. H., Weiner B., Tollerud E. J., Nadler E. O., Kallivayalil N., 2021, ApJ, 907, 85
- Martin G., Kaviraj S., Devriendt J. E. G., Dubois Y., Pichon C., 2018, MNRAS, 480, 2266

Martin G. et al., 2019, MNRAS, 485, 796 Martin G. et al., 2021, MNRAS, 500, 4937 Martin G. et al., 2022, MNRAS, 513, 1459 Massey R., Stoughton C., Leauthaud A., Rhodes J., Koekemoer A., Ellis R., Shaghoulian E., 2010, MNRAS, 401, 371 McCracken H. J. et al., 2012, A&A, 544, A156 Nakajima K. et al., 2022, ApJS, 262, 3 Nakajima K., Ouchi M., Isobe Y., Harikane Y., Zhang Y., Ono Y., Umeda H., Oguri M., 2023, ApJS, 269, 33 Onodera M. et al., 2012, ApJ, 755, 26 Polletta M. et al., 2007, ApJ, 663, 81 Popesso P. et al., 2023, MNRAS, 519, 1526 Poulain M. et al., 2021, MNRAS, 506, 5494 Prevot M. L., Lequeux J., Maurice E., Prevot L., Rocca-Volmerange B., 1984, A&A, 132, 389 Purkayastha S., Kanekar N., Chengalur J. N., Malhotra S., Rhoads J., Ghosh T., 2022, ApJ, 933, L11 Revaz Y., Jablonka P., 2018, A&A, 616, A96 Romano M. et al., 2021, A&A, 653, A111 Romano M. et al., 2023, A&A, 677, A44 Romano M. et al., 2024, A&A, 683, L9 Salim S. et al., 2016, ApJS, 227, 2 Salvato M. et al., 2009, ApJ, 690, 1250 Salvato M. et al., 2011, ApJ, 742, 61 Sawicki M. et al., 2019, MNRAS, 489, 5202 Schaap W. E., van de Weygaert R., 2000, A&A, 363, L29 Shivaei I. et al., 2022, ApJ, 928, 68 Sousbie T., 2011, MNRAS, 414, 350 Steinhardt C. L. et al., 2014, ApJ, 791, L25 Taniguchi Y. et al., 2007, ApJS, 172, 9 Taniguchi Y. et al., 2015, PASJ, 67, 104 Tanoglidis D. et al., 2021, ApJS, 252, 18 Thuruthipilly H., Junais, Pollo A., Sureshkumar U., Grespan M., Sawant P., Malek K., Zadrozny A., 2024, A&A, 682, A4 Tolstoy E., Hill V., Tosi M., 2009, ARA&A, 47, 371 Trujillo I. et al., 2021, A&A, 654, A40 Uzeirbegovic E. et al., 2024, MNRAS, 535, 3775 Venhola A. et al., 2018, A&A, 620, A165 Watkins A. E., Salo H., Kaviraj S., Collins C. A., Knapen J. H., Venhola A., Román J., 2023, MNRAS, 521, 2012 Weaver J. R. et al., 2022, ApJS, 258, 11 Weaver J. R. et al., 2023, A&A, 677, A184 Whitaker K. E., van Dokkum P. G., Brammer G., Franx M., 2012, ApJ, 754,

- L29
- Whitaker K. E. et al., 2017, ApJ, 838, 19
- Wright A. H. et al., 2017, MNRAS, 470, 283
- Yang H., Malhotra S., Rhoads J. E., Wang J., 2017, ApJ, 847, 38
- Zamojski M. A. et al., 2007, ApJS, 172, 468

APPENDIX A: IDENTIFICATION OF BLENDED SOURCES IN HSC IMAGES VIA VISUAL INSPECTION

As noted in Section 2.3, we use the higher resolution HST images to identify cases where the HSC photometry may have contributions from two separate sources. Figure. A1 shows an example of a galaxy where a single source is seen both in the HSC and HST images (top row) and an example of a system where the HSC centroid (shown using the red cross) lies between two objects in the HST image. These objects are blended in the HSC image. The colours of the pixels in the HSC image on either side of the blended object are slightly different, indicating that the HSC object is indeed likely to be a blend of two separate objects (as indicated by the HST image at this location). We remove objects similar to this one from our analysis.



Figure A1. Top row: an example of a galaxy where a single source is seen both in the *HST* (left) and HSC (right) images. The HSC centroid is shown using a red cross. Bottom row: an example of a system where the HSC centroid lies between two objects in the *HST* image. These objects are blended in the HSC image. The colour of the pixels in the HSC image on either side of the (blended) object are slightly different, indicating that the HSC object is indeed likely to be a blend of two separate objects.

This paper has been typeset from a TEX/LATEX file prepared by the author.