Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **541,** 1831–1850 (2025) https://doi.org/10.1093/mnras/staf1092
Advance Access publication 2025 July 4

Cosmic reflections I: the structural diversity of simulated and observed low-mass galaxy analogues

G. Martin [1,2,3★] A. E. Watkins [4] Y. Dubois [5] J. Devriendt [6] S. Kaviraj [4] D. Kim [7] K. Kraljic [8] I. Lazar [4] F. R. Pearce [1] S. Peirani [9,10] C. Pichon [5,11] A. Slyz[6] and S. K. Yi [12]

[1]*School of Physics and Astronomy, University of Nottingham, University Park, Nottingham NG7 2RD, UK*
[2]*Korea Astronomy and Space Science Institute, 776 Daedeokdae-ro, Yuseong-gu, Daejeon 34055, Korea*
[3]*Steward Observatory, University of Arizona, 933 North Cherry Ave, Tucson, AZ 85719, USA*
[4]*Centre for Astrophysics Research, University of Hertfordshire, College Lane, Hatfield AL10 9AB, UK*
[5]*Institut d'Astrophysique de Paris, UMR 7095, CNRS, Sorbonne Université, 98 bis boulevard Arago, 75014 Paris, France*
[6]*Department of Physics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH, UK*
[7]*Astronomy and Space Science Department, Chungnam National University, Daehak-ro 99, Yuseong-gu Daejeon 34134, Republic of Korea*
[8]*Observatoire Astronomique de Strasbourg, Université de Strasbourg, CNRS, UMR 7550, F-67000 Strasbourg, France*
[9]*ILANCE, CNRS, University of Tokyo International Research Laboratory, Kashiwa, Chiba 277-8582, Japan*
[10]*Kavli IPMU (WPI), UTIAS, The University of Tokyo, Kashiwa, Chiba 277-8583, Japan*
[11]*Kyung Hee University, Dept. of Astronomy & Space Science, Yongin-shi, Gyeonggi-do 17104, Republic of Korea*
[12]*Department of Astronomy, Yonsei University, 50 Yonsei-ro, Seodaemun-gu, Seoul 03722, Republic of Korea*

Accepted 2025 June 30. Received 2025 June 25; in original form 2025 February 11

ABSTRACT

Dwarf galaxies serve as powerful laboratories for investigating the underlying physics of galaxy evolution including the impact of baryonic feedback processes and external environmental influences. We compare the visual and structural properties of dwarf galaxies in ultra-deep HSC–SSP imaging of the COSMOS field with those measured from realistic HSC-like synthetic observations of dwarfs generated by the ILLUSTRIS TNG50 and NEWHORIZON simulations. Using Sérsic profile fitting and non-parametric morphological metrics (Gini, M_{20}, asymmetry, and concentration), we evaluate the diversity of structural properties in observed and simulated galaxies. Our analysis shows that NEWHORIZON and TNG50 galaxies lie at opposite extremes of observed structural trends: NEWHORIZON produces diffuse, extended galaxies with shallow Sérsic indices, while TNG50 yields compact, concentrated systems with steep indices. Both simulations reproduce observed structural trends more closely at higher stellar masses ($M_\star \sim 10^{9.5}$ M$_\odot$) but fail to capture the full diversity of COSMOS dwarfs at lower masses. Non-parametric metrics further show that NEWHORIZON galaxies exhibit more uneven, clumpy light distributions while TNG50 galaxies have smoother but excessively concentrated profiles. These structural differences reflect underlying differences in their physical prescriptions and are likely driven by differing approaches to ISM gas physics, supernova feedback and star formation in addition to differences in numerical resolution. Our findings highlight the unique power of low-mass galaxies to constrain differences in simulation physics, especially star formation and feedback prescriptions. Upcoming surveys from facilities like the Vera C. Rubin Observatory and *Euclid* will enable more rigorous comparisons with simulations, offering deeper insights into the physical processes shaping galaxy evolution.

Key words: Methods: data analysis – Methods: numerical – Methods: observational – galaxies: dwarf.

1 INTRODUCTION

The study of galaxy evolution has made significant advances in recent decades, largely propelled by extensive wide-area surveys such as the Sloan Digital Sky Survey (SDSS; York et al. 2000). A notable limitation of these surveys is their primary focus on detecting the brightest and most massive objects in the sky, with completeness rapidly decreasing for objects with surface brightnesses fainter than $\mu_r < 23$ mag arcsec^{-2} (Kniazev et al. 2004; Blanton et al. 2005; Driver et al. 2005). Both observational and theoretical investigations (e.g. Disney 1976; Dalcanton et al. 1997; Martin et al. 2019; Jackson et al. 2021; Kim et al. 2022) have postulated that they may only capture a fraction of the entire galaxy population, leaving us with an incomplete picture of galaxies and their evolution.

Notably absent from past wide-area surveys are dwarf galaxies (especially at distances beyond the local Universe) as well as more massive objects like the giant low-surface-brightness (LSB) galaxies Malin 1 (Bothun et al. 1987) and UGC 1382 (Hagen et al. 2016), as well as faint tidal structures found in the outskirts of galaxies and clusters (Kaviraj 2014; Montes et al. 2021; Martin et al. 2022).

★ E-mail: garreth.martin@nottingham.ac.uk

This issue is likely to be particularly pronounced in the low-mass regime ($M_\star < 10^{9.5}$ $M_\odot$), where galaxies are intrinsically faint and are often associated with relatively intense and intermittent bursts of star formation (Searle, Sargent & Bagnuolo 1973; Guo et al. 2016), which introduces a degree of stochasticity to the identification of low-mass galaxies with otherwise similar properties in shallower surveys (e.g. Jackson et al. 2021; Kaviraj et al. 2025). Furthermore, the shallower potential wells of these galaxies make them more vulnerable to processes that can result in them becoming increasingly diffuse (e.g. Di Cintio et al. 2017; Martin et al . 2019).

The detection of significant numbers of galaxies at the extremely faint and diffuse end of the low-surface-brightness population, now commonly referred to as ultra-diffuse galaxies (van Dokkum et al. 2015; UDGs), underscores the presence of a notable bias in our current ability to accurately sample and understand low-mass galaxy populations (e.g. Martínez-Delgado et al. 2016; Román et al. 2019; Prole et al. 2021). It is clear therefore that shallow imaging can introduce substantial biases in terms of which dwarf galaxy populations are recovered, limiting our understanding of galaxy populations and their evolution to a biased subset of the Universe.

Recent advancements have led to remarkable improvements in the sensitivity of wide-area survey instruments, enabling the routine detection of distant dwarf galaxies in various cosmic environments. Surveys now being conducted by next-generation ground- and space-based instruments including those undertaken by the Vera C. Rubin Observatory (Olivier, Seppala & Gilmore 2008; Ivezić et al. 2019), *JWST* (Gardner et al. 2006), and *Euclid* (Laureijs et al. 2011; Borlaff et al. 2021) are poised to revolutionize galaxy evolution studies. These surveys will provide an unprecedented opportunity to gain a detailed statistical understanding of the LSB Universe. In a matter of days, the Rubin Observatory will surpass the depth and fidelity of the SDSS's initial 8-year survey in the Southern Sky. Moreover, it will continue to produce even deeper imaging over the next decade, achieving a maximum limiting surface brightness of $\mu_r \sim 30.3$ mag arcsec^{-2}. Simultaneously, space-based, diffraction-limited observatories are already providing high-resolution observations of low-mass galaxies, reaching back as far as the early Universe (e.g. COSMOS-Web, PANORAMIC; Williams et al. 2021; Casey et al. 2022). These new observational data sets will be transformative, enabling the study of large, unbiased galaxy samples across the low- and high-mass regimes and significantly advancing our statistical understanding of galaxy populations.

In parallel, high-resolution cosmological simulations are beginning to play a crucial role in guiding our exploration of poorly understood corners of discovery space. By facilitating realistic synthetic observations through forward modelling of simulation data within realistic cosmological contexts, these simulations offer comprehensive predictions with fidelity matching the capabilities of new instruments. Although simulations with sufficient resolution to investigate the resolved properties of dwarf galaxies have traditionally been confined to single-halo zoom-ins (e.g. Guedes et al. 2011; Hopkins et al. 2014; Wetzel et al. 2016), larger simulations with box sizes spanning tens of Mpc are becoming more common (e.g. Tremmel et al. 2017; Nelson et al. 2019; Pillepich et al. 2019; Dubois et al. 2021; Feldmann et al. 2023).

By simulating relatively large contiguous volumes, it is possible to overcome several shortcomings associated with zoom simulations, including limited sample sizes, potential selection biases, and a restricted ability to probe cosmic structures and correlations over large scales. By encompassing more extensive volumes, these larger simulations offer a broader and more representative view of the properties of the Universe. Although the present coverage of such simulations is still relatively limited, the ongoing improvement in the size of high-resolution simulations promises to enable more direct comparisons between observed and simulated galaxies. For example, recent work by Dubois et al. (2021) and Kim et al. (2022) has demonstrated that a more realistic treatment of observational selection effects through forward modelling of simulation data can alleviate some of the observed discrepancies between theoretical and observed galaxy stellar mass functions. In a similar vein, high-resolution simulations in cosmological volumes are also able to resolve classical tensions between theory and observations at small physical scales (Bullock & Boylan-Kolchin 2017) e.g. the existence of planarity in the kinematics of satellites around massive galaxies (Ibata et al. 2013; Uzeirbegovic et al. 2024).

Cosmological simulations are typically calibrated to reproduce the properties of observed galaxy populations, but given their present observational incompleteness, this has not been possible in the low-mass regime. Dwarf galaxies, in contrast to massive galaxies, exhibit distinct evolutionary behaviour, dominated by stellar feedback and the influence of the local environment (Watkins et al. 2023). High-quality observations of the low-mass regime in combination with high-fidelity simulations can therefore provide important additional constraints on aspects of our galaxy evolution models such as stellar (Dekel & Silk 1986), AGN (Reines, Greene & Geha 2013; Kaviraj, Martin & Silk 2019; Davis et al. 2022), and UV (Haardt & Madau 1996) feedback, which are expected to exhibit significantly higher efficiency in the shallower haloes of low-mass galaxies.

In anticipation of data from larger-scale studies, we use currently available deep data to present a preview of what will soon be possible on a much larger scale. Here, we utilize ultra-deep HSC–SSP observations of the 2 square degree COSMOS field (Scoville et al. 2007) in combination with two intermediate volume, high-resolution cosmological simulations – ILLUSTRIS TNG50 (Nelson et al. 2019; Pillepich et al. 2019) and NEWHORIZON (Dubois et al. 2021). The focus of this paper is on measuring the morphologies of galaxies, which have been shown to correlate with their physical properties (e.g. Dressler 1980; Dressler et al. 1997; Strateva et al. 2001; Hogg et al. 2002; Bundy, Ellis & Conselice 2005; Conselice 2006; Skibba et al. 2009; Bluck et al. 2014; Whitaker et al. 2015; Uzeirbegovic, Geach & Kaviraj 2020; Uzeirbegovic, Martin & Kaviraj 2022; Jang et al. 2023), and thus probe underlying physical processes that drive their morphological evolution over cosmic time (Martin et al. 2018). This approach offers a means to constrain and enhance our understanding of these processes in the largely unexplored low-mass regime.

We begin in Section 2 with a detailed description of the observed and simulated data. Then, in Section 3, we describe the techniques we employ to generate realistic HSC–SSP-like images from the simulation data, as well as our methodology for assessing galaxy structure and comparing visual similarity. Finally, in Sections 4 and 5, we present a detailed comparison of the morphology of galaxies between the observed data and both simulations, discussing the implications that discrepancies between the simulations and observations may have for the accuracy of physical recipes implemented in each simulation.

Throughout this paper, we adopt separate lambda cold dark matter (ΛCDM) cosmologies consistent with the different values used by each simulation (see Section 2.1). For the observed COSMOS data, we adopt a Komatsu et al. (2011) cosmology with $\Omega_m = 0.272$, $\Omega_\Lambda = 0.728$, $\Omega_b = 0.045$ and $H_0 = 70.4$ km s^{-1} Mpc^{-1}.

2 DATA

In this paper, we present a comparative analysis of the structural properties of synthetic and observed dwarf galaxies. Our study incorporates realistic mock observations generated from the NEWHORIZON (Dubois et al. 2021) and ILLUSTRIS TNG50 (Nelson et al. 2019; Pillepich et al. 2019) simulations in addition to imaging and data products obtained from the COSMOS (Scoville et al. 2007) and HSC–SSP (Aihara et al. 2018) surveys. In the following sections, we provide an overview of each of these datasets.

2.1 Simulations

The theoretical component of our study makes use of two state-of-the-art cosmological simulations, NEWHORIZON and ILLUSTRIS TNG50. Despite simulating roughly similar volumes with similar fidelity, they diverge notably in their approaches to solving equations of (magneto)hydrodynamics and their sub-grid physical models.

Although both simulations implement models of black hole accretion and feedback, black hole growth is minimal for masses examined in this study in both simulations (Dubois et al. 2015, 2021; Peirani et al. 2024; Voit et al. 2024). We therefore neglect the role of active galactic nucleus (AGN) from this overview and future discussion regardless of observational and theoretical evidence of the potential role of AGN the evolution of dwarf galaxies (e.g. Kaviraj et al. 2019; Koudmani, Henden & Sijacki 2021; Davis et al. 2022).

Below, we give a concise summary of the key aspects of each simulation's approach.

2.1.1 *NEWHORIZON*

The NEWHORIZON simulation.[1] (Dubois et al. 2021) is a zoom-in of the (142 Mpc)^3 parent Horizon–AGN simulation (Dubois et al. 2014; Kaviraj et al. 2017). Within the original Horizon–AGN volume, a spherical volume with a diameter of 20 Mpc is defined, corresponding to a dark matter (DM) mass resolution and effective *initial* gas mass resolution of $m_{DM} = 1.2 \times 10^6$ M$_\odot$ and $m_{gas} = 2 \times 10^5$ M$_\odot$. The stellar mass resolution is 1.3×10^4 M$_\odot$ and the smallest allowed spatial refinement (resolution) is ~ 34 pc. Full details of the physics and galaxy formation model can be found in Dubois et al. (2021).

Cosmology: NEWHORIZON adopts a ΛCDM cosmology consistent with Komatsu et al. (2011) ($\Omega_\mathrm{m} = 0.272$, $\Omega_\Lambda = 0.728$, $\Omega_\mathrm{b} = 0.045$, $H_0 = 70.4$ km s^{-1} Mpc^{-1}).

Hydrodynamics: NEWHORIZON utilizes the adaptive mesh refinement (AMR) code RAMSES (Teyssier 2002) and gas is evolved with a second-order Godunov scheme and the approximate Harten–Lax–Van Leer-Contact (HLLC, Toro 1999) Riemann solver with linear interpolation of the cell-centred quantities at cell interfaces. 30–100 million leaf cells are used per level of refinement in the zoom-in region from level 12 to level 22 with the minimum physical size of cells kept approximately constant by adding an extra level of refinement at every doubling of the expansion factor.

Gas physics: Radiative cooling of primordial and metal-enriched gas occurs in the presence of a spatially uniform UV background beginning after redshift $z = 10$, following (Haardt & Madau 1996). Cooling is allowed down to $\sim 10^4$ K through collisional ionization, excitation, recombination, Bremsstrahlung, and Compton cooling. Metal-enriched gas can cool further down to 0.1 K based on Sutherland & Dopita (1993) above $\gtrsim 10^4$ K and from Dalgarno & McCray (1972) below $\lesssim 10^4$ K, allowing the simulation to partially resolve the multiphase nature of the ISM. The UV background is self-shielded in optically thick regions ($n_\mathrm{H} \gtrsim 0.01$ H cm^{-3}; Rosdahl & Blaizot 2012), with UV photo-heating rates reduced by a factor ofexp$(-n_\mathrm{H}/n_\mathrm{shield})$, where $n_\mathrm{shield} = 0.01$ H cm^{-3}.

Star formation: Star formation proceeds above a density threshold of $n_\mathrm{H} > 10$ H cm^{-3} following a Schmidt law with a variable efficiency related to the cloud turbulent Mach number and virial parameter (Kimm et al. 2017; Trebitsch et al. 2017; Trebitsch, Volonteri & Dubois 2020). It favours the rapid formation of stars in dense, gravitationally collapsing medium with compressible turbulence and results in potentially higher instantaneous star formation rates and more bursty star formation histories.

Feedback from massive stars: Feedback from Type II supernovae (SNe) proceeds assuming all stars with stellar masses greater than 6 M$_\odot$ explode instantaneously after 5 Myr. Each stellar particle is represented by a Chabrier (2003) initial mass function (IMF) with each SN explosion releasing kinetic energy of 10^{51} erg. To account for the cumulative effect of clustered SN explosions on total radial momentum (Thornton et al. 1998), the specific frequency of SNe is boosted by a factor of 2 to ~ 0.03 M$_\odot^{-1}$ (Kim, Ostriker & Raileanu 2017; Gentry et al. 2019).

A mechanical SN feedback model (Kimm & Cen 2014; Kimm et al. 2015) is employed. In the Sedov–Taylor energy-conserving phase, the assumed specific energy is injected to the gas, since hydrodynamics will naturally capture the expansion of the SN and impart the correct amount of radial momentum. In the momentum-conserving phase radial momentum is imparted directly according to (Thornton et al. 1998). This avoids artificially rapid radiative cooling caused by under-resolved cooling lengths, which would otherwise suppress the expansion of the SN bubble.

Other sources of stellar feedback such as stellar winds and Type Ia SNe are not implemented in the NEWHORIZON stellar feedback model.

Volume: The NEWHORIZON volume encompasses a single contiguous region of average density spanning approximately ~ 20 Mpc. This choice strikes a balance between sampling diverse environments and achieving high resolution, although it does not include extremely dense or rarefied regions. As a result, the volume effectively probes field and group environments, while falling short of cluster environments. The maximum halo mass found in the NEWHORIZON volume is $M_h \sim 10^{13}$ M$_\odot$.

2.1.2 *ILLUSTRIS TNG50*

The Next Generation Illustris[2] (ILLUSTRISTNG) is a suite of cosmological magnetohydrodynamical (MHD) simulations covering three different comoving volumes at varying resolutions. In this paper, we use of the TNG50 run (Nelson et al. 2019; Pillepich et al. 2019) which simulates a volume with 50 comoving Mpc on a side, with a stellar and gas particle resolution of 8.5×10^4 M$_\odot$ and a median spatial resolution for star-forming ISM gas of 100–140 pc. Full details of the physics and galaxy formation model can be found in Weinberger et al. (2017) and Pillepich et al. (2018).

Cosmology: TNG50 adopts a ΛCDM cosmology consistent with Planck Collaboration XIII (2016) ($\Omega_\mathrm{m} = 0.309$, $\Omega_\Lambda = 0.691$, $\Omega_\mathrm{b} = 0.049$, $H_0 = 67.7$ km s^{-1} Mpc^{-1}).

(Magneto)hydrodynamics: TNG50 uses the moving-mesh code AREPO (Springel 2010). AREPO uses a finite volume method on an

[1] https://new.horizon-simulation.org/

[2] https://tng-project.org/

unstructured, moving, Voronoi mesh, with a directionally unsplit second order Godunov scheme (Pakmor et al. 2016). TNG50 solves equations of idealized continuum MHD utilising cell-centred magnetic fields (Pakmor, Bauer & Springel 2011) combined with an approximate HLLD Riemann solver (Miyoshi & Kusano 2005).

Gas physics: Radiative cooling of primordial and metal-enriched gas occurs in the presence of a redshift-dependent, spatially uniform UV background, with corrections for self-shielding in the dense ISM (Katz, Hernquist & Weinberg 1992; Faucher-Giguère et al. 2009). Further redshift dependent cooling of metal-enriched gas from metal lines is allowed based on Smith, Sigurdsson & Abel (2008); Wiersma et al. (2009). TNG50 does not directly resolve the multiphase structure of the ISM, but rather treats it using an idealized model (Springel & Hernquist 2003).

Star formation: Star formation proceeds stochastically above a density threshold of $n_{\rm H} > 0.1\,{\rm H\,cm^{-3}}$ with constant efficiency following the Kennicutt–Schmidt relation.

Feedback from massive stars: Stellar feedback follows the Springel & Hernquist (2003) idealized multiphase effective equation of state model, in which feedback energy from Type Ia and Type II SNe directly heats the ambient hot phase and returns metal enriched gas to the ambient ISM assuming a Chabrier (2003) IMF. The model includes additional feedback from isotropically injected star formation driven kinetic winds including contributions from the asymptotic giant branch. The galactic-scale winds generated by stellar feedback are placed by hand and are decoupled from interacting hydrodynamically with the surrounding gas until they reach regions with densities significantly below the star formation threshold.

Volume: The TNG50 volume models a smaller region at a higher fidelity compared to the rest of the IllustrisTNG suite. Spanning approximately 50 Mpc, this region captures a relatively diverse range of environments, including voids and filaments, galaxy groups and poor clusters while excluding more extreme environments such as more massive clusters. The maximum halo mass found in the TNG50 volume is $M_{\rm h} \sim 10^{14}\,{\rm M_\odot}$.

2.2 Observations

For the observational component of our study, we make use of catalogues from the COSMOS survey as well as supplementary deep imaging from Hyper Suprime-Cam. The COSMOS field is extremely well covered by a wealth of photometric observations across the UV, optical and IR spectrum, allowing for precise determination of galaxy redshifts and physical properties.

Like TNG50 and NewHorizon, COSMOS probes relatively average environments. For the redshift range $0.05 < z < 0.3$ and galaxy stellar masses $M_\star > 10^{10.5}\,{\rm M_\odot}$, we calculate a galaxy number density of $0.0007\,{\rm Mpc^{-3}}$ for our COSMOS sample. This is smaller but still comparable to galaxy number densities obtained for larger fields at similar redshift ranges. For example, by integrating the galaxy stellar mass function for SDSS at $z < 0.05$ presented in Baldry, Glazebrook & Driver (2008), we obtain a value of $0.0017\,{\rm Mpc^{-3}}$. For the VIMOS VLT Deep Survey (VVDS) mass function at $0.05 < z < 0.4$ presented in Pozzetti et al. (2007), we obtain a more similar value of $0.0008\,{\rm Mpc^{-3}}$. In comparison, the number density of galaxies more massive than $10^{10.5}$ ${\rm M_\odot}$ at $z = 0.25$ is 0.0039 and $0.0029\,{\rm Mpc^{-3}}$ within the NewHorizon volume and TNG50 volume respectively, both slightly larger than the average number density obtained for SDSS.

2.2.1 HSC–SSP

We make use of `deepCoadd` i-band imaging from the third release of the Hyper Suprime-Cam Subaru Strategic Program (HSC–SSP DR3; Aihara et al. 2022)[3] Prior to the co-addition, a global third-order sky subtraction fit to 8k×8k ($\sim 20'$) superpixels followed by a second sky subtraction using 256 × 256 (~ 40 arcsec) superpixels is performed for each exposure.

We cross-match positions from the COSMOS2020 catalogue using the `pdr3_dud_rev` catalogue. The DR3 catalogue uses `calexp` frames for the detection and segmentation of sources, these are generated by performing an additional local sky subtraction using 128 × 128 superpixels on the `deepCoadd` image. Source detection proceeds following the same maximum likelihood detection method used by SDSS and is described in detail in Bosch et al. (2018). For imaging, we choose to use `deepCoadd` rather than `calexp` frames because these better preserve flux at scales larger than the local sky subtraction bin size. However, as discussed in the next section, this is not likely to be significant for our sample of galaxies.

In order to retain as complete a sample as possible, we restrict our study only to the central 1.5° of the COSMOS field where imaging depth exceeds $\mu(10\,{\rm arcsec} \times 10\,{\rm arcsec}, 3\sigma) = 31\,{\rm mag\,arcsec^{-2}}$ (see Section A for a more in-depth exploration).

2.2.2 COSMOS2020

We make use of data products from the COSMOS2020 CLASSIC catalogue (Weaver et al. 2022)[4]. COSMOS2020 source detections are performed using SExtractor (Bertin & Arnouts 1996) on a χ^2 combined $izYJHK_s$ detection image (Szalay, Connolly & Szokoly 1999) which incorporates deep i-band and z-band imaging from HSC–SSP DR2. The HSC–SSP i-band imaging provides exceptional depth compared to other imaging used in COSMOS2020, making it the primary factor in most source detections and therefore yielding very similar results to the `pdr3_dud_rev` catalogue.

The COSMOS2020 χ^2 detection images are constructed from HSC–SSP DR2 `calexp` frames, which use the same global sky subtraction method as the DR3 `deepCoadd` frames. These frames do not include the additional 128 × 128 superpixel local sky subtraction employed in the DR3 `calexp` frames which are used to generate the DR3 catalogues. Due to improved depth and image quality, we choose to make use of DR3 `deepCoadd` frames rather than attempt to precisely match the COSMOS2020 catalogue by using identical DR2 `calexp` frames. Although the additional local sky subtraction step may result in some differences in the data used to construct the COSMOS2020 and HSC–SSP DR3 object catalogues, including over-subtraction of the faint outskirts of large objects, the reduction of the two data sets is largely consistent at the scale of objects considered in this work, which are smaller than the 128 × 128 bin size used for the DR3 local sky subtraction step.

In our analyses, we employ LePhare (Arnouts et al. 1999; Ilbert et al. 2006; Arnouts & Ilbert 2011) photometric redshift, stellar mass, and rest-frame photometry estimates from the COSMOS2020 catalogue. COSMOS2020 utilizes extensive coverage from photometric bands spanning UV, optical and IR data from GALEX,

[3] https://hsc-release.mtk.nao.ac.jp/
[4] https://cosmos2020.calet.org/

CFHT, Subaru, UltraVISTA, and IRAC, which enables precise photometric redshift estimates that exceed a precision of 5 per cent, even for the faintest sources. Notably, for the brightest sources with $i < 21$, COSMOS2020 achieves photometric redshift performance that approaches the accuracy of spectroscopic estimates, with a precision better than 1 per cent (Weaver et al. 2022).

3 METHOD

3.1 Imaging

In the following sections, we detail the methodology employed to extract postage stamp images from the HSC–SSP survey and the generation of corresponding postage stamp images from the NEWHORIZON and TNG50 simulations.

3.1.1 Observed galaxies

We cut out g, r and i-band postage stamps from the HSC–SSP `deepCoadd` data via the HSC–SSP DAS cutout tool. Cutouts are centred on the coordinates specified in the COSMOS2020 catalogue with an angular size of 1 arcmin.

3.1.2 Synthetic galaxies

For both simulations, we select objects from a single snapshot corresponding to the median redshift of our COSMOS2020 sample ($z \sim 0.2$). While this does not take into account redshift evolution, the disparity in structural properties between the two simulations is considerably larger than the limited morphological changes that might occur between $z \sim 0.25$ and $z \sim 0.05$. Instead, to match the redshift distribution of the observed sample, we assign galaxies distances matching the redshift distribution of the observed sample. Our approach accounts for observational effects such as the reddening of the SED, cosmological dimming and evolution of angular size with redshift.

NEWHORIZON and TNG50 catalogues utilize different structure finders, namely (ADAPTA) HOP (Aubert, Pichon & Colombi 2004) and SUBFIND (Springel et al. 2001), respectively, for galaxy identification. To avoid selecting galaxy fragments we use the HOP rather than ADAPTAHOP structure finder for NEWHORIZON, which has the effect of keeping all stellar substructures connected to the same main substructure. To ensure that we are relatively agnostic to the structure finder used and that extended substructures are preserved, images are generated using every star particle lying within a 100 kpc box centred around the centroid of each galaxy as defined by their respective structure finder.

Images: We follow almost the same procedure as Martin et al. (2022), in this case to produce HSC-like, rather than Rubin-like, mock observations. We briefly outline our method below.

We begin by generating spectral energy distributions (SEDs) for each star particle extracted from within a 100 kpc box around the centroid of each galaxy using a grid of Bruzual & Charlot (2003) simple stellar population (SSP) models interpolated to the age and metallicity of each star particle and assuming a single Chabrier (2003) IMF. Dust attenuation is calculated for the SED of each stellar particle via a screen model which depends on the dust column density summed along the line of sight to each star particle, calculated assuming a gas-to-dust ratio of 0.4 (e.g. Draine et al. 2007). An $R = 3.1$ Milky Way dust grain model (Weingartner & Draine 2001) is then used to produce the final dust attenuated SED for each particle. As Watkins et al. (2025) show, low gas column densities mean that the precise dust model adopted has very limited influence on the broad-band photometry of galaxies in the mass range considered here. Adopting an SMC extinction curve and varying the dust-to-metal ratios between 0.01 and 1.0 introduces a typical variation in the central surface brightness of the order 0.01 mag arcsec^{-2}.

This approach accounts for the geometry of the spatial distribution of dust within and around the galaxy, but does not account for other effects such as scattering. However, our exact treatment of dust is unlikely to have any significant impact on galaxy structure, especially, at the low redshifts probed by this study, where gas-rich starbursts are rare (e.g. Atek et al. 2014). For the mass ranges considered in this study, we find that a majority of galaxies have very low dust column densities, which result in 95 per cent of galaxies having i-band magnitudes attenuated by less than 0.2 mag, with a majority of these being attenuated by a completely negligible amount. Additionally, due to the redshift distribution of our sample, for which more than 90 per cent of galaxies are found at redshifts $z > 0.1$, structures produced by dust are not well resolved in the vast majority of cases. Dust lanes are only expected to begin to appear at stellar masses of 10^9 M$_\odot$ (Holwerda et al. 2019), which is close to the end of the mass range considered in this study. Furthermore visual inspection of a subset of observed images from HSC–SSP in the mass and redshift range covered by our sample, does not yield any clear dust structures.

We finally create a 2D image by first redshifting and convolving each SED with the HSC i-band transmission functions (Kawanomoto et al. 2018) and then adaptively smoothing the particles in 3D to better represent the distribution of stellar mass in phase space and remove unrealistic variation between adjacent pixels due to resolution effects. The smoothed distribution is summed along one of the axes and binned into 0.168″ pixels to produce 2D g, r, and i-band flux maps, which are then convolved with the HSC PSF (Montes et al. 2021) and scaled so that the magnitude is given by $i = -2.5\log_{10}(f)$ where f is the flux in each pixel.

Source injection To enhance the realism of our synthetic images, we then inject them into real backgrounds drawn from the HSC–SSP data. This ensures that the distribution of depths in our synthetic data closely matches that of the observed data as well as incorporating a realistic distribution of nearby background objects, which could potentially introduce scatter or systematic biases into our measurements.

In the central region of the COSMOS field, we identify suitable sites for injecting synthetic sources. Using the ASTROPY package PHOTUTILS (Bradley et al. 2022), we select locations that are sufficiently distant from any detected pixels. To do this, we first estimate the 2D background using a median background estimator with 3σ clipping and then apply a detection threshold of 1.5 times the background RMS (1.5σ). This approach minimizes the risk of chance projections between real sources and our injected synthetic sources.

We then generate a detection map for multiple randomly selected regions within the central COSMOS HSC–SSP `deepCoadd` images. Each image is first convolved with a Gaussian kernel with a standard deviation matching the measured FWHM. We then apply the 1.5σ threshold to identify candidate injection sites, ensuring they are at least 20 pixels away from any detected pixels.

Each selected point serves as the centroid for injecting synthetic galaxies. Since the initial regions are randomly selected, we disregard any effects from correlated structure in the simulations.

The HSC–SSP `deepCoadd` images are scaled to a zero-point of 27 mag. After convolution, we rescale our synthetic images using the

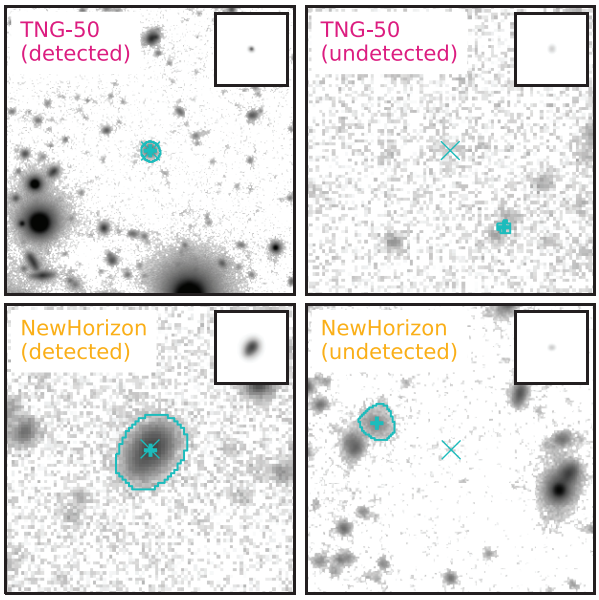

Figure 1. Examples of synthetic galaxy injections onto HSC–SSP `deepCoadd` data. The centroid of the injected galaxy is indicated with an 'X' while thicker crosses indicate the centroid of the nearest detected object, with contours showing the segmentation map for this object. The inset panel shows the injected object. All images have the same physical scale at the redshift of the injected galaxy, therefore the angular scale differs between images.

same `fluxmag0` parameter to ensure their fluxes match the image DNs of the HSC data (in units of nJy) before injecting them at the designated centroids.

Since our procedure involves injecting objects into pre-processed `deepCoadds`, the impact of the injected objects on sky subtraction is not accounted for. However, given their faintness and small sizes relative to the superpixels used for background modeling, we do not expect a significant effect on either sky subtraction or detectability.

3.2 Detection & segmentation of synthetic & observed galaxies

After injecting each synthetic object, we use PHOTUTILS to generate a new detection map following the same procedure described above. Once the detection map is created, we deblend it to produce a segmentation map using the PHOTUTILS `deblend_sources` procedure which utilizes a combination of multithresholding and watershed segmentation. For source deblending, we set a minimum galaxy area of 10 pixels, use 32 deblending levels, and adopt a minimum contrast of 0.001.

For each object identified in the segmentation map, we determine its centroid and compare it to the centroid of the injected galaxy. A synthetic object is considered detected if the segmentation map contains an object with a centroid located within either 10 pixels or one effective radius of the injection centroid. Since our selection criteria require no detected pixels within a 20-pixel radius of the injection centroid, the likelihood of a detection resulting from a chance projection is significantly reduced.

Fig.1 illustrates examples of both detections and non-detections for TNG50 and NEWHORIZON galaxies following their injection into HSC–SSP i-band data.

Rather than relying on the footprints provided by the HSC–SSP data release, we follow an almost identical procedure for detecting and segmenting observed galaxies from the HSC–SSP `deepCoadd`

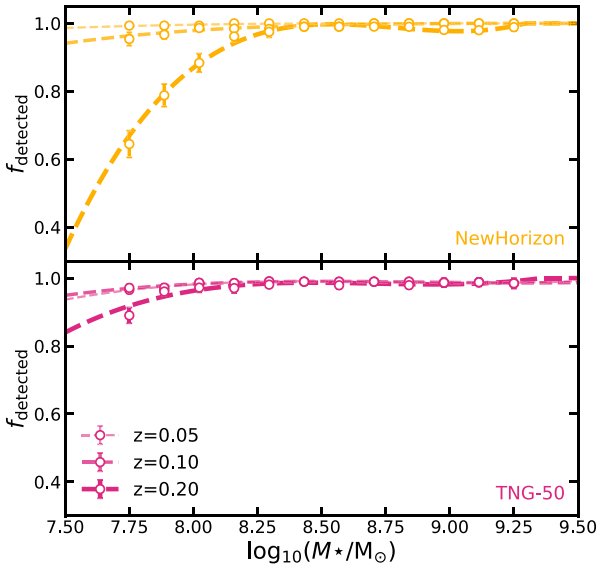

Figure 2. Fraction of galaxies with successful detections as a function of stellar mass following injection of the same sample of synthetic galaxies into HSC–SSP `deepCoadd` at three different redshifts indicated by the legend. Dashed lines show polynomial fits to the binned points. The top panel shows the results for NEWHORIZON and the bottom panel shows the results for TNG50.

images by again generating a detection map and performing source deblending using PHOTUTILS. To ensure that the object segmentation maps we generate using PHOTUTILS correspond well to the object footprints used to calculate the quantities in the COSMOS2020 catalogue, we require that the centroids from the COSMOS2020 catalogue match the centroids of the recovered sources to within 10 pixels.

3.3 Sample selection

We first select a sample of dwarf galaxies using the LEPHARE derived photometric redshifts and masses taken from the COSMOS2020 catalogue. Galaxies are selected only from the region of COSMOS with the deepest HSC i-band imaging (better than 31 mag arcsec^{-2}, Appendix A), and with the following properties:

(i) photometric redshift in the range $0.05 < z < 0.25$
(ii) fractional redshift error smaller than 10 per cent
(iii) $10^{7.5} < M_\star/{\rm M}_\odot < 10^{9.5}$
(iv) Within the central 1.5° of the COSMOS field.

yielding a sample of 1320 galaxies.

To evaluate the expected observational completeness of our sample. we conduct an analysis of detection rate of galaxies from each simulation injected into the HSC–SSP COSMOS field. Fig. 2 shows the fraction of detected galaxies as a function of stellar mass and redshift for both the NEWHORIZON and TNG50 simulations. In the case of both simulations, we find that nearly 100 per cent of galaxies are detected across the entire mass range of our sample up to a redshift of $z = 0.1$. However, it is worth noting that NEWHORIZON galaxies exhibit a noticeable decrease in completeness at higher redshifts (which is likely a result of their more diffuse structure as shown in Section 4.1). Specifically, at $z > 0.2$, NEWHORIZON galaxies begin to exhibit incompleteness for stellar masses below $10^8{\rm M}_\odot$, with completeness levels dropping to around 40 per cent for galaxies with a stellar mass of $10^{7.5}{\rm M}_\odot$.

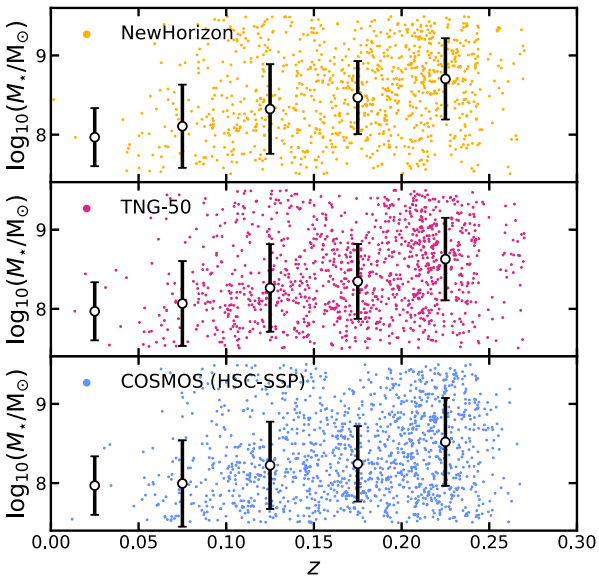

Figure 3. The stellar mass vs redshift distributions of the matched NEWHORIZON (top panel) and TNG50 (middle panel) distributions compared with the original COSMOS sample (bottom panel). Black error bars with white circles show the 1σ dispersion and median of each distribution for bins of redshft,.

Within the central region of the COSMOS field and within the mass and redshift ranges of our sample, completeness is expected to be high. However, at higher redshifts, we begin to observe some incompleteness. Notably, at low redshifts ($z < 0.1$), both simulations demonstrate exceptionally high completeness, suggesting that neither simulation predicts the existence of a significant population of galaxies that would remain undetected when accounting for selection effects. As we show in Section 4, the average Sérsic indices and effective radii of COSMOS galaxies fall between the ranges defined by NEWHORIZON and TNG50. Consequently, we anticipate the completeness of COSMOS galaxies to lie somewhere between the completeness levels seen in the two simulations (see e.g. Greco 2018; Greene et al. 2022, for discussion of the completeness of the HSC–SSP *wide* layer as a function of Sérsic parameters).

To replicate the observational biases present in the HSC–SSP data, we generate a redshift and mass matched sample of objects for both the NEWHORIZON and TNG50 simulations. We allow simulated galaxies to be drawn more than once, provided they are observed at a different redshift and in a different orientation. Both synthetic samples contain the same number of objects as the COSMOS sample after any undetected objects are rejected. Fig. 3 shows the mass and redshift distribution of these matched samples.

In order to understand any biases in our measurements, we additionally create 'rest-frame' images for the same mass-matched sample of NEWHORIZON and TNG50 galaxies in which the pixel scale of the images are set to a fixed physical size of 1 kpc arcsec^{-1}, equivalent to 0.168 kpc per pixel given the HSC pixel size. Stellar SEDs are also kept in the rest-frame but the the images are otherwise produced and processed identically.

3.4 Parametric and non-parametric galaxy structure measures

In this study, we employ non-parametric Gini, M_{20}, and CAS measures of galaxy structure, as well as single-component Sérsic fits, to analyse both simulated and observed galaxies (Conselice et al. 2003).

To calculate these parameters, we make use of the STATMORPH PYTHON package (Rodriguez-Gomez et al. 2019). STATMORPH is a tool designed for the analysis and characterization of structural properties of galaxies and has been previously employed to investigate both simulated and observed galaxies (e.g. Rodriguez-Gomez et al. 2019; Sazonova et al. 2020; Guzmán-Ortega et al. 2023; Kartaltepe et al. 2023). Simulation-based studies (Rodriguez-Gomez et al. 2019; Guzmán-Ortega et al. 2023) have focused on massive galaxies at low redshift ($z < 0.05$) using both the ILLUSTRIS TNG50 simulation and the larger volume, lower resolution TNG100 simulation. We adopt a slightly different approach to these studies in our analysis. As described earlier, we insert galaxies into observed backgrounds and then segment and perform measurements using the same methodology applied to the observed galaxies. This allows us to take into account observational biases, exclusively measuring the properties of galaxies that would be detectable in observational data.

Non-parametric measures are sensitive to imaging resolution and depth (e.g. Sazonova et al. 2024), making it challenging to compare the properties of objects observed at different redshifts. By selecting a sample with matched redshift and mass distributions, we mitigate this issue. However, individual measurements will still exhibit redshift dependence, as demonstrated by the rest-frame quantities presented in Appendix B.

In the subsequent sections, we provide a concise overview of the measures employed throughout the remainder of this paper. We discard all galaxies where the Sérsic index exceeds $n = 4$. As well as being rare in the nearby Universe in the low-mass regime (Su et al. 2021; Seo & Ann 2022; Watkins et al. 2022; Lazar et al. 2024a), many small, high Sérsic index objects will appear as point sources at the redshift and spatial resolution employed in this study and therefore likely to be poor fits.

3.4.1 *Sérsic profile fitting*

The Sérsic profile (Sérsic 1968) parametrizes the light profile of a galaxy. It encapsulates several important characteristics of the light profile, including the effective radius ($R_{\rm eff}$), the surface brightness at the effective radius ($\mu_{\rm eff}$), the ellipticity (e) and the Sérsic index (n), which parametrizes the central concentration of the profile.

To properly account for the impact of seeing, we take advantage of STATMORPH's ability to convolve the Sérsic model with a PSF while fitting. In our study, we use the same HSC i-band PSF obtained from Montes et al. (2021) which we previously used to generate our synthetic observations. Because there is some variation in the PSF across the HSC–SSP COSMOS field, we rescale the PSF model when fitting to our observed sample. For each object, we rescale the Montes et al. (2021) PSF to match the FWHM measured at the centre of the specific patch where each galaxy is located, as indicated in the `iseeing` column of the `patch_qa` table within the HSC–SSP catalogue.

3.4.2 *Gini-M_{20}*

The Gini-M_{20} classification system (Lotz, Primack & Madau 2004) is widely used in astronomy both to identify mergers or irregular galaxies and as a general measure of galaxy morphology.

Gini: The Gini statistic (Glasser 1962) quantifies the homogeneity of the light distribution within a galaxy by comparing its sorted cumulative distribution function to the expectation of an even flux distribution. In the context of this study, the Gini statistic is adapted from the original definition by Glasser (1962) to account for negative flux values, requiring a first-order correction (Lotz et al. 2004).

Lower Gini values correspond to galaxies with more uniform light distributions, such as late-type spirals with extended discs, while higher values indicate centrally concentrated systems, such as early-type galaxies or compact starbursts.

***M*$_{20}$** The M_{20} statistic (Lotz et al. 2004) measures the spatial variance of the brightest regions in a galaxy relative to the entire galaxy. It compares the intensity-weighted central second moment of the galaxy with the sum of the second moments of the pixels that containing the brightest 20 per cent of the total galaxy flux. Lower values of M_{20} correspond to more centrally concentrated galaxies, such as ellipticals or early-type spirals, where the brightest regions are closer to the centre. Higher values of M_{20} indicate more spatially extended bright regions, as seen in late-type spirals and irregular galaxies, often associated with star formation clumps or disturbed morphologies.

3.4.3 *Concentration, asymmetry and smoothness (CAS)*

Another commonly used non-parametric method for quantifying galaxy structure is based on the CAS (Concentration, Asymmetry, and Smoothness) statistics (Conselice et al. 2003). Our analysis is conducted within STATMORPH's default circular aperture size set at 1.5 Petrosian radii (R_P).

Concentration: The concentration parameter compares the radii enclosing 80 per cent and 20 per cent of a galaxy's total flux, reflecting the steepness of its surface brightness profile. Disc-like galaxies tend to have lower concentrations, while bulge-dominated and early-type systems have higher values. Mergers can span a wide range, with early-stage interactions appearing more diffuse and late-stage mergers becoming more centrally concentrated.

Asymmetry: The asymmetry parameter quantifies deviations from symmetry by comparing a galaxy's flux distribution to its 180 deg rotated counterpart, with a background correction applied. Regular elliptical and disc galaxies tend to be more symmetric, while interacting or merging systems, irregular galaxies, and those with prominent star-forming regions or dust lanes exhibit higher asymmetry.

Smoothness: The smoothness parameter measures the contribution of small-scale structures to a galaxy's light distribution. Smooth, well-ordered systems like ellipticals have lower smoothness, while clumpy or irregular galaxies, including mergers and star-forming systems, show higher values. However, due to the small angular sizes of most galaxies in our sample, we do not use smoothness, as reliable estimates are difficult to obtain.

In our final samples, we discard any objects flagged as having poor quality fits for either the Sérsic measurements or any of the other morphological measurements. We observe a comparable reduction in sample size in each data set: a decrease of 12 per cent, 12 per cent, and 8 per cent for NEWHORIZON, TNG50, and COSMOS, respectively. This similar reduction in both simulated and observed data suggests that there is no notable bias in terms of the quality of fits between them.

4 RESULTS

In this section, we present a comparison of the structural and visual properties of our matched synthetic and observed dwarf galaxy samples. We note that this study consists of a comparison of the statistical projected properties of galaxies across a relatively large redshift range which introduces variations in individual measurements due to factors such as seeing, redshifted galaxy SEDs, cosmological dimming, and evolution of the angular size distance.

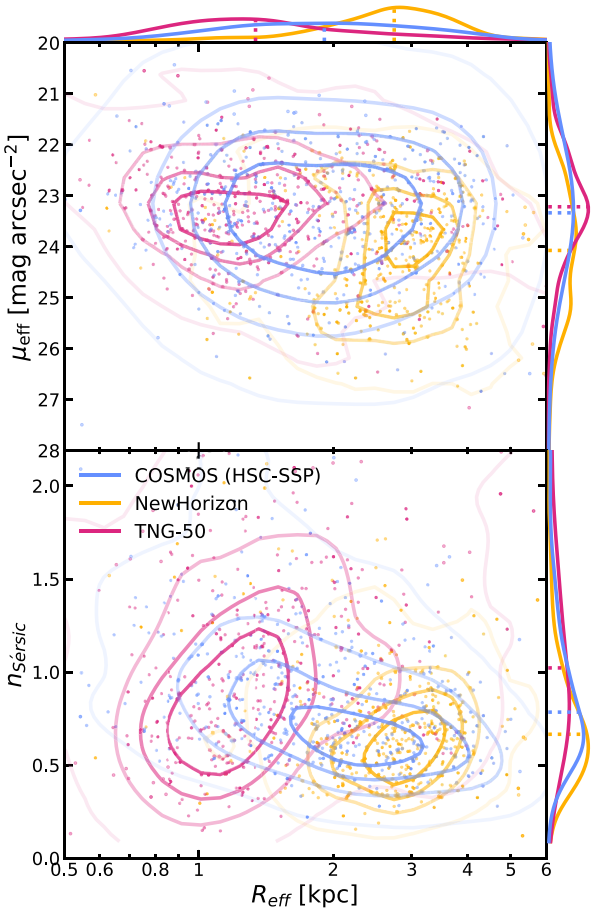

Figure 4. Contour plots showing the 2D distribution of surface brightness at the effective radius (top panel) and Sérsic index (bottom panel) as a function of effective radius for COSMOS (blue) with the redshift and mass-matched samples from NEWHORIZON (yellow) and TNG50 (red). Coloured points show a randomly selected sub-sample with the same colour scheme. The sides of each panel show the marginal distribution of each variable with dashed lines indicating the median values. See Fig. B1 for a plot showing the rest-frame quantities at fixed angular scale.

This is somewhat mitigated in the case of the Sérsic fits as the model is convolved with the PSF, but no corrections are applied to other quantities.

In Appendix B, we present rest-frame measurements of the simulated galaxies at a fixed angular scale. This facilitates more direct comparisons between the two simulations, which we also compare with a sample of COSMOS galaxies at $z < 0.1$. We note that, while we observe consistent qualitative trends between the observed- and rest-frame quantities of our Sérsic fits, there are some differences in some of the other quantities, driven primarily by the very compact sizes of some TNG50 galaxies. These differences and their causes are discussed in detail throughout this section.

4.1 Sérsic profile fitting

In this section, we compare the structural properties of simulated and observed galaxies estimated from single component Sérsic fits.

Fig.4 shows the 2D distribution of Sérsic parameters for COSMOS (blue) and the redshift and mass-matched samples from NEWHORIZON (yellow) and TNG50 (red). The top and bottom panels indicate the distribution of surface brightness at the effective radius ($\mu_{\rm eff}$)

and Sérsic index ($n_{\rm Sérsic}$) respectively as a function of effective radius ($R_{\rm eff}$) in the form of scatter and contour plots. The marginal distribution of each parameter is shown along the sides of each panel, with dashed lines indicating the median value of each distribution.

Except for the case of $\mu_{\rm eff}$, average values for NEWHORIZON and TNG50 galaxies lie at opposite extremes of the COSMOS distribution. NEWHORIZON dwarfs exhibit significantly larger sizes, evident from high $R_{\rm eff}$ and $n_{\rm Sérsic}$ values, while TNG50 dwarfs appear much more compact. At higher masses than those probed in this study, the size–mass relation of galaxies for both simulations has been shown to closely match observations (Dubois et al. 2021; Wang & Lilly 2023), indicating these differences are specific to the low-mass regime.

In contrast to COSMOS and TNG50, NEWHORIZON also shows a skewed distribution in $\mu_{\rm eff}$ that extends to low surface brightnesses. COSMOS galaxies exhibit brighter $\mu_{\rm eff}$ on average than both NEWHORIZON and TNG50 galaxies, although, as we discuss later, this is likely to be quite sensitive to both the assumptions made when producing the mock images and the chemical and star-forming properties of the galaxies. But given the generally high completeness of each sample, it is unlikely that there is a strong effect on the distribution of any of the other Sérsic parameters.

Fig. 5 shows the trend in the median value of each Sérsic parameter with stellar mass, with the addition of the projected ellipticity (e). For each parameter NEWHORIZON and COSMOS galaxies both exhibit similar trends with stellar mass (albeit with different normalization). TNG50 show strong trends in $R_{\rm eff}$, $n_{\rm Sérsic}$ and ellipticity. Galaxies appear significantly rounder at lower masses, but rise to similar values to those of NEWHORIZON and COSMOS at the high-mass end. In the rest-frame (Fig. B2), there is a less significant difference, likely arising from the fact that NEWHORIZON galaxies are physically larger so less susceptible to being smeared by the PSF. The same does not appear to be true for $n_{\rm Sérsic}$, which is still seen to evolve quite significantly both in the mass-matched sample and in the rest-frame. Finally, both the mass-matched and rest-frame $R_{\rm eff}$ for TNG50 begins to plateau at low mass ($M_\star \lesssim 10^8$ M$_\odot$). We note that, given the coarser resolution of TNG50 compared with NEWHORIZON (100-140 pc versus 30–40 pc) galaxies are sampled by relatively small number of resolution elements below an effective radius of 1 kpc.

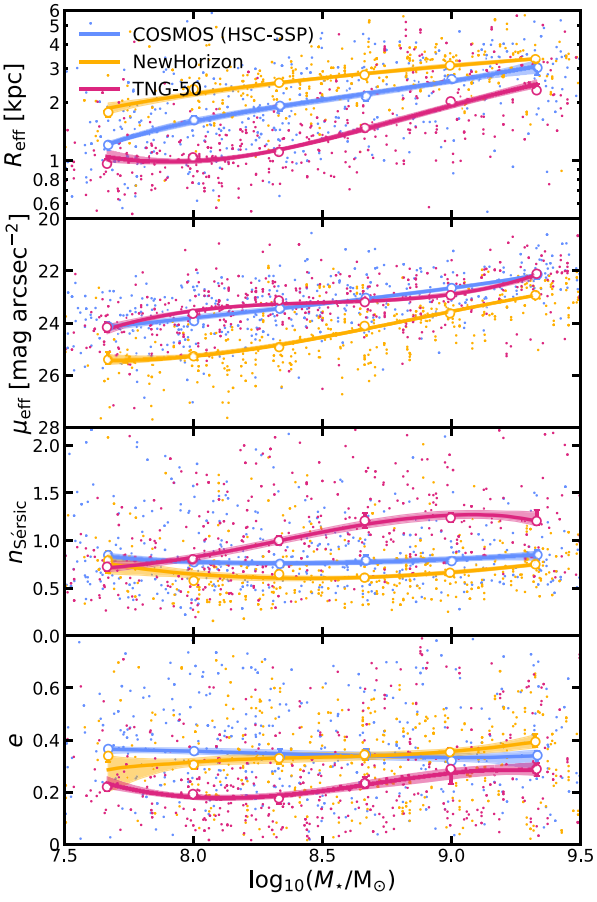

Figure 5. Plots showing the trend in the median effective radius, surface brightness at the effective radius, Sérsic index and projected ellipticity as a function of stellar mass for COSMOS (blue) and the redshift and mass-matched samples from NEWHORIZON (yellow) and TNG50 (red). Open circles with error bars show the median and error on the median for individual redshift bins, with filled regions indicating the 1σ uncertainty. Coloured points show a randomly selected sub-sample with the same colour scheme. See Fig. B2 for a plot showing the rest-frame quantities at fixed angular scale.

4.2 Gini and M_{20}

We also compare the non-parametric Gini and M_{20} (Lotz et al. 2004) measures. These metrics evaluate the homogeneity of the light distribution and the concentration of the brightest regions in relation to the centre of the light distribution, respectively.

Fig. 6 shows the 2D distribution of Gsini and M_{20} for COSMOS (blue) and the redshift and mass-matched samples from NEWHORIZON (yellow) and TNG50 (red). The marginal distribution of each parameter is along the edges of the plot, with dashed lines indicating the median value of each distribution.

NEWHORIZON and TNG50 share similar distributions although NEWHORIZON galaxies are considerably more skewed towards high Gini values and high M_{20} values indicating more uneven light distributions. COSMOS dwarfs, on the other hand, exhibit a narrow range of values concentrated towards higher Gini values and lower M_{20} values, indicating smoother, more centrally concentrated light distributions.

However, Fig. B3 reveals that, in the rest-frame, TNG50 dwarfs are shifted towards higher Gini values and lower M_{20} due to their intrinsically smaller sizes, which result in their more tightly concentrated central regions being more effectively smeared out by the PSF. Overall, TNG50 dwarfs are actually considerably more centrally concentrated than COSMOS galaxies, even compared with the COSMOS $z < 0.1$ sample. This better agrees with the results of our Sérsic fits which include a correction for the PSF and show that TNG50 galaxies have steeper Sérsic indices than COSMOS galaxies.

Examining the trend with mass in Fig. 7, COSMOS dwarfs show no notable change across the mass range, maintaining consistently high Gini values and low M_{20} values. Both TNG50 and NEWHORIZON galaxies display similar trends with stellar mass, with higher Gini and lower M_{20} values toward higher mass. This trend is qualitatively replicated in the rest-frame, except for the M_{20} of NEWHORIZON, which has flat trend with stellar mass.

4.3 Concentration and asymmetry

Finally, we compare the concentration and asymmetry parameters from the CAS (Conselice et al. 2003) system for simulated and observed galaxies. Concentration and asymmetry serve as metrics to assess the degree of central concentration and skewness in the light distribution, where higher asymmetry values and lower concentration levels signify galaxies that are more irregular and more likely

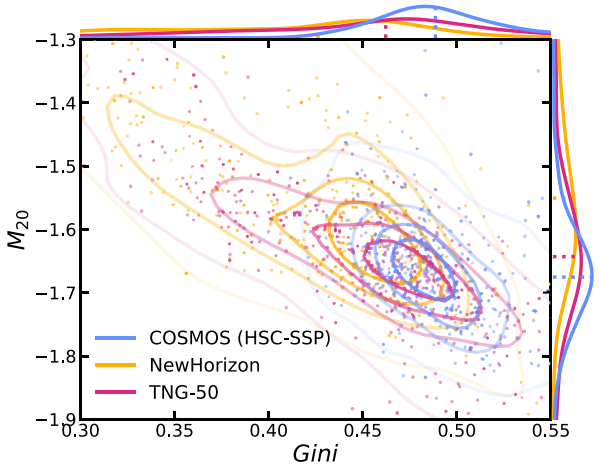

Figure 6. Contour plot showing the 2D distribution of Gini and M_{20} for COSMOS (blue) and the redshift and mass-matched samples from NEWHORIZON (yellow) and TNG50 (red). Coloured points show a randomly selected sub-sample with the same colour scheme. The sides of the plot show the marginal distribution of each variable with dashed lines indicating the median values. See Fig. B3 for a plot showing the rest-frame quantities at fixed angular scale.

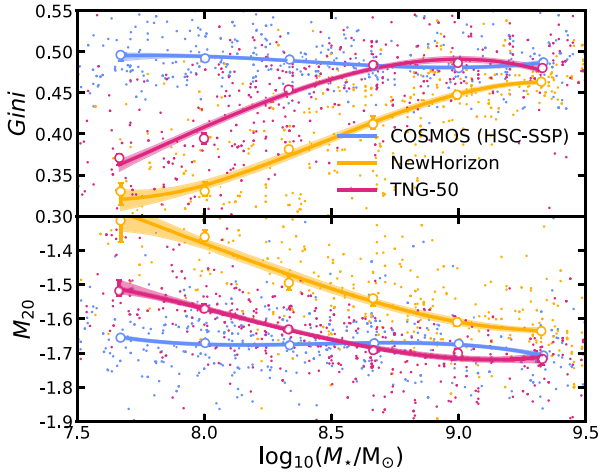

Figure 7. Plots showing the trend in the median Gini and M_{20} as a function of stellar mass for COSMOS (blue) and the redshift and mass-matched samples from NEWHORIZON (yellow) and TNG50 (red). Open circles with error bars show the median and error on the median for individual redshift bins, with filled regions indicating the 1σ uncertainty. Coloured points show a randomly selected sub-sample with the same colour scheme. See Fig. B4 for a plot showing the rest-frame quantities at fixed angular scale.

to be undergoing interactions. For the purposes of this study, we focus solely on concentration and asymmetry, while neglecting the smoothness parameter, as discussed in Section 3.4.3.

Fig. 8 shows the 2D distribution of concentration and asymmetry for COSMOS (blue) and the redshift and mass-matched samples from NEWHORIZON (yellow) and TNG50 (red). The marginal distribution of each parameter is shown along the sides of the plot, with dashed lines indicating the median value of each distribution.

NEWHORIZON dwarfs are both less concentrated and more asymmetric than the COSMOS dwarfs, indicating that they have less regular shapes and are more morphologically disturbed on average. TNG50 dwarfs have levels of asymmetry similar to COSMOS

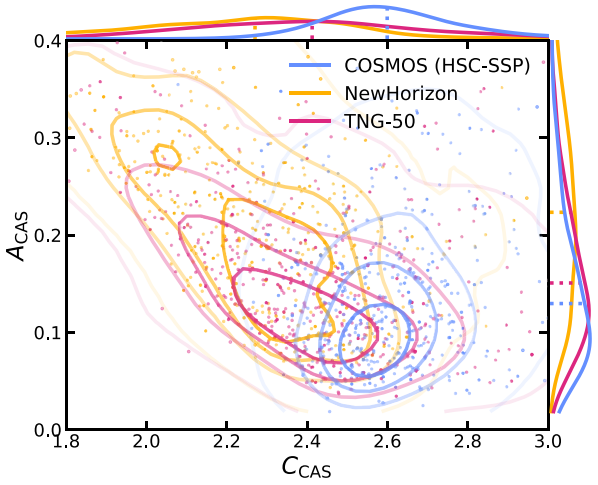

Figure 8. Contour plot showing the 2D distribution of concentration and asymmetry for COSMOS (blue) and the redshift and mass-matched samples from NEWHORIZON (yellow) and TNG50 (red). Coloured points show a randomly selected sub-sample with the same colour scheme. The sides of the plot show the marginal distribution of each variable with dashed lines indicating the median values. See Fig. B5 for a plot showing the rest-frame quantities at fixed angular scale.

dwarfs, but their concentrations are closer to those of NEWHORIZON dwarfs. However, when examining the same quantities in the rest-frame (Fig. B5), it is clear that TNG50 dwarfs are in fact considerably more concentrated than they appear in the observed-frame, likely even more so than the observed COSMOS dwarfs, whose low-redshift sample exhibit effective radii larger than the PSF FWHM. This again appears to be due to the bias resulting from their smaller sizes. For the NEWHORIZON dwarfs, the wider range of asymmetry values seen in the observed-frame is still borne out in the rest-frame distribution.

On average, NEWHORIZON galaxies have asymmetry parameters more consistent with being interacting systems (e.g. Conselice, Yang & Bluck 2009). Given the, on average, more rarefied environments inhabited by NEWHORIZON dwarfs and the expected decline in merger rate with halo mass (e.g. Fakhouri, Ma & Boylan-Kolchin 2010; Martin et al. 2021), it is unlikely that these trends are a genuine result of mergers. Instead, higher asymmetries are more likely due to internal processes which result in a clumpier light distribution.

There are a number of potential explanations for this, including differences in spatial resolution, ISM physics, star formation efficiencies and the way in which SN feedback is modelled and coupled to the ISM, which may result in the suppression stellar clumps and localized bursts of star formation. We discuss these in detail in Section 5.2.

Fig. 9 shows the trend in the median asymmetry and concentration values as a function on mass. COSMOS dwarfs exhibit limited changes across the mass range, while NEWHORIZON and TNG50 dwarfs appear more concentrated and more symmetrical towards higher masses. The same trends are observed in the rest-frame in the case of concentration, but there is no clear trend with stellar mass observed for asymmetry in either COSMOS or the simulations.

Fig. 10 provides a comprehensive summary of the parameter distributions discussed in Section 4. The coloured violin plots illustrate the distributions of each parameter, with yellow, red, and blue violins representing NEWHORIZON, TNG50, and COSMOS, respectively. To aid visualization, black box plots are overlaid on each violin, indicating the inter-quartile range, whiskers representing the extrema, and a white dot indicating the median value of the

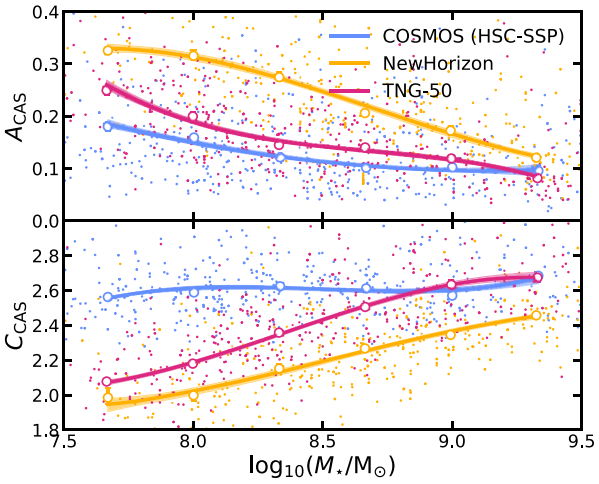

Figure 9. Plots showing the trend in the median asymmetry and concentration as a function of stellar mass for COSMOS (blue) and the redshift and mass-matched samples from NewHorizon (yellow) and TNG50 (red). Open circles with error bars show the median and error on the median for individual redshift bins, with filled regions indicating the 1σ uncertainty. Coloured points show a randomly selected sub-sample with the same colour scheme. See Fig. B6 for a plot showing the rest-frame quantities at fixed angular scale.

distribution. Additionally, a similar plot is presented in Fig. B7, showing rest-frame quantities at a fixed angular scale, alongside COSMOS dwarfs limited to $z < 0.1$.

Notably, both the median values and the overall distributions of NewHorizon, TNG50, and COSMOS dwarfs show significant differences across almost every parameter. These differences suggest substantial variations in the structure of galaxies produced by each simulation in comparison to each other and the observed galaxies. Although there are some differences in the values recovered for some parameters when viewed in the rest-frame (Fig. B7), large differences between the distributions of simulated and observed dwarfs persist across most parameters. As noted previously in this section, TNG50 dwarfs show significant shifts, particularly for the Gini, M_{20}, and CAS parameters, which are not PSF corrected, but these shifts typically do not result in improved agreement with the observations.

Our results in the low-mass regime show some areas of correspondence with the work of Eisert et al. (2024), who compare intermediate and high-mass TNG50 and TNG100 galaxies to counterparts in HSC observations. We also identify simulated galaxies with apparently unrealistic sizes, Sérsic profiles and structural parameters, although the ways in which our sample of low-mass TNG50 galaxies diverges from the observed sample differs compared with the higher mass sample of Eisert et al. (2024). Notably, we find that TNG50 dwarfs are too small compared to their observed counterparts, whereas Eisert et al. (2024) find that more massive galaxies tend to be too large, although anomalous examples in both our low-mass sample and Eisert et al. (2024)'s high mass sample tend to have Sérsic indices that are too high. Overall, Eisert et al. (2024) find better agreement between observed and simulated TNG50 galaxies. Similarly, other works such as Guzmán-Ortega et al. (2023) show a strong correspondence between the CAS and Gini-M_{20} structural parameters of high-mass TNG50 and KiDS (de Jong et al. 2013) observed galaxies as a function of mass. We also observe a similarly strong correspondence at the high end of our mass range, where our study has some overlap with Guzmán-Ortega et al. (2023). Overall, there appears to be significantly better correspondence in the structural properties of TNG50 and NewHorizon galaxies with their observed counterparts in the higher-mass regime. This correspondence begins to break down as upon entering the lower mass regime ($M_\star < 10^{9.5}$ $M_\odot$), which represents a region where the morphological behaviour and formation mechanisms of galaxies begin to transition (e.g. Lazar et al. 2024a).

5 DISCUSSION

We have shown that, despite covering a similar range to the observed galaxies in each of the morphological parameters measured in this paper, both NewHorizon and TNG50 produce dwarf galaxies whose overall distribution of structural properties are at significant odds with those measured in the COSMOS dwarfs. While TNG50 produces more centrally concentrated dwarfs with very small sizes, NewHorizon produces dwarfs at the opposite extreme, having very large sizes and small Sérsic indices.

In this section, we discuss possible limitations of our approach and possible physical drivers of the observed differences between the observed and simulated dwarfs.

5.1 Limitations

We begin by discussing some of the limits of our comparison between the simulations and between each simulation and the observed COSMOS galaxies.

5.1.1 Galaxy integrated fluxes and surface brightnesses

As seen in Fig. 4, NewHorizon galaxies have significantly fainter effective surface brightnesses on average than their observed counterparts. TNG50 galaxies exhibit similar effective surface brightnesses to COSMOS, but given the fact that they are significantly more compact when compared with their observed counterparts, this implies they still have fainter integrated brightnesses. While we do not rule out a genuine difference in the chemical or star-forming properties of galaxies driving this, there are several choices made when producing our synthetic images that also impact the brightness of the object.

(i) The choice of the IMF used to produce synthetic images influences both the mass-to-light ratio and the stellar mass loss prescription and can therefore significantly influence the integrated magnitudes of galaxies. Watkins et al. (2025) show that adopting a Salpeter (1955) IMF produces a decrease in the integrated magnitudes of the bluest galaxies of approximately 1 mag with a negligible effect on red galaxies. Given that both Chabrier (2003) and Salpeter (1955) are calibrated to the Milky Way, the adopted IMF may not be entirely appropriate or may vary significantly between galaxies and environments in lower mass regimes (e.g. Weidner & Kroupa 2005; Geha et al. 2013). In the case of this study we have tended to make assumptions that result in fainter magnitudes.

(ii) Similarly, our specific choice of SED template, in this case those of Bruzual & Charlot (2003), represent another source of uncertainty when comparing simulated galaxies to their observed counterparts because different SED models assume different stellar spectral libraries or stellar evolution tracks.

Given the differences observed in their structural properties, it is likely that there are genuine and significant physical differences between the chemical and star-forming properties of the simulated and observed dwarf galaxies. In the second part of this series of papers, we will revisit the subject of galaxy star formation histories and chemical enrichment.

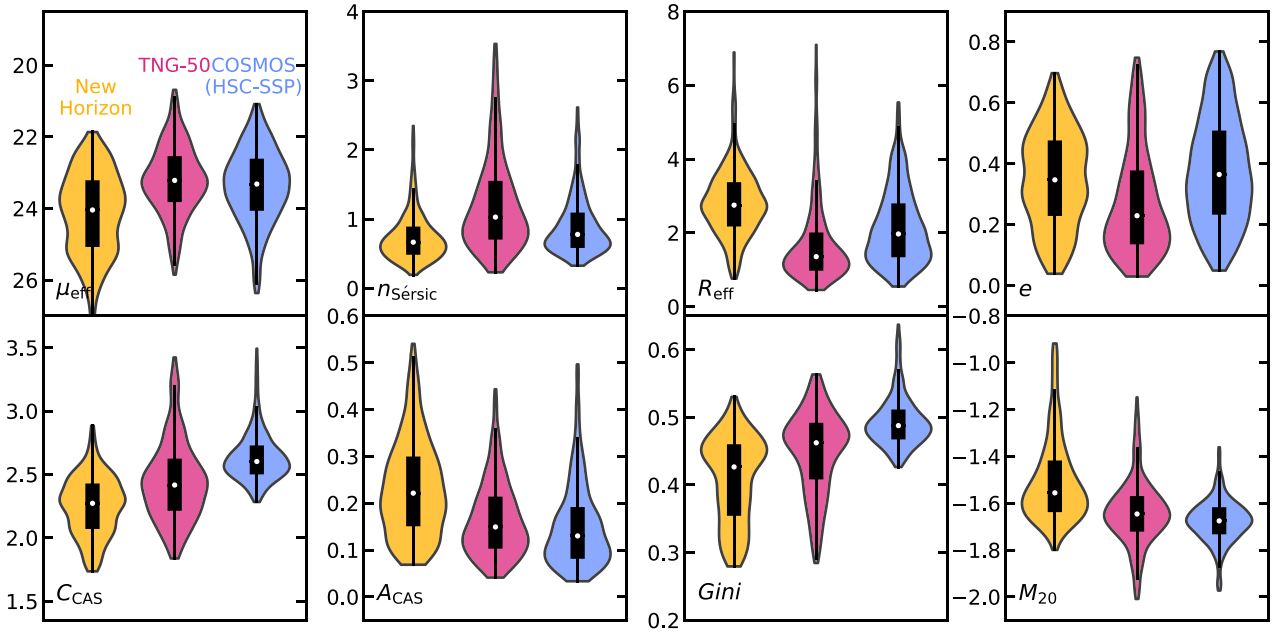

Figure 10. Violin plots summarizing the distribution of values discussed throughout Section 4 for NEWHORIZON (yellow), TNG50 (red), and COSMOS (blue). Black box plots overlaid over each violin indicate the inter-quartile range with whiskers representing the extrema, and a white dot indicating the median value of the distribution. See Fig. B7 for a plot showing the rest-frame quantities at fixed angular scale.

5.1.2 Biases due to the PSF

Due to the compact physical sizes of the galaxies under investigation in this paper, which depending on their redshift, can be close to the FWHM of the HSC PSF, it is important to consider the effect of seeing on our ability to spatially resolve them. In particular, this may lead to biases or significant uncertainties in recovered structural properties. As mentioned previously, we investigated the magnitude of this bias by examining the properties that we recover in the rest-frame at a fixed angular scale.

Despite the potential for bias in these measurements due to the smaller sizes of TNG50 galaxies, the fixed angular scale rest-frame quantities clearly demonstrate real and highly significant differences in the structural properties of galaxies between the two simulations. These plots are made available in Appendix B and, while showing similar qualitative trends, TNG50 galaxies in particular exhibit notably higher concentrations, Sérsic indices and notably smaller effective radii, indicating that many of them are compact enough that they become poorly resolved compared with the PSF at higher redshifts. For reference the equivalent physical scale of the average FWHM of the HSC i-band PSF at $z = 0.1$ and $z = 0.2$ is approximately 1.3 and 2.3 kpc, respectively, while the median effective radius of a TNG50 dwarf is only slightly larger than 1 kpc.

With upcoming large-area surveys such as the Legacy Survey of Space and Time (LSST; Ivezić et al. 2019; Watkins et al. 2024) and higher angular resolution space-based instruments like *Euclid* (Euclid Collaboration 2024), significantly larger resolved and mass complete samples of dwarf galaxies will soon be available. This will enable considerably more robust comparison with simulated low-mass galaxies.

5.2 Physical drivers

Finally, we discuss some possible physical drivers of the observed differences. A detailed investigation will be undertaken in the second part of this series of papers (Martin et al., in preparation). A more comprehensive discussion of environment and baryonic feedback specific to the NEWHORIZON simulation can be found in Watkins et al. (2025).

5.2.1 Environment

Environmental differences likely contribute to the observed discrepancies. For instance, Mercado et al. (2025) report a similar mismatch in the mass-size relation for low-mass galaxies in the FIREBOX simulation, attributing much of the size evolution to environmental effects, particularly their impact on dark matter halo masses. Previous studies using NEWHORIZON (Jackson et al. 2021) and its lower-resolution parent simulation Horizon-AGN (Martin et al. 2019) also indicate that the diffuse nature of low-mass galaxies arises from a combination of stellar feedback and environmental influences. Observationally, studies such as Privon et al. (2017) and Lazar et al. (2024b) suggest that the relationship between environment, star formation, and morphology differs significantly between low- and high-mass galaxies.

Environment is expected to influence galaxy properties through both large-scale tidal torques, which shape angular momentum buildup, and smaller-scale interactions with other galaxies, which become more frequent in denser environments. While these processes likely contribute to dwarf galaxy structure, COSMOS, TNG50, and NEWHORIZON primarily probe 'average' (field and group) environments, making it unlikely that environment alone fully explains the observed discrepancies.

As we will show in the second paper of this series (Martin et al., in preparation), restricting the TNG50 sample – which spans a larger box size and therefore covers somewhat more diverse environments – to the same local density range as NEWHORIZON reveals some evolution in the morphological and star-forming properties of the galaxies. However, these changes are not sufficient to bring TNG50

into clear agreement with either NEWHORIZON or COSMOS, suggesting that other factors play a more dominant role (e.g. Romeo, Agertz & Renaud 2020).

One such factor could be gas resolution, which has been shown to influence galaxy sizes at higher masses. Higher resolution allows for more accurate modelling of angular momentum loss on small scales, leading to more compact galaxies (Chabanier et al. 2020). However, the opposite effect is observed in our case, possibly due to the lower mass regime or the greater importance of other physical processes governing galaxy evolution.

NEWHORIZON dwarfs appear significantly more morphologically disturbed than their counterparts in COSMOS, exhibiting notably lower concentrations and higher asymmetries. TNG50 dwarfs display considerably higher concentrations to those in COSMOS and NEWHORIZON once the effect of PSF smearing has been taken into account. This could be explained by differences in environment leading to different frequencies of interactions, but could also be the result of physical differences in the galaxies themselves, such as differing halo or stellar mass profiles, which might influence how firmly material is retained within the potential well of the galaxy halo (e.g. Martin et al. 2019; Jackson et al. 2021; Martin et al. 2024). Given the considerably larger sizes and less concentrated stellar mass profiles of the NEWHORIZON dwarfs, it is plausible that this could lead to more pronounced irregularities and stronger tidal features, as material can be more easily moved or liberated from the galaxy during interactions. As discussed later, another likely driver of higher asymmetries may lie in the differences in implemented sub-grid physics, which could lead to burstier star formation and more clumpy stellar distributions.

5.2.2 *ISM, star formation, and feedback*

Beyond differences in numerical resolution, NEWHORIZON and TNG50 adopt significantly different treatments of the ISM and star formation physics, leading to contrasting galaxy properties. TNG50 assumes an ISM model that relies on an overpressurized medium driven by internal feedback effects with the Springel & Hernquist (2003) model. Their star formation follows a Schmidt law with a constant SF efficiency ($\simeq 6.5$ per cent) for gas densities larger than $n > 0.13\,\mathrm{cm}^{-3}$, and is combined to a hydrodynamical decoupling of the SN feedback from the star-forming gas densities so that the energy is released in diffuse gas.

In contrast, NEWHORIZON does not assume an effective equation of state for the ISM, allowing star formation to occur in denser gas ($n > 10\,\mathrm{cm}^{-3}$) with a variable efficiency that ranges from nearly zero to several tens of per cent (Dubois et al. 2021). Here, SN feedback is coupled to the surrounding gas in a local fashion, resulting in more effective feedback on cold star-forming gas.

Those different approaches to the ISM gas physics and stellar feedback have important consequences on the structure of the ISM: NEWHORIZON produces a multiphase ISM in which dense, star-forming clumps are efficiently disrupted by SNe, leading to a more dynamic, turbulent medium. In contrast, TNG50 generates a smoother ISM, where star-forming gas is more long-lived and less susceptible to episodic feedback-driven disruption.

The more bursty and variable star formation in NEWHORIZON likely drives stronger episodic outflows that preferentially eject central, low-angular-momentum gas (e.g. Azartash-Namin et al. 2024). The loss of this gas is likely to suppress star formation in the central regions and result in a more diffuse and extended stellar distribution. Such feedback-regulated gas redistribution also helps explain the more irregular morphologies and clumpy substructures seen in NEWHORIZON galaxies.

In contrast, the ISM model in TNG50 promotes smoother, more continuous feedback that lacks the violent, episodic ejections required to effectively disrupt central regions. Consequently, low-angular-momentum gas accumulates in the galaxy centres, fuelling central star formation and producing more compact, concentrated structures (e.g. De Almeida et al. 2024; Celiz et al. 2025). While strong stellar feedback can quench star formation, it has also been shown to lead to overly compact galaxies. Additionally, the inclusion of MHD in TNG50 may further enhance central compactness by regulating gas turbulence and outflows (e.g. Martin-Alvarez et al. 2023).

Although the models used in NEWHORIZON and TNG50 produce galaxy structural properties that differ significantly from observations, some integrated properties, such as the galaxy stellar mass function (e.g. Dubois et al. 2021), have been shown to be consistent. Wright et al. (2024) have further demonstrated that simulations with different feedback implementations can yield similar integrated properties through distinct mechanisms. This highlights that reproducing integrated properties alone does not ensure the underlying feedback processes are physically accurate. More detailed, resolved 2D properties – such as those presented in this paper and by Watkins et al. (2025) – therefore provide a more stringent test of the simulations' accuracy and greater power to constrain underlying physical mechanisms.

6 SUMMARY

In this paper, we have investigated the structural properties of dwarf galaxies by analysing data from both the NEWHORIZON and TNG50 simulations, in addition to ultra-deep HSC–SSP observations of the COSMOS field. Our approach involves first producing synthetic images of simulated galaxies matching the mass and redshift distribution of observed dwarfs in the COSMOS field and then injecting these synthetic images into realistic backgrounds taken from the HSC–SSP data. This methodology enables direct and robust comparison of the structural properties of observed and simulated dwarf galaxies. We use the Sérsic profile fitting and non-parametric, Gini, M_{20}, asymmetry and concentration parameters to compare the distribution of observed and simulated galaxies. We summarize our findings below.

(i) *Simulated and observed dwarf galaxies show distinct structural differences*: NEWHORIZON and TNG50 produce galaxies with structural properties that diverge significantly, lying to opposite extremes of the observed COSMOS dwarfs. NEWHORIZON produces diffuse, extended galaxies with shallow Sérsic indices, while TNG50 yields highly compact, concentrated galaxies with steep Sérsic indices. Observed COSMOS dwarfs span an intermediate range of structural properties, suggesting that neither simulation fully captures the structural diversity of real dwarf galaxies.

(ii) *Non-parametric measures also indicate discrepancies*: Gini, M_{20}, and CAS measurements show that NEWHORIZON galaxies have more uneven, and clumpy light distributions which are more likely to be linked to more bursty star formation processes than environmental processes. TNG50 galaxies exhibit smoother but excessively concentrated light profiles (in the rest-frame) compared with observed COSMOS dwarfs.

(iii) *Structural trends with stellar mass reveal differences between simulations and observations:* In NEWHORIZON and TNG50, galaxy

sizes, concentrations, and asymmetries evolve with mass but follow significantly different trends. NEWHORIZON dwarfs remain too diffuse even towards the high end of the mass range ($M_\star = 10^{9.5}$ $M_\odot$), although their Sérsic indices remain consistently shallow in agreement with the COSMOS dwarfs. TNG50 dwarfs exhibit a strong trend toward higher concentrations and steeper Sérsic indices towards higher masses, which aligns poorly with observations. COSMOS dwarfs, in contrast, show a much less pronounced trend with stellar mass, with relatively stable structural and morphological properties across the mass range.

(iv) *At the high end of the mass range* NEWHORIZON *and TNG50 galaxies begin to better reproduce the general structural trends seen in COSMOS dwarfs*: Most Sérsic and structural properties of NEWHORIZON and TNG50 dwarfs fall into better agreement with COSMOS at higher masses. TNG50 galaxies, however, remain somewhat too compact and retain Sérsic indices significantly steeper than observed COSMOS or NEWHORIZON dwarfs. These findings are consistent with previous studies (e.g. Eisert et al. 2024), which have shown that TNG50 tends to achieve better agreement with observations in the higher mass regime, contrasting with the larger discrepancies observed at lower masses in this study.

(v) *Observational effects alone cannot explain simulation discrepancies:* While observational factors like seeing introduce biases, repeating our analysis in the rest-frame confirms intrinsic differences in simulated galaxy properties. In particular, many TNG50 galaxies are smeared out by the PSF due to their extremely compact intrinsic sizes, meaning they are even more compact than they appear in the observed-frame, while NEWHORIZON galaxies remain diffuse and asymmetric even in the absence of observational biases.

(vi) *Differences in simulation physics offer valuable insights into the drivers of galaxy evolution:* The distinct structural properties of NEWHORIZON and TNG50 reflect the underlying differences in their physical models, including their ISM physics, star formation prescriptions and feedback implementations. The pronounced variations in the structural properties of simulated dwarfs in this mass regime underscore the heightened sensitivity of dwarf galaxy evolution to these processes. New observatories, such as the Vera C. Rubin Observatory and *Euclid*, will facilitate much more robust comparisons with simulations of low-mass galaxies, enabling stronger constraints on the physical mechanisms that shape their evolution.

We have shown that low-mass galaxy populations produced by two state-of-the-art simulations have structural properties that are significantly at odds with both each other and observed galaxies in the same mass regime. This disparity may offer insight into the underlying physics shaping galaxy evolution, particularly the physics of star formation and stellar feedback, to which galaxy properties are especially sensitive in this regime. In the second part of this series of papers (Martin et al., in preparation), we will investigate the correlation between galaxy morphology and galaxy star formation history in this regime.

ACKNOWLEDGEMENTS

GM is thankful to Taysun Kimm, Nina Hatch and Emmanuele Contini for fruitful discussions.

GM acknowledges support from the UK STFC under grant ST/X000982/1.

AEW acknowledges support from the STFC under grant ST/X001318/1.

SKY acknowledges support from the National Research Foundation of Korea (2020R1A2C3003769; RS-2022-NR070872).

DK acknowledges support from the National Research Foundation of Korea (NRF) grant funded by the Korean government(MSIT) (no. NRF-2022R1C1C2004506).

SK, IL, and AEW acknowledge support from the STFC (grant numbers ST/Y001257/1 and ST/X001318/1). SK also acknowledges a Senior Research Fellowship from Worcester College Oxford.

Part of this work was carried out by GM during a Balzan Visiting Fellowship held at the Department of Physics, University of Oxford.

The NEWHORIZON simulation was undertaken with HPC resources of CINES under the allocations c2016047637, A0020407637 and A0070402192 by Genci, KSC-2017-G2-0003 by KISTI, and as a 'Grand Challenge' project granted by GENCI on the AMD Rome extension of the Joliot Curie supercomputer at TGCC. A large data transfer was supported by KREONET which is managed and operated by KISTI.

The ILLUSTRISTNG simulations were undertaken with compute time awarded by the Gauss Centre for Supercomputing (GCS) under GCS Large-Scale Projects GCS-ILLU and GCS-DWAR on the GCS share of the supercomputer Hazel Hen at the High Performance Computing Center Stuttgart (HLRS), as well as on the machines of the Max Planck Computing and Data Facility (MPCDF) in Garching, Germany.

This work has made use of the Infinity cluster on which the NEWHORIZON simulation was post-processed, hosted by the Institut d'Astrophysique de Paris. We warmly thank S. Rouberol for running it smoothly.

This research is part of the Spin(e) ANR-13-BS05-0005 http://cosmicorigin.org), Segal ANR-19-CE31-0017 (http://secular-evolution.org).

This research made use of PHOTUTILS, an ASTROPY package for detection and photometry of astronomical sources (Bradley et al. 2022).

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. We thank the Rubin Observatory for making their code available as free software at http://pipelines.lsst.io/.

This paper is based in part 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 cultural, historical and natural significance in Hawaii.

DATA AVAILABILITY

The simulation data analysed in this paper were provided by the Horizon collaboration and the ILLUSTRIS project. NEWHORIZON data will be shared on request to the corresponding author, with the permission of the Horizon collaboration or may be requested from https://new.horizon-simulation.org/data.html. ILLUSTRIS TNG50 data are publicly available and can be accessed at https://tng-project.org/.

The third public release of HSC–SSP data is publicly available and can be accessed at https://hsc-release.mtk.nao.ac.jp/.

REFERENCES

Aihara H. et al., 2018, PASJ, 70, S4
Aihara H. et al., 2022, PASJ, 74, 247
Arnouts S., Cristiani S., Moscardini L., Matarrese S., Lucchin F., Fontana A., Giallongo E., 1999, MNRAS, 310, 540
Arnouts S., Ilbert O., 2011, Astrophysics Source Code Library, record ascl: 1108.009
Atek H. et al., 2014, ApJ, 789, 96
Aubert D., Pichon C., Colombi S., 2004, MNRAS, 352, 376
Azartash-Namin B. et al., 2024, ApJ, 970, 40
Baldry I. K., Glazebrook K., Driver S. P., 2008, MNRAS, 388, 945
Bertin E., Arnouts S., 1996, A&AS, 117, 393
Blanton M. R., Lupton R. H., Schlegel D. J., Strauss M. A., Brinkmann J., Fukugita M., Loveday J., 2005, ApJ, 631, 208
Bluck A. F. L., Mendel J. T., Ellison S. L., Moreno J., Simard L., Patton D. R., Starkenburg E., 2014, MNRAS, 441, 599
Borlaff A. S. et al., 2021, A&A, 657, 21
Bosch J. et al., 2018, PASJ, 70, S5
Bothun G. D., Impey C. D., Malin D. F., Mould J. R., 1987, AJ, 94, 23
Bradley L. et al., 2022, astropy/photutils: 1.5.0, Zenodo, https://doi.org/10.5281/zenodo.6825092
Bruzual G., Charlot S., 2003, MNRAS, 344, 1000
Bullock J. S., Boylan-Kolchin M., 2017, ARA&A, 55, 343
Bundy K., Ellis R. S., Conselice C. J., 2005, ApJ, 625, 621
Casey C. M. et al., 2022, ApJ, 954, 32
Celiz B. M., Navarro J. F., Abadi M. G., Springel V., 2025. Astronomy & Astrophysics, 699, A12
Chabanier S. et al., 2020, A&A, 643, 12
Chabrier G., 2003, PASP, 115, 763
Conselice C. J., 2006, MNRAS, 373, 1389
Conselice C. J., Bershady M. A., Dickinson M., Papovich C., 2003, AJ, 126, 1183
Conselice C. J., Yang C., Bluck A. F. L., 2009, MNRAS, 394, 1956
Dalcanton J. J., Spergel D. N., Gunn J. E., Schmidt M., Schneider D. P., 1997, AJ, 114, 635
Dalgarno A., McCray R. A., 1972, ARA&A, 10, 375
Davis F. et al., 2022, MNRAS, 511, 4109
De Almeida A. P., Mamon G. A., Dekel A., Lima Neto G. B., 2024, A&A, 687, A131
de Jong J. T. A., Verdoes Kleijn G. A., Kuijken K. H., Valentijn E. A., 2013, Exp. Astron., 35, 25
Dekel A., Silk J., 1986, ApJ, 303, 39
Di Cintio A., Brook C. B., Dutton A. A., Macciò A. V., Obreja A., Dekel A., 2017, MNRAS, 466, L1
Disney M. J., 1976, Nature, 263, 573
Draine B. T. et al., 2007, ApJ, 663, 866
Dressler A. et al., 1997, ApJ, 490, 577
Dressler A., 1980, ApJ, 236, 351
Driver S. P., Liske J., Cross N. J. G., De Propris R., Allen P. D., 2005, MNRAS, 360, 81
Dubois Y. et al., 2014, MNRAS, 444, 1453
Dubois Y., Volonteri M., Silk J., Devriendt J., Slyz A., Teyssier R., 2015, MNRAS, 452, 1502
Dubois Y. et al., 2021, A&A, 651, A109
Eisert L., Bottrell C., Pillepich A., Shimakawa R., Rodriguez-Gomez V., Nelson D., Angeloudi E., Huertas-Company M., 2024, MNRAS, 528, 7411
Euclid Collaboration, 2024, A&A, 697, 94
Fakhouri O., Ma C.-P., Boylan-Kolchin M., 2010, MNRAS, 406, 2267
Faucher-Giguère C.-A., Lidz A., Zaldarriaga M., Hernquist L., 2009, ApJ, 703, 1416
Feldmann R. et al., 2023, MNRAS, 522, 3831
Gardner J. P. et al., 2006, Space Sci. Rev., 123, 485
Geha M. et al., 2013, ApJ, 771, 29
Gentry E. S., Krumholz M. R., Madau P., Lupi A., 2019, MNRAS, 483, 3647
Glasser G. J., 1962, J. Am. Stat. Assoc., 57, 648
Greco J. P., 2018, PhD thesis, Princeton University, New Jersey
Greene J. E. et al., 2022, ApJ, 933, 150
Guedes J., Callegari S., Madau P., Mayer L., 2011, ApJ, 742, 76
Guo Y. et al., 2016, ApJ, 833, 37
Guzmán-Ortega A., Rodriguez-Gomez V., Snyder G. F., Chamberlain K., Hernquist L., 2023, MNRAS, 519, 4920
Haardt F., Madau P., 1996, ApJ, 461, 20
Hagen L. M. Z. et al., 2016, ApJ, 826, 210
Hogg D. W. et al., 2002, AJ, 124, 646
Holwerda B. W. et al., 2019, AJ, 158, 103
Hopkins P. F., Kereš D., Oñorbe J., Faucher-Giguère C.-A., Quataert E., Murray N., Bullock J. S., 2014, MNRAS, 445, 581
Ibata R. A. et al., 2013, Nature, 493, 62
Ilbert O. et al., 2006, A&A, 457, 841
Ivezić Ž. et al., 2019, ApJ, 873, 111
Jackson R. A. et al., 2021, MNRAS, 502, 4262
Jang J. K. et al., 2023, ApJ, 950, 4
Kartaltepe J.et al., 2023, ApJ, 946, 17
Katz N., Hernquist L., Weinberg D. H., 1992, ApJ, 399, L109
Kaviraj S. et al., 2017, MNRAS, 467, 4739
Kaviraj S., 2014, MNRAS, 440, 2944
Kaviraj S., Lazar I., Watkins A. E., Laigle C., Martin G., Jackson R. A., 2025, MNRAS, 538, 153
Kaviraj S., Martin G., Silk J., 2019, MNRAS, 489, L12
Kawanomoto S. et al., 2018, PASJ, 70, 66
Kim C.-G., Ostriker E. C., Raileanu R., 2017, ApJ, 834, 25
Kim J. et al., 2022, ApJ, 951, 24
Kimm T., Cen R., 2014, ApJ, 788, 121
Kimm T., Cen R., Devriendt J., Dubois Y., Slyz A., 2015, MNRAS, 451, 2900
Kimm T., Katz H., Haehnelt M., Rosdahl J., Devriendt J., Slyz A., 2017, MNRAS, 466, 4826
Kniazev A. Y., Grebel E. K., Pustilnik S. A., Pramskij A. G., Kniazeva T. F., Prada F., Harbeck D., 2004, AJ, 127, 704
Komatsu E. et al., 2011, ApJS, 192, 18
Koudmani S., Henden N. A., Sijacki D., 2021, MNRAS, 503, 3568
Laureijs R. et al., 2011, preprint (arXiv:1110.3193)
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
Lotz J. M., Primack J., Madau P., 2004, AJ, 128, 163
Martin G. et al., 2019, MNRAS, 485, 796
Martin G. et al., 2021, MNRAS, 500, 4937
Martin G. et al., 2022, MNRAS, 513, 1459
Martin G., Kaviraj S., Devriendt J. E. G., Dubois Y., Pichon C., 2018, MNRAS, 480, 2266
Martin G., Pearce F. R., Hatch N. A., Contreras-Santos A., Knebe A., Cui W., 2024, MNRAS, 535, 2375
Martin-Alvarez S., Sijacki D., Haehnelt M. G., Farcy M., Dubois Y., Belokurov V., Rosdahl J., Lopez-Rodriguez E., 2023, MNRAS, 525, 3806
Martínez-Delgado D. et al., 2016, AJ, 151, 96
Mercado F. J. et al., 2025, ApJ, 983, 14
Miyoshi T., Kusano K., 2005, J. Comput. Phys., 208, 315
Montes M., Brough S., Owers M. S., Santucci G., 2021, ApJ, 910, 45
Nelson D. et al., 2019, MNRAS, 490, 3234

Olivier S. S., Seppala L., Gilmore K., 2008, in Atad-Ettedgui E., Lemke D., Proc. SPIE Conf. Ser. Vol. 983, Advanced Optical and Mechanical Technologies in Telescopes and Instrumentation.. SPIE, Bellingham, p. 70182G
Pakmor R., Bauer A., Springel V., 2011, MNRAS, 418, 1392
Pakmor R., Springel V., Bauer A., Mocz P., Munoz D. J., Ohlmann S. T., Schaal K., Zhu C., 2016, MNRAS, 455, 1134
Peirani S. et al., 2024, A&A, 686, A233
Pillepich A. et al., 2018, MNRAS, 475, 648
Pillepich A. et al., 2019, MNRAS, 490, 3196
Planck Collaboration XIII, 2016, A&A, 594, A13
Pozzetti L. et al., 2007, A&A, 474, 443
Privon G. C. et al., 2017, ApJ, 846, 74
Prole D. J., van der Burg R. F. J., Hilker M., Spitler L. R., 2021, MNRAS, 500, 2049
Reines A. E., Greene J. E., Geha M., 2013, ApJ, 775, 116
Rodriguez-Gomez V. et al., 2019, MNRAS, 483, 4140
Román J., Beasley M. A., Ruiz-Lara T., Valls-Gabaud D., 2019, MNRAS, 486, 823
Román J., Trujillo I., Montes M., 2020, A&A, 644, A42
Romeo A. B., Agertz O., Renaud F., 2020, MNRAS, 499, 5656
Rosdahl J., Blaizot J., 2012, MNRAS, 423, 344
Salpeter E. E., 1955, ApJ, 121, 161
Sazonova E. et al., 2020, ApJ, 899, 85
Sazonova E. et al., 2024, Open J. Astrophys., 7
Scoville N. et al., 2007, ApJS, 172, 1
Searle L., Sargent W. L. W., Bagnuolo W. G., 1973, ApJ, 179, 427
Seo M., Ann H. B., 2022, MNRAS, 514, 5853
Sérsic J. L., 1968, Atlas de Galaxias Australes, Observatorio Astonomico, Universidad Nacional de Cordoba, Argentina
Skibba R. A. et al., 2009, MNRAS, 399, 966
Smith B., Sigurdsson S., Abel T., 2008, MNRAS, 385, 1443
Springel V., 2010, MNRAS, 401, 791
Springel V., Hernquist L., 2003, MNRAS, 339, 289
Springel V., White S. D. M., Tormen G., Kauffmann G., 2001, MNRAS, 328, 726
Strateva I. et al., 2001, AJ, 122, 1861
Su A. H. et al., 2021, A&A, 647, A100
Sutherland R. S., Dopita M. A., 1993, ApJS, 88, 253
Szalay A. S., Connolly A. J., Szokoly G. P., 1999, AJ, 117, 68
Teyssier R., 2002, A&A, 385, 337
Thornton K., Gaudlitz M., Janka H. T., Steinmetz M., 1998, ApJ, 500, 95
Toro E., 1999, Riemann Solvers and Numerical Methods for Fluid Dynamics. Springer, Heidelberg
Trebitsch M., Blaizot J., Rosdahl J., Devriendt J., Slyz A., 2017, MNRAS, 470, 224
Trebitsch M., Volonteri M., Dubois Y., 2020, MNRAS, 494, 3453
Tremmel M., Karcher M., Governato F., Volonteri M., Quinn T. R., Pontzen A., Anderson L., Bellovary J., 2017, MNRAS, 470, 1121
Uzeirbegovic E. et al., 2024, MNRAS, 535, 3775
Uzeirbegovic E., Geach J. E., Kaviraj S., 2020, MNRAS, 498, 4021
Uzeirbegovic E., Martin G., Kaviraj S., 2022, MNRAS, 510, 3849
van Dokkum P. G., Abraham R., Merritt A., Zhang J., Geha M., Conroy C., 2015, ApJ, 798 L45
Voit G. M., Oppenheimer B. D., Bell E. F., Terrazas B., Donahue M., 2024, ApJ, 960, 28
Wang E., Lilly S. J., 2023, ApJ, 955, 55
Watkins A. E. et al., 2022, A&A, 660, A69
Watkins A. E., Salo H., Kaviraj S., Collins C. A., Knapen J. H., Venhola A., Román J., 2023, MNRAS, 521, 2012
Watkins A. E., Kaviraj S., Collins C. C., Knapen J. H., Kelvin L. S., Duc P.-A., Román J., Mihos J. C., 2024, MNRAS, 528, 4289
Watkins A., Martin G., Kaviraj S., Collins C., Dubois Y., Kraljic K., Pichon C., Yi S. K., 2025, MNRAS, 537, 3499
Weaver J. R. et al., 2022, ApJS, 258, 11
Weidner C., Kroupa P., 2005, in Jerjen H., Binggeli B., eds, IAU Colloq. 198: Near-fields cosmology with dwarf elliptical galaxies. Cambridge Univ. Press, Cambridge, p. 130
Weinberger R. et al., 2017, MNRAS, 465, 3291
Weingartner J. C., Draine B. T., 2001, ApJ, 548, 296
Wetzel A. R., Hopkins P. F., Kim J.-h., Faucher-Giguère C.-A., Kereš D., Quataert E., 2016, ApJ, 827, L23
Whitaker K. E. et al., 2015, ApJ, 811, L12
Wiersma R. P. C., Schaye J., Theuns T., Dalla Vecchia C., Tornatore L., 2009, MNRAS, 399, 574
Williams C. C. et al., 2021, PANORAMIC—A Pure Parallel Wide Area Legacy Imaging Survey at 1–5 Micron, JWST Proposal. Cycle 1, ID. #2514
Wright R. J., Somerville R. S., Lagos C. d. P., Schaller M., Davé R., Anglés-Alcázar D., Genel S., 2024, MNRAS, 532, 3417
York D. G. et al., 2000, AJ, 120, 1579

APPENDIX A: DEPTH OF HSC–SSP COSMOS FIELD

In this section, we present estimates of the *i*-band depth of the HSC–SSP `deepCoadd` data in the Cosmos field.

A 256 × 256 pixel bin third-order sky correction task is performed on every exposure before co-addition. This can lead to systematics in background subtraction (Watkins et al. 2024, and references therein).

objects smaller than this scale (which is the case for all of the objects we consider here) should not suffer from over-subtraction.

We split the central pointing of the Cosmos field into a 9 × 9 element grid, measuring the limiting surface brightness within each square by first masking all pixels with detections and then further splitting each square into a grid with elements of 256 × 256 pixels (the same scale on which the sky correction is performed) and calculating the median standard deviation across each grid element.

We then convert to a limiting surface brightness according to the following equation from (Román, Trujillo & Montes 2020):

$$\mu_{\rm lim}(n\sigma, \Omega \times \Omega) = -2.5 \log_{10}\left(\frac{n\sigma}{pix \times \Omega}\right) + Z_p\,, \tag{A1}$$

where σ is the measured standard deviation of the image counts, Ω is the box length in arcseconds, pix is the pixel size in arcseconds and $Z_{\rm p}$ is the zero-point (which is set to 27 mag for all `deepCoadd` images).(10x10, 3 sigma)

Fig. A1 illustrates the measured COSMOS depth within the central 1.5° field of view. The *i*-band surface brightness limit, $\mu_i^{\rm lim}(3\sigma, 10'' \times 10'')$, exceeds 31 mag arcsec^{-2}, with mean and median values of 31.35 and 31.41 mag arcsec^{-2}, respectively.

To evaluate object recovery under realistic conditions, we inject synthetic galaxies – sourced from the NewHorizon and TNG50 data sets – into HSC–SSP `deepCoadd` images. These galaxies are inserted into empty regions of the sky to minimize systematic effects from source confusion. Specifically, injection positions are chosen within the central COSMOS field such that they are at least 20 pixels away from any detected source. This separation is sufficiently large to avoid confusion with nearby objects, yet remains well below the 256 × 256 pixel bin size used for sky correction, ensuring that potential systematics from sky background estimation near other sources are captured. We note that galaxies projected close to extended or heavily clustered sources might still be affected; however, such cases are expected to increase scatter in our measurements rather than introduce systematic biases.

After injecting the synthetic galaxies, we rerun the detection and segmentation algorithm to produce an updated segmentation map that includes the injected objects. A detection is deemed successful if the centroid of the closest detected object lies within 10 pixels of the injected galaxy's centre (i.e. half the minimum injection separation).

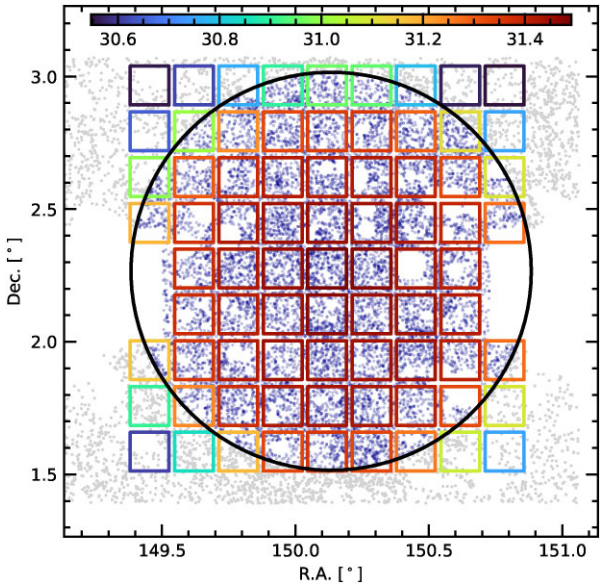

Figure A1. Scatter plot showing the coordinates of COSMOS2020 galaxies within the central 1.5° of the COSMOS field (bounded by a black circle). The sample used in this study is indicated in blue, while objects lying outside of the central 1.5° are shown in grey. Overplotted is a 9 × 9 grid, where the colour of the lines indicates the median depth within the region bounded by each grid element. Colours correspond to the colour bar at the top of the plot and indicate the depth as a limiting surface brightness $\mu_i^{\rm lim}(3\sigma$, 10 arcsec × 10 arcsec measured in mag arcsec^{-2}.

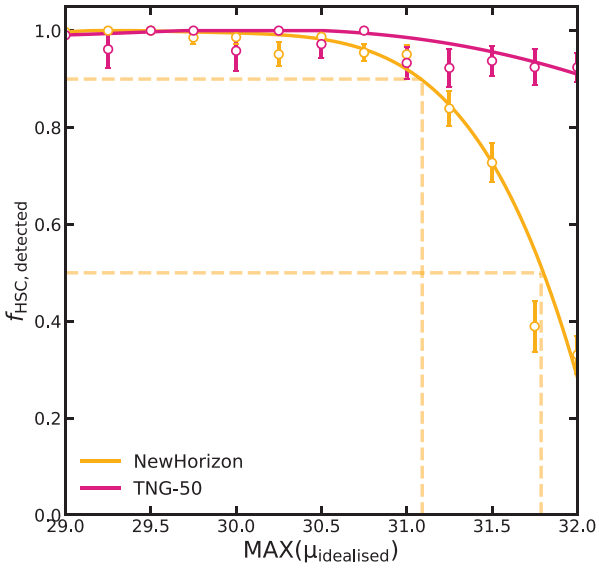

Figure A2. Detection efficiency of synthetic galaxy injections in HSC–SSP `deepCoadd` images as a function of the idealized limiting surface brightness, MAX($\mu_{\rm lim,idealised}$).

To estimate the equivalent limiting surface brightness, we also perform a separate set of detections under idealized conditions – using only Gaussian random noise – across a grid of limiting surface brightnesses between 29 and 32 mag arcsec^{-2}. We define MAX($\mu_{\rm lim,idealised}$) as the faintest limiting surface brightness at which a given object remains detectable after noise injection, and $f_{\rm HSC,detected}$ as the fraction of objects injected into the HSC `deepCoadd` imaging that are detected for a given bin of MAX($\mu_{\rm lim,idealised}$).

As shown in Fig. A2, we find that the NEWHORIZON sample is 90 per cent and 50 per cent complete up to limiting surface brightnesses, $\mu_i^{\rm lim}(3\sigma$, 10 arcsec × 10 arcsec, of 31.1 and 31.8 mag arcsec^{-2}, respectively. In the case of TNG50 galaxies, which are significantly more compact, the sample is over 90 per cent complete beyond 32 mag arcsec^{-2}. These values appear consistent with our previous limiting surface brightness estimate of $\sim$ 31.4 mag arcsec−2. Note that we would not expect exact correspondence, as the limiting surface brightness measured directly from the background noise and the limiting surface brightness at which an object is detected are not equivalent measurements.

APPENDIX B: REST-FRAME QUANTITIES

In this section, we present 'rest-frame' versions of the plots from Section 4. Sérsic, CAS and gini-M_{20} parameters are measured in the rest-frame with a fixed angular scale of 1 kpc arcsec^{-1}.

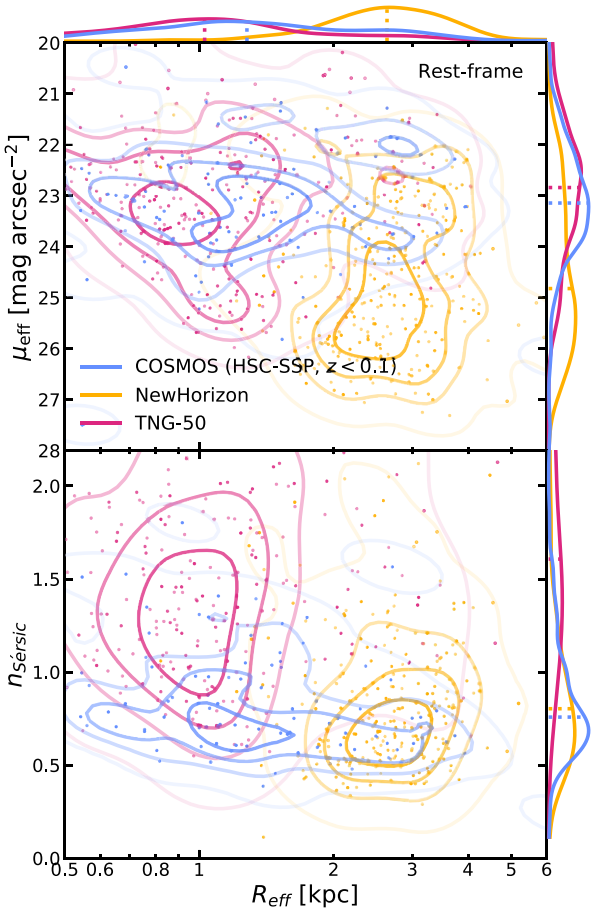

Figure B1. Contour plots showing the 2D distribution of the rest-frame measured surface brightness at the effective radius (top panel) and Sérsic index (bottom panel) as a function of effective radius for the redshift and mass-matched samples from NEWHORIZON (yellow) and TNG50 (red) with a $z < 0.1$ restricted sub-sample of COSMOS (blue). Coloured points show a randomly selected sub-sample with the same colour scheme. The sides of each panel show the marginal distribution of each variable with dashed lines indicating the median values.

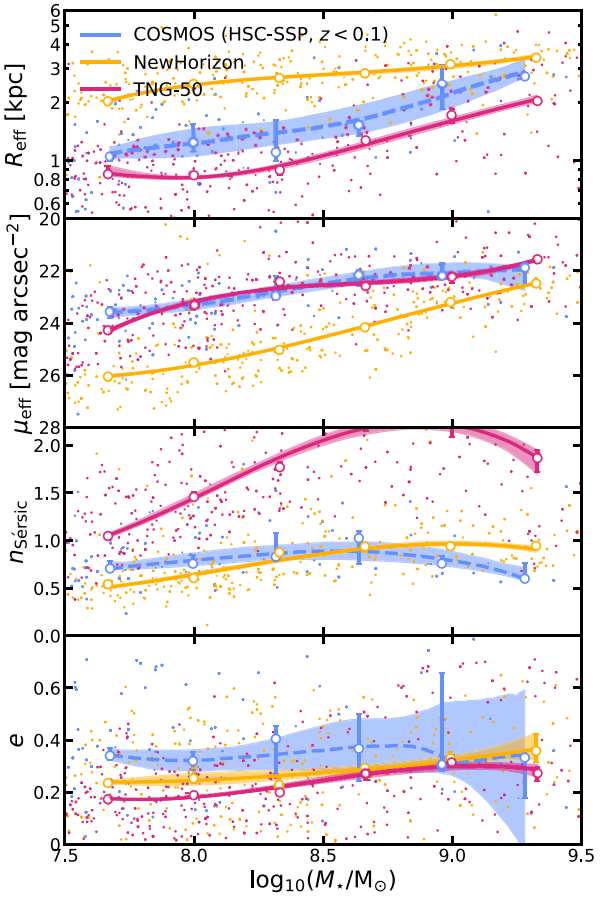

Figure B2. The evolution of the median rest-frame effective radius, surface brightness at the effective radius, Sérsic index and projected ellipticity as a function of stellar mass for a $z < 0.1$ restricted sub-sample of COSMOS (blue) and the redshift and mass-matched samples from NEWHORIZON (yellow) and TNG50 (red). Open circles with error bars show the median and error on the median for individual redshift bins, with filled regions indicating the 1σ uncertainty. Coloured points show a randomly selected sub-sample with the same colour scheme.

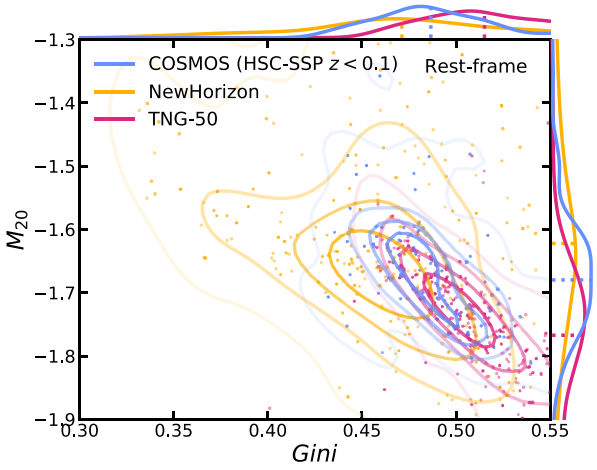

Figure B3. Contour plot showing the 2D distribution of the rest-frame measured Gini and M_{20} for the redshift and mass-matched samples from NEWHORIZON (yellow) and TNG50 (red) with a $z < 0.1$ restricted sub-sample of COSMOS (blue). Coloured points show a randomly selected sub-sample with the same colour scheme. The sides of the plot show the marginal distribution of each variable with dashed lines indicating the median values.

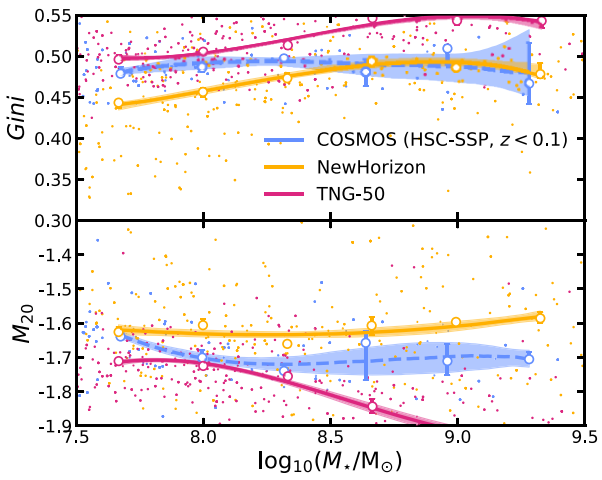

Figure B4. Plots showing the evolution of the median rest-frame measured gini and M_{20} as a function of stellar mass for a $z < 0.1$ restricted sub-sample of COSMOS and the redshift and mass-matched samples from NEWHORIZON (yellow) and TNG50 (red). Open circles with error bars show the median and error on the median for individual redshift bins, with filled regions indicating the 1σ uncertainty. Coloured points show a randomly selected sub-sample with the same colour scheme.

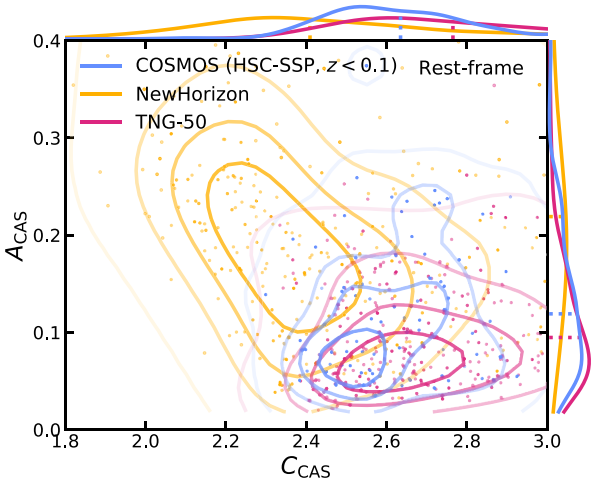

Figure B5. Contour plot showing the 2D distribution of the rest-frame measured concentration and asymmetry for the redshift and mass-matched samples from NEWHORIZON (yellow) and TNG50 (red) with a $z < 0.1$ restricted sub-sample of COSMOS (blue). Coloured points show a randomly selected sub-sample with the same colour scheme. The sides of the plot show the marginal distribution of each variable with dashed lines indicating the median values.

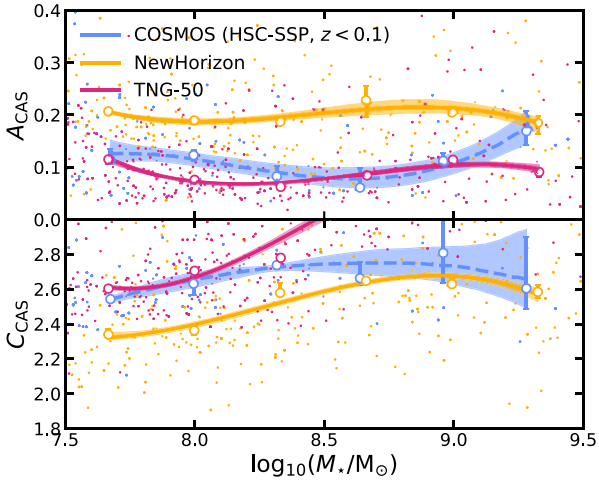

Figure B6. Plots showing the evolution of the median rest-frame measured asymmetry and concentration as a function of stellar mass for a $z < 0.1$ restricted sub-sample of COSMOS and the redshift and mass-matched samples from NEWHORIZON (yellow) and TNG50 (red). Open circles with error bars show the median and error on the median for individual redshift bins, with filled regions indicating the 1σ uncertainty. Coloured points show a randomly selected sub-sample with the same colour scheme.

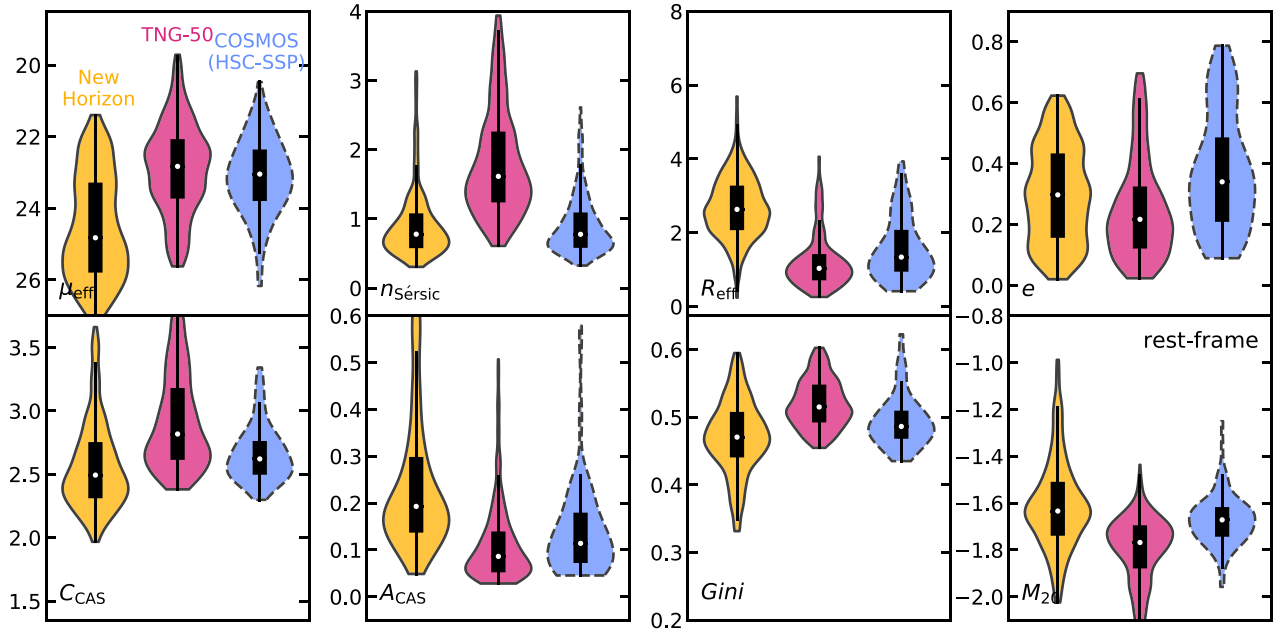

Figure B7. Violin plots summarizing the distribution of rest-frame values discussed throughout Section 4 for NEWHORIZON (yellow), TNG50 (red), and a $z < 0.1$ restricted sub-sample of COSMOS (blue). Black box plots overlaid over each violin indicate the inter-quartile range with whiskers representing the extrema, and a white dot indicating the median value of the distribution.

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