OBSERVED COSMIC EVOLUTION OF GALAXY STELLAR MASS AND DARK MATTER HALO MASS

A dissertation by

Muhammad Kumail Zaidi

Submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Physics: Astrophysics

May 2025

Chair of Committee and PhD advisor: Prof. Danilo Marchesini

Committee Members: Prof. Anna Sajina

Prof. David Martin

Prof. Hugh Gallagher

Prof. Kyoung-Soo Lee

ABSTRACT

Galaxies evolve within dark matter halos, and their co-evolution shapes the cosmic-web structure in the Universe, and vice versa. Understanding the connection between galaxies and their host dark matter halos, therefore, is paramount in uncovering the physics of galaxy formation and evolution. To that end, this dissertation undertakes laying down the observed cosmic evolution in the relationship between the fundamental quantities of galaxies and dark matter halos, the galaxy stellar mass to dark matter halo mass relationship (SHMR). For such an endeavor, large samples of cosmos are necessary to mitigate cosmic variance in the measurements that lead to the constraining of the SHMR. Additionally, to probe a significant fraction of the cosmic historical evolution of SHMR, deep catalogs of galaxies are needed, which probe the cosmos out to high redshifts. We started by constructing a photometric catalog−FENIKS UDS catalog, covering ~ 0.8 deg^2, assembling data from several surveys covering a wide wavelength range in aggregate from optical (0.38 µm) to NIR (4.5 µm). This catalog construction resulted in a publication; all the catalog products−photometric redshifts, stellar populations properties (stellar mass, star formation rate, etc.) were made public. The FENIKS UDS catalog was then combined with the UltraVISTA DR3 catalog (COSMOS field) of comparable depth and area and constructed in the same way as the FENIKS UDS catalog to probe the SHMR. With the combined effective area of ~ 1.61 deg^2 at hand, we measured the galaxy abundance and clustering out to a redshift of ~ 5 in 9 redshift bins and mass complete samples within each bin, and used the Halo Occupation Distribution (HOD) modeling to translate these measurements into SHMR. However, the range in stellar mass with increasing redshift shrinks due to the depleting complete-

ness of fainter/less massive galaxies, limiting the clustering and abundance measurements to high mass galaxies with increasing redshift. Naturally, this leads to much weaker HOD constraining power with increasing redshift. To mitigate this, we introduce a novel HOD fitting technique in which we leverage the strong constraining power of the low redshift data to help constrain the high redshift SHMR better by simply realizing that the HOD parameters should evolve smoothly with redshift. We call this the smooth-z modeling technique, in which we jointly fit data in all redshift bins at once while the redshift evolution in the HOD parameters is laced through the use of continuity (smoothing) priors. This greatly and unprecedentedly reduces the uncertainties in the SHMR at high redshift ($z \gtrsim 2$) compared to previous HOD-based analyses, establishing a concrete picture of the evolution in SHMR out to roughly the first billionth year of the cosmos ($z \sim 5$). Stellar mass remains tightly coupled to halo mass out to $z{\sim}5$, with the halo mass with peak integrated star-forming efficiency (SFE), M_h^{peak}, staying consistently around $10^{12.3}$ $\mathrm{M}_\odot$ as opposed to some of the previous work speculating an increase or a decrease in M_h^{peak} at $z \gtrsim 2$.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

FIGURE Page

CHAPTER 1

INTRODUCTION

1.1 Cosmology to Galaxies

The beautiful cosmic web structure we see in the Universe today (see Figure 1.1) was orchestrated through an intricate gravitational dance of the contents of the Universe, at least according to the currently standard paradigm of the cold dark matter with dark energy $-\Lambda$CDM. It all started with the quantum fluctuations in the otherwise smooth density field of the Universe after the Big Bang, the imprint of which is observable today in the cosmic microwave background (CMB) radiation pattern (see Figure 1.2) 13.8 billion years later. CMB was first discovered by Arno Penzias and Robert Wilson (Penzias & Wilson 1965) by accident in 1964 as a mysterious isotropic background noise while studying radio signals for satellite communications at Bell Telephone Laboratories in Holmdel, New Jersey. Interestingly, a small fraction of the static hiss (white noise) picked up by the old CRT television sets when tuned to an unused channel is also actually CMB.

The emission of the cosmic microwave background (CMB) 380,000 years after the Big Bang marks the first time the Universe had cooled enough, due to expansion, for free electrons and protons to combine into neutral hydrogen, an event termed the recombination. With fewer free electrons available to scatter photons, light was finally able to travel freely (photon decoupling), creating the nearly isotropic background radiation we observe today as the CMB. However, dark matter, which does not interact with photons, could begin forming clumps even prior to photon decoupling, creating gravitational potential wells onto which baryonic matter could ultimately funnel in leading

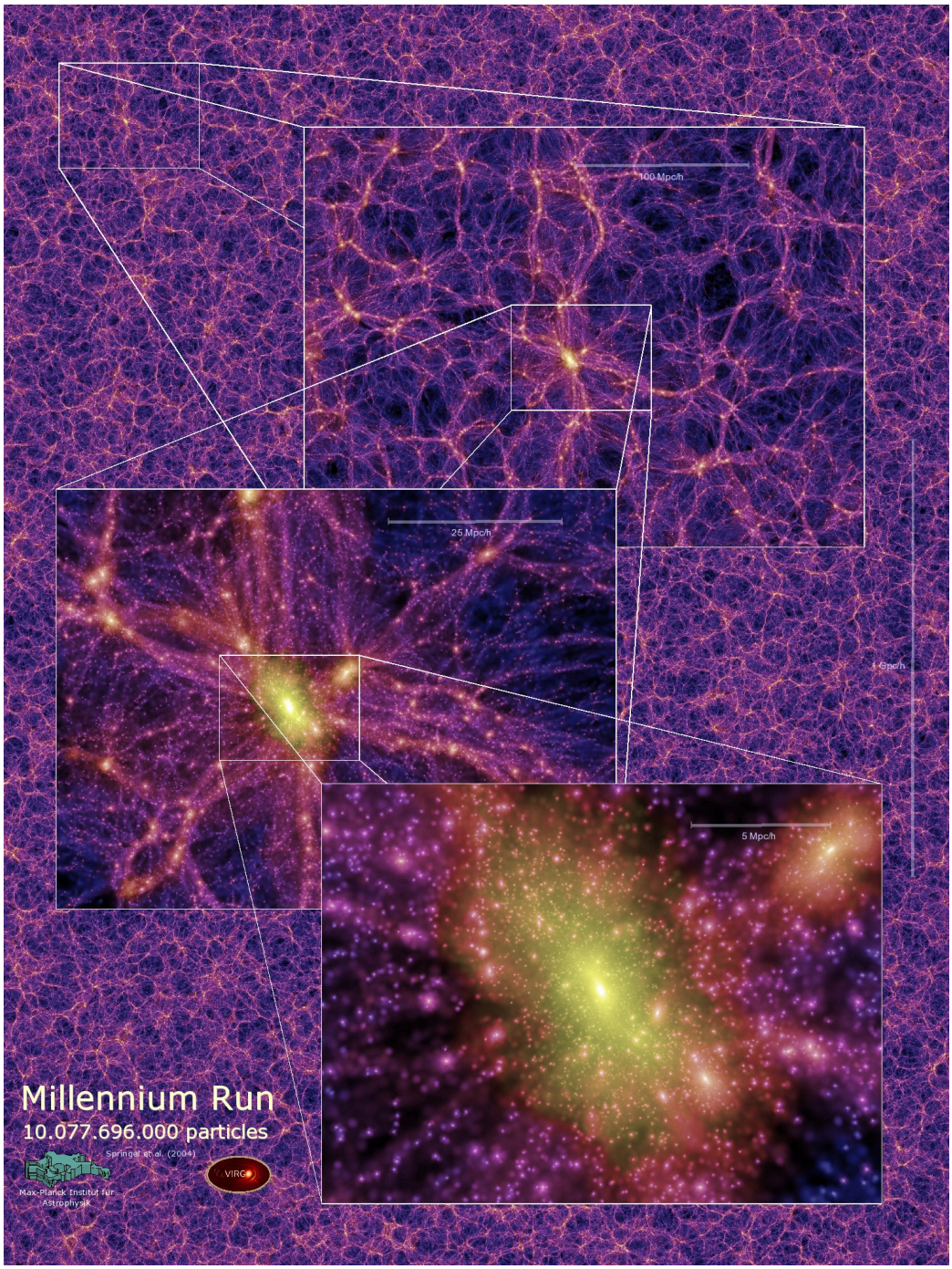

Figure 1.1: The cosmic web structure from the Millennium simulation (Springel et al. 2005). Projected matter density in a 15 Mega-parsecs (Mpc)/h thick slice at redshift=0 is shown in the main panel, with four inset panels showing zoom-ins each by a factor of 4. The scale is indicated through a white line in each panel.

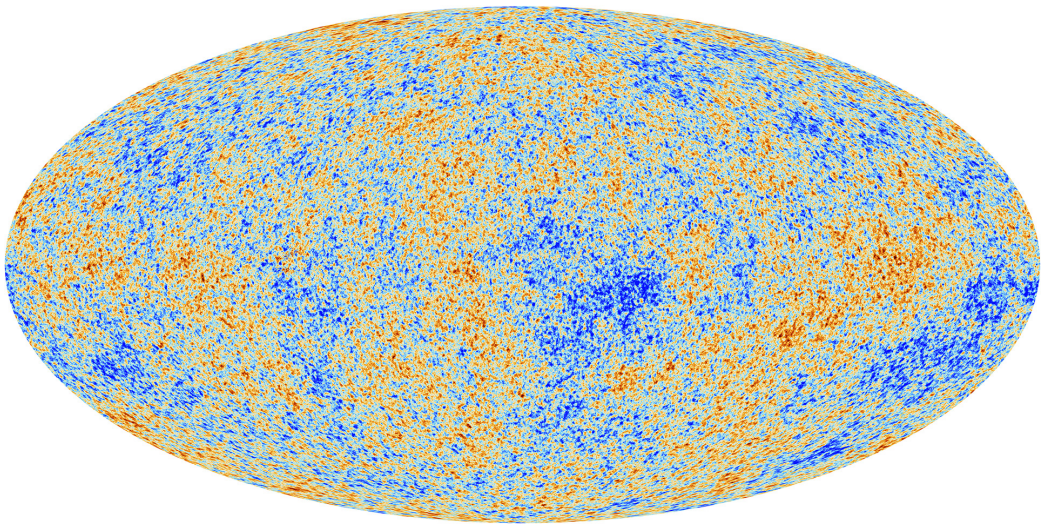

Figure 1.2: "The anisotropies of the Cosmic microwave background (CMB) as observed by Planck. The CMB is a snapshot of the oldest light in our Universe, imprinted on the sky when the Universe was just 380,000 years old. It shows tiny temperature fluctuations that correspond to regions of slightly different densities, representing the seeds of all future structure: the stars and galaxies of today." −Image and caption credit: ESA and the Planck Collaboration

.

to star formation. Dark matter also dominates over baryonic matter by a factor of more than ~ 5. Without dark matter, baryons alone would have taken much longer to form the intricate cosmic web structure we observe today, shaped over 13.8 billion years since the Big Bang.

Over time, regions with slightly more matter pulled more matter in due to gravity, making initially slightly over-dense regions grow in density, while the initially slightly under-dense regions became more and more sparse over time. Each density peak, after achieving a certain threshold of over-density, is able to decouple from the expansion of the Universe, ultimately forming self-gravitating, virialized structures, called dark matter halos. Definitions vary in the literature, but typically in dark matter-only simulations showing structure formation, an over-density by a factor of ~ 200, compared to

average background density, is considered to be a threshold to form dark matter halos, based on the spherical top-hat collapse models[1] (Eke et al. 1996). Within these collapsed halos, cold streams of hydrogen gas funnel in, reaching density thresholds to trigger nuclear fusion. Stars are born. These stars, and the soup of their surrounding gas, form rotating luminous structures called galaxies, residing at the centers of much larger and invisible dark matter halos.

As the Universe evolves, dark matter halos merge due to gravity, growing in size and mass—heavier halos eat up lighter halos. As a result, galaxies within merging halos also merge, acquiring stars (*ex-situ* star-formation) and gas, growing in size and mass. Additionally, based on the availability of cold and dense gas, galaxies continue to form stars internally (*in-situ* star formation). This availability of cold and dense gas is crucial for star formation and is not unaffected by the various feedback effects mentioned below that take place throughout the rest of cosmic history, regulating star formation.

Massive stars run out of their fuel fast and die with supernova explosions, spewing out heavier elements up to Iron formed in their cores back into their environment. Supernova explosions, along with stellar winds from massive stars, heat up the surrounding gas, suppressing the availability of cold gas, and in turn lowering the efficiency of star formation in low-mass galaxies. On the other hand, ultimately, black holes at the centers of galaxies accumulate enough mass by the time their host galaxies have also become massive. The swirling accretion disks and/or the jets emanating from their central region can heat the surrounding gas and/or push the gas out of galaxies and prevent further accretion, halting star formation. As regulated star formation

[1]https://ned.ipac.caltech.edu/level5/Sept11/Benson/Benson2.html

continues, the Universe becomes more and more structured, guided by the same gravitational principle: Larger galaxies and their halos attract smaller galaxies and their halos, forming clusters. The result is a highly structured cosmic web as seen in Figure 1.1, with regions of extreme over-densities, called nodes and filaments, and regions with very little matter, called voids.

1.2 The Galaxy-Halo connection

Due to the dominance of dark matter in terms in terms of mass and scale over baryonic matter, the structure formation is dominated by the assembly of dark matter halos. In essence, dark matter halos provide cocoons within which galaxies form and evolve. Not only that, but it is through the dark matter halos that these galaxies connect to and are influenced by the large-scale structure of the Universe. Therefore, to understand the full physics of galaxy formation, it is imperative to understand its connection to the assembly of dark matter halos.

As the cosmic web we see today is the result of the assembly histories of dark matter halos and the galaxies within, hidden within its imprint are the clues to galaxy formation. The first step in this is to map this cosmic web, or in other words, the spatial distribution of galaxies within the Universe. This is where large-scale surveys come into play. Not only do they map the configuration of this structure today, but also back in time. As light travels at a finite speed, galaxies that are far away appear as they did in earlier times. This allows us to map the evolution of cosmic structure with time, allowing us to trace the evolutionary history of galaxies and ascertain the galaxy formation physics at play.

While galaxy evolution depends on the complex baryonic physics of gas cooling, star formation, for instance, the evolution of dark matter halos depends only on the cosmological model. By focusing on the spatial distribution of galaxies which depends on the cosmology-only dependent dark matter halos, we can indirectly probe the galaxy formation physics. The results thereof are therefore incredibly helpful in disentangling the different models of galaxy formation physics.

One of the earliest works on the measurements of the spatial distribution of galaxies came from Totsuji & Kihara (1969) based on the brightest galaxies captured on photographic plates of size $6^\circ \times 6^\circ$ by Shane & Wirtanen (1967) at the Lick Observatory. Peebles (1973) expanded on earlier studies to present a comprehensive framework to analyze the statistics of the spatial distribution of objects in the extragalactic catalogs based on the power spectrum. Soon, it was realized that the measurements of the clustering of galaxies, which is connected to the clustering of halos, should enable probing the halo masses, as clustering of halos depends strongly on halos mass (Bahcall & Soneira (1983); Kaiser (1984); Bardeen et al. (1986); Mo & White (1996)).

1.3 Current understanding of the Galaxy-Halo Connection

Since the turn of the 21^{st} century, two strands have seen significant development, enabling more and more refined understanding of the galaxy-halo connection. On the one hand, advances in computational power enabled N-body cosmological simulations showing the evolution of cosmic structure

starting from dark matter particles (Springel et al. 2005; Klypin et al. 2011; Riebe et al. 2013; Vogelsberger et al. 2014; Schaye et al. 2015), improving in size and resolution. Additionally, galaxy formation models could be applied onto these dark matter simulations to produce observables that could be compared directly with the real observations (e.g., Springel et al. (2018)). On the other hand, advances in large survey telescopes allowed robust mapping of the galaxy distributions, using spectroscopy, over large areas of sky and depth. Spectroscopic (or redshift) surveys measure the full spectral energy distributions (SEDs) of galaxies, allowing for robust redshift estimation and stellar population properties, far surpassing the photometric surveys in that regard (see §1.5 for more details on photometric surveys). The earliest revolutionary spectroscopic survey was the Two-degree Field Galaxy Redshift Survey − 2dFGRS (Colless et al. 2001) , which measured redshifts of about 250,000 galaxies, followed by the landmark Sloan Digital Sky Survey − SDSS (York et al. 2000) , capturing millions of galaxies over multiple phases.

A myriad of studies have explored the galaxy-halo connection utilizing the advancements both in the simulations and the observations mentioned above. By far, the most focused upon aspect has been the connection between the galaxy stellar mass and its host host dark matter halo mass, the stellar-to-halo mass relationship (SHMR) as it is thought to capture the fundamental essence of the galaxy-halo connection, and is also easier to constrain.

While laying down a complete multi-variate connection between galaxy properties (stellar mass, star formation rate, morphology, etc. and halo properties (halo mass, halo formation time, halo angular momentum, etc.) across cosmic time is the long-term goal of the field, this dissertation focuses on the fundamental facet of the galaxy-halo connection: the dependence of a

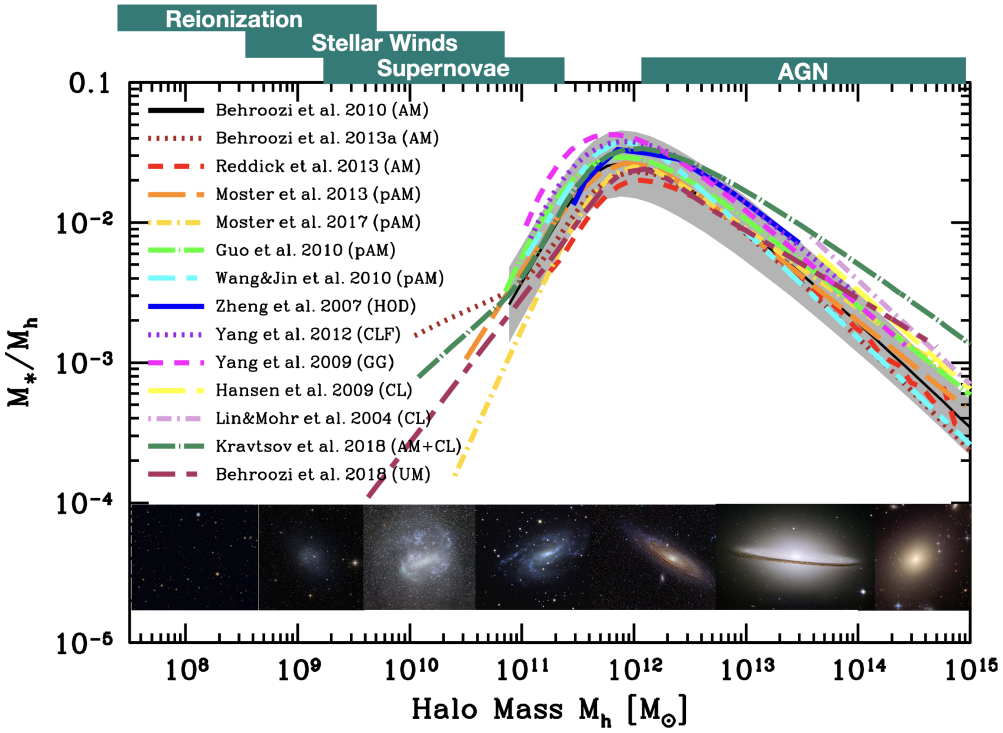

Figure 1.3: M_*/M_h as a function of M_h of central galaxies in the local Universe (z=0) derived using various methods as indicated in the legend, and described in Wechsler & Tinker (2018). It shows a characteristic peak at $M_h \sim 10^{12} M_\odot$ indicating the most efficient conversion of baryonic matter into stellar matter. The candidate physical processes responsible for suppression in integrated star forming efficiency (SFE) at different halo mass scales are indicated at the top, while the typical galaxies hosted by these halos are shown at the bottom. *Figure taken from Wechsler & Tinker (2018).*

galaxy's stellar mass on the mass of its host halo.

There is broad, overwhelming agreement in the shape of the SHMR at z = 0. Usually it is plotted as a ratio of stellar mass to halo mass, (M_*/M_h) as a function of halo mass, M_h, as it informs how efficient the halo of a given mass has been in converting baryons into stars integrated over its lifetime. The ratio, M_*/M_h, peaks at about $M_h \sim 10^{12}$ $M_\odot$ (which we call $M_h{}^{peak}$) in the local Universe (see Figure 1.3). $M_h{}^{peak}$ can be converted into integrated star forming efficiency (SFE) by dividing it by the baryonic fraction of the

Universe, $f_b = 0.158$ (Planck Collaboration et al. 2020), i.e., (M_*/M_h) / f_b. It turns out that even at the ${M_h}^{peak}$, the SFE never exceeds ~ 20 % in any of the studies, highlighting that star formation, the act of converting available gas into stars, is indeed a very inefficient process, in a cosmically averaged sense. Prior to the work presented in this dissertation, there was general consensus in the increase of ${M_h}^{peak}$ out to $z \sim 2$, by ~ 0.3 dex (Leauthaud et al. 2012; Moster et al. 2018; Behroozi et al. 2019; Shuntov et al. 2022), and it was unclear if it continues to rise, flatten, or decline beyond $z \sim 2$ (Ishikawa et al. 2017; Moster et al. 2018; Legrand et al. 2019; Shuntov et al. 2022). However, most of these results are either based primarily on or rely heavily on the singular COSMOS field. As shown in chapter 3, with the unprecedented use of the homogeneously prepared catalogs of the UDS and COSMOS fields, both contributing roughly equal sky coverage, and a novel modeling framework, we establish that the ${M_h}^{peak}$ stays consistent with no evolution throughout $0.2 < z < 4.5$. This indicates that the star formation in the Universe is even more tightly coupled to the halo mass than previously thought.

It is an ongoing quest to understand the even lower integrated SFE of halos of masses on either side of $M_h \sim 10^{12}$ $M_\odot$. In less massive galaxies and halos, currently the consensus is that small-scale feedback effects such as supernovae and stellar winds heat the surrounding gas, suppressing star formation (e.g., Hopkins et al. (2012)). In the more massive galaxies and halos, large-scale active galactic nuclei (AGN) jets powered by the central black holes push the gas out of the galaxies, heat it, and prevent it from falling back on, thus halting star formation (e.g,. Silk & Rees 1998; Kondapally et al. 2023).

1.4 Constraining the Galaxy-Halo Connection−Galaxy Stellar Mass to Halo Mass Relationship (SHMR)

In the literature, a wide range of approaches are present to connect galaxy stellar masses to their host halo masses. Full hydrodynamical simulations such as IllustrisTNG (Weinberger et al. 2017) evolve gas, stars, dark matter, and sometimes black holes self-consistently within a cosmological box, assuming galaxy formation physics informed via an amalgamation of lessons learned through observations of galaxies and their feedback cycles. Then there are semi-analytical approaches like DARK SAGE (Stevens et al. 2024) that leverage dark matter halo merger trees from dark matter-only simulations and use analytical equations to approximate galaxy formation processes atop those halo merger trees. On the other hand, empirical modeling avoids assuming any galaxy formation physics altogether, and instead, utilizes a combination of measurements such as galaxy clustering, stellar mass functions, etc., to constrain parameterized models of, for instance, star formation in halos (Behroozi et al. 2019). Finally, there are other approaches which, in addition to not assuming any galaxy formation physics like in empirical models, do not suffer from the biases of parameterized galaxy formation present in empirical models.

The simplest of such approaches is abundance matching (e.g., Marinoni & Hudson (2002)). In this technique, the abundance of halos (volume number density of halos per halo mass) as a function of halo mass, called the halo mass function (HMF), is first determined from the dark matter-only simulations run assuming a set of cosmological parameters. Resulting halos are then rank ordered based on their masses and then matched to the rank or-

dered galaxies from the observed galaxy abundance (volume number density of galaxies per stellar mass), or the galaxy stellar mass function (SMF). Such an assignment of galaxies to halos produces the SHMR, albeit fraught with major shortcomings discussed below that can be overcome with the measurements of galaxy clustering, the spatial distribution of galaxies in the cosmic web. As mentioned in section 1.2, it was realized early on that galaxies are biased tracers of the underlying dark matter halos, and therefore, the spatial distribution of galaxies provides an invaluable probe into the halo mass.

In abundance matching, it is unclear how to assign galaxies as central and/or satellites within halos and/or sub-halos. The same abundance of galaxies or SMF can produce varied relative distributions of centrals and satellites within the same halos. However, the relative central/satellite distributions within halos leave an imprint on the small-scale galaxy clustering amplitude, which can be leveraged to help constrain the halo mass. Not only that, but since halo clustering increases with mass as high mass halos are more biased, a highly clustered galaxy population indicates a higher host halo mass, therefore, clustering provides an independent probe on the halo mass. In addition, the scatter between halo mass and stellar mass is also highly sensitive to the clustering of galaxies. The same SMF, with different scatter, produces a different clustering signal. Given all the above, clustering becomes crucial in addition to the abundance of galaxies (SMF) in connecting halo mass to galaxy stellar mass.

The halo occupation distribution−HOD modeling framework allows leveraging clustering and number density measurements, and without assuming any galaxy formation physics, connects galaxy stellar mass to dark matter halo mass. The HOD model simply dictates how many galaxies a halo above

a given threshold (usually a stellar mass or a luminosity threshold) hosts, on average. For a given set of HOD parameters and the resulting population of galaxies within halos, a unique combination of clustering and number density appears. Therefore, for a given observed clustering and number density, HOD parameters can be determined by matching the observations with HOD model predictions.

1.5 Constraining SHMR with large extragalactic photometric surveys

It is ideal to constrain the HOD models utilizing measurements based on spectroscopic surveys that capture the distribution of galaxies in 3-D in detail, due to accurate redshifts. Currently existing spectroscopic samples such as SDSS Abazajian et al. (2009) have been vital in uncovering the galaxy-halo connection, especially the SHMR, utilizing the HOD approach (e.g., Moster et al. (2010)). While they are great for constraining the galaxy-halo connection at low redshift ($z \lesssim 0.7$), any inference about the halo or galaxy assembly at high redshift is still extrapolated. Some spectroscopic surveys do probe higher redshifts, out to $z \sim 2$, but only for small and biased samples, e.g., KBSS Steidel et al. (2014). Such samples are not ideal for conducting complete statistical studies of the galaxy-halo connection, but rather are designed to target specific galaxy populations.

Only now, spectroscopic surveys that probe higher redshifts and are complete down to a stellar mass are being initiated. For instance, the Prime Focus Spectrograph (PFS) survey is starting to collect its first light. PFS is going to be the most extensive high redshift survey ever conducted, exquisitely

charting the large-scale structure of our Universe in 3D over ~ 12 deg^2 just with its low-z ($z < 1.7$) component. While the PFS *low-z* data is a treasure trove for understanding how galaxy assembly intertwines with halo assembly since $z \sim 2$, an era when star formation in the Universe peaked, there is still time in its collection and processing to be ready for science. Therefore, the extraction of the evolution of the galaxy-halo connection over a significant portion of cosmic history, at least until now, could only be done through deep photometric surveys.

Advances in ground-based 8-m class telescopes with adaptive optics have allowed conducting photometric surveys that not only cover large sky areas sampling the cosmos adequately, but are deep, capturing galaxy populations as they evolve with cosmic time. A photometric survey involves pointing the telescope at a desired patch of sky using different wavelength filters, mostly broad-band and less commonly, medium or narrow-band. These resulting multiple wavelength fluxes serve as data points for fitting spectral energy distributions (SEDs), as opposed to continuous spectra obtained in spectroscopic surveys. Therefore, to confidently model the SEDs, wide wavelength coverage in addition to enough filters is crucial. Accurately modeling an SED allows the understanding how much the SED has shifted in wavelength due to redshift, hence allowing for accurate estimation of the redshift. Galaxy redshifts estimated using photometry are called photometric redshifts, which are less reliable than spectroscopic redshifts, but with wide wavelength coverage and fine sampling, the resulting photometric redshifts can reach high accuracies (below $1-5\%$). With accurate SED modeling of galaxies, stellar population properties, such as stellar mass, star formation rates, etc., can also be reliably estimated. Catalogs with high-quality photometric redshifts

and stellar population properties are therefore immensely powerful in extragalactic science due to their generally much better depth. In addition, photometric surveys are generally less expensive and less time-consuming to conduct, making them invaluable.

1.6 Large extragalactic photometric surveys

In the last couple of decades, several large and high redshift photometric campaigns have been conducted with ground-based optical and near-infrared (NIR) telescopes, resulting in abundant literature within the field of extragalactic astrophysics. A few of the most notable surveys are the following:

i- NEWFIRM Medium Band Survey−NMBS (van Dokkum et al. 2009) was conducted with the Kitt Peak 4m Mayall Telescope covering 5 medium-band NIR filters−J_1, J_2, J_3, H_1, and H_2 designed to detect the Balmer and 4000Å breaks of $1.5 \lesssim z \lesssim 3.5$ galaxies with high confidence due to their fine sampling and narrower width as opposed to broad-band filters. This ultimately led to a more accurate SED modeling of galaxies, and as a result, the photo-z's and stellar population properties. This survey covered a total of ~ 0.5 deg^2 within the COSMOS and AEGIS fields.

ii- United Kingdom Infrared Telescope (UKIRT) was used to conduct the Ultra-Deep Survey (UDS) (Lawrence et al. 2007) utilizing 3 broad-band NIR filters−J, H, and K covering ~ 0.8 deg^2 within the UDS field, aiming to measure the growth of structure since $z = 3$ (Lawrence et al. 2007).

iii- Another NIR survey, UltraVISTA (McCracken et al. 2012) was conducted with the Visible and Infrared Survey Telescope for Astronomy (VISTA) in 4 broad-bands−Y, J, H, K_s and a narrow-band−NB118 in the COSMOS

field covering a large 1.5 deg^2. Its 1^{st} data release appeared in February 2012, and the 5^{th} data release was in May 2023, substantially improving the initial survey depth, allowing probing of the galaxy evolution and the evolution of structure out to $z \sim 5$.

iv- An optical survey which has recently concluded was conducted with the Hyper Suprime-Cam Subaru Strategic Program−HSC SSP (Aihara et al. 2019) at the Subaru Telescope in the five broad-band− g, r, i, z, y, and three narrow-bands − NB387, NB816, and NB921 bands covering 26 deg^2 across four fields (E-COSMOS, XMM-LSS, ELAIS-N1, DEEP2-3) called the Deep layer and 4 deg^2 embedded within them called the UltraDeep layer in COSMOS (within E-COSMOS) and SXDS (within XMM-LSS) fields.

v- Then there is the Canada-France-Hawaii Telescope (CFHT) Large Area U-band Deep Survey (CLAUDS) (Sawicki et al. 2019) conducted on the bluer end of the optical wavelengths, just blue-ward of the optical g band, covering $\sim$ 18 deg^2 designed. The CLAUDS was designed to supplement the lacking U-band (u^* and u) photometry to the HSC-SSP Deep and UltraDeep fields with comparable depth.

1.7 Extensively multi-wavelength photometric catalogs

All of the aforementioned surveys and many others have been integral in the progress of our understanding of the evolution of galaxies. However, a single survey's wavelength range is limited, and its wavelength sampling, which can be noticed from above, typically ranges between ∼2 to ∼8. The limited range and sparse sampling are usually not robust enough for accurate SED modeling. Therefore, multiple surveys usually need to be combined to

cover a wide range in wavelength, and a greater number of filters finely sampling the SEDs, to confidently estimate photo-z's with their errors down to a few percent. This need leads to a myriad of photometric catalogs in the field that combine data from multiple surveys like the ones mentioned above to achieve robust cataloging of galaxies, in terms of their photo-z's and their stellar population properties, e.g., UltraVISTA (Muzzin et al. 2013) and COSMOS2020 (Weaver et al. 2022). This dissertation describes the construction of one such catalog−FENIKS catalog (Zaidi et al. (2024); see chapter 2), which combines ground-based (and space-based *Spitzer* IRAC) data in 24 bands spread over multiple surveys.

Such extensive multi-wavelength cataloging is essential to understand the evolution of the galaxy-halo connection since high redshift, especially in the absence of high redshift spectroscopic surveys. Additionally, beyond combining multi-wavelength data from various surveys in one field, multi-wavelength data across multiple fields can be combined to expand the total cosmic area and volume probed. This is vital in mitigating cosmic variance while ascertaining the galaxy-halo connection from the cosmic structure. Such is the topic of Chapter 3.

However, careful attention needs to be given to avoid any systematic discrepancies going from raw data to final science-ready catalogs of galaxies across multiple fields, again, a feature of the analysis mentioned in chapter 3. The homogeneous photometric catalog construction while combining data form multiple fields/surveys is crucial for understanding the galaxy-halo connection out to high redshift right now, but even in the future when the spectroscopic surveys at high-z will become commonplace as photometric surveys will always tend to be cheaper to conduct and guide the follow-up

spectroscopic efforts.

1.8 This Dissertation

In this dissertation, I will discern the relationship between galaxies and their dark matter halos through careful utilization of multiple large-scale surveys conducted at various filter wavelengths (photometry). In chapter 2, I will walk you through the construction of a photometric galaxy catalog starting from multi-wavelength large-scale surveys. The wide range and high sampling of galaxy SEDs allowed robust characterizations of an array of galaxy properties. Such catalogs, which contain the spatial distribution, redshifts, and stellar mass of galaxies in addition to a myriad of other properties, are the bedrock on which the galaxy-halo connection can be gleaned from the cosmic web structure. In chapter 3, using HOD modeling, I constrain the evolution in the fundamental facet of the galaxy-halo connection, which is the relationship between the galaxy stellar mass and its host dark matter halo mass out to the first billionth year ($z \sim 5$) of the Universe. Finally, in chapter 4, I mention my ongoing effort to extend the SHMR out to the cosmic dawn ($z\sim10$), thanks to enough accumulation of sky coverage in three years since everyone's favorite − *James Webb Space Telescope* started conducting surveys deep enough to capture the earliest phases of galaxies' lives.

CHAPTER 2

THE FENIKS SURVEY: MULTI-WAVELENGTH PHOTOMETRIC CATALOG IN THE UDS FIELD, AND CATALOGS OF PHOTOMETRIC REDSHIFTS AND STELLAR POPULATION PROPERTIES

2.1 Introduction

Over the last 15 years, the construction of near-infrared (NIR) detected multi-wavelength photometric survey catalogs has become progressively more sophisticated (MUSYC, Quadri et al. 2007; NMBS, Whitaker et al. 2011; UVISTA-DR1, Muzzin et al. 2013; ZFOURGE, Straatman et al. 2016; COSMOS2015, Laigle et al. 2016; COSMOS2020, Weaver et al. 2022).

The high-quality, point-spread function (PSF) matched multi-wavelength photometry is essential for any investigation of galaxy evolution. Besides allowing for large photometric statistical analysis, they enable the identification of robust targets for detailed follow-up studies, e.g., spectroscopy for confirmation and stellar continuum and emission line studies, longer wavelength investigations (e.g., with ALMA), or space-based imaging and spectroscopy.

This has become increasingly important in the era of JWST, as the previously unprobed regions of parameter space (e.g., low-mass galaxies and objects in the epoch of re-ionization) by ground-based facilities can now be explored. For instance, rest-frame optical spectroscopic studies of massive

Reprinted with permission from "The FENIKS Survey: Multi-wavelength Photometric Catalog in the UDS Field, and Catalogs of Photometric Redshifts and Stellar Population Properties" by Kumail Zaidi et al. 2024, ApJ, 969, 84, doi: 10.3847/1538-4357/ad45fa

dusty star-forming galaxies at $z > 4$ are now possible with JWST. To facilitate such endeavors, here we present a K-band detected, PSF-matched multi-wavelength (from the u-band to the mid-IR) photometric catalog for the F2 Extragalactic NearIR K-Split Survey (FENIKS). The FENIKS survey is currently being carried out using two medium-band filters, K_{blue} (λ_{eff} = 2.06μm) and K_{red} (λ_{eff} = 2.31μm) (see Esdaile et al. 2021 for more details) recently added to the `FLAMINGOS-2 (F2)` instrument (Eikenberry et al. 2008, 2012; Gomez et al. 2012) atop the 8.1-meter Gemini South Telescope in Chile. The medium K_{blue}/K_{red} filters, which straddle the K-band, increase the photometric redshift accuracy with better sampling of either side of the Balmer/4000Å breaks of the $z > 4$ galaxies.

The footprint of the catalog and the primary data comes from the UKIDSS UDS data release 11 (UDS DR11, hereafter) for which other catalogs already exist (Almaini et al. in prep.; Hartley et al. in prep.)[1] The UKIDSS project is defined in Lawrence et al. (2007). Further details on the UDS can be found in Almaini et al. (in prep). UKIDSS uses the UKIRT Wide Field Camera (WFCAM; Casali et al. (2007)). The photometric system is described in Hewett et al. (2006), and the calibration is described in Hodgkin et al. (2009). The pipeline processing and science archive are described in Irwin et al (in prep) and Hambly et al. (2008). Besides the data in the broad-band filters of UDS DR11 (J, H, K) and the FENIKS medium bands, we also included ancillary data from other surveys: CLAUDS (MegaCam uS-band; Sawicki et al. 2019), SXDS (B, V, R, i, z; Furusawa et al. 2008), HSC (g, r, i, z, y, NB0816, NB0921; Aihara et al. 2019), VIDEO (H, K_s, Y, z; Jarvis et al. 2013), *Spitzer*-IRAC (ch1-4; Mauduit et al. 2012; Lacy et al.

[1] https://www.nottingham.ac.uk/~ppzoa/DR11/

2021), totaling data in 26 bands over a surveyed area of ~ 0.9 deg^2. Our catalog, containing about half the area but $\sim$0.7(0.3) mag deeper than the UltraVISTA K-band 'deep'('ultra-deep) regions, is nicely complementary to the COSMOS2020 catalog (Weaver et al. 2022). This makes the FENIKS catalog the deepest wide-field ground-based NIR dataset and importantly, it probes a different line of sight, vital in reducing the uncertainties due to cosmic variance in measured quantities (e.g., number densities).

The photometric catalog, along with the associated catalogs of photometric redshifts and stellar population properties, are made publicly available[2], enabling scientific investigations over a wide range in redshift ($0.2 < z < 8.0$, i.e., covering about 95% of cosmic history). The combination of very deep NIR imaging over a wide area allows for the study of both distant faint/low-mass galaxies, as well as the population of rare distant bright/massive galaxies. The ultra-deep K-band photometry in combination with high-quality photometric redshifts enables complete studies of galaxies out to $z \approx 5$, as the selection in the rest-frame optical reduces biases against evolved and/or dusty star-forming galaxies.

The paper is structured as follows: in §3.2, we describe the different surveys from which the multi-wavelength data have been combined to build the catalog; in §2.3, we present the procedure for the construction of the catalog in detail, as well as define the included quantities/parameters in the catalog and how they can be used to make a typical selection of galaxies; §2.4 presents the modeling of the spectral energy distributions (SEDs) to derive

[2]The Version 1 (v1) release of the FENIKS UDS catalogs containing photometric catalog, the redshift catalog and the catalogs of stellar population properties as described in this paper can be found at https://www.zaidikumail.com/feniks-uds-catalogs and through Zenodo at 10.5281/zenodo.11002299

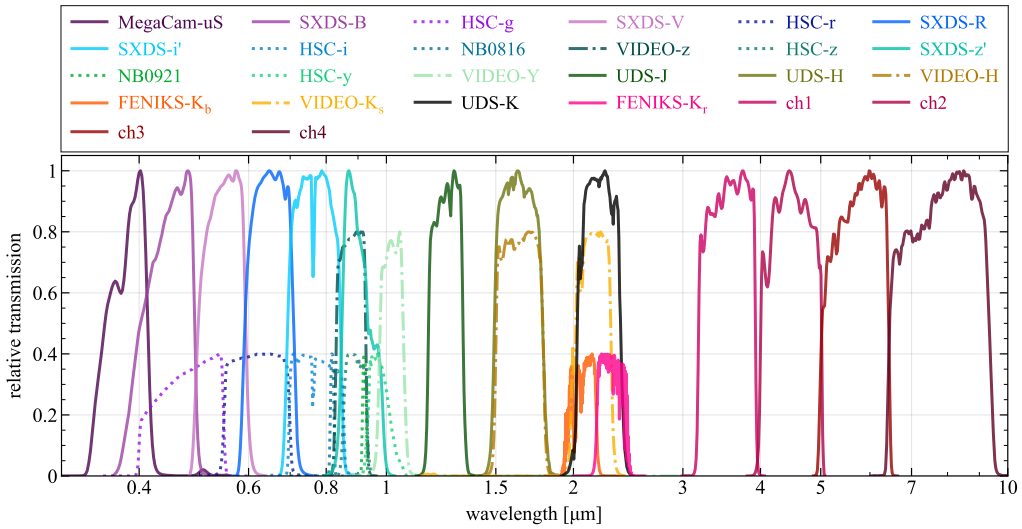

Figure 2.1: Filter transmission curves of different bands. For clarity, some filters are normalized to have a maximum relative transmission less than 1: FENIKS (solid) K_{blue} and K_{red} peak at 0.4; HSC (dotted) g, r, i, z, y, NB0816, and NB0921 have peaks at 0.4; VIDEO (dash-dotted) z, Y, H, and K_s peak at 0.8. All the other bands (solid) peak at 1. The detection UDS K-band (see §2.3.2 for details) is shown in solid black.

photometric redshifts and stellar population properties; we conclude with a summary in §3.7.

Throughout this paper, we assume a ΛCDM cosmology with $H_0 = 70$ km s^{-1} Mpc^{-1}, $\Omega_{m,0} = 0.3$, and $\Omega_{\Lambda,0} = 0.7$. The magnitudes and colors are presented in the AB magnitude system (Oke & Gunn 1983). The stellar population parameters are derived assuming the Chabrier initial mass function (IMF) (Chabrier 2003).

2.2 Data and Overview

Our catalog is based on the final release (UDS DR11)[3] of the J, H, and K data over the full UDS field which covers an area of ∼0.9 deg^2. Along with

[3] https://www.nottingham.ac.uk/astronomy/UDS/data/dr11.html

these primary data, ancillary data have been combined from various deep surveys to produce a PSF-matched photometric catalog of 26 bands, whose transmission curves are shown in Fig. 2.1, the footprint shown in Fig. 2.2, and the metadata is shown in Table 2.1.

In the optical, we employ the deep imaging available in B, V, R, i', z' from the SubaruXMM-Newton Deep Survey (SXDS), which covers the UDS field with five SuprimeCam pointings from the 8.2m Subaru Telescope (Furusawa et al. 2008). We utilize the version of these mosaics generated by C. Simpson (Gemini) to complement the UDS DR11 release which benefits from a considerably better zero-point (ZP, hereafter) scaling and background subtraction compared to the initial release[4]. We target the shortest wavelength in our catalog using the available data in the uS filter obtained with MegaCam on the Canada-France Hawaii Telescope (CFHT) as part of the CLAUDS survey (Sawicki et al. 2019).

Additionally, we took advantage of the overlap with the public data release 2 (PDR2) of the Hyper Suprime-Cam (HSC) Subaru Strategic Program (SSP) (Aihara et al. 2019), also conducted by the Subaru Telescope. The PDR2 provides multi-band data in a large area of over 300 deg^2 including deep/ultra-deep and wide regions out of which the tract number 8523 overlaps with our UDS field in g, r, i, z, Y, $NB0921$, and $NB0816$ bands, all of which we included in our catalog. We noticed that the astrometry of the PDR2 was offset from UDS DR11 in both the RA and DEC direction, with median offsets of $\sim$ 0.1 arcsec . Furthermore, these offsets were not uniform across the image. To remedy this, we re-mapped the astrometry of the relevant part of PDR2 images using the centroids of stars relative to the

[4] https://www.nottingham.ac.uk/~ppzdm/research.html

corresponding stars in UDS DR11, resulting in residual offsets of the order of milli-arcseconds.

Available data from the data release 4 (DR4) of the VISTA Deep Extragalactic Observations (VIDEO; Jarvis et al. 2013) in the NIR filters Z, Y, H, K_s were also included. VIDEO is a deep NIR imaging survey covering $\sim$12 deg^2 in three fields XMM1, XMM2, and XMM3 out of which XMM1 fully encapsulates the UDS DR11 pointing.

Finally, in the mid-IR, we took advantage of the data from the SERVS (Mauduit et al. 2012) and DeepDrill (Lacy et al. 2021) surveys using the *Spitzer Space Telescope* Infrared Array Camera (IRAC) in ch1 and ch2. For IRAC ch3 and ch4, we used the imaging data from the [SpUDS survey]https://www.ipac.caltech.edu/doi/irsa/10.26131/IRSA403.

Due to the delays in the FENIKS data collection in the medium K_{blue} and K_{red} bands at the Gemini Telescope, we have not incorporated the FENIKS medium-band photometry in our catalog yet. As soon as the data collection is complete, we will release an updated version of the publicly released catalogs including the FENIKS photometry along with the updated photometric redshifts and stellar population properties.

2.3 Catalog construction

2.3.1 PSF matching of the optical-NIR bands

The optical-NIR images in the 20 filters between MegaCam-uS (0.38μm) and the UDS K (2.1μm) bands suffer from atmospheric seeing resulting in varying image resolutions that manifest in inhomogeneous point spread functions

(PSFs). As a result, aperture photometry cannot be performed consistently across the bands as a fixed aperture size would not necessarily correspond to the same physical region of the source. To ensure consistent aperture photometry, we need to homogenize the PSFs among all the optical-NIR bands. To do so, we first determined the lowest resolution band to which all the other bands should be PSF-matched. As a side note, the IRAC images are heavily blended and have much broader PSFs than any of the optical-NIR bands, and therefore they require a different approach to obtain consistent aperture photometry which we describe in §2.3.5.

In practice, the median full-width-at-half-maximum (FWHM) of each band is determined by fitting a 2D Gaussian or Moffat profile on high signal-to-noise ratio stars and taking the mean or median of the resulting distribution. Then, the band with the largest FWHM is considered to be the worst resolution band. However, any variability in the PSF shapes of the different bands implies that the FWHMs resulting from simple Moffat or Gaussian fits are sometimes not appropriate to reliably identify the band with the worst image quality, especially if multiple bands have very similar FWHMs. Therefore, we adopted a much simpler approach assuming no functional form for the PSF shape. Specifically, for each band, we extracted the median growth curve using the normalized circular star stamps with a diameter equal to 32 pixels (8.6 arcsec), large enough to contain all of the stars' fluxes, and the corresponding median 50% and 75% encircled light radii, r_{50}, and r_{75}, respectively. The band with the worst resolution is MegaCam-uS ($r_{50} \sim$ 0.63 arcsec ; FWHM $\sim$ 1.00 arcsec) which is slightly worse than the second worst resolution band, VIDEO-Y ($r_{50} \sim 0.60$ arcsec ; FWHM $\sim$ 0.87 arcsec). To avoid the extra loss of signal we would have had if all optical-NIR bands

were PSF-matched to MegaCam-uS instead of VIDEO-Y, we decided to PSF match all the optical-NIR bands (except MegaCam-uS) to VIDEO-Y.

In §2.3.3, we describe how we differently included PSF-matched aperture photometry for the MegaCam-uS band.

The FWHMs of stars are found to vary between +/-2% and +/-4% across the field, depending on the band, which is expected as the image footprint is large enough ($\sim$0.9 deg^2). Therefore, instead of having one single model PSF for the whole image, we subdivided our images into grids, determined the model PSFs for each patch, and then PSF matched the images patch by patch. To determine model PSFs, both in the low-resolution image (VIDEO-Y) and high-resolution images, we utilized unsaturated, high S/N stars in the local area, enough in number to sample the local PSF appropriately. The density of the PSF grid was based on the number of total stars selected in each high-resolution image. Once the PSF grid for a high-resolution band and the corresponding grid for VIDEO-Y is developed, we fit the Fourier transform of the model PSFs with a series of Gaussian-weighted Hermite Polynomials as presented in Skelton et al. (2014) and Shipley et al. (2018) to obtain the corresponding convolution kernel grid. The convolution kernels were then used to PSF match the whole image patch by patch, and then stitched together using the code TOPH[5].

Figure 2.3 shows the ratio of the high-resolution image median growth curves to the VIDEO-Y's median growth curve before (top) and after (bottom) PSF matching. The median growth curves were derived from the same cleaned star stamps utilized in the PSF matching. After PSF matching, the median growth curves of all bands agree very well, within 1%, at all radii.

[5] https://github.com/jacqdanso/TOPH

Table 2.1: Metadata of the filters used in the catalog

Instrument/Telescope/Survey	Filter	λ_{eff}^{a} [Å]	A^{b} [mag]	ZP offset^{c}	$3\sigma/5\sigma$ depth^{d} [AB]	$r_{50}{}^{e}$ [arcsec]
MegaCam/CFHT/CLAUDS	uS	3828	0.085	1.068	27.62 / 27.07	0.625
Suprime-Cam/	B	4448	0.075	1.011	28.14 / 27.58	0.554
Subaru XMM-Newton	V	5470	0.058	0.997	27.86 / 27.31	0.540
survey (SXDS)	R	6505	0.045	1.004	27.60 / 27.05	0.547
	i'	7671	0.035	1.023	27.41 / 26.86	0.560
	z'	9028	0.027	1.049	26.59 / 26.03	0.538
HSC/	g	4798	0.069	0.989	27.70 / 27.14	0.477
Subaru/	r	6218	0.048	0.956	27.25 / 26.69	0.508
SSP	i	7727	0.034	0.986	27.08 / 26.52	0.443
(PDR2)	*NB0816*	8177	0.032	1.001	26.44 / 25.89	0.412
	z	8908	0.027	0.941	26.40 / 25.85	0.472
	NB0921	9213	0.026	0.966	26.18 / 25.62	0.536
	y	9775	0.024	0.930	25.62 / 25.06	0.447
VISTA/	z	8787	0.028	0.937	26.02 / 25.46	0.592
VIDEO	Y	10217	0.022	0.951	25.49 / 24.94	0.593
(DR4)	H	16433	0.010	1.038	24.39 / 23.84	0.531
	Ks	21503	0.007	1.010	24.32 / 23.76	0.534
UKIRT/	J	12502	0.015	1.002	26.15 / 25.60	0.531
UDS	H	16360	0.010	1.002	25.53 / 24.98	0.539
(DR11)	K	22060	0.006	1.000	25.90 / 25.35	0.503
IRAC/*Spitzer*	ch1	35200	0.004	1.002	24.74 / 24.19	-
(SERVS/DeepDrill)	ch2	44600	0.003	1.017	24.31 / 23.75	-
IRAC/*Spitzer*	ch3	56600	0.003	0.870	22.72 / 22.17	-
(SpUDS)	ch4	76800	0.003	0.908	22.62 / 22.07	-

[a]Effective wavelength; [b]Galactic dust extinction from Schlafly & Finkbeiner (2011) calculated using the online tool at `https://irsa.ipac.caltech.edu/applications/DUST/`; [c]ZP offsets calculated using `eazy-py` with respect to the K-band - this multiplicative factor was applied to the photometry to derive photometric redshifts and stellar population properties; [d]3σ and 5σ depths of the PSF-matched images scaled to total using the final PSF-matched growth curve, except for MegaCam-*uS*. For MegaCam-*uS*, the depths were calculated using the non-PSF-matched image and scaled to the total using the corresponding growth curve; [e]median 50% encircled light radius derived using the stars from the non-PSF-matched images to assess image quality.

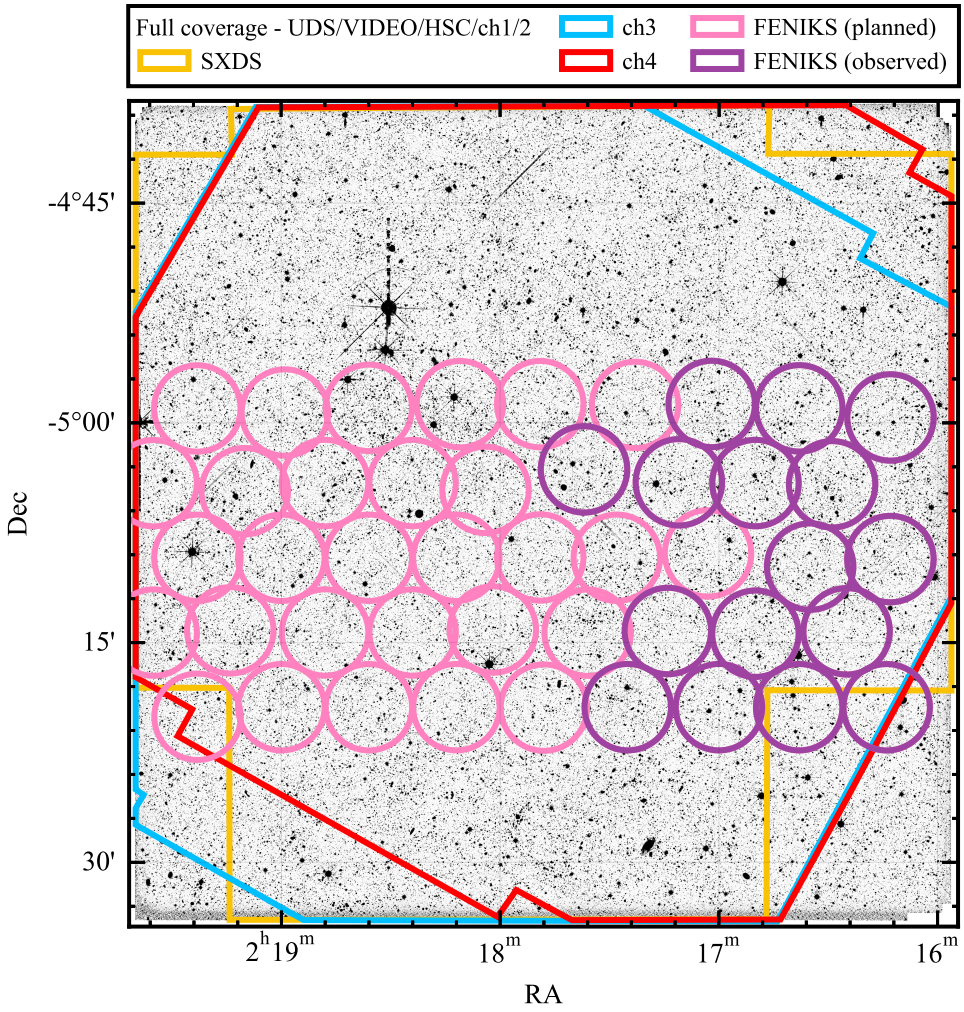

Figure 2.2: The K-band (detection band) image overlaid with the footprints of different filters or surveys. The UDS, VIDEO, HSC, and IRAC ch1/2 bands (16 out of 26 bands utilized in the catalog) have full coverage across the shown K-band image. SXDS coverage is shown in yellow, IRAC ch3 in blue, IRAC ch4 in red, observed FENIKS K_{blue}/K_{red} in purple, and planned FENIKS K_{blue}/K_{red} in pink.

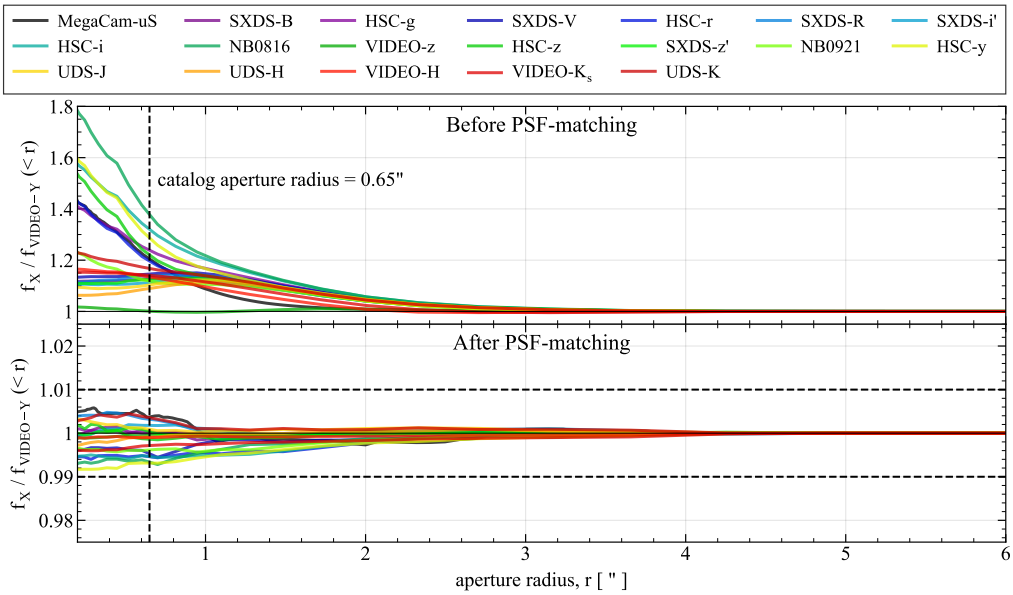

Figure 2.3: Flux ratios (f_X / $f_{VIDEO-Y}$) of the median growth curves of the stars in the individual bands, X, and VIDEO-Y before (*top panel*) and after PSF matching (*bottom panel*), except for MegaCam-uS. The PSF-matching for MegaCam-uS was treated differently in which UDS-K was PSF-matched to MegaCam-uS (see §2.3.3), therefore, we plot (f_{UDS-K} / $f_{MegaCam-uS}$). These median growth curves are derived using the same cleaned star stamps utilized in the PSF matching. The vertical dashed line in both panels represents the catalog aperture radius (color aperture) of 0.65 arcsec (aperture diameter of 1.3 arcsec). The median growth curves of the PSF-matched bands agree within 1% at all radii, as shown in the bottom panel.

2.3.2 K-band Detection

The UDS DR11 K-band, with a total 5-sigma depth of 25.3 AB, was used as the source detection band. We used `SExtractor` (Bertin & Arnouts 1996) to perform the detection using the sky-subtracted UDS K-band image along with the associated weight map. A series of tests were performed to fine-tune the critical `SExtractor` detection parameters, namely `DETECT_MINAREA`, `DETECT_THRESH`, and `DEBLEND_MINCONT`.

The best combination of detection parameters was identified by inspecting the resulting segmentation map and the detected sources. Firstly, we ensured that the object centroids were located inside the pixel with peak flux value, and that resolved close pairs were not detected as single objects. Furthermore, occasionally a faint few pixels (usually noise) get assigned the same object ID as the brighter source in the vicinity, even if those faint pixels are not physically connected to the brighter source (quite common in `SExtractor`) i.e., a big 'island' (brighter source) and a small 'island' (a faint few pixels) get assigned the same object ID. This is regardless of whether the brighter source is a point source or not. We optimized the detection parameters by minimizing such occurrences, however, there still remained many such stray faint pixels assigned to the neighboring brighter sources. We set such pixels to a value of '0' in the segmentation map *a posteriori*, meaning that those pixels are not assigned to any object. Finally, we arrived at the best set of detection parameters ensuring that no real object was undetected. The final set of chosen parameters is shown in Table 2.2.

Table 2.2: Adopted SExtractor detection and photometry parameters

Parameter	Value
ANALYSIS_THRESH	0.9
BACK_FILTERSIZE	5
BACK_SIZE	64
BACK_TYPE	MANUAL
BACK_VALUE	0
CLEAN	Y
CLEAN_PARAM	Y
DETECT_MINAREA	7
DEBLEND_MINCONT	$1x10^{-7}$
DEBLEND_NTHRESH	32
DETECT_TYPE	CCD
FILTER	Y
FILTER_NAME	gauss_3.0_7x7.conv
MASK_TYPE	CORRECT
PHOT_APERTURES	2.235, 3.726, 4.844, 5.589, 7.452,11.177, 14.903, 18.629
PHOT_AUTOPARAMS	2.5, 3.5
PIXEL_SCALE	0.2684
SATUR_LEVEL	150000
STARNNW_NAME	default.nnw
THRESH_TYPE	RELATIVE
WEIGHT_TYPE	MAP_WEIGHT,NONE

2.3.3 Aperture photometry

To obtain aperture photometry, we first determined the optimal aperture diameter which would maximize the signal-to-noise ratio (S/N). To do so, similar to the extraction of growth curves before PSF-matching above using bright and unsaturated stars, we extracted the growth curves after PSF-matching using the same diameter of 32 pixels (8.6 arcsec) used for normalization of the circular apertures. Then, using the resulting median growth curves in each band, we calculated the corresponding S/N as a function of aperture radius, r, as follows:

$$S/N_X(r) = \frac{\tilde{g}_X(r)}{\sqrt{\tilde{g}_X(r) + 4\pi r^2}}, \tag{2.1}$$

where $\tilde{g}_X$ denotes the median PSF-matched growth curve of any band (denoted by X), and S/N$_X$ is the corresponding S/N curve. The median of the aperture diameters corresponding to the peak S/N in all bands - D=1.3 arcsec - was adopted as the optimal (color) aperture diameter to be used to derive all color fluxes f_X^{col} in each band X. We ran `SExtractor` in dual mode using the parameter values shown in Table 2.2 to detect sources and extract the photometry for the detected objects in the PSF-matched bands within the optimal color aperture diameter.

It should be noted that as mentioned in §2.3.1, we do not PSF-match MegaCam-uS to VIDEO-Y to obtain its optimal aperture photometry like the rest of the optical-NIR bands. Instead, following Quadri et al. (2007), we take advantage of the assumption that the (u - K) color within D=1.3 arcsec at VIDEO-Y's resolution, $(u-K)|^{@Y-res}_{D=1.3"}$, is the same as (u - K) color within D=1.5 arcsec at MegaCam-uS's resolution, $(u-K)|^{@u-res}_{D=1.5"}$. This allows for easy extraction of the optimal color aperture photometry for the

MegaCam-uS without the need to PSF-match it to VIDEO-Y. In Figure 2.3, we show the growth curve of the K-band PSF-matched to MegaCam-uS (black curve labelled 'MegaCam-uS'). As investigated by Quadri et al. (2007), this assumption works reliably for distant galaxies that our catalog is optimized for. This is expected because distant galaxies are barely resolved by ground-based imaging in the color aperture sizes considered here, hence any discrepancies in colors measured using different aperture sizes and/or different resolutions will be negligible (Quadri et al. 2007).

Furthermore, we measured the total flux for the objects in the K-band utilizing the `SExtractor` output FLUX_AUTO - the flux within a flexible elliptical aperture (Kron 1980). Although FLUX_AUTO is a good approximation of the total flux, it is still an underestimation of it. We therefore applied an auto-to-total correction using the PSF-matched K-band growth curve, $\tilde{g}_K$, as shown in Eq. 2.2, where r_{Kron} is the circularized Kron radius evaluated using the `SExtractor`'s KRON_RADIUS, A_IMAGE, and B_IMAGE, $f_{tot,K}$ is the total flux in the K-band, and $f_{AUTO,K}$ is the `SExtractor`'s FLUX_AUTO in K-band.

$$f_{tot,K} = f_{AUTO,K} \times \frac{1}{\tilde{g}_K(r_{Kron})} \tag{2.2}$$

This correction is smaller for large Kron radii, and in general, is on the order of $\sim 10\%$. We note that these corrections utilize $\tilde{g}_K$ which is derived from point sources, and therefore, might not be accurately applicable to extended faint sources outside of the AUTO aperture. However, any resulting bias will be sub-dominant in the presence of large flux uncertainties associated with extended faint sources (Quadri et al. 2007). We show the comparison of our total photometry with other catalogs in Appendix 2.7.

2.3.4 Errors on photometry

The errors on the aperture fluxes have contributions from background noise and Poisson noise. The error due to the background noise, σ_{rms}, depends on the extent of the correlation between the number of pixels, N, within the adopted apertures. If the pixels are completely uncorrelated, then $\sigma_{rms} \propto \sqrt{N}$. On the other hand, if the pixels are perfectly correlated, $\sigma_{rms} \propto N$. In reality, the correlation of the pixels is neither perfect nor absent but somewhere in between (Labbé et al. 2003; Quadri et al. 2007; Whitaker et al. 2011; Straatman et al. 2016; Papovich et al. 2016). On small scales, pixels are correlated due to interpolations performed on the local PSF. On large scales, there are various contributions, such as i) imperfect background subtraction, ii) the glow of the bright stars extending beyond the masked regions, and iii) faint, undetected sources.

Therefore, we determined the background noise for each band empirically by randomly placing 5000 empty apertures of diameter 1.3 arcsec (the color aperture diameter as described in §2.3.3), avoiding the pixels occupied by the sources in the segmentation map. Then, we fitted Gaussian functions to the resulting distributions and calculated the standard deviation, $\sigma_{rms,X}$ for each PSF-matched X-band. The 3-σ and 5-σ total depths listed in Table 2.1 are calculated using this method. Finally, we calculate the total error budget for fluxes in the color aperture by adding the contribution from background noise and Poisson noise in quadrature. The error due to background dominates for the faint sources whereas the Poisson error dominates for the bright sources (Whitaker et al. 2011).

We similarly calculated the errors on the auto fluxes (fluxes enclosed by the Kron radii) in the K-band. As the Kron radius differs for each object,

we model the background noise on auto fluxes, $\sigma_{auto,K}$ as a function of area, A, enclosed by the Kron radius as:

$$\sigma_{auto,K}(A) = \alpha A^{\beta/2}, \tag{2.3}$$

where α and β are constants that were empirically determined. We fit for α and β by randomly placing 5000 empty apertures for each area based on the range of Kron radii available. We then calculated the error on the K-band auto fluxes in a similar way to how the errors in color apertures were calculated above by adding the contributions from background noise and the Possion error in quadrature. All fluxes and associated errors were then further scaled to an AB ZP of 25, and corrected for Milky Way extinction as listed in Table 2.1.

2.3.5 IRAC photometry

The resolution of the IRAC images is significantly worse than the optical-NIR images. Therefore, we chose not to PSF match the optical-NIR bands to the IRAC PSF. That would have resulted in a serious blending of objects and an unnecessary loss of vital depth. However, we still needed to derive consistent photometry in color apertures from the IRAC images. To that end, we use an algorithm designed specifically to clean the blended photometry of the heavily confused images (Labbe et al. 2006).

The brighter sources impacting the blending in the IRAC images are also typically bright in the K-band. Therefore, the K-band image is used as the high-resolution template to deblend the IRAC photometry. In practice, we use the K-band image PSF-matched to the VIDEO-Y image and, hereafter, whenever the K-band image is mentioned, it is assumed that it is the

PSF-matched to the VIDEO-Y, if not otherwise explicitly indicated. The segmentation map of the K-band image is used to identify the position and extent of the sources in the IRAC images. Then using bright, unsaturated, and isolated stars in the K and IRAC images, convolution kernels are built to convolve the K-band image to the IRAC band resolution. A model is then fit to the K-band image convolved to the IRAC band resolution. Finally, for each object, instead of performing the photometry on the modeled image as the fit is never perfect, it is directly performed on the original IRAC image after the models of the neighbors have been subtracted out (see Wuyts 2007 for a more detailed description of this method). The adopted approach and the deblending code have been previously used in, e.g., Marchesini et al. 2010 (MUSYC), Whitaker et al. 2011 (NMBS), Skelton et al. (2014) (3D-HST), Straatman et al. 2016 (ZFOURGE), Muzzin et al. 2013 (UltraVISTA DR1).

We started by measuring the deblended IRAC fluxes in an aperture of diameter D=3 arcsec . The extracted deblended IRAC fluxes are then corrected to match the resolution of the template K-band (i.e., the resolution of the VIDEO-Y image). Finally, we obtained the IRAC fluxes in the color aperture, D=1.3 arcsec , at the resolution of VIDEO-Y image to conform with the rest of the catalog by assuming that the K-IRAC color in D=1.3 arcsec is the same as K-IRAC color in D=3 arcsec . Therefore, the resulting IRAC fluxes in the color aperture homogeneously matching the color aperture fluxes of the optical-NIR bands, $f_{D=1.3",IRAC|_{Y-res}}$, is calculated as follows:

$$f_{D=1.3",IRAC|_{Y-res}} = f_{D=3",IRAC|_{Y-res}} \times \frac{f_{D=1.3",K|_{Y-res}}}{f_{D=3",K|_{Y-res}}}, \tag{2.4}$$

where, $f_{D=3",IRAC|_{Y-res}}$ is the deblended IRAC flux within D=3 arcsec at

the resolution of the VIDEO-Y band, $f_{D=1.3",K|_{Y-res}}$ is the flux within D=1.3 arcsec in the K-band PSF-matched to the VIDEO-Y image, and $f_{D=3",K|_{Y-res}}$ is the flux within D=3 arcsec measured on the K-band image PSF-matched to the VIDEO-Y image after the removal of the neighboring sources. We applied the same method to obtain catalog fluxes for all of the IRAC bands (ch1-ch4).

2.3.6 Star classification

First, we flagged our catalog sources as stars using their positions in the observed $u-J$ versus $J-K$ color-color space (uJK diagram). Stars segregate themselves from extended sources in such color-color diagrams which target similar combinations of wavelengths (see e.g. Straatman et al. (2016) who use B, J, and K filters). We split our uJK diagram (top-left panel of Figure 2.4), containing only sources with K-band $S/N > 3$ and $K_{tot} < 25.75$, with dashed lines defined by Equation 2.5. These dashed lines were marked where we see a clear division between the stellar sequence at the bottom and extended sources at the top. We classify sources as stars if they fall below the dashed lines. We indicate the brightness in the K-band of so-selected stars through purple-yellow shading; the distribution of these stars is then shown in the bottom-right panel of Figure 2.4 with the same color shading.

$$(J-K) < 0.1727(u-J)-0.6546 \;\; and (J-K) < 0.0176(u-J)-0.1112 \quad (2.5)$$

Then, to complement the star selection described above at the bright end, we again considered sources with K-band $S/N > 3$ and plotted them in the log (`FLUX_RADIUS`) vs K_{tot} diagram. This method of selecting stars

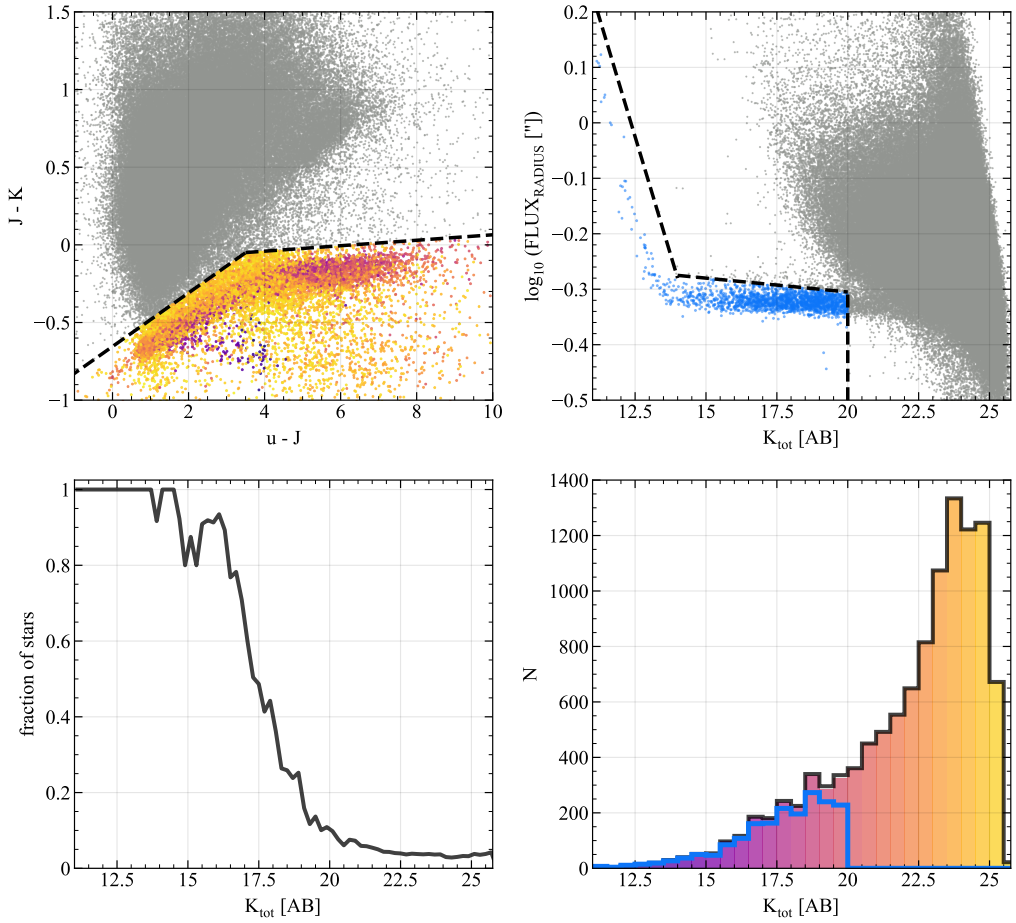

Figure 2.4: Star selection using the uJK, and the K_{tot} versus `SExtractor`'s `FLUX_RADIUS` diagrams. *Top-left:* Observed $u - J$ versus $J - K$ colors. The stars are selected as those objects falling below the plotted dashed lines (defined by Eq. 2.5), and are shown as points with a yellow-purple color gradient indicating the K-band total magnitudes, as reflected in the histogram in the bottom-right panel. All other objects are shown in gray. *Top-right:* Additional star selection at the bright end ($K_{tot} < 20$) using the K_{tot} versus `SExtractor`'s `FLUX_RADIUS` diagram. The objects falling below the dashed lines (defined by Eq. 2.6) are classified as stars and shown in blue. Their K_{tot} distribution is also shown in the bottom-right panel in blue. All other objects are again shown in gray. *Bottom-left:* The fraction of the stars in the catalog, selected using both the methods above (`star`=1 in the catalog), as a function of K_{tot} magnitude. The K_{tot} distribution of this black curve is indicated in the bottom-right panel. *Bottom-right:* The K_{tot} distribution of stars selected using the uJK diagram (in purple-yellow), the K_{tot} versus `SExtractor`'s `FLUX_RADIUS` diagrams (in blue, applicable only for $K_{tot} < 20$ AB), and the combined final selection of stars (in black; `star`= 1 in the catalog).

works well for bright sources, as it identifies point-like bright objects, i.e., most likely stars (Shipley et al. 2018) with a clear separation at $K_{tot} \lesssim 20$ AB as shown in the top-right panel of Figure 2.4. Here, the `FLUX_RADIUS` is the half-light radius provided by `SExtractor` i.e., the radius at which 50% of the `FLUX_AUTO` is enclosed. The stars are selected as falling below the dashed lines as defined by Eq. 2.6 and having $K_{tot} < 20$ AB. As can be seen, the stars form a clear sequence shown in blue; their K_{tot} distribution is further plotted in the bottom-right panel of Figure 2.4.

$$\log(\texttt{FLUX\_RADIUS}) < -0.17\,K_{tot} - 2.1 \;\; and \log(\texttt{FLUX\_RADIUS}) < -0.005\,K_{tot} - 0.205 \tag{2.6}$$

We finally combined the stars selected from both methods and flagged them as `star` = 1 in our photometric catalog, and plotted (in black) their fraction in the overall catalog as a function of K_{tot} and their K_{tot} distribution in the bottom-left and bottom-right panels of Figure 2.4, respectively. The two methods of selecting stars using the $u - J$ versus $J - K$ color-color diagram and the $\log(\texttt{FLUX\_RADIUS})$ vs K_{tot} diagram agree well where they overlap at $K_{tot} < 20$ AB. This can be seen by the fact that the black histogram (showing all stars; `star`=1) almost perfectly overlaps the boundary of the purple-yellow histogram (showing uJK-selected stars), meaning that the blue histogram (showing the $\log(\texttt{FLUX\_RADIUS})$ vs K_{tot}-selected stars) do not introduce too many additional stars except at $19 < K_{tot}$ [AB] < 20.

2.3.7 Masking bright stars and other artifacts

In general, the photometry of the galaxies can be contaminated if they lie close enough to artifacts created by the foreground bright Milky Way stars. Some of the artifacts that bright stars can introduce, depending on the telescope design, are diffraction spikes, blooming (leakage of the flux along the horizontal axis of the star), halos, ghosts, and others (see, e.g., Aihara et al. 2019, 2022 for a detailed explanation of such artifacts). The size and extent of these features are mostly predictable as they usually depend on the brightness of the stars. Therefore, we created masks based on brightness-based empirical relationships for such artifacts, and hand-drew masking regions when needed.

For UDS DR11 bands (J, H, and K), we derive empirical relationships of the sizes for circular patches to mask halos around bright stars and rectangular patches to mask diffraction spikes around bright stars. As such features only become visible for stars brighter than $\mathrm{K}_{tot} < 16$ AB, we do not apply these masks for dimmer stars. Moreover, in the UDS DR11 bands, stars also produce ghost patterns which are very faint, small, and roughly circular residual repeating patterns along the vertical and horizontal axes of the stars. For this, a mask was created for the DR8 data release by Quadri et al. (2008) which we adapted and extended to the footprint of the UDS DR11 data images. Some other artifacts, e.g., cosmic rays, random streaks generated by the detectors in the UDS DR11 images, or image borders with no data or severely contaminated data, do not behave predictably. For such artifacts, we drew the masking regions manually. It should also be noted that any area masked in the K-band image was also masked in all of the other bands as we only consider the photometry of the K-band detected and unmasked

objects. The objects detected in the K-band image and masked are denoted by the flag `det_contam` = 1 in our photometric catalog, and `det_contam` = 0 otherwise.

For the HSC bands (g, r, i, z, y, $NB0816$, and $NB0921$), we started from the list of bright objects (mag $<$ 18 AB in each band), identified for masking, provided with the third public data release - PDR3 (Aihara et al. 2022). However, we did not use the corresponding masks as they seem to be too aggressive at least within the tract 8523 applicable for our catalog, and in applying these masks, we would lose a lot of valuable data. Therefore, we self-determine the empirical relationships of brightness vs. size of the halo around bright objects, and brightness vs. length/width of the rectangles for horizontal blooming around bright objects. We masked circular patches around all bright objects with mag $<$ 18 AB for each band as in Aihara et al. (2022), and masked horizontal rectangular patches around objects brighter than 14.2 AB, where blooming becomes visible. Furthermore, only for the HSC-y band, do we see visible vertical diffraction spikes for objects brighter than 14.2 AB, which we masked using rectangular patches whose sizes were again determined empirically.

For VIDEO-Y and VIDEO-z, to mask the halos around bright objects, we utilized the brightness vs. radius relationships of HSC-y and HSC-z, respectively. As these VIDEO and HSC bands have similar wavelengths, the same objects will have similar intrinsic brightness between them. However, the PSF changes between the HSC bands and VIDEO bands so the brightness vs. radius relationships of HSC-y and HSC-z will not be applicable directly. We find that scaling them down by a factor of ~ 0.5 works well for VIDEO-Y and VIDEO-z. On the other hand, for VIDEO-H and VIDEO-K_s, we derived

their empirical relationships to mask halos around bright stars for objects brighter than the total magnitudes of 16 AB in both bands. Besides, artifacts such as diffraction spikes, ghosts, etc. within VIDEO-$zYHK_s$ bands do not appear predictable, and therefore, we mask such appearances manually.

For MegaCam$-uS$, we once again masked halos around bright stars using an empirically determined relationship for stars with a total magnitude less than 19 AB, and append them with hand-drawn masks of the bad regions that could not be easily predicted.

2.3.8 *K*-band Completeness

The completeness of our catalog is limited by the sensitivity in the K-band, the band used to perform the detection of sources, and the deblending capabilities of `SExtractor` with our chosen parameter settings. We estimated the detection completeness for point sources as a function of K-band magnitude by calculating the recovery fraction of injected simulated point sources into the K-band image. To simulate the point sources, we selected a bright, unsaturated, and isolated star randomly from our K-band image and scaled its flux up or down before injecting 2000 copies of it at random locations in the image. We repeated this for the magnitude range $19 < K_{tot} < 28$ with a step of 0.1 mag. Then, we ran `SExtractor` with the exact same settings as used for our catalog detection and as shown in Table 2.2, and calculated the recovery fraction as a function of K-band magnitude. We performed this test in two different ways: 1) not allowing the injected sources to fall on previously detected sources with the help of the segmentation map, to avoid blending of objects and maximize the recovery fraction (teal curve in

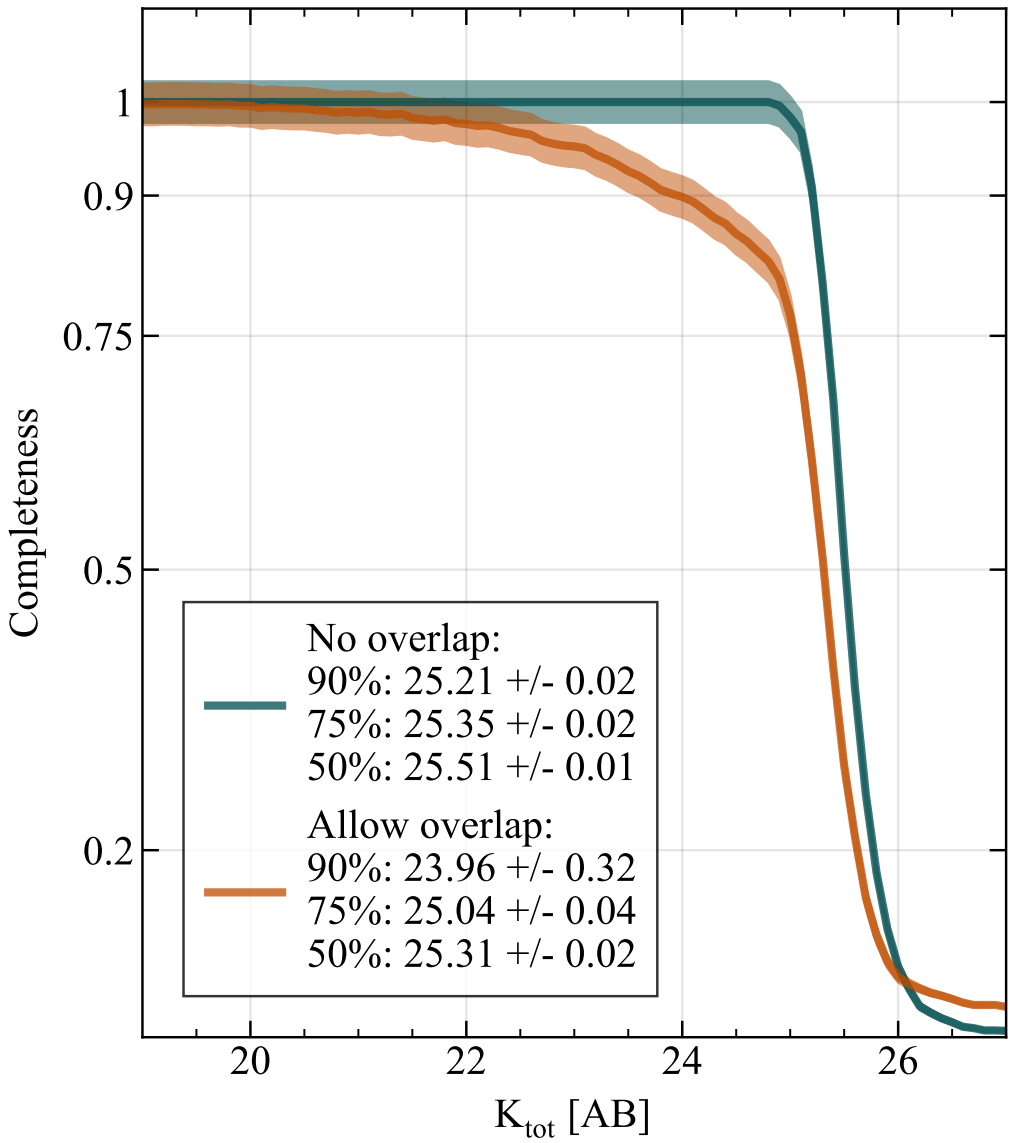

Figure 2.5: K-band detection completeness as a function of K-band total magnitude derived from the recovery of injected stars both allowing for overlap (blending) in orange and not allowing for overlap in teal. The K-band magnitudes corresponding to the 50%, 75%, and 90% detection completeness are also indicated for both cases.

Fig. 2.5); 2) allowing for the injected sources to fall on the image completely randomly, hence allowing for the overlap with previously detected sources to estimate a more realistic recovery fraction (dark orange curve in Fig. 2.5).

The 90% detection completeness of our catalog, when no overlap is allowed, is $K_{tot} = 25.21$ +/- 0.02 AB. On the other hand, a more conservative estimate for the detection completeness, when the overlap of sources is allowed, is found to be $K_{tot} = 23.96$ +/- 0.32 AB, as shown in Figure 5. These errors (and those shown in Fig. 5) are the Poisson errors from counting the recovered number of injected stars. Throughout the rest of the paper, we will adopt the upper 1-σ envelop of the completeness curves shown in Figure 5, specifically $K_{tot} = 25.2$ AB (no overlap allowed) and $K_{tot} = 24.3$ AB (allowing overlap) for the 90% detection completeness.

The completeness-corrected galaxy number counts as a function of K-band magnitude are shown in Figure 2.6, along with the number counts from other similar surveys, finding general good agreement. Here, we corrected our galaxy number counts using the detection completeness curve evaluated while allowing for the overlap of sources, excluding fainter magnitudes with less than $\sim 70\%$ completeness.

2.3.9 Catalog Format/Typical selection for galaxies

We present the various quantities associated with each parameter included in the photometric catalog in detail in Table 2.3. To select a robust and complete sample of galaxies with reliable photometry for an extragalactic study where completeness is important, we recommend at least applying the following selection cuts: `use_phot` $= 1$, $0.2 < z < 8.0$, $\log(M_*/M_\odot) > 8$,

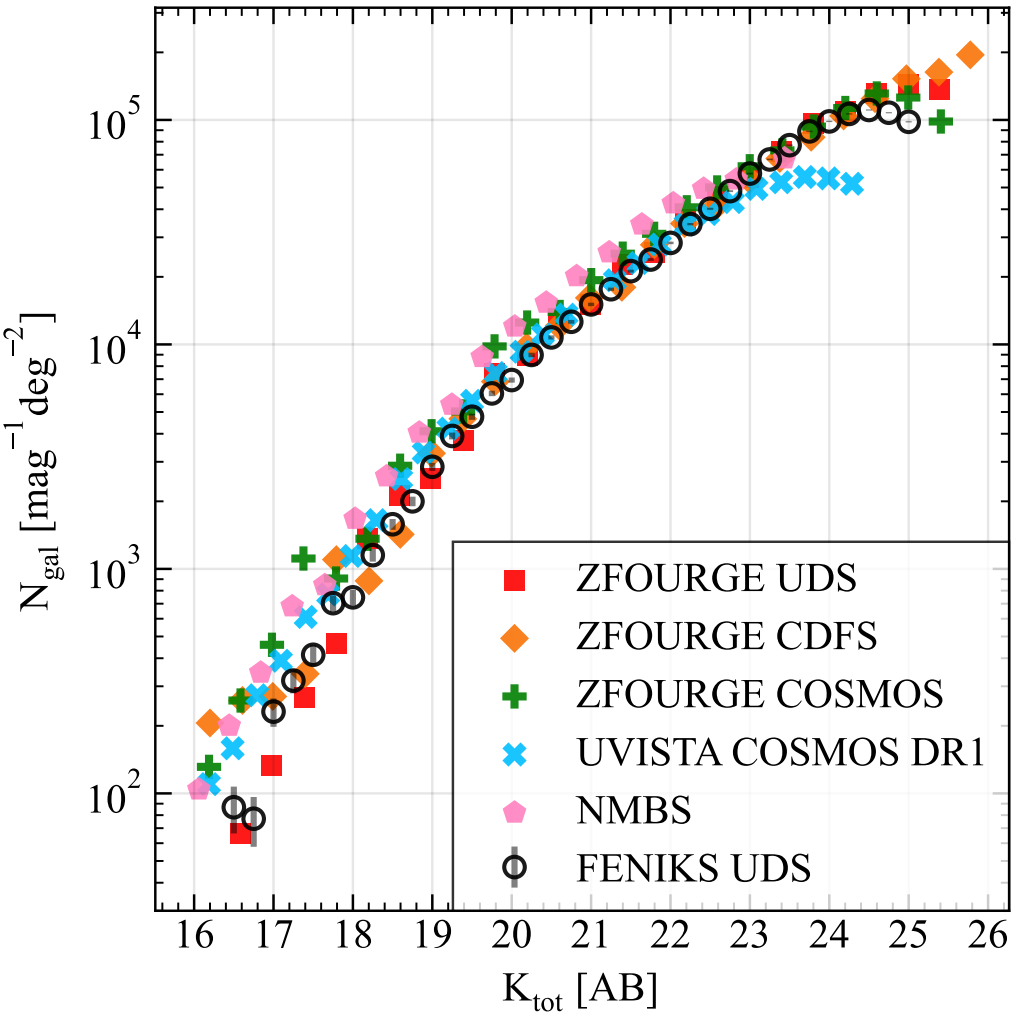

Figure 2.6: Completeness-corrected K-band number counts of galaxies in our catalog (black empty circles) plotted as a function of K-band total magnitude, and compared with similar measurements from the literature. For our catalog, we only use galaxies with reliable photometry (`use_phot` = 1). The adopted completeness correction is estimated using the detection completeness derived allowing for the overlap of sources (orange curve in Fig. 2.5). Furthermore, we omit plotting the number counts for magnitudes corresponding to less than $\sim 70\%$ completeness.

and $K_{tot} < 24.3$ (the 90% completeness level from the upper 1-sigma envelop of the completeness curve estimated when allowing for overlap - the upper envelope of the dark orange curve in Fig. 2.5). This results in 136, 235 sources. Using a more relaxed estimate of the completeness - $K_{tot} < 25.2$ (the 90% completeness level from the upper 1-sigma envelop of the completeness curve estimated when no overlap is allowed - the upper envelope of the teal curve in Fig. 2.5), while keeping all other cuts the same gives 188, 564 sources.

2.4 Redshifts and Stellar Population Parameters

2.4.1 Photometric Redshifts by `eazy-py`

We derived the photometric redshifts using the software `eazy-py` (Brammer 2021), a python implementation of `EAZY` (Brammer et al. 2008). It uses a linear combination of spectral energy distribution (SED) templates to fit the photometric data points for catalog sources. To fit our SEDs, we chose the Flexible Stellar Population Synthesis (`FSPS`) templates of Conroy et al. (2009) and Conroy & Gunn (2010), a set of 12 templates denoted as `tweak_fsps_QSF_12_v3` available among the template sets provided by `eazy-py`, designed to target the various types of stellar populations. However, any stellar population synthesis model is hard to perfectly calibrate (e.g. due to errors in stellar evolution tracks, missing spectral features, etc.). Moreover, the vast range of star formation histories and dust extinction in galaxies within a cosmic epoch, but also over time, are difficult to fully reproduce with any given set of templates (Brammer et al. 2008). Therefore, to mitigate such uncertainties, `eazy-py` utilizes a wavelength-dependent tem-

plate error function that captures these errors. Besides, `eazy-py` allows for the use of an apparent magnitude prior as a function of redshift, which we employed on our detection UDS-K band, that assigns a lower probability of finding a bright galaxy at a high redshift than at a low redshift.

Due to imperfect background subtraction, uncertainties in the PSF matching, and other effects, the zero-points (ZPs) of each PSF-matched band may need small corrections. Following Straatman et al. (2016) and Shipley et al. (2018), we determined these ZP corrections iteratively within `eazy-py`, using only sources with non-grism spectroscopic redshifts, i.e., sources with `use_phot`=1, `z_spec`>0, and `z_grism`=0, adjusting the individual band ZPs at each iteration to minimize the residuals between the observed photometry and the best-fit template photometry (see Straatman et al. 2016 for a detailed description on the ZP correction determination). The ZP corrections so determined, as multiplicative factors to scale the flux, are listed in Table 2.1.

To assess the quality of photometric redshifts, we compare them with the available spectroscopic redshifts from the following surveys: UDSz survey[6] (Bradshaw et al. 2013; McLure et al. 2013; Maltby et al. in prep.), 3D-HST survey (Momcheva et al. 2016; Brammer et al. 2012), SDSS-IV DR15 (Blanton et al. 2017), 6DF Galaxy Survey (Jones et al. 2009), VIPERS PDR2 (Scodeggio et al. 2018), PRIMUS (Coil et al. 2011; Cool et al. 2013), IMACS (Kelson et al. 2014), and others (Lacy et al. 2013; Le Fèvre et al. 2013; Garcet et al. 2007; Lacy et al. 2007; Stalin et al. 2010; Yamada et al. 2005; Ouchi et al. 2008; Ono et al. 2010; Sargsyan & Weedman 2009; Childress et al. 2017; Masters et al. 2019; McLure et al. 2018). All of these spectroscopic redshifts

[6] https://www.nottingham.ac.uk/astronomy/UDS/UDSz/

Table 2.3: Contents of the Photometric catalog

Column	Description	units/range
ID	Object identifier	-
X, Y	X and Y image coordinates. Pixel scale = 0.2684 arcsec/pixel	pixel
RA, DEC	Right Ascension and Declination	J2000 deg
fcol_X, ecol_X	color aperture flux and error. ZP = 25 AB	-
w_X	"weight": relative coverage	0 - 1
fauto_Kuds, eauto_Kuds	Flux and error in the UDS K-band within the Kron-like elliptical aperture. ZP = 25 AB	-
ftot_Kuds, etot_Kuds	Total flux and error in the UDS K-band. ZP = 25 AB	-
auto_to_tot_corr	correction factor to multiply fauto_Kuds or eauto_Kuds to get the total flux or error	> 1
fD3_Kuds, eD3_Kuds	Flux and error in the UDS K-band within the aperture diameter of 3 arcsec . ZP = 25 AB	-
fD5_Kuds, eD5_Kuds	Flux and error in the UDS K-band within the aperture diameter of 5 arcsec . ZP = 25 AB	-
Kronradius_Kuds	circularized Kron radius	arcsec
aper	Diameter of the color aperture	arcsec
aper_tot	Diameter of the AUTO FLUX aperture, the Kron-like aperture	arcsec
aper_to_tot_corr	correction factor to multiply fcol_X or ecol_X to get the total flux or error	> 1
r50_Kuds	Half-light radius in the UDS K-band given by SExtractor	arcsec
ellipticity_Kuds	Ellipticity in the UDS K-band as output by SExtractor	-
PA_Kuds	Position Angle (east of north) in the UDS K-band as output by SExtractor	deg
flags_Kuds	SExtractor flags in the UDS K-band as output by SExtractor	-
class_star	Star/Galaxy classifier of SExtractor	0 - 1
star	Binary flag for stars derived as described in §2.3.6; star=0 means that the object is classified as a star	0 or 1
gaia_star	Binary flag for stars classified in the Gaia catalog[a] with $> 95\%$ confidence; not included in the star flag above	0 or 1
opt_nir_maxflags	maximum of the FLAGS output by SExtractor in each band	
det_contam	Binary flag indicating contamination in the detection band (UDS-K)	0 or 1
use_phot	Binary flag indicating galaxies with reliable photometry having star = 0, det_contam = 0, S/N (calculated using color aperture flux/error in the UDS K-band) > 3, and nusefilt (parameter in the eazy-py output catalog quoting the number of filters with data) > 7	0 or 1
z_grism	Binary flag indicating if the corresponding z_spec came from grism spectroscopy	0 or 1
z_spec	Spectroscopic redshift	-

[a] https://www.cosmos.esa.int/web/gaia/dr3

are included in the catalog, labeled as `z_spec`. The redshifts based on grism spectroscopy from the 3D-HST survey (flag `z_best_s` = 2) additionally have their `z_grism` flag set to 1 (see Table 2.3). As shown in Figure 2.7, we see excellent agreement between z_{phot} and z_{spec}. The normalized median absolute deviation, $\sigma_{\rm NMAD} \sim 1.48$ x MAD of $\frac{z_{phot}-z_{spec}}{1+z_{spec}}$ ($\frac{\Delta z}{1+z_{spec}}$, hereafter), is typically quoted as a measure of the agreement between z_{phot} and z_{spec}. For $z_{spec} < 2$, we find $\sigma_{\rm NMAD} = 0.0112$, the median in $\frac{\Delta z}{1+z_{spec}}$ of 0.0014, and the catastrophic outlier percentage ($\frac{\Delta z}{1+z_{spec}} > 0.15$) to be $\sim 1.80\%$. At high redshift, for $z_{spec} > 2$, we find $\sigma_{\rm NMAD} = 0.0505$, the median in $\frac{\Delta z}{1+z_{spec}}$ of 0.0306, and the catastrophic outlier percentage ($\frac{\Delta z}{1+z_{spec}} > 0.15$) to be $\sim 9.38\%$.

2.4.2 Photometric redshift errors using close-pairs

Close-pairs method

The distribution of spectroscopic redshifts is dominated by low-redshift (median $z_{spec} \sim 0.87$) and bright sources (median $K_{tot} \sim 21.5$). Therefore, estimating the photometric redshift errors from the comparison with spectroscopic redshifts does not provide a reliable assessment of the quality of photometric redshifts at high redshifts.

To reliably estimate photometric redshift uncertainties over the full ranges of photometric redshifts, magnitudes, and stellar mass, we adopted the method involving galaxy pairs, first developed by Quadri & Williams (2010). This method relies on the fact that the galaxies that lie in close proximity on the sky (close pairs) actually have a high probability of being physically associated, meaning that they lie at the same redshift. This is simply because

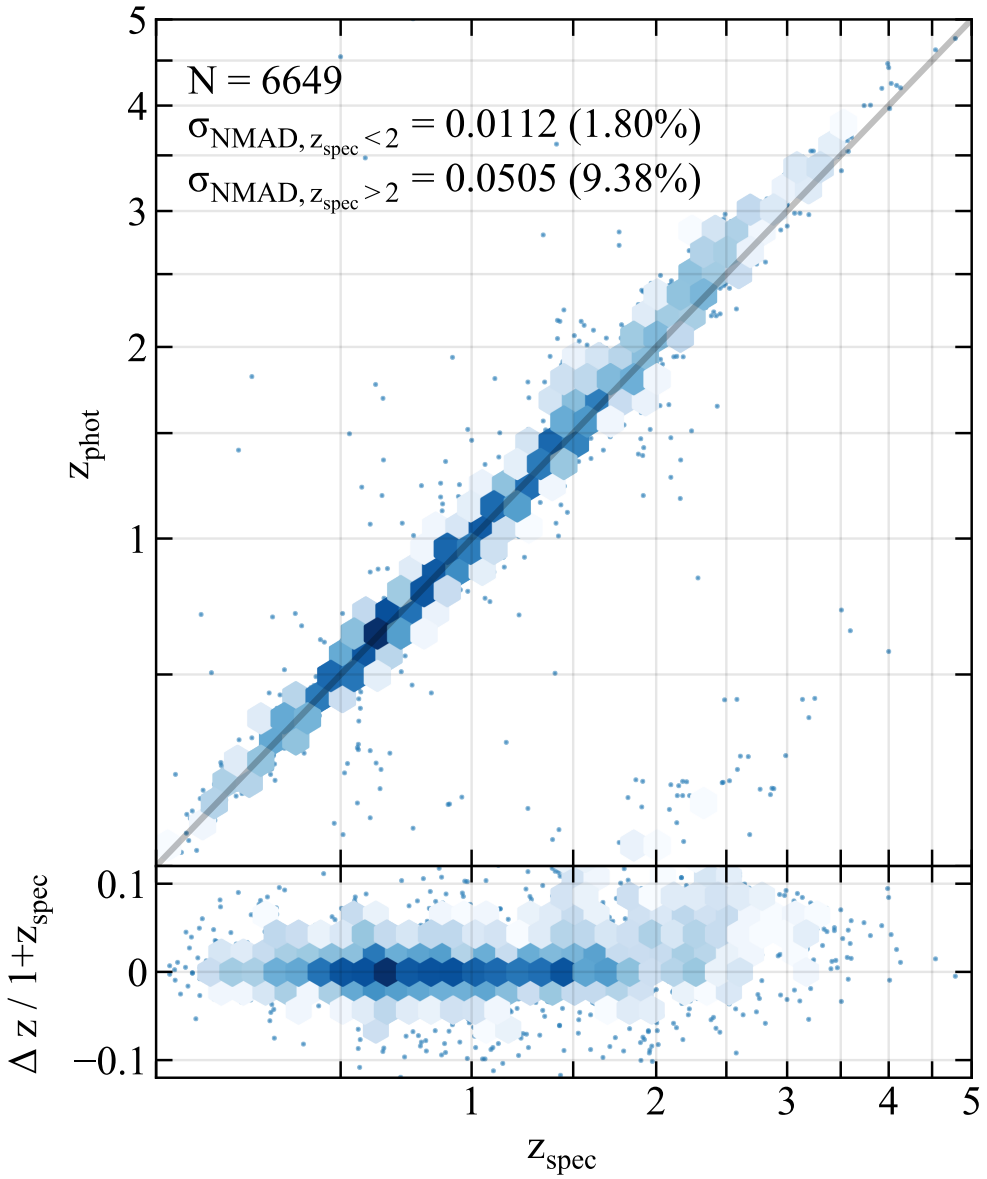

Figure 2.7: *Top:* `eazy-py` derived photometric redshifts (z_{phot}), vs spectroscopic redshifts (z_{spec}) of objects with `use_phot` = 1 and $K_{tot} < 25.2$ AB. The $\sigma_{\rm NMAD}$ of objects with $z_{spec} < 2$ and $z_{spec} > 2$ are indicated along with the corresponding outlier ($\frac{\Delta z}{1+z_{spec}} > 0.15$) percentages in parenthesis. *Bottom:* $\Delta z/(1+z_{spec})$ as a function of z_{spec}. In both panels, the darkness of the blue hexagon bins indicates the density of the objects. If the number of objects within the hexagon bins is less than 5, the individual objects are shown as blue scatter points.

galaxies are not distributed randomly in the universe, but they rather tend to cluster together. There are of course contributions to close pairs due to random projections on the sky, but they can be accounted for statistically.

In practice, we first determined all close pairs in our data catalog. If a true pair lying at the same redshift is considered, then the mean of their photometric redshifts gives a good estimate of their true redshift. Therefore, their error can then be computed by $\Delta z_{pair}/(1 + z_{mean})$. We estimated this error by fitting a Gaussian to the distribution of $\Delta z_{pair}/(1 + z_{mean})$ after subtracting the contribution from random pairs. To compute the contribution to this distribution from the random projections, we first randomized the positions in our data catalog leaving redshifts the same. Then, similar to finding pairs in the data catalog, we found the pairs in the random catalog. For example, to derive the photometric redshift errors of all the galaxies within a specified redshift interval and brighter than a given K-band magnitude limit, we selected all the galaxies that satisfy these criteria and their neighbors between 2.5 arcsec and 15 arcsec from the list of pairs evaluated earlier and then obtained their distribution of $\Delta z_{pair}/(1 + z_{mean})$. Then we applied the same selection and methodology to the random catalog, followed by removing pairs randomly until the number of galaxies in the data catalog becomes exactly equal to the number of galaxies in the random catalog.

Finally, we subtracted the resulting distribution of $\Delta z_{pair}/(1 + z_{mean})$ of the random pairs from the original distribution which contains both random and true pairs. We repeated this process a few times to obtain an average distribution of $\Delta z_{pair}/(1 + z_{mean})$ of the true pairs. The broadness of this distribution indicates the error on the photometric redshifts, with narrower distributions reflecting smaller errors and vice versa. We quantified

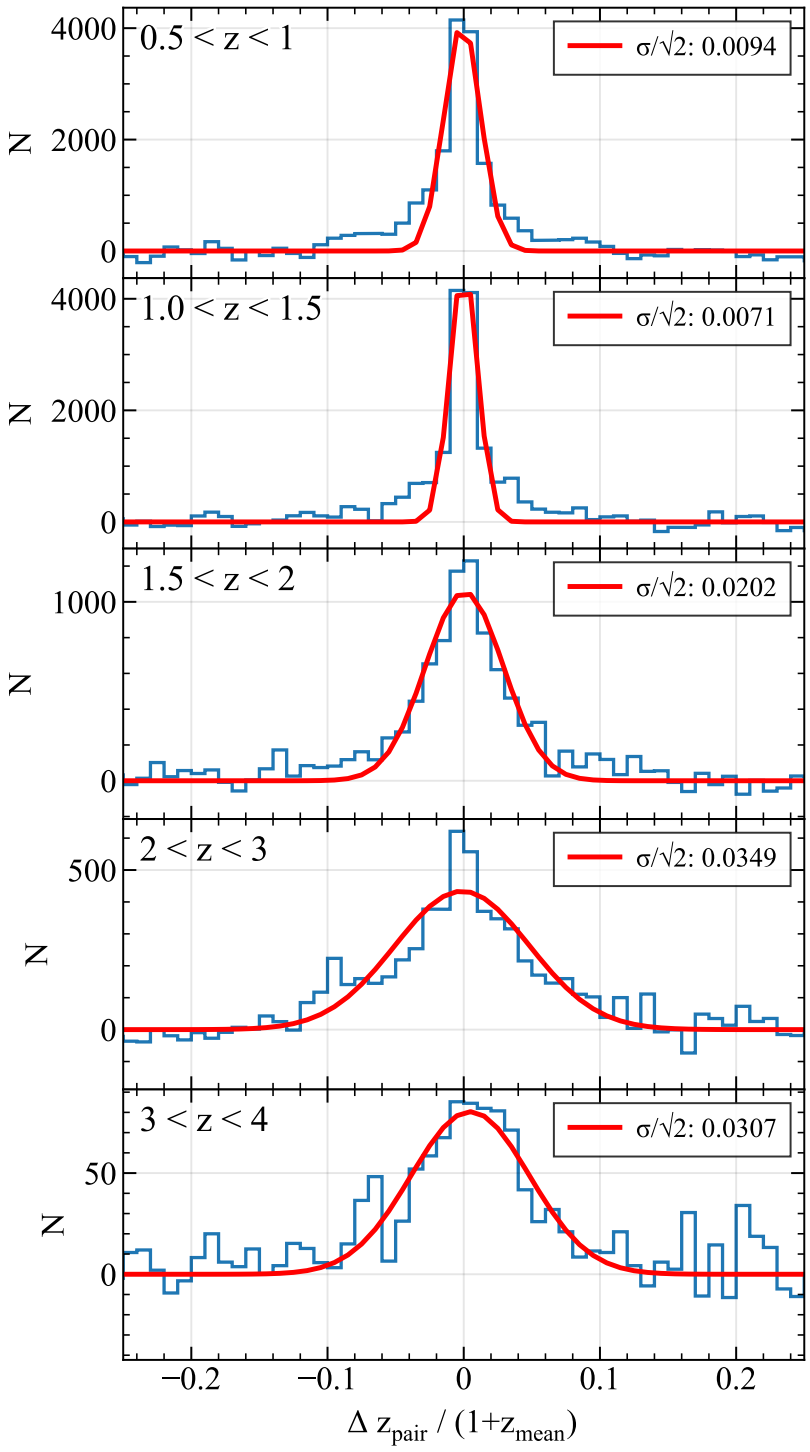

Figure 2.8: The estimated distributions of $\Delta z_{pair}/(1 + z_{mean})$ for the true close pairs (in blue) in bins of redshifts for galaxies with $K_{tot} < 23.3$ after removing contributions from the random close pairs. The fitted Gaussians are shown in red and their corresponding σ divided by $\sqrt{2}$ values, which are the errors on the photometric redshifts, are also indicated (see §2.4.2).

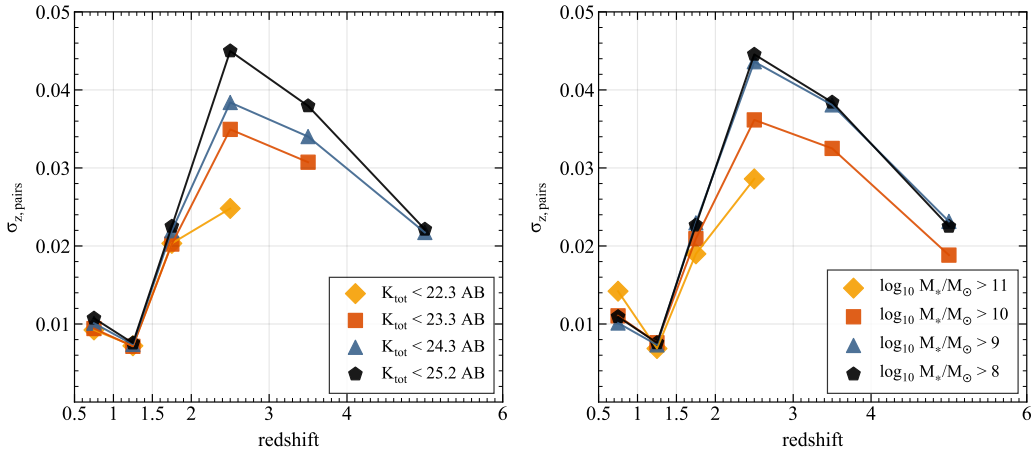

Figure 2.9: *Left:* Photometric redshift errors from the close-pair analysis as a function of redshift for different K-band magnitude limit samples. *Right:* Same as the left panel but for different stellar mass limit samples.

the photometric uncertainty by fitting a Gaussian to this distribution and determining its standard deviation, σ. Since this distribution comes from pairs, we divide the σ by $\sqrt{2}$ to obtain the average error per galaxy. We show an example of such distributions in Figure 2.8 in bins of redshift for a sample with $K_{tot} < 23.3$. In the redshift interval, $0.5 < z < 1.0$, the average uncertainty in photometric redshifts is less than 1%, which progressively increases to $\sim 3.3\%$ at $3 < z < 4$.

Validation of the close-pairs method

We also performed this analysis for different K-band magnitude limited samples and stellar mass threshold samples. The results of this analysis are shown in Figure 2.9, where $\sigma_{z,pairs}$ is the average photometric uncertainty per galaxy, equal to $\sigma/\sqrt{2}$ as described in §2.4.2 above, where σ is the standard deviation of the Gaussian fit on the corresponding $\Delta z_{pair}/(1 + z_{mean})$ distribution. The trends in the photometric redshift errors as a function of

redshift are generally similar for K-band bright (faint) samples of galaxies and high (low) stellar mass samples, with photometric redshift errors being larger for faint and low-mass galaxies. Moreover, for both samples, the errors in z_{phot} peak at $2 < z < 3$ across all the sub-samples, and it is found in the range of 2.5-4.5%. This peak can be explained by the absence of the NIR medium-band filters in our catalog which, if present, would allow for a better sampling of the SEDs around the rest-frame optical Balmer/4000Å breaks in $z \sim 2 - 3$ galaxies.

Figure 2.10 shows the comparison of the photometric redshift uncertainties estimated using the three different methods, i.e., the comparison of photometric redshifts with spectroscopic redshifts where available, the close-pair analysis, and the median errors as inferred from the `eazy-py` redshift posterior probabilities. Because the sample of galaxies with spectroscopic redshifts is limited to relatively bright ($K_{tot} < 23.3$) galaxies, the same K-band magnitude cut was adopted when deriving the photometric redshifts uncertainties with the close-pair analysis and the `eazy-py` redshift posterior probabilities. The typical uncertainty from the `eazy-py` redshift probability distributions was derived by calculating the median of $(z_{u68} - z_{l68})/[2(1 + z_{phot})]$ for all galaxies in a given redshift interval and with $K_{tot} < 23.3$, where z_{u68} and z_{l68} are the 16-th and 84-th percentiles of the `eazy-py` redshift probability distributions. As shown in Figure 2.10, the z_{phot} uncertainties derived from the comparison with the spectroscopic redshifts are generally larger than the uncertainties derived from the close-pair analysis, whereas the z_{phot} errors from the `eazy-py` redshift probability functions are the smallest, potentially indicating a tendency by `eazy-py` to underestimate photometric redshift errors. However, we note that in the redshift interval $z < 1.5$, where most of

the spectroscopic redshifts are available, the z_{phot} errors estimated from the three methods agree very well.

2.4.3 Redshift Distribution

The accuracy of photometric redshifts derived by `eazy-py` allows us to capture the large-scale structure of the Universe, as clearly visible in the top panel of Figure 2.11, with the sequence of dark-colored stripes showing high-density peaks, and the light-colored stripes showing density troughs. The jaggedness of the histograms in the bottom panel of Figure 2.11 also reflects the large-scale structure.

2.4.4 Stellar Population Parameters

FAST

We use the FAST code (Kriek et al. 2009) to derive stellar population parameters such as stellar mass, star formation rate (SFR), stellar age, and dust obscuration. We adopted the stellar population synthesis library of Bruzual & Charlot (2003) and assumed a delayed exponentially declining star-formation history (SFH $\propto t\ e^{-t/\tau}$), where t is the time since the beginning of star formation in the range $7 < \log(t[yr]) < t_{Univ}$, with t_{Univ} the age of the Universe at the redshift of the galaxy, and τ is the e-folding star-formation timescale in the range $7 < \log(\tau[yr]) < 10$. We assumed the Chabrier (2003) IMF, the Calzetti et al. (2000) dust attenuation law with the visual attenuation, A_V, allowed to vary in the range $0 < A_V$ [AB mag]< 10, and solar stellar metallicity. We used the spectroscopic redshifts when avail-

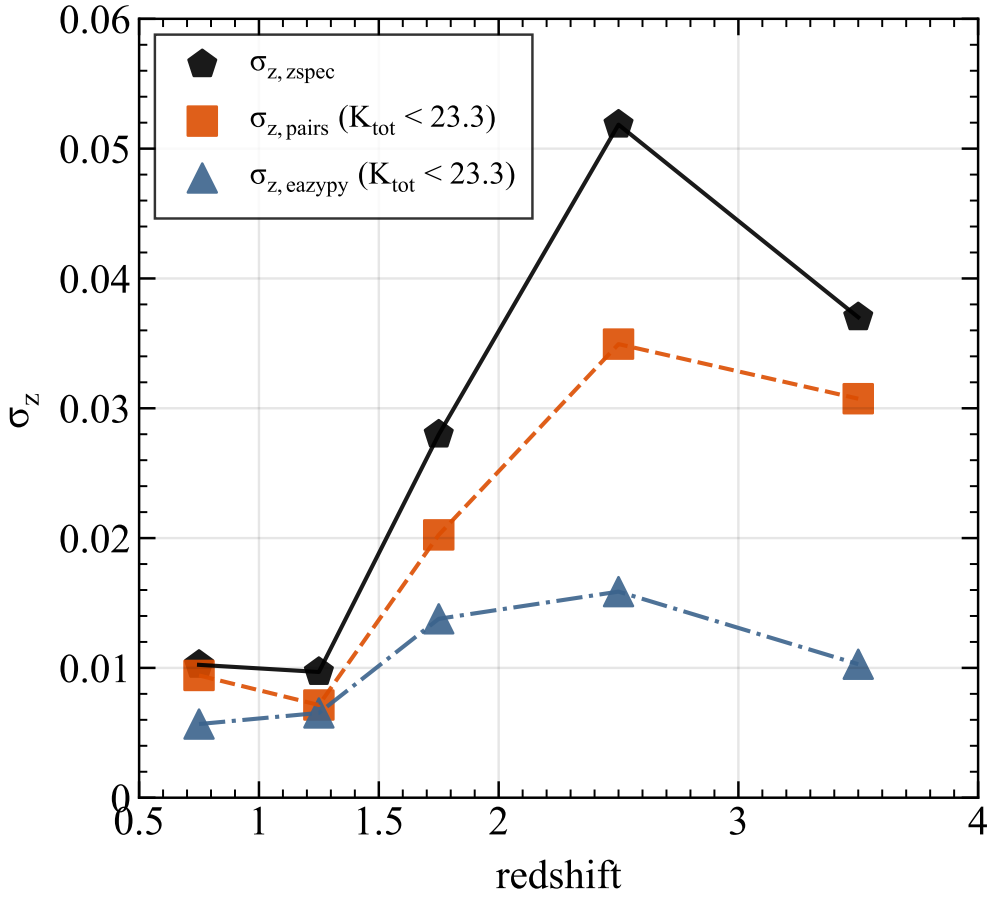

Figure 2.10: Comparison of the z_{phot} errors estimated using three different methods. The errors derived from the comparison to the spectroscopic redshifts are shown in black; the errors derived from the close-pair analysis in orange; and the median errors from the `eazy-py` redshift posterior probabilities in blue. The vertical grid lines correspond to the boundaries of redshift bins within which the close-pair errors and the `eazy-py` errors were calculated. To match the K-band magnitude limit of the sample with spectroscopic redshifts, the z_{phot} errors using the close-pair analysis and from `eazy-py` were derived using only galaxies with $K_{tot} < 23.3$.

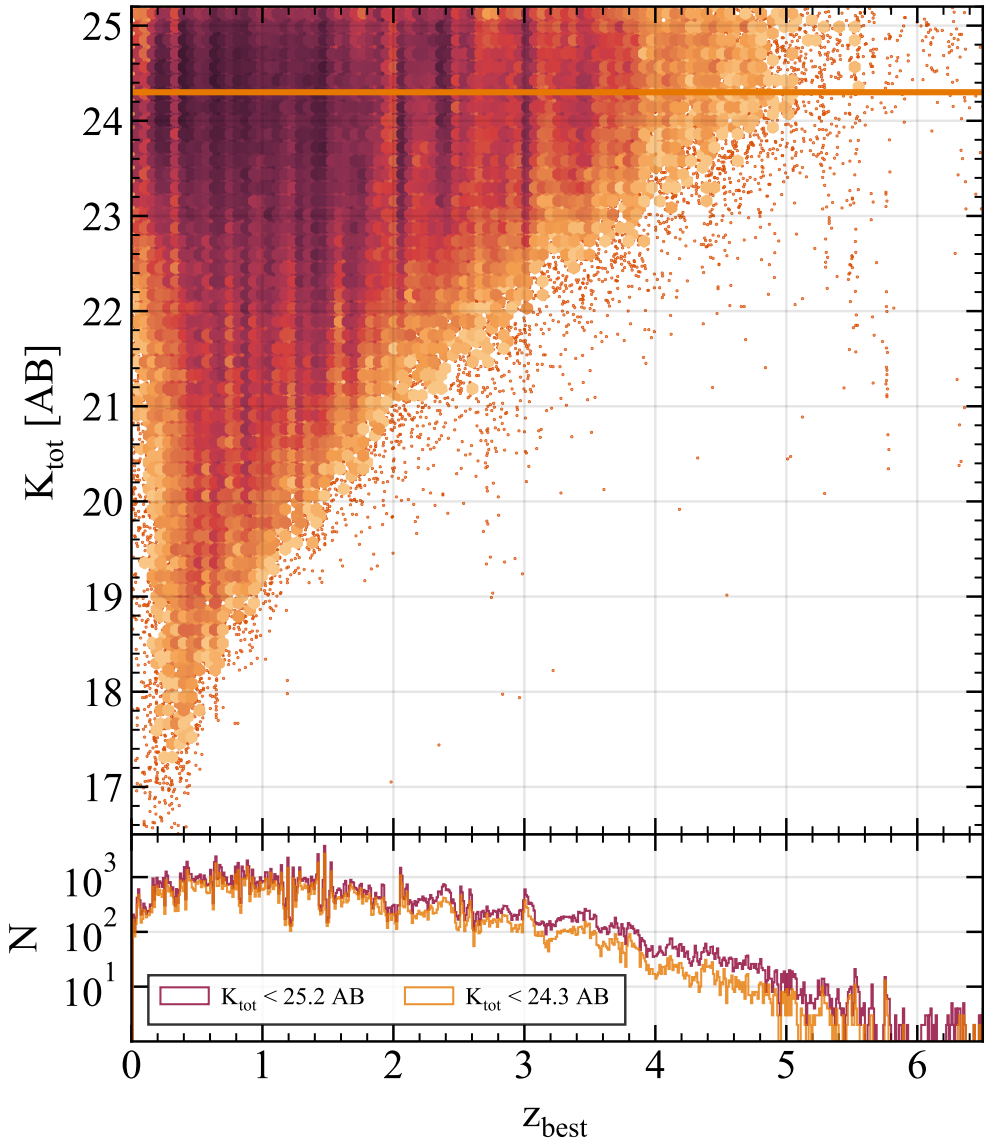

Figure 2.11: *Top:* Hexagonal density plot showing the K-band magnitude vs. the best redshift (z_{spec} if available, otherwise z_{phot}). Galaxies are selected with `use_phot` $= 1$ and $K_{tot} < 25.2$. For reference, the horizontal light orange line at $K_{tot} = 24.3$ indicates the adopted 90% completeness limit when overlapping of sources is allowed. Individual scatter points are shown where the hexagons contain less than 5 galaxies. *Bottom:* The redshift distribution of galaxies adopting the two different limits in K-band magnitude.

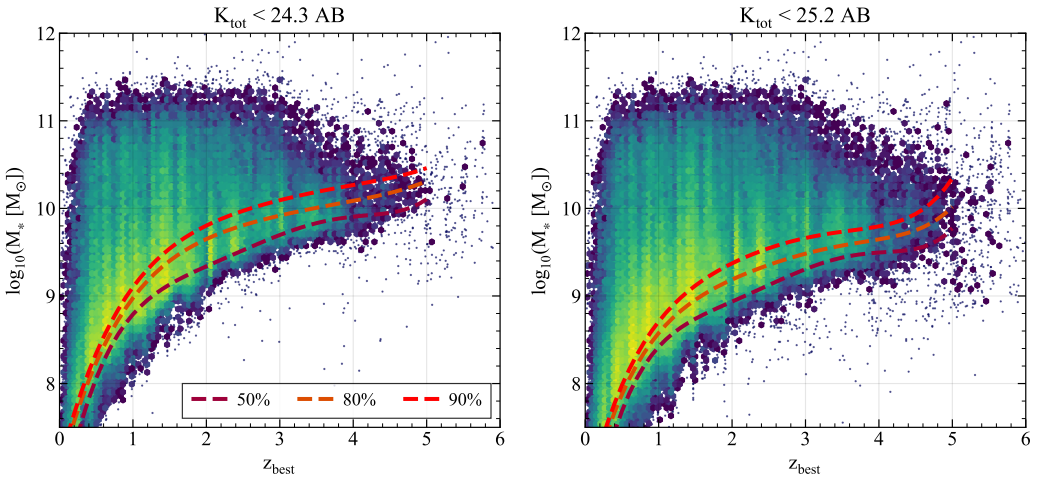

Figure 2.12: Density maps showing the FAST-derived stellar masses as a function of the best available redshift, z_{best} (z_{spec} if available, otherwise z_{phot}) for the two K-band 90% completeness limits. Bright yellow indicates high density which turns into green as the density decreases and eventually into dark blue showing the lowest densities. The left panel corresponds to the adopted 90% K-band completeness limit of 24.3 AB derived allowing for source overlapping; the right panel corresponds to the adopted 90% K-band completeness limit of 25.2 AB derived using the no overlap method (see §2.3.8 for the description on the two methods). The 50%, 80%, and 90% stellar mass completeness curves as a function of redshift are also plotted as dashed brown, orange, and red curves, respectively.

able, otherwise, we adopted the photometric redshifts derived using `eazy-py` (see §2.4.1).

In Figure 2.12, the stellar mass of the galaxies is plotted vs the redshift, along with the 50%, 80%, and 90% stellar mass completeness curves as a function of redshift. The stellar mass completeness was derived using the method described in Pozzetti et al. (2010) for the two different K-band survey limits (90% detection completeness in the K-band; see §2.3.8. As can be seen from Figure 2.12, our catalog is complete at the 50%(90%) level down to $\log(M_\star/M_\odot)$=10(10.4) at $z = 5$.

Dense Basis

We also derived the stellar population properties using Dense Basis (Iyer et al. 2019), an SED fitting code that makes use of non-parametric SFHs, which differs from codes like FAST that use only a singular function to fit the SFHs (see §2.4.4). Dense Basis is designed to extract smooth SFHs with no functional form (non-parametric SFHs) utilizing the Gaussian Process (GP; Iyer et al. (2019)). Singular parameterizations of SFH are unable to capture different stellar populations that reside in galaxies, and therefore, are known to give biased estimates of the derived quantities from the SED fitting (Mowla et al. 2022). They are known to underestimate the stellar masses (Iyer & Gawiser 2017; Leja et al. 2019; Lower et al. 2020) as can be seen in the left panel of Figure 2.13 which shows the comparison between the FAST-derived and Dense Basis-derived stellar masses for our catalog.

As with `eazy-py`, (see §2.4.1), in Dense Basis, we use the `FSPS` templates of Conroy et al. (2009) and Conroy & Gunn (2010), along with the nebular emission lines in `CLOUDY` (Ferland et al. 2017; Byler et al. 2017). Following Chworowsky et al. (2023), and based on the quality of available photometry, we constrained the shape of the non-parametric SFHs using three lookback times (t_{25}, t_{50}, t_{75}) at which 25%, 50%, and 75% of the stellar mass have been assembled, respectively. Furthermore, we assumed the Chabrier (2003) IMF, and the Calzetti et al. (2000) dust attenuation law just as we did in our FAST run described in §2.4.4. Although, unlike FAST, we imposed an exponential prior on the dust attenuation with the range $0 < A_V$[AB mag]< 4. Besides that, we assigned a flat prior on the specific star formation rate (sSFR=SFR/M$_\star$) with bounds $-14 <$ sSFR [yr^{-1}] < -7, and a flat prior in log-space on metallicity within $0.01 < Z/Z_\odot < 2.0$.

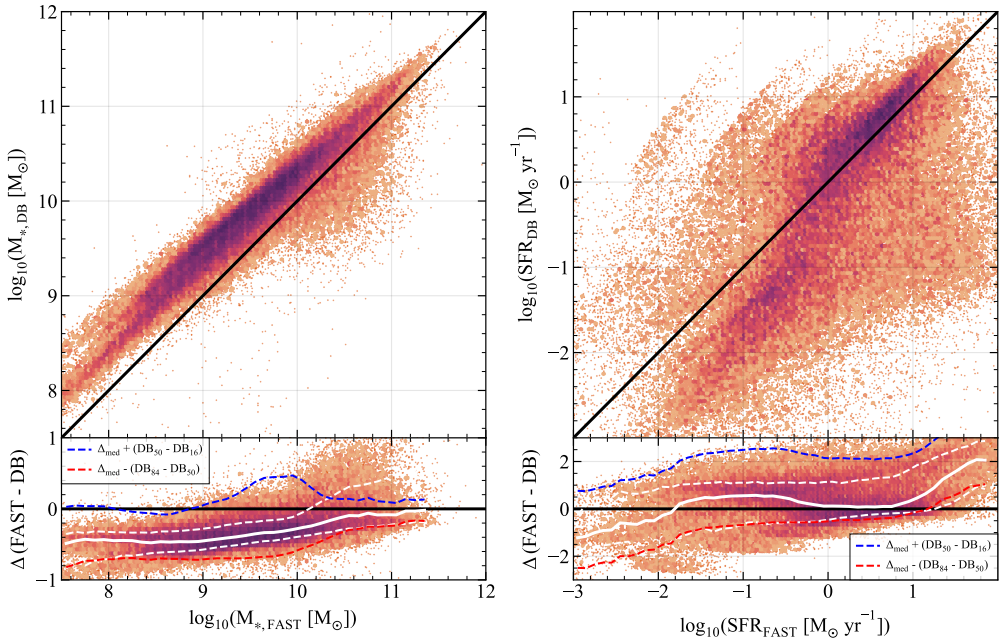

Figure 2.13: **Left:** Comparison of stellar masses derived using FAST and Dense Basis. *Top:* Density map showing the median posterior stellar masses from Dense Basis as a function of best-fit stellar masses from FAST. To guide the eye, the 1-1 line is shown in black. *Bottom:* The offset in stellar masses between the two codes as a function of FAST stellar masses. The white solid lines show the running median calculated in FAST stellar mass bins of 0.1 dex and the white dashed lines show the corresponding $16^{th}/84^{th}$ percentiles. We overplot red and blue dashed lines showing how the white solid line would shift if the 16^{th} (blue) or 84^{th}(red) percentile Dense Basis stellar masses were used instead of the 50^{th}. This highlights how the uncertainties in the Dense Basis stellar masses compare to the systematic offset between the FAST and Dense Basis stellar masses. The red dashed line is derived by subtracting the running median of the difference of the 84^{th} and 50^{th} percentiles of the Dense Basis stellar mass posteriors. Similarly, the blue dashed line is obtained by adding the running median to the difference of the 50^{th} and 16^{th} percentiles of the Dense Basis stellar mass posteriors. As mentioned in the text, codes like FAST that assume parametric SFHs tend to underestimate the stellar masses as seen here. **Right:** Comparison of SFRs from FAST and Dense Basis shown in the same way as in the left panel.

2.4.5 Selecting Quiescent Galaxies

Using UVJ diagrams

The rest-frame $V - J$ versus $U - V$ color-color diagram (UVJ diagram; hereafter) has been used to separate star-forming from quiescent galaxies, a method proposed by Williams et al. (2009) and reliably utilized by many (e.g., Whitaker et al. 2011; Muzzin et al. 2013; Straatman et al. 2016) at least out to $z \approx 3$ (see below for a more appropriate method for $z > 3$). We derived the rest-frame magnitudes in the U, V, and J filters using `eazy-py` by integrating the bandpass fluxes from the best-fitted templates of the redshifted SEDs. In Figure 2.14, we show the resulting UVJ diagrams, each row representing a color-coding according to a different quantity - number counts, stellar mass, sSFR, and dust attenuation A_V, respectively - while the different columns correspond to six redshift intervals from $z = 0.2$ to $z = 6$.

The bi-modality of star-forming and quiescent galaxies is evident in the first two panels of the first row of Figure 2.14 with the star-forming galaxies mostly sitting at the bluer, lower-left section of the UVJ diagram, and the quiescent galaxies occupying the redder, quiescent wedge at the top-left. The second row of Figure 2.14 shows the trend of increasing stellar masses from the bluest corner (lower-left) to the reddest corner (top-right), implying that more massive galaxies tend to be redder in color (either older or dustier). The ability of the of UVJ diagram to separate star-forming from quiescent galaxies (at least out to $z \sim 3$ is nicely shown by the third row of Figure 2.14, where the sSFR progressively decreases as galaxies move into the quiescent wedge of the UVJ diagram. Finally, the bottom row of Figure 2.14 shows how dust obscuration A_V increases from the bottom-left to the top-right

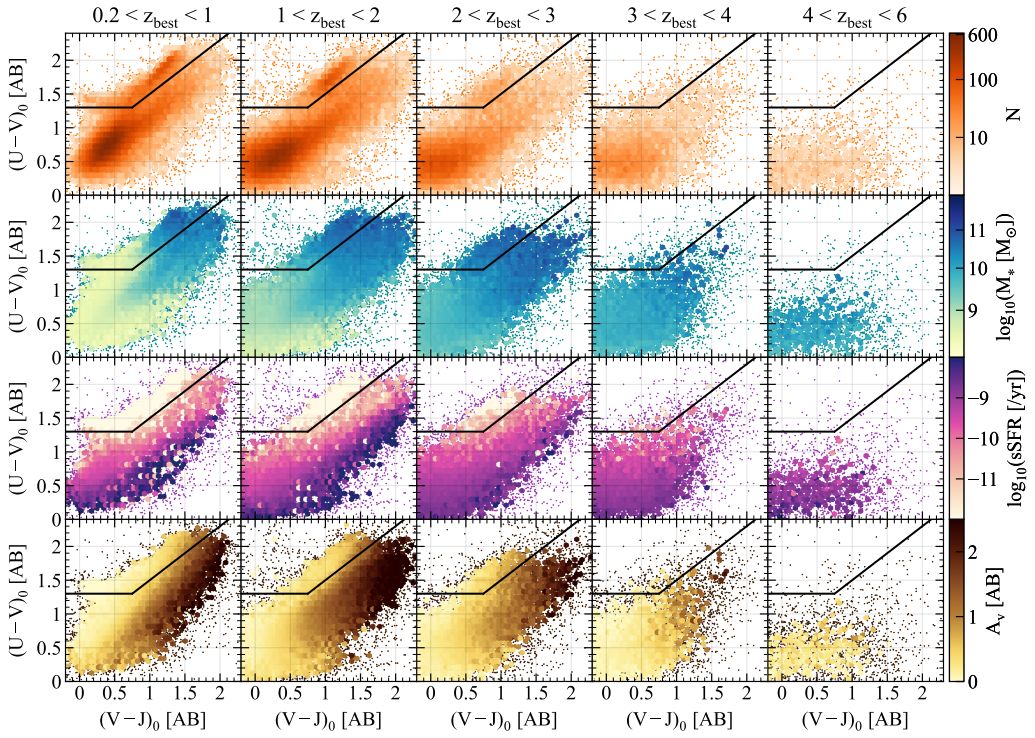

Figure 2.14: `eazy-py` derived rest-frame $V-J$ versus $U-V$ colors of galaxies in bins of redshifts (columns), color-coded by the number of objects in the first row, followed by the FAST-derived properties - the stellar mass, the sSFR, and the dust attenuation in the V-band (A_V) in the second, third and the fourth rows, respectively.

of the UVJ diagram, especially outside of the quiescent galaxies' wedge. We note that the properties of the population of galaxies in the reddest corner of the UVJ diagram are degenerate, as their red colors can be caused by their stellar population being predominantly old and quiescent, or dusty star-forming. This is also the reason why the photometric redshifts of this population of galaxies are the least reliable (e.g., Straatman et al. 2016).

Using (ugi_s) diagrams

Although the UVJ diagram has been quite successful in selecting quiescent galaxies at $z \lesssim 3$, it has been known to produce contaminated samples

at higher redshifts with up to ∼30% galaxies undergoing considerable star-formation activity as reported by recent spectroscopic surveys (e.g., Schreiber et al. 2018; Forrest et al. 2020). To remedy this, Antwi-Danso et al. (2023) recently introduced a rest-frame $g_s - i_s$ versus $u_s - g_s$ color-color diagram ($(ugi)_s$ diagram), a selection scheme for quiescent galaxies at high redshift ($z > 3$) based on the synthetic top-hat u_s, g_s, and i_s filters centered at 2900 Å 4500 Å and 7500 Å, respectively. We refer the reader to Antwi-Danso et al. (2023) for a detailed discussion on the design of the $(ugi)_s$ filters. Briefly, the u_s and g_s filters have narrower widths and are separated by 1600Å which means that they produce a stronger u_s - g_s color due to the Balmer break. Moreover, unlike the UVJ filters, the $(ugi)_s$ filters avoid wavelength regions with strong emission lines for $z \gtrsim 3$ galaxies, avoiding artificially boosting the photometry and contaminating the quiescent selection. Furthermore, the rest-frame J filter of the UVJ diagram is probed, at $z \gtrsim 4$, by *Spitzer*-IRAC ch3 and ch4 images, which are usually not available or shallow, leading to somewhat unreliable rest-frame $V - J$ colors. The antidote to this is the rest-frame i_s-band of the $(ugi)_s$ diagram, which overlaps with the shorter wavelengths IRAC ch1 and ch2 at $z \gtrsim 4$, where much deeper data than from ch3 and ch4 are usually available. Therefore, this leads to more accurate $g_s - i_s$ colors, helping to break the degeneracy between dusty star-forming and quiescent galaxies. Figure 2.15 shows the $(ugi)_s$ diagrams for our catalog similar to UVJ diagrams in Figure 2.14.

Example SEDs from the catalog

Based on the UVJ and $(ugi)_s$ diagrams, we select a handful of galaxies from our catalog and show their SEDs and modeling as output by `eazy-py` in

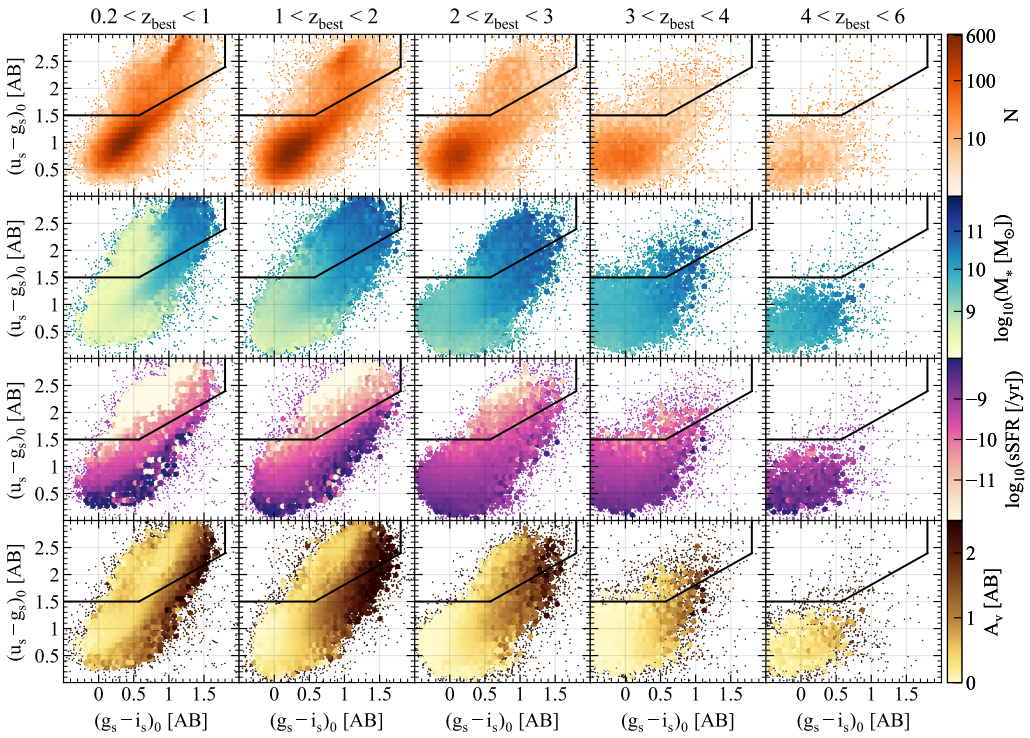

Figure 2.15: `eazy-py` derived rest-frame $g_s - i_s$ versus $u_s - g_s$ colors of galaxies in bins of redshifts (columns). The color-coding is the same as in Fig. 2.14

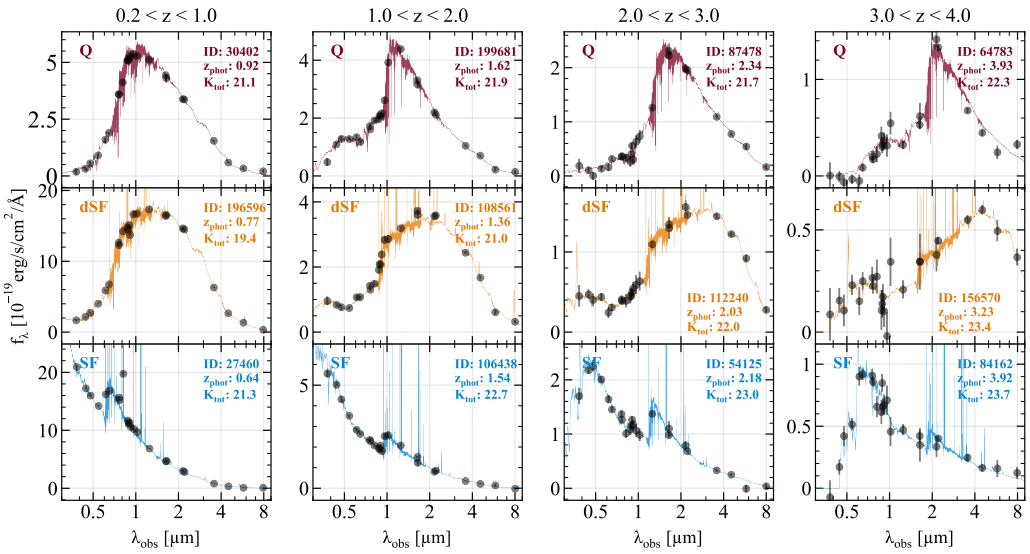

Figure 2.16: Example SED modeling of different types of galaxies at $0.2 < z < 4$ using `eazy-py`. The SED fits of quiescent galaxies (Q) are shown in dark red *(top row)*, dusty star-forming galaxies (dSF) in orange *(middle row)*, and the star-forming galaxies (SF) in blue *(bottom row)*, as identified using the UVJ and $(ugi)_s$ diagrams (described in §2.4.5). To identify dSF galaxies *(middle row)*, we used the wedge in the top-right corner of the UVJ diagrams as defined in Marsan et al. (2022). Finally, the photometric data points with error bars are shown in black.

Figure 2.16. The four redshift bins shown (four columns) correspond to the first four redshift bins (columns) shown in Figures 2.14 and 2.15, and each row showcases the SEDs of quiescent (Q), dusty star-forming (dSF), and star-forming (SF) galaxies, respectively. We classify galaxies as dSF if they were found on the right side of the line represented by $U - J = 3.2$ (Marsan et al. 2022), splitting the SF region of the UVJ diagrams.

2.5 Summary

We present a deep, PSF-matched multi-wavelength photometric catalog based on the final data release of the UDS survey (UDS DR11) covering an area

of $\sim$0.9 deg^2. We combine ancillary data from multiple surveys: CLAUDS, SXDS, HSC, VIDEO, DeepDrill, SERVS, and SpUDS to produce a high-quality PSF-matched multi-wavelength catalog with photometry in 24 bands. The additional photometry from the K-split medium bands (K_{blue} and K_{red}) of the FENIKS survey will soon be added to the catalog as soon as the data collection is completed. Many helpful flags indicating star classification, contaminated objects, etc. are also provided enabling the selection of robust stellar-mass complete samples of galaxies at $z < 6$.

In addition to the multi-wavelength photometric catalog, we also provide the catalogs of high-fidelity photometric redshifts derived using `eazy-py` (Brammer 2021) and a detailed assessment of the photometric redshift errors using three different methods: `eazy-py` posteriors, comparison with spectroscopic redshifts, and the z-pair analysis (see §2.4.2). Furthermore, the stellar population parameters were derived using FAST (Kriek et al. 2009) and Dense Basis (Iyer et al. 2019) (see §2.4.4 for their comparisons). All of the catalogs and the associated products such as the masks for bad regions (bright stars and other artifacts) in each band are provided publicly to enable research in the greater astronomy community.

2.6 Acknowledgments

We dedicate this paper to the memory of our dear friend and esteemed colleague Mario Nonino.

KZ acknowledges the John F. Burlingamme Graduate Fellowship in the Physics & Astronomy Department at Tufts University for support during the latter part of this project. KZ and DM acknowledge support from the

National Science Foundation under grant AST-2009442. Furthermore, CP and JAD acknowledge support from the National Science Foundation under grant AST-2009632

We extend our gratitude to the staff at the UK Infra-Red Telescope (UKIRT) for their tireless efforts in ensuring the success of the UDS project. We also wish to recognize and acknowledge the very significant cultural role and reverence that the summit of Mauna Kea has within the indigenous Hawaiian community. UKIDSS was undertaken on UKIRT while it was operated by the Joint Astronomy Centre on behalf of the Science and Technology Facilities Council of the UK, which also provided support for the Cambridge Astronomical Survey Unit (CASU) and the Edinburgh Wide Field Astronomy Unit (WFAU) that generated and served the wide field infrared public surveys from UKIRT.

Based on data products from observations made with ESO Telescopes at the La Silla Paranal Observatory as part of the VISTA Deep Extragalactic Observations (VIDEO) survey, under programme ID 179.AL2006 (PI: Jarvis).

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.

Based in part on data from the CFHT Large Area U-band Deep Survey (CLAUDS), which is a collaboration between astronomers from Canada, France, and China described in Sawicki et al. (2019, [MNRAS 489, 5202]). CLAUDS data products can be accessed from https://www.clauds.net. CLAUDS is based on observations obtained with MegaPrime/ MegaCam, a joint project of CFHT and CEA/DAPNIA, at the CFHT which is operated by the National Research Council (NRC) of Canada, the Institut National des Science de l'Univers of the Centre National de la Recherche Scientifique (CNRS) of France, and the University of Hawaii. CLAUDS uses data obtained in part through the Telescope Access Program (TAP), which has been funded by the National Astronomical Observatories, Chinese Academy of Sciences, and the Special Fund for Astronomy from the Ministry of Finance of China.

This paper makes use of software developed for the Large Synoptic Survey Telescope. We thank the LSST Project for making their code available as free software at http://dm.lsst.org

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 National Astronomical Observatory of Japan. Data analysis was in part carried out with the cooperation of Center for Computational Astrophysics (CfCA), National Astronomical Observatory of Japan. The Subaru Telescope is honored and grateful for the opportunity of observing the Universe from Maunakea, which has the cultural, historical and natural significance in Hawaii.

We acknowledge ESO Programme 180.A-0776 (PI: Almaini) which produced the publicly available spectroscopic redshifts of the UDSz survey.

This work is based on observations taken by the 3D-HST Treasury Program (HST-GO-12177 and HST-GO-12328) with the NASA/ESA Hubble Space Telescope, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS5-26555.

Funding for the Sloan Digital Sky Survey IV has been provided by the Alfred P. Sloan Foundation, the U.S. Department of Energy Office of Science, and the Participating Institutions. SDSS acknowledges support and resources from the Center for High-Performance Computing at the University of Utah. The SDSS website is www.sdss4.org.

SDSS is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS Collaboration including the Brazilian Participation Group, the Carnegie Institution for Science, Carnegie Mellon University, Center for Astrophysics — Harvard & Smithsonian (CfA), the Chilean Participation Group, the French Participation Group, Instituto de Astrofsica de Canarias, The Johns Hopkins University, Kavli Institute for the Physics and Mathematics of the Universe (IPMU) / University of Tokyo, the Korean Participation Group, Lawrence Berkeley National Laboratory, Leibniz Institut fr Astrophysik Potsdam (AIP), Max-Planck-Institut fr Astronomie (MPIA Heidelberg), Max-Planck-Institut fr Astrophysik (MPA Garching), Max-Planck-Institut fr Extraterrestrische Physik (MPE), National Astronomical Observatories of China, New Mexico State University, New York University, University of Notre Dame, Observatrio Nacional / MCTI, The Ohio State University, Pennsylvania State University, Shanghai Astronomical Observatory, United Kingdom Participation Group,

Universidad Nacional Autnoma de Mxico, University of Arizona, University of Colorado Boulder, University of Oxford, University of Portsmouth, University of Utah, University of Virginia, University of Washington, University of Wisconsin, Vanderbilt University, and Yale University.

Funding for PRIMUS is provided by NSF (AST-0607701, AST-0908246, AST-0908442, AST-0908354) and NASA (Spitzer-1356708, 08-ADP08-0019, NNX09AC95G).

`Astropy` (Astropy Collaboration et al. 2022), `Source Extractor` (Bertin & Arnouts 1996), `eazy-py` (Brammer 2021), `numpy` (Harris et al. 2020)

2.7 Appendix A: Comparison of FENIKS photometry with other catalogs

2.7.1 A1 Comparison with the ZFOURGE photometric catalog

Here we assess the quality of our photometry against the optical through mid-IR photometric catalog produced as a result of the FourStar Galaxy Evolution (ZFOURGE) survey (Straatman et al. 2016). The ZFOURGE survey was conducted using medium-band filters of the FourStar imager (Persson et al. 2013) ranging from $1 - 1.8\mu m$ and consisted of three 11 arcmin x 11 arcmin pointings including one in the UDS field. We compare the total, galactic extinction corrected, and ZP-corrected photometry from our catalog with the same from the ZFOURGE catalog in Fig. 2.17.

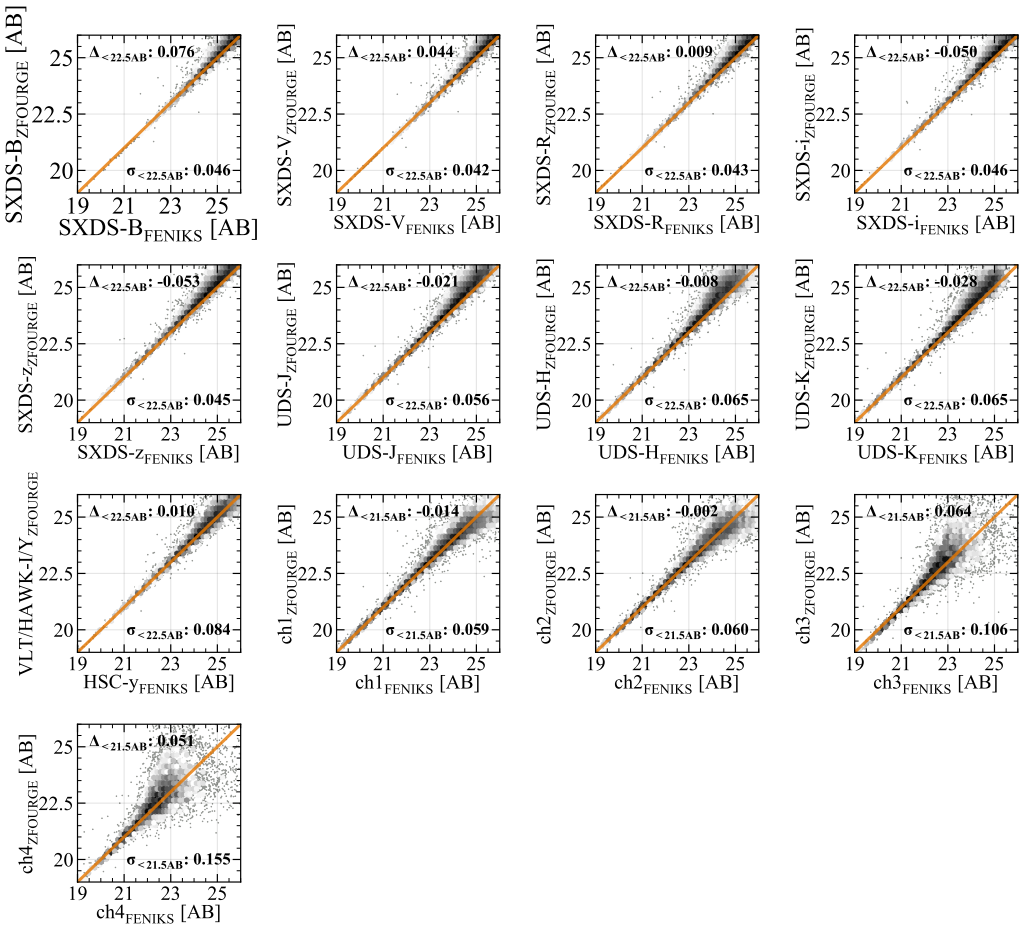

Figure 2.17: Comparison of total magnitudes in bands common between our catalog and ZFOURGE, except for the Y-band. The ZFOURGE's Y-band photometry comes from VLT/HAWK-I (Fontana et al. 2014). The median offsets for sources brighter than 22.5 AB mag for non-IRAC bands and 21.5 AB for IRAC bands (last four panels) are shown on the top-left of each panel. The corresponding 1-σ values are shown on the bottom-right of each panel.

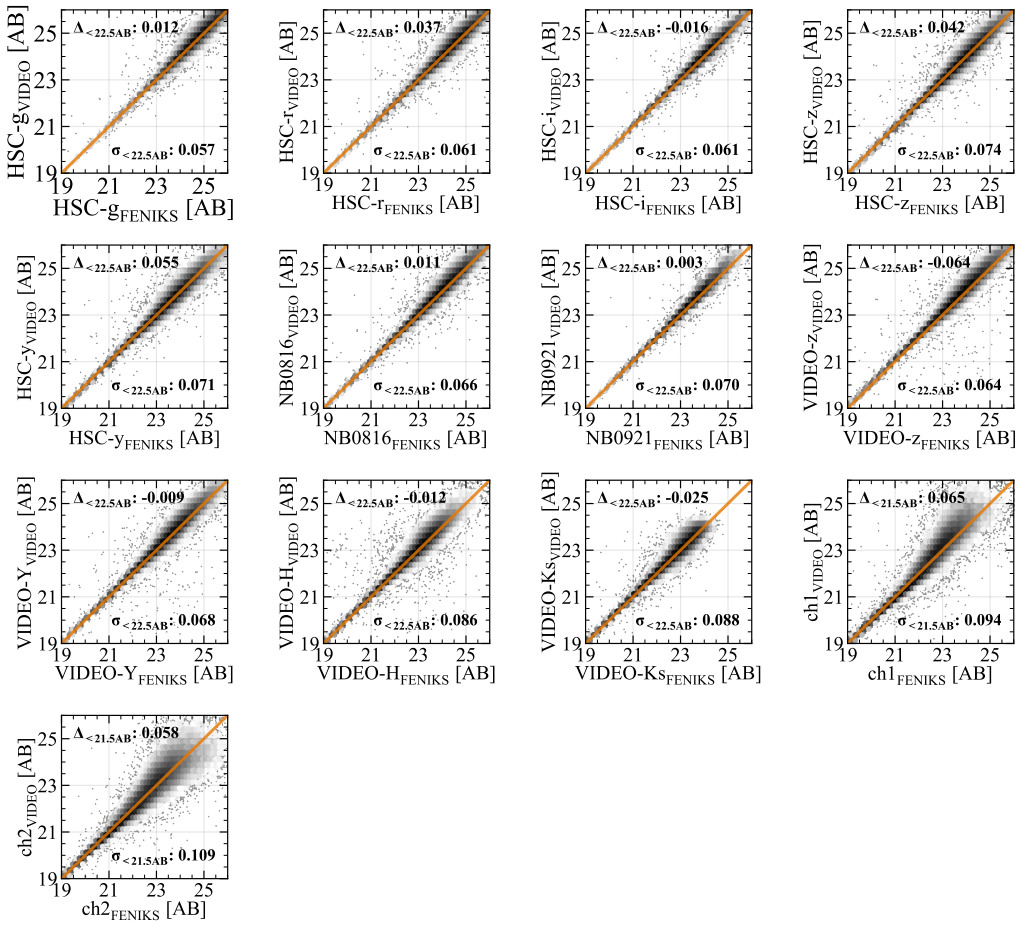

Figure 2.18: Comparison of total magnitudes in bands common between our catalog and the VIDEO catalog constructed for the MAGAZ3NE Survey (Forrest et al. 2020) shown as in Figure 2.17.

2.7.2 A2 Comparison with the VIDEO photometric catalog (MAGAZE3NE Survey)

Here we assess the quality of our photometry against the VIDEO catalogs constructed for the MAGAZ3NE survey (Forrest et al. 2020), by Annunziatella et al. (2024; in prep.), using the same technique as described in Muzzin et al. (2013). We compare the total, galactic extinction corrected, and ZP-corrected photometry from our catalog with theirs in Fig. 2.18.

CHAPTER 3

THE FENIKS SURVEY: STELLAR-HALO MASS RELATIONSHIP OF CENTRAL AND SATELLITE GALAXIES IN UDS AND COSMOS AT $0.2 < z < 4.5$

3.1 Introduction

In the standard cold-dark matter with dark energy (ΛCDM) model, structure in the Universe forms hierarchically. Small seeds of primordial over-densities in the structure of the Universe, as evidenced through the cosmic microwave background radiation (CMB), grow larger over time. Since dark matter accounts for $\sim 84\%$ of the matter in the Universe (Planck Collaboration et al. 2020), through gravity, it guides the bulk of this evolution. Regions over-dense in dark matter grow in size, accumulate enough mass to become self-gravitating halos, and decouple from the expansion of the Universe caused by dark energy. Within the potential wells of these halos, cold streams of gas funnel in, forming cold and dense clumps eventually reaching pressure, temperature, and density thresholds to trigger nuclear fusion - star formation and galaxy formation (White & Rees 1978; Fall & Efstathiou 1980; Springel et al. 2006).

Over time, halos grow in size and mass through merging and so do their resident galaxies. The evolution of dark matter halos and galaxies are intricately linked in this manner. To understand the influence of the aforementioned structure formation on galaxy formation and vice versa, we need to

Reprinted with permission from "The FENIKS Survey: Stellar-Halo Mass Relationship of Central and Satellite Galaxies in UDS and COSMOS at $0.2 < z < 4.5$" by Kumail Zaidi et al. 2024 doi: 10.48550/arXiv.2411.04256

understand how galaxy properties like stellar mass are dictated by their host dark matter halo properties like mass, the so-called galaxy-halo connection (see Wechsler & Tinker (2018) for a review) and vice versa. Our first hint into the galaxy-halo connection comes by comparing the abundance of halos − Halo mass functions (HMFs) from N-body simulations to the observed abundance of galaxies − Galaxy Stellar mass Functions (SMFs) (e.g. see Legrand et al. (2019) for a relatively recent example). Through this approach, it was realized early that galaxy formation is indeed a very inefficient process (e.g. Conroy & Wechsler (2009)). The SMFs have a much flatter slope at low mass than the HMFs, indicating inefficient star formation, and on the other hand, a much sharper decline at higher masses than the HMFs again indicating inefficient star formation.

Another facet of this inefficient star formation is the observationally constrained Stellar-Halo mass relationship (SHMR) itself (Mandelbaum et al. 2006; Moster et al. 2010). The resulting stellar-to-halo mass ratio (M_*/M_h) at a given epoch is a strong indicator of how efficient the halos with the current mass, M_h have been throughout their history in converting available baryons into stars ("in-situ" star formation), and in the accretion of stellar matter through mergers ("ex-situ" star formation). M_*/M_h can be easily converted into this SFE by dividing it with the available baryonic fraction of the Universe, $f_b = \Omega_b/\Omega_m = 0.158$ (Planck Collaboration et al. 2020), where Ω_b and Ω_m are the baryonic and total matter (baryons+dark matter) density fractions.

Numerous low and high redshift observational data based studies−Leauthaud et al. (2012) out to $z \sim 1$, Coupon et al. (2012) out to $z \sim 1.2$, Wake et al. (2011) at $1 < z < 2$, McCracken et al. (2015) out to $z \sim 2$, Hatfield et al.

(2016) at $0.5 < z < 1.7$, Cowley et al. (2018) at $1.5 \lesssim z \lesssim 3$, Legrand et al. (2019) out to $z \sim 5$, Shuntov et al. (2022) out to $z \sim 5$, Behroozi et al. (2013) out to $z \sim 8$, and Behroozi et al. (2019) out to $z \sim 10$, either directly show or support that SFE remains low at all redshifts - on the order of a few percent. It peaks at most at ~ 20 % at $M_h \sim 10^{12}$ $M_\odot$, a mass which evolves mildly with redshift. At lower and higher halo masses, SFE drops sharply due to various feedback processes that suppress star formation. At the low mass end, it is thought to be stellar winds that expel or heat the surrounding gas and hot bubbles created by supernovae explosions that limit star formation (Hopkins et al. 2012). At the high mass end, active galactic nuclei (AGN) feedback specifically in 'radio mode' (radiatively inefficient, low accretion rate) is thought to shut down star formation (Silk & Rees 1998; Croton et al. 2006; Kondapally et al. 2023; Scharré et al. 2024). This suppression in star formation at lower and higher halo masses relative to $\sim 10^{12}$ $M_\odot$ also becomes visible in the deviation of the slopes at low/high mass of observed SMFs from the cosmological simulation-derived HMFs, as mentioned above.

As the observationally constrained SHMR is a result of the amalgamation of physical processes at various mass scales, it provides strong priors for the cutting-edge semi-analytic models such as Dark SAGE (Stevens et al. 2016, 2023) and L-GALAXIES (Henriques et al. 2015, 2020), as well as suites of state-of-the-art hydrodynamical simulations such as IllustrisTNG (Weinberger et al. 2017; Pillepich et al. 2018) and the FIRE project (Wetzel et al. 2023; Hopkins et al. 2023). This unison of observations and simulations has been incredibly helpful in advancing our understanding of the galaxy-halo connection (Somerville & Primack 1999; Cole et al. 2000; Springel et al. 2005; Bower et al. 2006; Croton et al. 2006; De Lucia & Blaizot 2007; Guo

et al. 2011; Vogelsberger et al. 2014; Schaye et al. 2015; Pillepich et al. 2018). While the hydrodynamical simulations and semi-analytical models are incredibly helpful in informing the baseline understanding of the galaxy-halo connection, their conclusions are heavily dependent on assumptions related to gas dynamics, treatment of supernovae, and the implementation of the chosen *sub-grid* recipes beyond the resolution scale of the simulation, among others.

A more observation-dependent approach to constraining the relationship between assembly histories of dark matter halos and their resident galaxies is the so-called empirical modeling where physical priors come directly from observations. Several observational probes of galaxy formation such as star-formation rates, SMFs, quenched fractions, galaxy clustering, etc. are utilized in a self-consistent way to inform the creation of mock universes by populating halo merger trees (with galaxies) from dark matter-only simulations (Behroozi et al. 2013; Moster et al. 2018; Behroozi et al. 2019). Even though empirical models do not assume any models of galaxy formation, they adopt parameterizations to capture, e.g. star formation in halos based on halo properties and then further parameterize their redshift evolution, they too remain quite model-dependent. Moreover, the inter-observational systematics, resulting from combining large amounts of data from several surveys with differently calibrated data reduction pipelines, are bound to skew the results.

Perhaps the simplest approach to connecting galaxies to their host halos is to match their abundances – 'abundance matching' (Marinoni & Hudson 2002; Conroy & Wechsler 2009; Behroozi et al. 2010). In this method, the halos are usually rank-ordered by their mass and assigned the rank-ordered

galaxies by stellar mass (or luminosity) individually. The idea is simple; the more massive a galaxy, the more massive its host. However, just matching abundances ignores the galaxy-halo connection information hidden in the structure on all scales which can be measured via clustering. Clustering on small scales contains information on how the satellite galaxies are distributed within a halo and on large scales how both the central and satellites are distributed. Furthermore, since clustering is directly related to halo mass, it is an indirect probe on the halo mass, which combined with the information contained in SMFs, provides tighter constraints on the galaxy-halo connection (Behroozi et al. 2010; Coupon et al. 2015). Finally, if we keep the abundances of stellar mass and halo mass unchanged but tweak the scatter in stellar mass at fixed halo mass, we will assign the galaxies to halos differently resulting in a change in clustering (Wechsler & Tinker 2018). Therefore, measurements of clustering, beyond helping constrain the halo mass, allow constraining the scatter in stellar mass (at fixed halo mass) better.

The observations of both galaxy clustering and abundance can be leveraged to break degeneracies as mentioned above to get a robust estimate of SHMR within the framework of Halo occupation distribution (HOD) modeling. The HOD model simply populates halos of different masses with different numbers of galaxies, on average, and is split between contributions from central and satellite galaxies in the traditional HOD models (Zheng et al. 2005, 2007; Zehavi et al. 2005, 2011; Leauthaud et al. 2011; Zu & Mandelbaum 2015). Besides having centrals and satellites, galaxies can be further split into contributions from star-forming and passive galaxies as in the Tinker et al. (2013) and Zu & Mandelbaum (2016) HOD models.

A chosen HOD model in combination with an adopted HMF from simula-

tions and a few more ingredients (halo density profile, halo mass-concentration relation, etc.) can be used to model a power spectrum that would be produced by a distribution of galaxies on the sky, or in other words − galaxy clustering. The HOD parameters that produce this clustering and the abundance can then be fine-tuned by fitting against the observations of galaxy clustering and the highly constraining galaxy abundance (SMFs). A constrained HOD model then gives the SHMR and many other interesting quantities such as the fraction of galaxies that are satellites as opposed to centrals in a halo of a given mass. HOD model fitted at different redshifts allows us to trace the assembly of central and satellite galaxies and since it does not assume any galaxy formation physics but is purely observation-dependent in that regard, it gives us clues into unknown astrophysical phenomena that might be at play. This is especially true at high redshift where our understanding of overall galaxy formation remains quite extrapolated.

Besides being constrained by the observations of galaxy clustering and abundance, the HOD model can additionally incorporate constraints from galaxy-galaxy lensing observations. For instance, Leauthaud et al. (2012) combined all of the three probes: galaxy clustering, galaxy abundance, and galaxy-galaxy lensing to constrain the HOD, and thus SHMR out to z$\sim$1 using data in the COSMOS field (Scoville et al. 2007a), although, lensing observations prove difficult at z>1.

To set the stage for our paper, below we highlight a few recent contributions in the literature that constrain the SHMR and the corresponding M_h^{peak} through observations of clustering and abundance under the framework of HOD:

Ishikawa et al. (2017) measured the SHMR by fitting the Zheng et al.

(2005) HOD with clustering and abundance of u-, g-, and r- dropout galaxies at z $\sim$3, 4 and 5, respectively. They find that M_h^{peak} climbs up steadily from $10^{11.92\pm0.097}M_\odot$ at z $\sim$ 5 to $10^{12.25\pm0.053}M_\odot$ at z $\sim$ 3, opposite of the expectation from *downsizing* (Cowie et al. 1996) in which case the M_h^{peak} should decline with time. However, this was in agreement with the empirical modeling (Behroozi et al. 2013).

Using the Zheng et al. (2007) HOD, Cowley et al. (2018) probed the SHMR at $1.5 \lesssim z \lesssim 5$ leveraging the *Spitzer* IRAC 3.6 and 4.5μm data (SMUVS survey (Ashby et al. 2018) within the UltraVISTA ultra-deep stripes (McCracken et al. 2012) of the COSMOS field (Scoville et al. 2007a). They were the first to use a stellar mass-selected sample to constrain the M_h^{peak} at z $\sim$ 2.5 which was found to be $10^{12.5^{+0.10}_{-0.08}}M_\odot$, a mild increase from lower redshifts, consistent with *downsizing*.

Most recently, using the Leauthaud et al. (2011, hereafter L11) HOD, Shuntov et al. (2022) constrained the SHMR out to z $\sim$ 5 by utilizing a large homogeneously prepared catalog – COSMOS2020 (Weaver et al. 2022) covering an effective area of 1.27 deg^2 in a single field. For the first time with a stellar mass-selected sample, they extended the redshift frontier for M_h^{peak} out to z $\sim$ 5, showing a drop of $\sim$ 1 dex going from $10^{12.94^{+0.36}_{-0.34}}$ $M_\odot$ at z $\sim$ 5 to $10^{12.07^{+0.07}_{-0.07}}M_\odot$ at z $\sim$ 0.35, a strong *downsizing* (Cowie et al. 1996) signal indeed. However, their uncertainty on M_h^{peak} at z $\sim$ 5 spans a large range of $\sim$ 0.8 dex. This is primarily because the sample remains complete only for the massive galaxies ($M_* \gtrsim 10^{10}$ $M_\odot$) at high-z, thus providing no constraints on the low-mass end of the SHMR.

In this paper, we constrain the relationship between galaxy stellar mass and their host dark matter halo mass using HOD modeling. Out to z $\sim$ 5, this

study is the first to use a large total effective area of ~ 1.61 deg^2 evenly split between two fields: UDS ~ 0.79 deg^2 and COSMOS ~ 0.82 deg^2 which greatly reduces the effect of cosmic variance in measurements. The availability of deep optical to mid-IR photometry in both fields leads to robust redshifts and stellar population properties allowing accurate measurement of galaxy clustering and abundance. More importantly, the photometric catalogs in both fields are homogeneously prepared so they are free from systematic biases affecting the interpreted SHMR (Behroozi et al. 2013; Moster et al. 2018; Behroozi et al. 2019). Much of the previous work in observationally deriving SHMR is based on the single COSMOS field (e.g. Cowley et al. 2018; Legrand et al. 2019; Shuntov et al. 2022). We show that by adding the UDS field, the resulting SHMR changes drastically, indicating that the scatter in SHMR might be dependent on the environment.

Another novel aspect of this work is to jointly fit the galaxy clustering and abundance data in the entire redshift range of $0.2 < z < 4.5$ by imposing smooth continuity priors on the redshift evolution of the HOD parameters. Tying together the z-bins in this way allows us to constrain the evolution of SHMR with higher confidence, especially at high redshift $z \gtrsim 2$ where the sample is complete only for the most massive galaxies ($M_* \gtrsim 10^{10}$ $M_\odot$). This would not have been possible if, following the usual procedure in the literature (e.g. Shuntov et al. (2022)), we had solely relied on the weaker constraining power of the individual z-bins at high redshift. Finally, we trace the SFE of halos as they grow in mass along their merger trees in the TNG300-1-Dark simulation (Pillepich et al. 2018).

The paper is structured as follows: In Section 3.2, we present the photometric catalogs utilized in this study and the sample selection. Section 3.3

describes the methodology used to measure galaxy clustering and abundance (SMF) ultimately used to constrain the HOD models. The adopted HOD modeling along with our new approach of fitting individual z-bins connected through smooth continuity priors is detailed in Section 3.4. This is followed by our SHMR results presented in Section 3.5. In section 3.6, we discuss the intuition gained based on the results about the assembly of centrals and satellites and their host halos. We conclude with a summary in Section 3.7.

Throughout this paper, we adopt a flat ΛCDM cosmology with the following parameter values at z = 0 (denoted by the subscript '0'): Hubble constant, $H_0 = 70$ km s^{-1} Mpc^{-1}; matter density, $\Omega_{m,0} = 0.3$ with a baryon fraction of 15.8% (Planck Collaboration et al. 2020) giving baryon density, $\Omega_{b,0} = 0.0474$; dark energy density, $\Omega_{\Lambda,0} = 0.7$; cosmic microwave background (CMB) temperature, $T_{cmb,0} = 2.7255\ K$. The masses are always quoted with the implicit dependence on the dimensionless scaled Hubble constant, $h = 0.7$ (complying with the H_0 given above): $M_\odot h^{-1}$ for dark matter halo masses and $M_\odot h^{-2}$ for the stellar masses. Furthermore, in the derivation of stellar population properties, we assumed the Chabrier initial mass function (IMF; Chabrier (2003)) and the Calzetti dust attenuation law (Calzetti et al. 2000).

3.2 Data

3.2.1 COSMOS and UDS catalogs

For this study, we utilize two homogeneously prepared catalogs in the COSMOS (UVISTA (Muzzin et al. 2013) DR3, ultra-deep stripes) and UDS (Zaidi et al. 2024) fields. These catalogs are based on deep, multi-wavelength

Optical-NIR (IRAC) photometry in 50 and 24 filters for COSMOS and UDS, respectively.

The photometric redshifts are derived using the SED template fitting code `EAZY` (Brammer et al. 2008) for COSMOS and its python implementation `eazy-py` for UDS. The photo-z's agree with the available spectroscopic redshifts well. Within the redshift span probed in this study – $0.2 < z < 4.5$, for COSMOS, the normalized median absolute deviation, $\sigma_{\rm NMAD} = 0.00505$, and the catastrophic outlier percentage ($\frac{\Delta z}{1+z_{spec}} > 0.15$) is $\sim 1.60\%$. For the same redshift span, for UDS, $\sigma_{\rm NMAD} = 0.01242$, while the catastrophic outlier percentage is $\sim 2.38\%$. The details on the spectroscopic redshifts considered for these analyses are mentioned in Zaidi et al. (2024) for UDS and Muzzin et al. (2013) for COSMOS. As spectroscopic redshifts are dominated by bright objects, we also addressed this bias in Zaidi et al. (2024) for UDS by estimating the photo-z errors for the full catalog using the close-pair method (Quadri & Williams 2010). This method leverages the fact that galaxies that are close together on sky have a high likelihood of being at the same redshift, as galaxies tend to cluster together. Unavoidably, some contribution to close-pairs also come from random projections from different redshifts, however, since they are random, their contribution can be easily estimated and subtracted out to leave physically associated close-pairs. The dispersion in the photo-z distributions of these true close pairs then allow estimating the error on the photo-z's. Using this method on the UDS catalog in Zaidi et al. (2024) , we found the photo-z errors to be within $\sim 0.7-4.5$ % depending on redshift, stellar mass and K-band magnitude considered.

The stellar masses along with other stellar population parameters were derived using the Dense Basis code (Iyer & Gawiser 2017; Iyer et al. 2019)

which uses non-parametric star formation histories (SFHs) that are able to robustly capture different types of stellar populations giving unbiased estimates of stellar population properties as opposed to the codes implementing a singular functional form for SFH (Mowla et al. 2022). As shown in Zaidi et al. (2024), the `FAST` code (Kriek et al. 2009) implementing a singular functional form for star-formation history, a delayed-exponential, provides systematically lower stellar masses by ~ 0.4 dex at $M_* \lesssim 10^{10} M_\odot$. Therefore, we also report the effect of this stellar mass offset in §3.5.4. In both cases, we assumed the Chabrier (2003) IMF and the Calzetti et al. (2000) dust attenuation law.

3.2.2 Sample selection

To measure the clustering and the SMF reliably, we select galaxies with robust photometry meaning that they are i- identified as a galaxy (not a star), ii- do not fall on contaminated pixels which can be due to the proximity to nearby bright stars or other image artifacts, and are brighter than the 90% completeness limit in the detection band (see below for details). This selection was achieved using the masks generated for bright stars/artifacts/bad pixels and the following catalog cuts. For COSMOS, `use` $= 1$, `star` $= 0$, `contamination` $= 0$, `nan_contam` < 3, `K_flag` ≤ 3, $15 <$ total mag_{K_s} [AB] $<$ 24.5 (90% completeness), and similarly for UDS, `use_phot` $= 1$, `flag_Kuds` ≤ 3, total mag_K [AB] < 24.3 (90% completeness). Additionally for UDS, we also avoid noisy data found along the four sides of the image. After such exclusions, we are left with the effective areas of 2954.51 arcmin^2 (~ 0.82 deg^2) in COSMOS and 2828.25 arcmin^2 (~ 0.79 deg^2) in UDS which equals

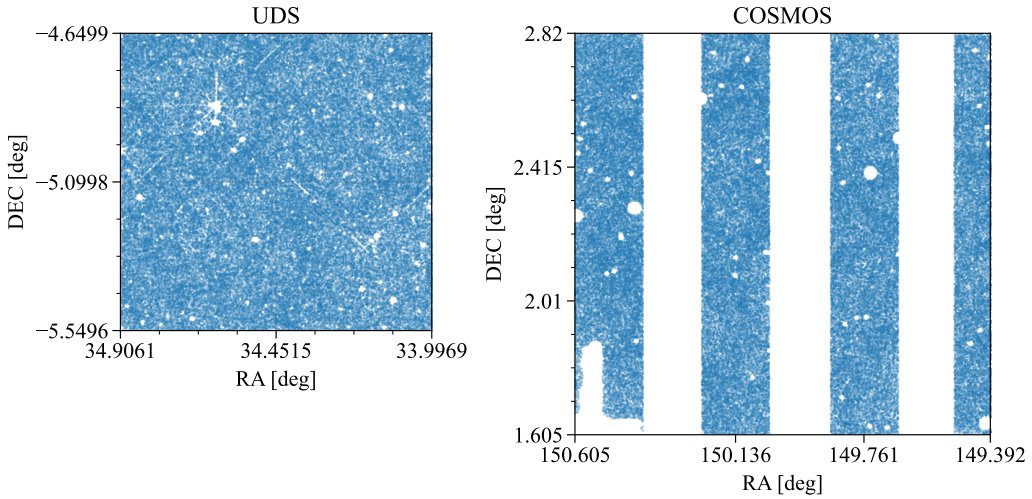

Figure 3.1: The final sample of galaxies selected in UDS (N = 143,549) and COSMOS (N = 166,826) is shown in blue points. The irregularly shaped white patches within both fields represent masked areas due to bright stars and other artifacts. In COSMOS, only ultra-deep stripes of the UltraVISTA DR3 survey are utilized, while the shallower deep stripes, visible in white, are excluded. The relative dimensions of the panels highlight the relative RA/DEC span of the two fields. The effective areas of UDS and COSMOS are $\sim$ 0.79 deg^2 and $\sim$ 0.82 deg^2, respectively. This amounts to a total effective area of $\sim$ 1.61 deg^2.

a combined effective area of ~ 1.61 deg^2.

We determined the completeness for both COSMOS and UDS by calculating the recovered detections of injected stars in the images as a function of magnitudes in the detection bands, K_s and K for COSMOS and UDS, respectively. Since we estimated the completeness by injecting stars which are point sources, instead of galaxies which are extended sources, it is valid to ponder if the true completeness should be worse than the one estimated here. However, as we are dealing with ground-based high-redshift catalog in which galaxies are not much more resolved than point sources with few low surface brightness galaxies, such a simplification should not lead to highly biased estimates of the completeness of galaxies. We injected the stars at random locations on the images allowing for overlap with other sources to mimic the original images where some overlap is natural. The 90% completeness in the detection K_s-band for COSMOS is 24.5 AB whereas the 90% completeness in the detection K-band for UDS is 24.3 AB.

Then, we used these estimates of 90% K_s/K completeness limits to derive the stellar mass completeness for both fields separately following the method outlined in Juneau et al. (2005) and Pozzetti et al. (2010). We scaled the stellar masses of galaxies as a function of redshift to the masses they would have if they were at the limiting magnitudes mentioned above. This gives a representative distribution of mass-to-light (M/L) ratios for our two fields at the limiting magnitudes. We chose the 80^{th} percentile of this distribution as a function of redshift as the stellar-mass limit (80% stellar mass completeness) above which most galaxies are observable.

Finally, we restrict our combined sample (UDS+COSMOS) in this study to the higher of the two 80% stellar mass completeness limits in UDS and

COSMOS which lie very close to each other as shown in Fig.3.2. Our final sample, divided into 9 z-bins within $0.2 < z < 4.5$ and further into stellar mass threshold (M_*^t) samples, is marked with black lines ("ladders") in Figure 3.2 and listed in Table 3.1. Note that the sub-samples are cumulative as opposed to binned in stellar mass, i.e. each stellar mass threshold sample contains all the galaxies within the stellar mass threshold samples below it.

Besides, the complete stellar mass indeed evolves within each z-bin, as visible by the non-horizontal completeness curve. Therefore, to fix our lowest stellar mass threshold within each z-bin, we chose it to be at the intersection of the completeness curve with the highest redshift end of any z-bin. These are the bottom right corners of the black "ladders" within each z-bin in Figure 3.2. By doing so, we ensure that any z-bin sample is at least 80% complete throughout the redshift thickness of its bin. Within each z-bin, the samples become more and more complete as the redshift decreases.

Now, we address the consequence of our choice of the lower limit in stellar mass completeness − 80% instead of 95%. By choosing an 80% completeness criteria, we miss some galaxies with the highest M_*/L_* ratios (old and/or dusty galaxies) lying close to the bottom-right corners of the lowest mass threshold sample within each z-bin. We determined that the average 95% completeness curve is higher in stellar mass than the average 80% completeness by the following amounts (evaluated at the right edge of three z-bins): $\sim$ 0.15 dex at $z = 0.5$, and $\sim$ 0.1 dex at $z = 2.25$ and $z = 4.5$. This means that our lowest stellar mass threshold samples are up to 20% incomplete predominantly in the bottom-right corners of the bottom $\sim 0.1-0.15$ dex windows in stellar mass, depending on the z-bin. Since this incompleteness only affects the bottom-right corner of the lowest stellar mass threshold samples, and

we jointly model the all the stellar mass threshold samples using the L11 HOD model, we expect the consequence of this choice on the final results to be subservient to the errors on photo-z's and stellar mass. Additionally, it should be noted that as the dynamical range in stellar mass of our sample shrinks with redshift due to worsening completeness, our HOD constraints are increasingly informed via only massive galaxies going from low to high redshift. We highlight the stellar masses directly probed by our analysis in the figures showing results where appropriate.

3.3 Measurements

3.3.1 Two-point angular correlation function

We compute the two-point angular correlation function, $w_{obs}(\theta)$ for the stellar mass threshold samples which is essentially the excess probability above Poisson of finding a galaxy-galaxy pair as a function of angular scale, θ. To do so, we use the popular Landy & Szalay (1993) estimator shown in Eq.3.1 within the python-based pair counter engine, `CORRFUNC`[1] (Sinha & Garrison 2020) implemented within the python utility wrapper, `pycorr`[2]. In Eq.3.1, DD, DR, and RR, indicate the number of data-data, data-random, and random-random pairs. All of these pairs are normalized by the total number of possible pairs, as shown in Eq.3.1 where, $n_{\rm D}$ and $n_{\rm R}$ indicate the number of points in the data and random catalogs, respectively.

[1] https://github.com/manodeep/Corrfunc
[2] https://github.com/cosmodesi/pycorr

Table 3.1: Sub-samples used in this study

z-bin	$\log_{10}(M_*^t[M_\odot])$	N_{COSMOS}	N_{UDS}	N_{total}
$0.2 < z < 0.5$	8.6	12386	9681	22067
	9.0	8359	6068	14427
	9.5	4988	3460	8448
	10.0	2963	2056	5019
	10.5	1524	1032	2556
$0.5 < z < 1.0$	9.3	25173	21628	46801
	9.6	19056	14648	33704
	10.1	11357	7481	18838
	10.6	5861	3625	9486
	11.0	2333	1212	3545
$1.0 < z < 1.25$	9.4	9119	9385	18504
	9.8	6234	5640	11874
	10.2	3905	2967	6872
	10.6	2306	1504	3810
	10.8	1644	951	2595
$1.25 < z < 1.5$	9.6	7598	11131	18729
	9.9	5493	7304	12797
	10.2	3746	4381	8127
	10.5	2388	2423	4811
	10.7	1663	1547	3210
$1.5 < z < 1.75$	9.6	4720	6341	11061
	10.1	2556	3448	6004
	10.5	1278	1622	2900
	10.7	843	999	1842
$1.75 < z < 2.25$	9.8	7067	7647	14714
	10.1	4942	4714	9656
	10.3	3606	3143	6749
	10.5	2506	1890	4396
$2.25 < z < 2.75$	9.9	4689	4664	9353
	10.2	2695	2685	5380
	10.5	1211	1227	2438
$2.75 < z < 3.5$	10.1	2984	4839	7823
	10.2	2348	3587	5935
	10.4	1217	1650	2867
$3.5 < z < 4.5$	10.3	589	1603	2192
	10.35	514	1280	1794

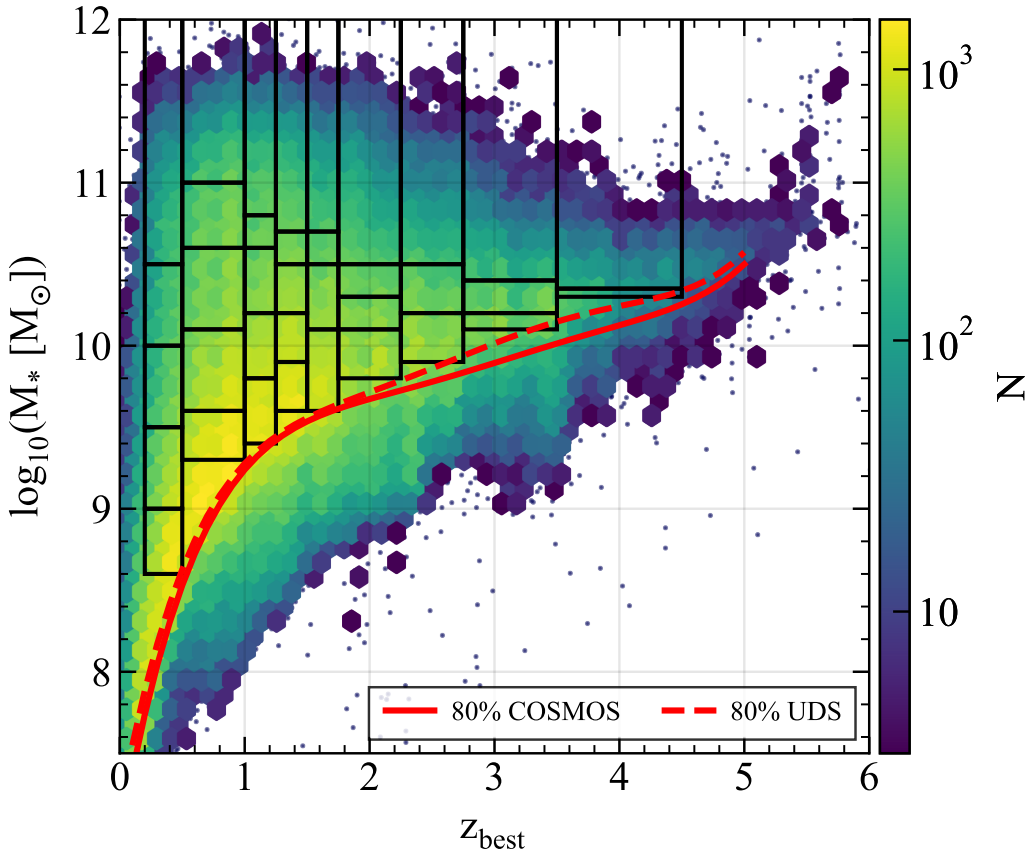

Figure 3.2: Distribution of the stellar mass versus redshift of the galaxies in COSMOS and UDS along with the selected sub-samples (black ladders) for this study. The rungs of the ladders show the mass thresholds (for clustering) and mass bins (for SMF) of the sub-samples. The right edge of the bottom rung of the ladders at each redshift corresponds to the higher of the solid red and dashed red curves, representing the 80% stellar mass completeness of COSMOS and UDS, respectively.

$$w_{obs}(\theta) = \frac{\left[\frac{DD(\theta)}{n_{\rm D}^2} - 2\frac{DR(\theta)}{n_{\rm D}n_{\rm R}} + \frac{RR(\theta)}{n_{\rm R}^2}\right]}{\left[\frac{RR(\theta)}{n_{\rm R}^2}\right]}. \tag{3.1}$$

The random catalog is created using exactly the same survey geometry mapped out by the galaxies considered in this study avoiding areas around bright stars, etc. Moreover, the number of points in the random catalogs are at least ~ 65 times more than the galaxies in any stellar mass threshold sample ensuring that $w_{obs}(\theta)$ is properly measured. We measure the $w_{obs}(\theta)$ in 13 angular separation bins within $-3 < log_{10}(\theta[\rm deg]) < -0.6$ which corresponds to a maximum separation of just over 0.25 deg. As a result we probe a maximum transverse co-moving scale of ~ 6 Mpc in our lowest z-bin which increases to ~ 32 Mpc for our highest z-bin.

We estimate the errors on $w_{obs}(\theta)$ by calculating the square root of the diagonal elements of the covariance matrix capturing the covariance among all angular scales. To estimate the covariance matrix, we use the "delete-one" jackknife method (Wall & Jenkins 2012) wherein we subdivide the whole region of our sample footprint into 50 equal-sized patches. Then for each mass threshold sample, we compute the correlation functions for 50 realizations each time with a unique patch removed. Finally, for each sub-sample, the covariance matrix, C_{ij} is computed as follows:

$$C_{ij} = \frac{N_r - 1}{N_r} \sum_{r=1}^{N_r} \left(w_r(\theta_i) - \tilde{w}\right)^T \left(w_r(\theta_j) - \tilde{w}\right), \tag{3.2}$$

where N_r is the number of realizations, $\tilde{w}$ is the average correlation function of all realizations and i and j indices label the angular separation bins.

As we measure $w_{obs}(\theta)$ from limited-sized patches on the sky (total effective area ~ 1.61 deg^2), it is bound to be underestimated by a constant

Figure 3.3: Redshift distribution, N(z) for each of the 9 bins used in this study generated by stacking the galaxy redshift probability distributions given by `eazy-py`.

integral constraint (IC) factor (Groth & Peebles 1977). The IC correction scales inversely with the field size and directly with the clustering strength as it is equal to the fractional variance on galaxy counts (Wake et al. 2011). We describe how we deal with this correction in §3.4.6.

3.3.2 Redshift Distributions

In order to convert the modeled two-point correlation function in three dimensions $\xi(r)$ into angular correlation functions, $\mathrm{w}_{model}(\theta)$ (as discussed in §3.4.6), we need the redshift distributions, N(z), for each of the z-bins in our sample. We derived them by stacking the `eazy-py` output redshift probability distributions, p(z)'s of galaxies above the lowest stellar mass threshold within each z-bin, ensuring that the $16^{th}/84^{th}$ percentiles of p(z)'s lie within the limits of the considered z-bin. The final N(z)'s are shown in Figure 3.3 where the large-scale structure can be easily gleaned through their jagged nature−sharp spikes and dips.

3.3.3 Stellar Mass Function

We derive the SMFs in the UDS and COSMOS fields above the lowest stellar masses threshold at each z-bin as described in §3.2.2, defined in Table 3.1 and shown in Figure 3.2. As we only compute SMFs for essentially complete samples, we do not need to implement the usual $1/V_{max}$ technique (Schmidt 1968) or similar, useful when not all of the galaxies within a stellar mass bin are observable within the full volume spanned by a given z-bin and survey area.

The Poisson errors on the SMFs were estimated using Gehrels (1986) specially useful for small number counts at the high mass end. We also calculated cosmic variance errors for each of the two fields separately following Moster et al. (2011) and then combined them using the volume-weighted average technique described in the same paper. Briefly, to calculate the cosmic variance in the individual fields, we used their Equation 5: σ_v^2 (M_*, z) = b^2 (M_*, z) σ_{dm}^2 (z), where b the galaxy bias that depends on stellar mass and redshift mentioned therein, whereas σ_{dm}^2 is the cosmic variance error on dark matter based on redshift and the dimensions of the survey area calculated using the public QUICKCV code (Newman & Moster 2014). We derived the cosmic variance for the combined sample from the two fields using the individual field's cosmic variance and survey area as in Equation 9 of Moster et al. (2011). Finally, the total uncertainties on the SMFs were calculated by adding the Poisson errors ($\sigma_{Poisson}$), errors due to cosmic variance (σ_{CV}), and a fixed 3% error attributed to the uncertainty in the survey area (σ_{Area}) in quadrature. We show the individual and combined SMFs of the two fields in Appendix A. Additionally, we also tabulate the combined SMFs and the associated percentage errors due to all three sources mentioned above in Ap-

pendix A.

3.4 Modeling the HOD

We derive the SHMR under the framework of HOD which models the average number of galaxies selected above a stellar mass threshold as a function of halo mass. There are many implementations of it in the literature (e.g. Zheng et al. 2005, 2007; Zehavi et al. 2005, 2011; Leauthaud et al. 2011; Zu & Mandelbaum 2015) most of which at least split the contribution from the central galaxy and satellite galaxies to the average total occupation number at a given halo mass. The central galaxy is the most massive galaxy dominating the potential well of a halo and depending on the halo mass, it might be surrounded by one or more lower-mass satellite galaxies.

The clustering measurements are an amalgamation of galaxy-galaxy (satellite-satellite or central-satellite) pairs within the *same* halos and therefore dominant at smaller angular scales, and the galaxy-galaxy (central-central or central-satellite) pairs between different halos that contribute more at larger scales. These are the so-called '1-halo' (pairs from the *same* halos) and '2-halo' (pairs from *different* halos) contributions to the total clustering which result in a break in total clustering at a certain angular scale where the clustering strength of '1-halo' drops and '2-halo' picks up (Zehavi et al. 2004).

3.4.1 SHMR of Central Galaxies

In this study, we utilize the HOD prescribed by L11 which starts by assuming a functional form for SHMR, $f_{\rm SHMR}(M_h)$ which gives the mean-log stellar mass, $\langle \log_{10}(M_*(M_h)) \rangle$ for central galaxies as a function of a fixed halo

mass, M_h. L11 adapted the $f_{\rm SHMR}(M_h)$ from Behroozi et al. (2010). For convenience, $f_{\rm SHMR}(M_h)$ can be inversely defined as follows:

$$log_{10}\left(f^{-1}_{\rm SHMR}(M_*)\right) = log_{10}(M_h) =$$

$$\log_{10}(M_1) + \beta log_{10}\left(\frac{M_*}{M_{*,0}}\right) + \frac{\left(\frac{M_*}{M_{*,0}}\right)^{\delta}}{1+\left(\frac{M_*}{M_{*,0}}\right)^{-\gamma}} - \frac{1}{2}, \quad (3.3)$$

where M_1 governs the characteristic halo mass, the higher its value, the more massive the host halo for a given stellar mass; $M_{*,0}$, in turn, governs the characteristic stellar mass, such that increasing it results in less massive host halos assigned to a given stellar mass; β controls the power-law slope at the low-mass end; whereas δ controls the sub-exponential slope on the high-mass end; finally, γ dictates how sharp the transition is between the low-mass and high-mass regimes. The effect of varying each of these parameters on the shape of SHMR can also be seen through Figure 1 in L11.

3.4.2 Scatter in Stellar Mass at a fixed Halo mass

At a fixed halo mass there is still expected to be a scatter in the stellar mass, which arises from the intrinsic scatter in the Universe and the scatter introduced by measurement errors due to SED modeling, photo-z uncertainties, and the errors in the observed magnitudes (Behroozi et al. 2010; L11). L11 captures this stochasticity by assuming a log-normal distribution describing the probability distribution function of $log_{10}(M_*)$ at a fixed halo mass. The standard deviation of this distribution, σ_{logM_*}, the log-normal scatter in the stellar mass at a fixed halo mass is then fitted as a free parameter.

The data at the high mass end mainly tends to constrains the σ_{logM_*} (Leauthaud et al. 2012). It is because firstly, the low mass data leaves a

weaker imprint on the data (Leauthaud et al. 2012), and secondly, the low mass data below the completeness limit is excluded in our analysis as opposed to e.g. Zu & Mandelbaum (2015), where a modified modeling technique was employed to incorporate the mass regime where the data is incomplete. Therefore, the inclusion of halo mass dependence in σ_{logM_*} in our case does not matter in any meaningful way. In fact, Leauthaud et al. (2012) considered two models, one in which σ_{logM_*} is a constant and another in which it varies as a function of stellar mass. They found no meaningful differences in the resulting HOD parameter fit and the SHMR model. As a consequence, we have assumed σ_{logM_*} to be independent of halo mass in this work, i.e. σ_{logM_*} is considered to be constant for all halo masses within a z-bin.

3.4.3 Mean Occupation of Central Galaxies

To compute the mean occupation of central galaxies, we can integrate the conditional SMF for centrals, $\Phi_{cen}(M_*|M_h)$ which is the number of centrals with stellar mass within $M_* \pm dM_*/2$, given a fixed halo mass. For instance, for a sample of centrals with a stellar mass threshold, M_*^t, the mean occupation can be calculated as follows:

$$\langle N_{cen}(M_h|M_*^t)\rangle = \int_{M_*^t}^{\infty} \Phi_{cen}(M_*|M_h) dM_* \tag{3.4}$$

In L11, $\Phi_{cen}(M_*|M_h)$ is defined as log-normal distribution (see their Equation 1) which means that $\langle N_{cen}(M_h|M_*^t)\rangle$ can be derived analytically. The final expression for $\langle N_{cen}(M_h|M_*^t)\rangle$ using $f_{\rm SHMR}(M_h)$ and a constant σ_{logM_*} then becomes (see L11 for the shape of $\langle N_{cen}(M_h|M_*^t)\rangle$ at different M_*^t):

$$\langle N_{cen}(M_h|M_*^t)\rangle = \frac{1}{2}\left[1 - {\rm erf}\left(\frac{log_{10}(M_*^t) - log_{10}(f_{\rm SHMR}(M_h))}{\sqrt{2}\sigma_{logM_*}}\right)\right], \tag{3.5}$$

where the error function, erf is defined by

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt. \tag{3.6}$$

The alternate commonly employed expression for the mean occupation function of centrals in the literature, e.g. as in Zheng et al. (2005), can be derived from Equation 3.5 if: i- $f_{\mathrm{SHMR}}(M_h)$ is assumed to be a power-law at all masses, and ii- a minimum halo mass, M_{min}, which corresponds to M_*^t, is introduced, i.e. $M_{min} \equiv f_{\mathrm{SHMR}}^{-1}(M_*^t)$ (L11). Even though this common version agrees well with the Eq. 3.5 at low masses ($\mathrm{M}_{min} \lesssim 10^{12}$ $\mathrm{M}_\odot$), the difference between the two grows between $10\% - 40\%$ at higher masses (L11) as the L11 SHMR deviates from the power-law and becomes sub-exponential at high masses. As observations confirm that SHMR rises faster than a power-law at high masses, it is crucial to use an expression for $\langle \mathrm{N}_{cen} \rangle$ that explicitly incorporates a shift from power-law to sub-exponential at higher masses.

3.4.4 Mean Occupation of Satellite Galaxies

The model for the mean occupation of satellites is expected to follow a power-law at high masses with an exponential sharp cut-off at low masses (Berlind et al. 2003; Kravtsov et al. 2004; Zheng et al. 2005). Therefore, the expression for $\langle N_{sat} \rangle$ can be written as in Eq. 3.7 where the α_{sat} is the power-law slope with M_{sat} as the power-law amplitude. M_{cut} defines the scale for the exponential cut-off. Moreover, halos are not expected to have satellite galaxies before having a central galaxy. This is ensured by scaling to the satellite

occupation to $\langle N_{cen} \rangle$.

$$\langle N_{sat}(M_h|M_*^t) \rangle = \langle N_{cen}(M_h|M_*^t) \rangle \left(\frac{M_h}{M_{sat}} \right)^{\alpha_{sat}} \exp\left(\frac{-M_{cut}}{M_h} \right). \tag{3.7}$$

As it turns out according to the observational studies, M_{sat} is proportional to M_{min}, where M_{min} is the inverse of the SHMR, i.e. $M_{min} \equiv f^{-1}_{\rm SHMR}(M_*^t)$, where $f^{-1}_{\rm SHMR}(M_*^t)$ is assumed to be a power-law (L11). For added flexibility, instead of M_{sat} and M_{cut} simply scaling with $f^{-1}_{\rm SHMR}(M_*^t)$, L11 defines M_{sat} and M_{cut} as power-law functions of $f^{-1}_{\rm SHMR}(M_*^t)$ as shown in Eq. 3.8 and Eq. 3.9, respectively.

$$\frac{M_{sat}}{10^{12} M_\odot} = B_{sat} \left(\frac{f^{-1}_{\rm SHMR}(M_*^t)}{10^{12} M_\odot} \right)^{\beta_{sat}}, \tag{3.8}$$

$$\frac{M_{cut}}{10^{12} M_\odot} = B_{cut} \left(\frac{f^{-1}_{\rm SHMR}(M_*^t)}{10^{12} M_\odot} \right)^{\beta_{cut}}. \tag{3.9}$$

Furthermore, we can easily calculate the fraction of stellar mass locked up in satellites, f_{sat} by dividing the mean number density of satellites, $\tilde{n}_{g,\,sat}$ with the mean total (central+satellites) number density, $\tilde{n_g}$, as shown in Equation 3.10 below. Here, the mean number densities are calculated by integrating the HMF, dn/dM_h along with the desired mean occupation function.

$$f_{sat} = \frac{\tilde{n}_{g,\,sat}}{\tilde{n_g}} = \frac{\int dM_h \frac{dn}{dM_h} \langle N_{sat}(M_h|M_*^t) \rangle}{\int dM_h \frac{dn}{dM_h} \langle N_{total}(M_h|M_*^t) \rangle} \tag{3.10}$$

3.4.5 Total Stellar Mass in Halos

The L11 HOD model also allows us to conveniently calculate the total stellar mass in galaxies as a function of halo mass and the contributions from central and satellite galaxies. The total stellar mass within a stellar mass bin can

be calculated by integrating the conditional SMFs of centrals and satellites multiplied by M_* with the stellar mass range as the limits of integration as shown in the second line of Equation 3.11. Then, using the fact that integrating the conditional SMF of centrals/satellites gives the occupation number of centrals/satellites and the integration by parts rule, we can further simplify the integration as shown in the third and fourth lines of Equation 3.11:

$$M_*^{tot}\left(M_h|M_*^{t_1}, M_*^{t_2}\right) =$$

$$\int_{M_*^{t_1}}^{M_*^{t_2}} \left[\Phi_c(M_*|M_h) + \Phi_s(M_*|M_h)\right] M_* dM_* =$$

$$\int_{M_*^{t_1}}^{M_*^{t_2}} \langle N_{cen}\left(M_h|M_*\right)\rangle dM_* - \left[\langle N_{cen}\left(M_h|M_*\right)\rangle M_*\right]_{M_*^{t_1}}^{M_*^{t_2}}$$

$$+ \int_{M_*^{t_1}}^{M_*^{t_2}} \langle N_{sat}\left(M_h|M_*\right)\rangle dM_* - \left[\langle N_{sat}\left(M_h|M_*\right)\rangle M_*\right]_{M_*^{t_1}}^{M_*^{t_2}}, \quad (3.11)$$

where the first term (third line) reflects the contribution from the central galaxies and the second term (fourth line) corresponds to the satellite galaxies.

3.4.6 Two-point angular correlation function from the HOD

To derive the modeled two-point angular correlation function, $w_{\rm halomod}(\theta)$ within the `halomod` code (Murray et al. 2021), we start from the model galaxy real-space correlation function, $\xi_{gg}(r)$, the derivation of which is described in detail in Murray et al. (2021). Briefly, $\xi_{gg}(r)$ is connected to the galaxy power spectrum, $P_{gg}(k)$ via order-1/2 *Hankel* transform:

$$\xi_{gg}(r) = \frac{1}{2\pi^2}\int_0^\infty P_{gg}(k)k^2 j_0(kr)dk, \tag{3.12}$$

with j_0 being the zeroth-order spherical Bessel function, $j_0(x) = \frac{sinx}{x}$. $P_{gg}(k)$ is further divided into contributions from the 1-halo (1h) and 2-halo (2h) pairs: $P_{gg}(k) = P_{gg}^{1h}(k) + P_{gg}^{2h}(k)$, and correspondingly, $\xi_{gg}(r) = [1 + \xi_{gg}^{1h}(r)] + \xi_{gg}^{2h}(r)$. $P_{gg}^{1h}(k)$ and $P_{gg}^{2h}(k)$ are defined as follows:

$$P_{gg}^{1h}(k) = \int dM_h \frac{dn}{dM_h} u(k|M_h)\Big[\langle N_{sat}\rangle^2 u(k|M_h) + 2\langle N_{sat}N_{cen}\rangle\Big], \tag{3.13}$$

$$P_{gg}^{2h}(k) = P_m(k)\left[\int dM_h \frac{dn}{dM_h} u(k|M_h)\frac{\langle N_{total}\rangle}{\tilde{n_g}} b(M_h)\right]^2. \tag{3.14}$$

In Equations 3.13 and 3.14, $u(k|M_h)$ is the halo profile's mass-normalized Fourier transform; $P_m(k)$ is the matter power spectrum; $\tilde{n_g}$ is the mean number density of galaxies given calculated as in Equation 3.10; and $b(M_h)$ is the scale-independent halo bias from Tinker et al. (2010).

From $\xi(r)$, we derive the angular correlation function within the `halomod` code (Murray et al. 2021) using the Limber's equation (Limber 1954) as presented in Blake et al. (2008):

$$w_{\text{halomod}}(\theta) = 2\int_0^\infty dx f^2(x)\int_0^\infty du\xi(r = \sqrt{u^2 + x^2\theta^2}) \tag{3.15}$$

$$f(x) = \frac{p(z)}{dx/dz(z)}, \tag{3.16}$$

where $p(z)$ is the redshift distribution which we deduce from the stacking of the individual galaxy's redshift probability density functions within a z-bin.

x is the comoving radial distance from the median redshift, and $f(x)$ defines the overall radial distribution.

As mentioned in §3.3.1, the measured angular correlation function, $w_{obs}(\theta)$ is underestimated by a constant additive IC factor. Instead of adding this correction to $w_{obs}(\theta)$, we subtract it from $w_{\rm halomod}(\theta)$ to get the model angular correlation function, $w_{model}(\theta)$, i.e. $w_{model}(\theta) = w_{\rm halomod}(\theta)$ - IC. We estimate IC as in Infante (1994) and Roche & Eales (1999):

$$\mathrm{IC} = \frac{\sum_i w_{\rm halomod}(\theta_i) RR(\theta_i)}{\sum_i RR(\theta_i)}, \tag{3.17}$$

where we run the θ_i out to much larger separations ($\sim$ 4 deg) than while measuring/modeling the angular correlation function itself (only out to $\sim$ 0.25 deg) to ensure that we have counted all of the RR pairs. While fitting we can now compare $w_{obs}(\theta)$ and $w_{model}(\theta)$ directly (see §3.4.9) as both of them are lower by the same IC factor.

3.4.7 Stellar Mass Functions from the HOD

The SMF corresponding to a chosen set of HOD parameter values can be calculated using the occupation function and the HMF, dn/dM_h. Now, Equation 3.5 and Equation 3.7 define the occupation functions with respect to a stellar mass threshold, however, the SMF is a series of values of number density per unit volume in bins of stellar mass. This means that we need the occupation function defined in bins of stellar masses instead of in thresholds of stellar mass. Fortunately, the occupation function within a stellar mass bin is simply the difference between the two stellar mass threshold-defined

occupation functions at the mass thresholds $M_*^{t_1}$ and $M_*^{t_2}$ defining the bin edges:

$$\langle N\left(M_h|M_*^{t_1}, M_*^{t_2}\right)\rangle = \langle N\left(M_h|M_*^{t_1}\right)\rangle - \langle N\left(M_h|M_*^{t_1}\right)\rangle, \tag{3.18}$$

where, $\langle N\rangle$ represents the mean occupation function of the centrals, satellites, or total depending on whichever is desired. For a stellar mass bin defined by $\Delta log_{10}M_* = M_*^{t_2} - M_*^{t_1}$, the modelled SMF, $\Phi_{model}(M_*^{t_1}, M_*^{t_2})$ can then be calculated as:

$$\Phi_{model}(M_*^{t_1}, M_*^{t_2})\Delta log_{10}M_* = \int_0^\infty \left(\langle N\left(M_h|M_*^{t_1}, M_*^{t_2}\right)\rangle \frac{dn}{dM_h}\right) dM_h, \tag{3.19}$$

where again $\langle N\rangle$ can be the mean occupation function of the centrals, satellites, or total depending on whichever SMF is desired.

3.4.8 Redshift evolution of the HOD parameters

As can be seen in Figure 3.2, the dynamical range in the stellar mass shrinks dramatically with redshift due to the widening range of incomplete stellar masses. The lowest stellar mass considered within the highest z-bin is $10^{10.3}$ $M_\odot$ about 2 orders of magnitude higher than considered at the lowest z-bin. As a result, it becomes challenging to constrain the low-mass slope (controlled by β) of the SHMR at high-z. Due to the small dynamical range in stellar mass, the scatter in stellar mass at a fixed halo mass, σ_{logM_*} also proves harder to constrain.

Furthermore, even though the total effective survey area used in this study is large, it is not large enough to capture a high number of massive

galaxies at all redshifts. Therefore, at all redshifts, it remains a struggle to constrain the δ and γ parameters which control the slope of the SHMR at the massive end and the transition regime between low-mass and high-mass, respectively.

This motivated us to consider an alternate approach to fitting the HODs. All of the 11 HOD parameters at each z-bin can be better constrained if it is realized that the clustering and SMFs in the neighboring z-bins are not disconnected from each other but rather evolve from one to another. Consequently, the HOD parameters that produce the clustering and abundance should also evolve smoothly, rather than having sharp jumps between the neighboring z-bins. We implement this understanding by penalizing sharp changes in the HOD parameter values between the adjacent z-bins through the use of continuity ("smoothing") priors.

Consequently, instead of fitting the clustering and abundance data for each of the individual z-bins ("discrete-z" model hereafter), we jointly fit the data in all redshifts at once with the HOD evolving smoothly with redshift due to the continuity priors ("smooth-z" model hereafter). This approach allows us to leverage the wealth of information contained within the clustering and abundance measurements at lower redshift spanning wider dynamical ranges in stellar mass, by connecting them to higher redshift via smooth redshift evolution of the HOD parameters.

To implement this smoothing, we use Gaussian priors that peak when there is no change in a given HOD parameter between the adjacent z-bins. For the HOD parameters that are fitted in log-space, $M_{*,0}$, M_1, and σ_{logM_*}, we impose the Gaussian continuity prior (centered at '0') on the difference of the HOD parameter values between the adjacent z-bins, e.g. $\Delta \equiv M_1(z_x) -$

Table 3.2: σ of the gaussian priors on the change in HOD parameters between adjacent bins

HOD parameter	range	σ
$M_{*,0}$	$10 - 12$	0.065
M_1	$11.8 - 14$	0.12
β	$0.1 - 0.9$	0.06
δ	$0.1 - 4$	0.08
γ	$0 - 8$	0.12
σ_{logM_*}	$0.001 - 0.5$	0.025
α_{sat}	$0.4 - 2$	0.07
β_{sat}	$0.01 - 4$	0.09
B_{sat}	$0 - 17$	0.065
β_{cut}	$-1 - 8$	0.145
B_{cut}	$0 - 14$	0.145

$M_1(z_{x+1})$ as follows:

$$\text{prior}(\Delta) = \frac{e^{-\frac{1}{2}\left(\frac{\Delta}{\sigma}\right)^2}}{\sigma\sqrt{2\pi}} \tag{3.20}$$

where x denotes any x^{th} z-bin. Similarly, for all of the 8 other HOD parameters that are not fitted in log space, we impose the continuity prior (centered at '1') on the ratio of the HOD parameter values between the adjacent z-bins, e.g. $\Delta \equiv \frac{\beta(z_x)}{\beta(z_{x+1})}$ as follows:

$$\text{prior}(\Delta) = \frac{e^{-\frac{1}{2}\left(\frac{\Delta-1}{\sigma}\right)^2}}{\sigma\sqrt{2\pi}} \tag{3.21}$$

where σ is the width of the Gaussian continuity prior decided such that the maximum range obtained with the "discrete-z" fits is allowed to occur over the whole redshift range. Their values of σ along with the bounds on all the HOD parameters are tabulated in Table 3.2.

3.4.9 Fitting the HOD

Our primary results come from the "smooth-z" model in which we jointly fit the HOD parameters at the 9 z-bins (described in 3.4.8) by minimizing the following χ^2:

$$\chi^2 = \sum_{z}^{Z} \chi^2_w + \chi^2_\Phi \tag{3.22}$$

where

$$\chi^2_w = \sum_{i}^{N_{M^t_*}} (w_{obs,\ i} - w_{model,\ i})^{\mathrm{T}}\ C^{-1}_{obs,\ i}\ (w_{obs,\ i} - w_{model,\ i}) \tag{3.23}$$

and

$$\chi^2_\Phi = \sum_{j}^{N_{M^b_*}} \left(\frac{log_{10}(\Phi_{obs,\ j}) - log_{10}(\Phi_{model,\ j})}{log_{10}(\sigma_{\Phi_{obs,\ j}})} \right)^2. \tag{3.24}$$

In Equation 3.23 above, the summing index i runs over $N_{M^t_*}$ stellar-mass threshold samples (shown in Table 3.1), and $\mathrm{C}^{-1}_{obs,\ i}$ is the inverse covariance matrix calculated using Equation 3.2. In Equation 3.24 the summing index j runs over $N_{M^b_*}$ stellar-mass bins used to evaluate the SMFs and $\sigma_{\Phi_{obs,\ j}}$ is the total error on the observed SMF, Φ_{obs} including contributions both from the Poisson statistics and the cosmic variance as described in §3.3.3. Finally, as we jointly fit all redshifts simultaneously, the index z in the outer sums in both terms runs over all the Z (= 9) z-bins.

In practice, we maximize the log-likelihood, ln($\mathcal{L}$) = -χ^2/2 which is the same as minimizing the χ^2 using the python package `emcee` (Foreman-Mackey et al. 2013), which implements the affine-invariant Markov chain Monte Carlo (MCMC) ensemble sampler of Goodman & Weare (2010). As we jointly fit the HODs for all the z-bins, we have a total of 99 parameters = 11 HOD parameters at each z-bin × 9 z-bins to be fit. We run 300 walkers to fit them

until the resulting chains are converged. We consider the chains converged if they pass the typical convergence criteria that their auto-correlation time, τ (number of iterations before the chain 'forgets' where it started from) is at least 50 times smaller than the total number of iterations, and that the change in chains is less than $\sim 1\%$. After rejecting a certain number of iterations as 'burn-in', we retrieved the marginalized posterior distributions from the chains. The medians of the posteriors are considered the best-fit values and the errors on them come from the $16^{th}/84^{th}$ percentiles.

To show the improvement in the HOD parameter fits due to the novel "smooth-z" model, we also perform the HOD fitting in the standard way ("discrete-z") and compare the results in §3.5.2. For the "discrete-z" model, the χ^2 formulation remains the same as in Equation 3.22 except for the outer sums running over all the z-bins which no longer apply. Furthermore, for the "discrete-z" model as it fits each z-bin separately, we now only have 11 HOD parameters per fit. We fit them using 60 walkers until the chains are converged according to the same criteria as mentioned above.

3.5 Results

3.5.1 Evolution of the measured clustering and abundance

In Figure 3.4, we show our measurements (empty circles with error bars) of the two-point angular correlation function, $w_{obs}(\theta)$ for all of the stellar mass threshold sub-samples within each z-bin as defined in Table 3.1. The clustering of different stellar mass threshold samples is highlighted using different

shades of the same color within each z-bin with darker shades representing higher mass thresholds as indicated in the legends. The strength of clustering increases with stellar mass on all scales which is unsurprising because more massive galaxies tend to reside in denser environments. We can also see a break from a power-law in the clustering amplitude between small scales and large scales due to the contributions from the pairs within the same halos at small scales ('1-halo') and from the pairs between different halos ('2-halo') at large scales as described in §3.4. In each panel, we highlight the scale where the '2-halo' contribution to the real-space correlation function drops below that of '1-halo' with a black marker. This break can be seen to be occurring at different scales in different z-bins because of how the co-moving transverse distance projected on the same angular scale on the sky changes with redshift due to the changing rate of expansion of the Universe.

The measured SMFs which are jointly fit with clustering to get the HODs are shown in the inset panels via filled circles and error bars. The evolution in SMF with redshift is also highlighted by the lowest z-bin SMF over-plotted as a dotted magenta curve in all the other panels. The lowest z-bin SMF has the characteristic 'knee' depicting a sharp drop in the number densities of the high-mass galaxies. This 'knee' is barely noticeable, if at all within the highest z-bin but gets more and more pronounced with decreasing redshift, indicative of increasing strength of feedback effects suppressing star formation in high-mass halos/galaxies (Silk & Rees 1998; Benson et al. 2003). Moreover, the steepening of the low mass slope with increasing redshift as reported in the literature (Davidzon et al. 2017; Weaver et al. 2023) can be seen. The over-plotted lowest z-bin SMF in all the other panels also highlights how the dynamic range in stellar mass shrinks with redshift going from over 3

decades in stellar mass at z $\sim$ 0.35 (lowest z-bin) to only about a decade at z $\sim$ 4 (highest z-bin). This considerably weakens the SHMR constraining power of the high-z data (clustering+SMF) in isolation. However, the use of "smooth-z" modeling in this paper connecting high-z with low-z using continuity priors remedies this to a large extent as shown in the following §3.5.2.

3.5.2 "Discrete-z" versus "smooth-z" model fits

We introduced a novel approach to fitting the HOD model in §3.4.8 wherein we jointly fit clustering and SMF data within all of the 9 z-bins connected via smooth evolution ("smooth-z" model) of the HOD parameters that produce the observed data.

In Figure 3.5, we show the increased constraining power of the "smooth-z" model over the standard "discrete-z" model in which the data in the individual z-bins are fitted with no regard to the adjacent z-bins. For each of the 11 HOD parameters, the yellow scatter points show the medians of the posteriors for the "discrete-z" model, whereas the corresponding error bars show the range spanned by their $16^{th}/84^{th}$ percentiles. As hinted in §3.4.8, with the "discrete-z" model, it proved a challenge to constrain the parameters sensitive to the data from the high stellar mass end of the SHMR at all redshifts and low-stellar mass end of the SHMR with increasing redshift.

At the high-mass end, it is the δ and γ that control the rise in the SHMR in the high-mass regime and the transition from low-mass to high-mass, respectively. Within most z-bins, the error-bars span a wide range within the bounds on $\delta \in [0.1,\ 4]$ and $\gamma \in [0,\ 8]$ meaning that they are very weakly

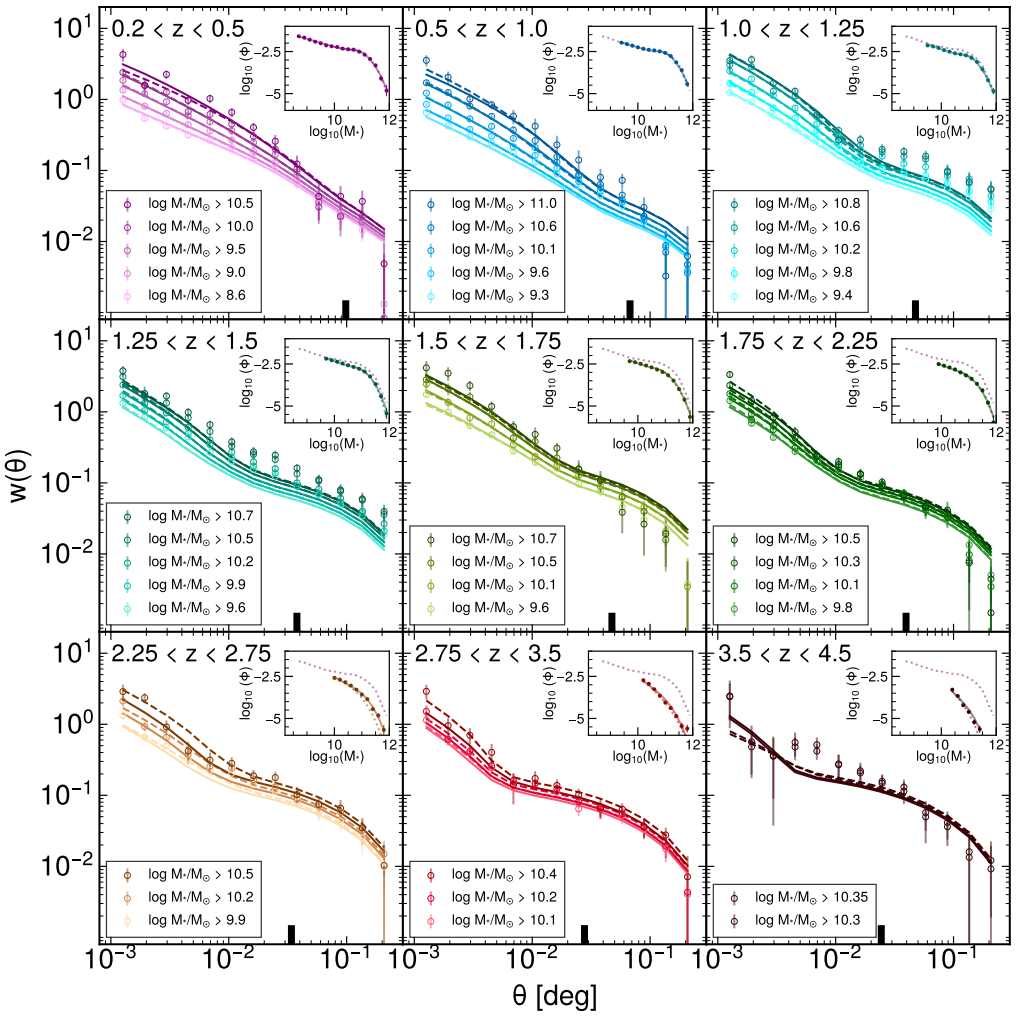

Figure 3.4: Clustering (main panels) and SMF (inset panels) measurements in the whole UDS+COSMOS sample along with the best fits of the "smooth-z" model as a function of redshift. The empty circles in the main panels show the clustering measurements and their associated uncertainty, whereas the solid curves show the clustering produced by the best-fit "smooth-z" model, our main result, whereas the dashed curves show the results from the "discreet-z" model for comparison. Different mass threshold samples are indicated via color shading as indicated in the legends, with darker shades showing higher mass threshold samples. The breaks in scale, where the '2-halo' contribution to $w(\theta)$ overtakes that of '1-halo', are marked in black at the bottom of each panel. In the inset panels, the SMF measurements are shown via filled points and error-bars. The lowest redshift bin SMF is also shown as a dotted magenta curve in all the other inset panels to highlight the redshift evolution in the measured SMF. The abundance produced by the best-fit "smooth-z" model is shown via solid curves, our main result, and the "discreet-z" results are shown via dashed curves for comparison.

constrained by the data when only considered one z-bin at a time. The weak constraining of the high-mass sensitive parameters δ and γ can be attributed to large uncertainties in the SMFs at the high-mass end (and the corresponding contributions to clustering) due to not capturing enough rare massive galaxies. Even though, our dataset covers a large effective survey area of ~ 1.61 deg^2, pivotal in capturing a good number of rare massive galaxies, it is still not large enough to capture a large number of rare ultra-massive galaxies ($M_* \gtrsim 10^{11}$ $M_\odot$) to pin down the SMF at the high mass end with confidence on par with the intermediate stellar masses. Therefore, it makes sense that the HOD parameters sensitive to the high-mass population, δ and γ cannot be constrained better, especially in the "discrete-z" model.

At the low-mass end at high-z, it is challenging to constrain the β parameter, which controls the low-mass slope of the SHMR, due to the shrinking dynamical range in complete stellar masses with redshift. Moreover, at all redshifts due to the reasons mentioned above, the scatter in stellar mass at fixed halo mass, σ_{logM_*} also struggles to be constrained better than the accuracy of $\Delta \sim 0.1$ dex which remains unhelpful in informing the strength of the stellar mass-halo mass connection. A small scatter would mean that the halo mass assembly is tightly linked to the stellar mass assembly in the Universe. In contrast, a large scatter would hint at the stellar mass dependence on other halo properties beyond its mass, a possible indication of galaxy assembly bias perhaps. Therefore, it is crucial to constrain σ_{logM_*} better.

The "smooth-z" model results are shown via the blue line (median) and the light blue shaded region ($16^{th}/84^{th}$ percentiles) in Figure 3.5. The three HOD parameters that specially struggled to converge in the "discrete-z" model as described above, are now remarkably well constrained at all red-

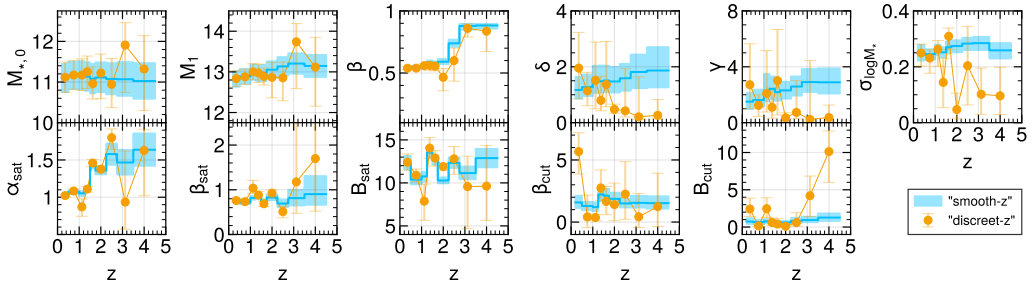

Figure 3.5: HOD parameter fits derived through the "discrete-z" model are shown in yellow scatter points and the ones derived using the novel "smooth-z" model are shown via the blue line. The errors indicated through yellow bars for "discrete-z" and blue-shaded envelope for "smooth-z" encompass the posteriors' $16-84$ percentile range.

shifts.

Figure 3.4 shows that there are subtle differences between the clustering and SMF fits from the "discrete-z" (dashed curves) and "smooth-z" (solid curves) models. Yet the "smooth-z" fits are the result of a continuous redshift evolution of the HOD parameters, which is likely preferred by nature, as opposed to the discontinuous evolution in the "discrete-z" model. Quantitatively, the reduced best-fit χ^2 for the"smooth-z" fit, $\chi^2_{red,smooth-z} = 2.46$, where the degrees of freedom, DOF $= 472 = 571$ (# of data points within all z-bins) - 99 (11×9 HOD parameters within all z-bins). Within "discrete-z" model, if we add all the raw best-fit χ^2 from the individual z-bin fits, we can then divide the total by the same DOF as in "smooth-z", to get $\chi^2_{red,discrete-z}$[3] $= 2.13$, which can be compared directly to $\chi^2_{red,smooth-z}$.

[3]the $\chi^2_{red,discrete-z}$ calculated using the median models is actually larger since some parameters are so poorly constrained in high z-bins that the median is not that close to the best fit

3.5.3 Redshift evolution of the mean SHMR for centrals

The fitted SHMR at each z-bin also allows us to paint a continuous picture of the evolution of SFE from $z = 4.5$ to $z = 0.2$. Using the median fitted values of the central HOD parameters in equation 3.3 at each redshift, we plot the stellar-halo mass ratios (M_*/M_h) as a function of the halo mass in Figure 3.6, and show the $16^{th}/84^{th}$ percentiles based on the MCMC run via lightly shaded envelopes. The M_*/M_h of a halo directly indicates how efficient the halo has been, throughout its lifetime, in converting baryonic matter into stellar matter. The SFE, given by dividing M_*/M_h by the baryonic fraction of the universe, $f_b = 15.8\%$ (Planck Collaboration et al. 2020), is indicated on the right in Figure 3.6.

Furthermore, the lowest and highest halo masses probed within each redshift bin are marked via vertical dashed lines, with color once again indicating redshift as in the legend. The lowest halo mass probed is given by the mean central SHMR at the lowest stellar mass threshold within each z-bin. On the other side, the highest halo mass probed is considered where the mean number of galaxies (derived using the fitted HOD at the respective redshift) drops below 0.6 within the volume covered. Even though we extrapolate our models beyond the halo masses directly probed by our analysis, we alert the reader to peruse them with caution.

The low-mass slope starts relatively flat in the highest z-bins and steepens the most through the $2.25 < z < 2.75$ (7^{th}) and $1.75 < z < 2.25$ (6^{th}) z-bins. This is also evident in the decrease in β with decreasing redshift seen in Figure 3.5; higher values indicate flatter slopes. At the low mass end, smaller-scale

feedback effects such as supernovae, stellar winds, etc. are thought to heat up or expel their surrounding gas and suppress star formation (Hopkins et al. 2012). Therefore, the redshift evolution of the low-mass slope hints at the strengthening of such feedback effects with time.

On the other hand, $M_{*,meancen}/M_h$ declines at the high-mass end at all redshifts. The active galactic nuclei (AGN) feedback is purported to be responsible for curbing star formation in massive galaxies (Silk & Rees 1998; Benson et al. 2003), predominantly in 'jet-mode' (Silk & Rees 1998; Croton et al. 2006; Kondapally et al. 2023; Scharré et al. 2024), explaining the decline in $M_{*,meancen}/M_h$ at the high mass end. Moreover, the high-mass slope remains remarkably consistent throughout the z-bins. This indicates that the strength of AGN feedback beyond the peak halo mass is strongly coupled to the halo mass and depends weakly, if at all, on the age of the Universe.

The combination of the low and high-mass slopes result in a narrow range in halo mass for peak star formation at each redshift, $M_h^{peak} \sim 10^{12.2} - 10^{12.4}$ $M_\odot$, reflecting SFE $\sim 13 - 20\%$. This value of $M_h^{peak} \sim 10^{12}$ $M_\odot$ is well supported by a series of observational studies (Behroozi et al. 2013, 2019; Ishikawa et al. 2017; Cowley et al. 2018; Moster et al. 2018; Shuntov et al. 2022). It is important to note that even though the adopted L11 model allows for a transition regime between low mass and high mass, it does not necessitate a peak halo mass; it naturally comes out of the fitting.

3.5.4 Redshift evolution of the peak halo mass, M_h^{peak}

To better understand the evolution of the feedback effects suppressing star formation, we now look at the evolution of the halo mass with *peak* integrated

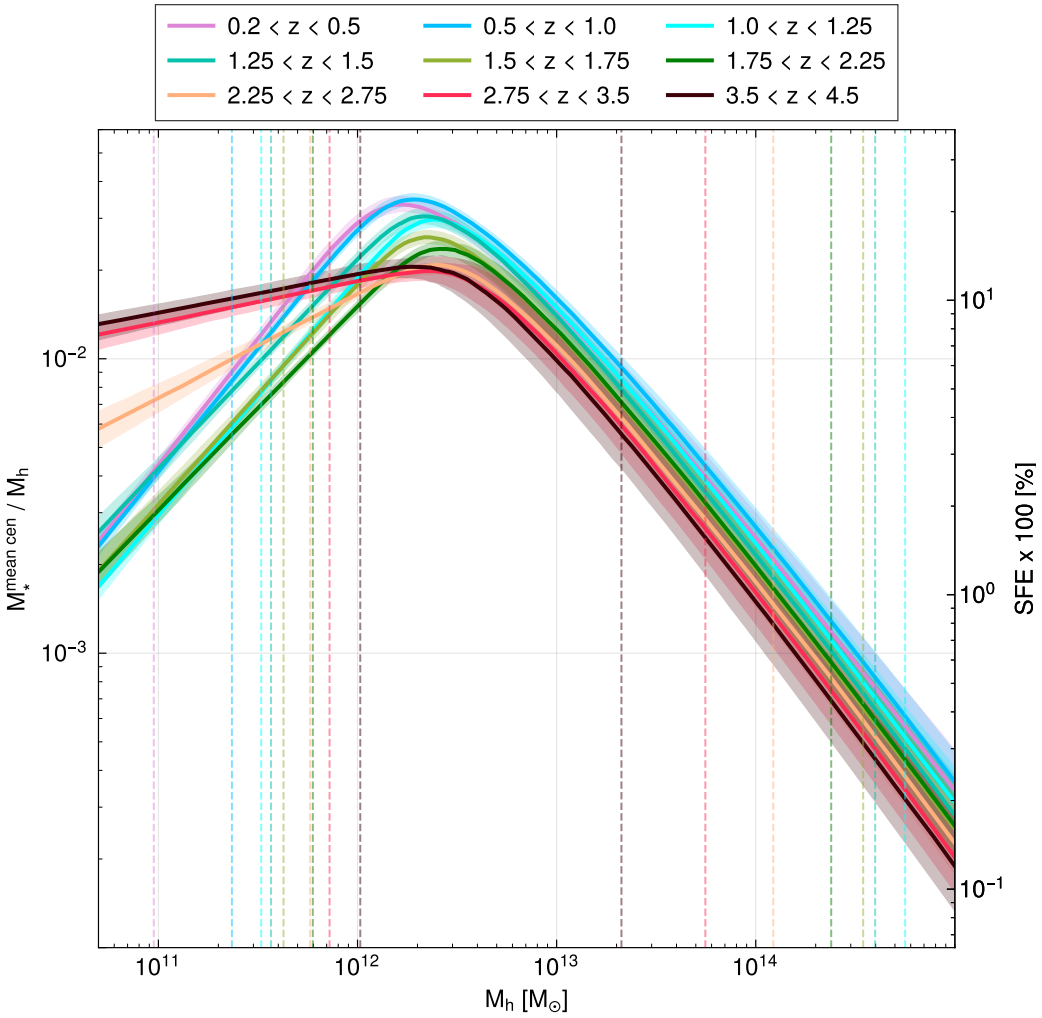

Figure 3.6: Mean stellar-to-halo mass ratio of central galaxies, $M_{*,\mathrm{mean\,cen}}/M_h$ as a function of halo mass for each z-bin. The SFE ($= (M_*^{\mathrm{mean\,cen}}/M_h)/f_b$) in percentages is indicated on the right-most ordinates. The colored vertical dashed lines (two for each z-bin) indicate the lowest and highest halo mass probed by our analysis, as described in §3.5.3, where the color indicates the z-bin as in the legend. Note: The highest halo mass end line (vertical dashed magenta) of the first z-bin is missing as it lies at the halo mass beyond the right-most extent of the x-axis.

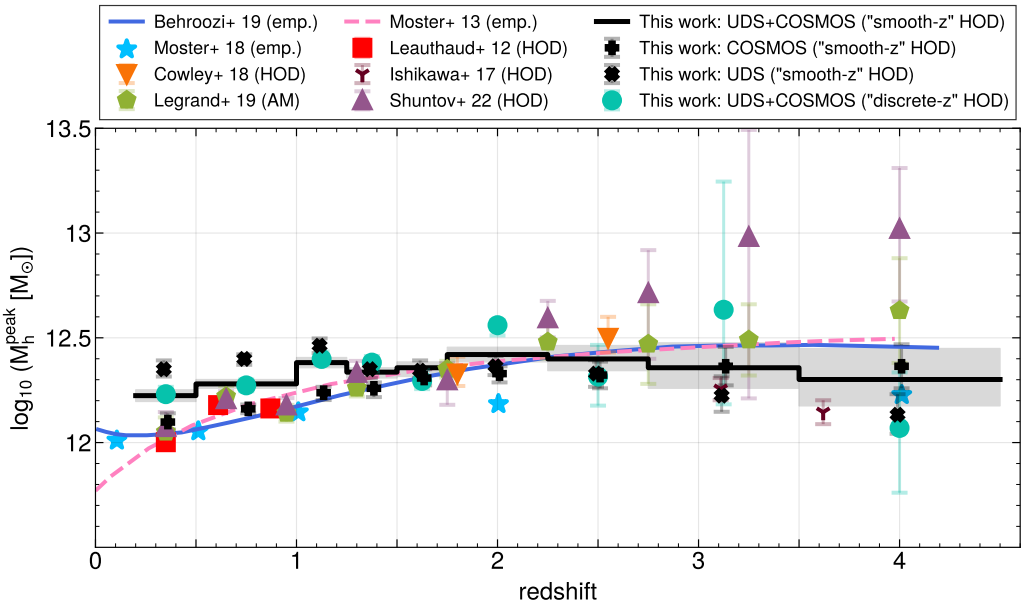

Figure 3.7: Redshift evolution of the halo mass, M_h^{peak} at which the $M_*^{\text{mean cen}}/M_h$ ratio as shown in Figure 3.6 peaks representing the peak SFE (integrated over halo's lifetime). The black line and the gray shaded region shows our main result (using "smooth-z") and the corresponding 1σ uncertainty, respectively, for the full UDS+COSMOS sample. We also indicate the median M_h^{peak} and 1σ errors using only UDS (black $\times$'s) and COSMOS (black +'s), (both with "smooth-z") highlighting the cosmic variance. A compilation of other observationally constrained M_h^{peak} values in the literature using different methods - Abundance matching (AM), empirical modeling (emp.), and HOD (equivalent to our "discrete-z" HOD model), are also shown. Finally, to highlight the decrease in error on M_h^{peak} going from "discrete-z" to "smooth-z", specially at $z > 3$, we also include our "discrete-z" UDS+COSMOS results as filled turquoise circles.

SFE, M_h^{peak} in Figure 3.7. M_h^{peak} are the peaks in Figure 3.6. Using the combined sample (UDS+COSMOS), we show the evolution of M_h^{peak} in Figure 3.7 using a black line, with the gray shaded region showing the 1σ uncertainty evaluated using the MCMC chains. We also show a compilation of results from the literature indicating the method they used to derive M_h^{peak} in the legend, whether it was empirical modeling (emp), abundance matching (AM), or HOD like in our case. In our case, there is a *downsizing* (Cowie et al. 1996) trend since $z = 1.5$, meaning that the most efficient halo mass (M_h^{peak}) declines from $10^{12.37^{+0.03}_{-0.03}}$ $M_\odot$ at z $\sim$ 1.4 to $10^{12.22^{+0.03}_{-0.04}}$ $M_\odot$ in the lowest z-bin (z $\sim$ 0.35), exhibiting a decline of 0.11 dex. Beyond $z = 1.5$, M_h^{peak} is consistent with no evolution, within the errors. The overall weak evolution in M_h^{peak} with redshift shows that halo mass assembly in halos is tightly linked to the stellar mass assembly.

The relatively small error (0.2$-$0.3 dex) on M_h^{peak} within the highest two z-bins (2.75 $<$ z $<$ 4.5) can be attributed to the new "smooth-z" model (introduced in §3.4.8) used in this study. This model, which connects the HOD parameters across redshift discouraging any sharp inter-z-bin 'jumps', leverages the high constraining power of the low-z data with a much wider dynamical range in stellar mass than at high-z. As a result, we could shrink the uncertainty on $\log_{10}(M_h^{peak}$ [$M_\odot$]) at high-z by up to a factor of $\sim$ 6.5 compared to Shuntov et al. (2022), who applied the same HOD (L11) as we did on a similarly large effective area of $\sim$ 1.27 deg^2, but fitted the individual z-bins separately as in our "discrete-z" model. For direct comparison, we also show our "discrete-z" M_h^{peak} results with filled turquoise circles in Figure 3.7, where similar to Shuntov et al. (2022), our M_h^{peak} values have much larger uncertainties compared to that from "smooth-z" within the highest two z-

bins ($2.75 < z < 4.5$). The smaller errors in our case compared to Shuntov et al. (2022) could be attributed to our slightly larger effective area (~ 1.61 deg^2 as opposed to ~ 1.27 deg^2) and due to the reduced impact of cosmic variance by utilizing two fields instead of one.

Overall, our results generally agree with the observationally constrained M_h^{peak} in the literature as can be seen in Figure 3.7. However, at $z < 1.5$, where the wider dynamical range in stellar masses, than at higher redshifts, provides stronger constraining power for all studies, there seems to be a systematic offset between our M_h^{peak} and others': Our M_h^{peak} remains consistently higher by $\sim 0.10-0.15$ dex at $z < 1.5$. We checked if this discrepancy washes away using the stellar masses from derived using `FAST` instead of Dense Basis, but that is not the case: At $z < 1.5$, M_h^{peak} values derived using the stellar masses from the two codes have a median offset of ~ 0.018 and a mean offset of 0.006. Below, we further delve into our investigation of this systematic discrepancy.

Effect of cosmic variance on the redshift evolution of the peak halo mass at z < 1.5

All of the previous M_h^{peak} results in the literature at these redshifts are based either purely on the COSMOS field (Leauthaud et al. 2012; Cowley et al. 2018; Legrand et al. 2019; Weaver et al. 2022) or rely heavily on the COSMOS field despite having data from other fields in the case of Behroozi et al. (2019)[4]

[4]Besides COSMOS-based SMFs, at $0.05 < z < 1$, Behroozi et al. (2019) utilize the SMFs from Moustakas et al. (2013) which is based on 5 PRIMUS fields (CDFS, COSMOS, ELAIS-S1, XMM-SXDS, XMM-CFHTLS) covering a large total area of 5.5 deg^2. However, due to the photometry problems in Moustakas et al. (2013) as discussed in the appendix (C2) of Behroozi et al. (2019), they exclude the constraints from Moustakas et al. (2013) on the massive end ($M_* > 10^{11}$ $M_\odot$) beyond $z > 0.2$. This is also the mass regime where the COSMOS SMF deviates the most from UDS SMF (COSMOS SMF is much over-dense) as shown in Fig.

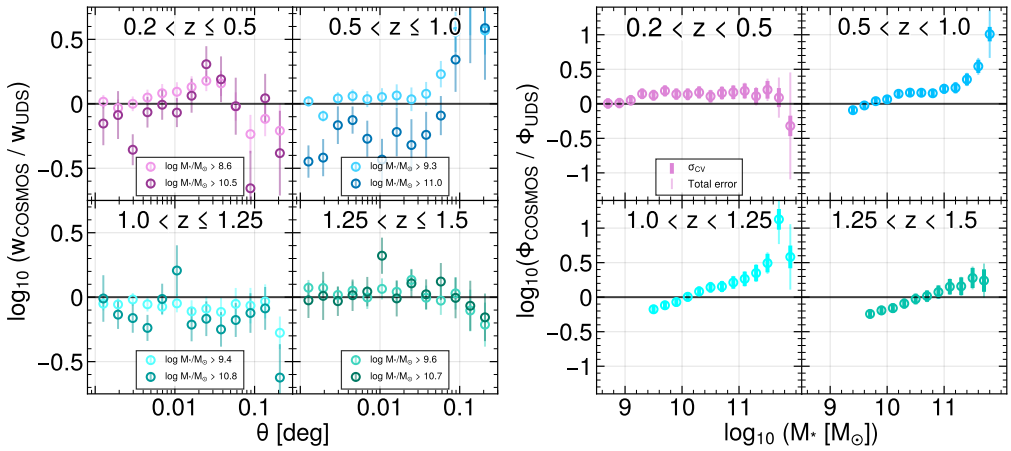

Figure 3.8: Left: Measured clustering offsets between the UDS and COSMOS fields at z $<$ 1.5. Each of the four panels (four z-bins as indicated at each panel's top) shows the offset in the two-point angular correlation functions corrected for integral constraints. Within each panel (z-bin), the offset is shown for the lowest and the highest stellar mass threshold samples (as described in Table 3.1) with the legend indicating the thresholds. Right: Measured SMF offsets between the UDS and COSMOS fields in the four z-bins indicated at each panel's top. Here, we highlight the error budget attributed to cosmic variance (σ_{CV}) with thicker error bars, whereas the total error including Poisson error ($\sigma_{Poisson}$), and the error due to uncertainty in survey area (σ_{Area}) is shown with thinner error bars.

and Moster et al. (2018).

One of the key advantages of this work is the introduction of the UDS as second field. This difference seems a likely explanation for the offset in our $z < 1.5$ M_h^{peak} results from those in the literature. To test this possibility we reanalyzed the data from COSMOS and UDS fields alone to see if our COSMOS-only M_h^{peak} agrees with the literature. If the M_h^{peak} resulting from the clustering and SMF measurements in our COSMOS dataset fitted with the HOD model agreed well with the COSMOS-dominated results in the literature, it would be a clear indication that our UDS+COSMOS M_h^{peak} is driven upward due to the addition of UDS field. This is exactly what we see in Figure 3.7, where our COSMOS-only M_h^{peak} results, shown with the black '+', lie almost exactly atop the M_h^{peak} points from other studies. Our UDS-only M_h^{peak} results shown with the black '×' are much higher, which as speculated, drives the UDS+COSMOS M_h^{peak} upward compared to the COSMOS-only case.

We delved further into the cause of these field to field differences in the clustering and SMF in Figure 3.8. In the left part of the figure, we show the clustering offsets for the lowest and highest mass threshold samples within each of our four z-bins out to $z = 1.5$. Clustering does not change significantly between the two fields except for the highest mass threshold ($M_* > 10^{11} M_\odot$) sample at $0.5 < z < 1.0$. Therefore, it is hard to infer any resulting change on M_h^{peak}.

However, the SMF offset shown in the right part of Figure 3.8 shows that the COSMOS SMFs are systematically over-dense (higher normalization) for most of the stellar masses out to $z = 1.5$. Below, we elaborate on the

3.8.

convincing evidence for the generally over-dense nature of the COSMOS SMF compared to the UDS SMF:

• $0.2 < z < 0.5$: At $M_* \gtrsim 10^{9.2}$ $M_\odot$, the COSMOS SMF offset remains consistently over-dense by $\sim 0.1-0.2$ dex, except for the high-mass end, where the offset diminishes within errors.

• $0.5 < z < 1.0$: At $M_* \gtrsim 10^{10}$ $M_\odot$, the COSMOS SMF is again over-dense by ~ 0.2 dex until $M_* \sim 10^{11.2}$ $M_\odot$, beyond which it abruptly increases in over-density to reach an offset of ~ 1 dex at $M_* \sim 10^{11.8} M_\odot$.

• $1.0 < z < 1.25$: At $M_* \sim 10^{10.3}$ $M_\odot$, the COSMOS SMF starts to become over-dense steadily until $M_* \sim 10^{11.5}$, before rising sharply to reach an offset of ~ 1.1 dex at $M_* \sim 10^{11.7} M_\odot$.

• $1.25 < z < 1.5$: Only at $M_* \gtrsim 10^{10.9}$ $M_\odot$, the COSMOS SMF starts to become over-dense, steadily reaching a maximum offset of ~ 0.3 dex at $M_* \sim 10^{11.5} M_\odot$.

COSMOS is found to be under-dense than UDS only within $1.0 < z < 1.25$ and $1.25 < z < 1.5$ bins at intermediate masses at $M_* \lesssim 10^{9.9}$ $M_\odot$ and $M_* \lesssim 10^{10.3}$ $M_\odot$, respectively.

The over-dense nature of COSMOS is not surprising as several massive structures, out of which five are $M_* > 10^{13} M_\odot$ have long been reported (Scoville et al. 2007b) in the field at $z < 1.1$. There is a super-structure at $z \sim 0.8-0.9$ (Darvish et al. 2014). McLeod et al. (2021) have also shown that UDS SMFs are less dense than COSMOS SMFs at $0.25 \leq z < 1.25$.

The effect this has on our HOD-modeling-derived M_h^{peak} can be understood by considering the SMFs and the adopted HMF. Due to the higher normalization of COSMOS, more galaxies need to be assigned to halos. The additional galaxies in COSMOS can be accommodated in two ways: (i) be

hosted as satellites in high mass halos, thus increasing the slope of satellite occupation function, α_{sat}, OR (ii) by decreasing the halo mass cutoff to host central galaxies, which allows less massive halos to also host galaxies.

According to the first way, the clustering should increase on all scales but we do not see that in Figure 3.8. Therefore, the second option is more likely - the halo mass cutoff to host central galaxies lowers to accommodate more galaxies as centrals in COSMOS. Inevidently, this lowers the average halo mass essentially captured by the characteristic halo mass, the M_1 parameter in the L11 HOD. Since M_1 in COSMOS is lower, the effect of which is to shift the SHMR leftwards (towards lower halo masses), the M_h^{peak} also shifts to lower halo masses as we see in Figure 3.7.

We see a clear signature of the effect of the SMFs requiring the HOD to put galaxies at lower halo masses in these z ranges in Figure 3.4. The best fit $w(\theta)$ models are all offset systematically low compared to the data in the $1.0 < z < 1.25$ and the $1.25 < z < 1.5$ bins, particularly on large scales.

3.5.5 Redshift evolution of the total (central+satellite) SHMR

Beyond halo masses of about $\sim 10^{12.5}$ $M_\odot$, depending on redshift, the contribution of satellites to the total stellar mass budget of the halo starts to become noticeable. The L11 HOD model used in this study allows us to easily calculate the contribution of central and satellites to the total stellar mass locked up in a halo of a given mass through Equation 3.11, where we essentially integrate the conditional SMF down to a lower limit on stellar mass.

Similar to Leauthaud et al. (2012) and Shuntov et al. (2022), we integrate down to a stellar mass, $M_* = 10^9$ $M_\odot$ and show the resulting total stellar-to-halo mass ratios, M_*^{tot}/M_h (solid curves) split between centrals (dashed curves) and satellites (dash-dotted curves) in Figure 3.9. We also estimate the 1σ errors on $M_{*,\,total}/M_h$ using the MCMC samples and show them as shaded regions. Furthermore, as in Figure 3.6 (and described in §3.5.3), we shade the halo masses not directly probed in our analysis in gray.

Below $M_h \sim 10^{12}$ $M_\odot$, the satellite contribution remains subdominant compared to the central galaxy. Beyond $M_h \sim 10^{12}$ $M_\odot$, the central galaxies account for a smaller and smaller fraction of the total stellar mass in halos, and the satellites take over beyond $M_h \sim 10^{13.6}$ $M_\odot$ (group/cluster scale halos), with a slight dependence on redshift. Interestingly, within the halo masses probed, the peak total SHMR is dictated almost entirely by centrals and occurs around the M_h^{peak} values in §3.5.4 ($\sim 10^{12.3}$ $M_\odot$). This is unsurprising as M_h^{peak} just represents the peak of $M_*^{mean\,cen}/M_h$ which is the mean of $M_*^{total\,cen}/M_h$.

Although, beyond $M_h \sim 10^{13}$ $M_\odot$ at $z > 1.5$, there is a hint of an upturn in M_*^{tot}/M_h which could be due to satellites on their way to merge with the central galaxy. M_*^{tot}/M_h also rises in the 2^{nd} z-bin ($z \sim 0.75$) relative to the 3^{rd} z-bin ($z \sim 1.1$), and then falls back again in the lowest z-bin ($z \sim 0.35$). This indicates that at $z \sim 0.75$, there is an uptick in the satellite accretion onto the most massive galaxies, and by $z \sim 0.35$ those satellites have mostly merged with the massive central galaxies. We discuss this in §3.6.2 as well.

So far for these total stellar mass calculations we have used a lower stellar mass limit of the integration of $M_* = 10^9 M_\odot$. Ideally, we would like to extend the lower limit down to dwarf galaxies ($10^7 - 10^8$ $M_\odot$) and it is possible to

extrapolate our fitted HOD models to such low stellar masses. However, our data does not provide constraints below the stellar mass completeness limit, which is already above 10^9 $M_\odot$ for all but our lowest redshift samples. However, as has previously been noted (Leauthaud et al. 2012; Shuntov et al. 2022), as long as the low mass slope of the satellite conditional mass function (seeFigure C1 in Appendix §C) is sufficiently shallow, lowering the integration limit does not change the satellite contribution significantly. This is the case for all but our highest z-bin.

3.5.6 Redshift evolution of the M_*^{tot}/M_h at fixed halo masses

We present the evolution of M_*^{tot}/M_h (or integrated SFE) in halos of fixed masses - M_h^{peak} (z), $10^{11.5}$, 10^{12}, $10^{12.5}$, and $10^{13.5}$ $M_\odot$ in Figure 3.10. M_h^{peak} (z) halos are always the most efficient by definition which closely trace the $10^{12.5}M_\odot$ halos (in red) indicating that the M_h^{peak} does not deviate much from $\sim 10^{12.5}M_\odot$ with redshift, as already described in §3.5.4. At $z < 2$, the slightly lower $M_h = 10^{12}M_\odot$ (in green), and the lowest $M_h = 10^{11.5}M_\odot$ (in blue), continue to rise in SFE with decreasing redshift. However, both remain consistently lower in SFE than the $M_h = 10^{12.5}M_\odot$ (or M_h^{peak}) halos by factors of ~ 1.5 and ~ 4, respectively.

On the other side of M_h^{peak}, at $M_h = 10^{13.5}M_\odot$ (in yellow), a reverse trend is seen where the SFE continues to decrease with decreasing redshift (barring a slight uptick at $z \sim 0.75$), indicating that the high mass halos do bulk of their star formation early on and become more and more quiescent over time.

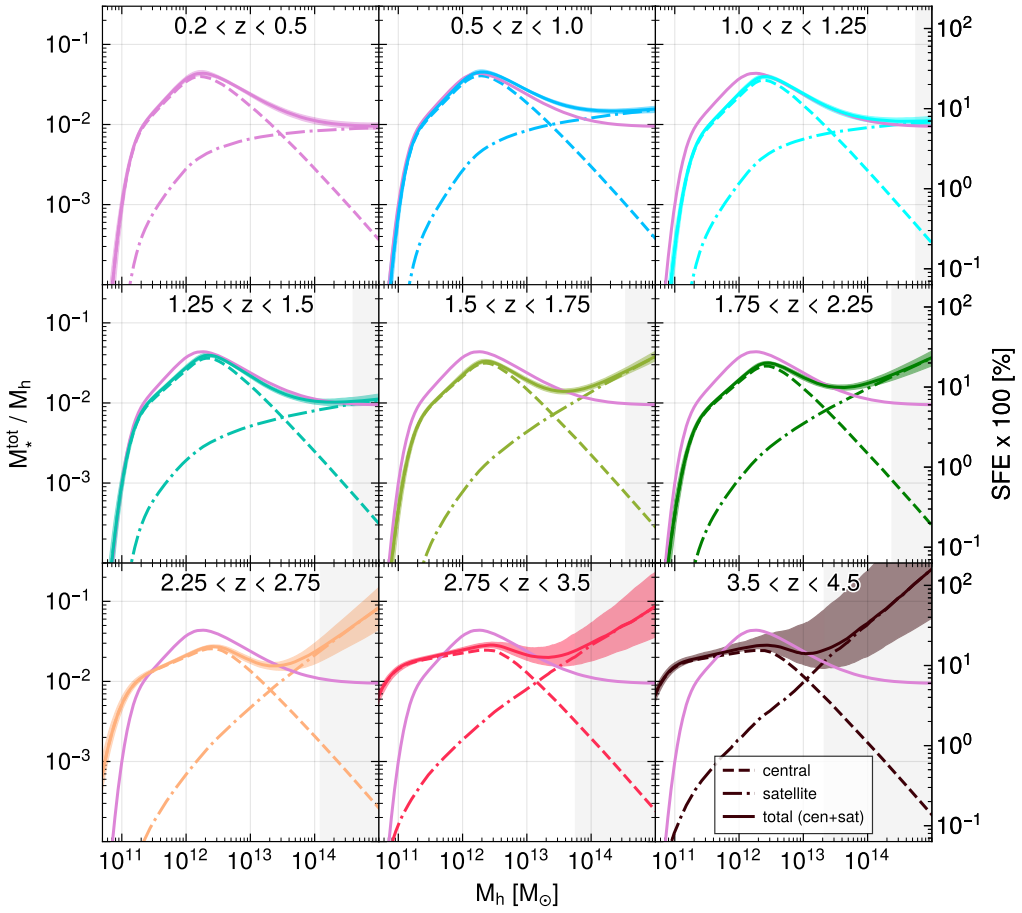

Figure 3.9: Total (central+satellites) stellar mass locked up in halos divided by the halo mass, $M_{*,total}/M_h$ ratio as a function of halo mass within each z-bin (each panel). The solid-colored curve shows the median $M_{*,total}/M_h$ and the shaded envelope shows encapsulates 1-σ error. The dashed curves show the contribution of central galaxies to the $M_{*,total}/M_h$, whereas the dot-dashed curves show the contribution from satellite galaxies. The magenta-colored curve showing the $M_{*,total}/M_h$ of the first z-bin (top-left panel) is over-plotted in all the other panels to highlight its evolution with redshift. As in Figure 3.6, the right-most ordinates indicate the SFE ($= (M_*^{\rm mean\,cen}/M_h)/f_b$) in percentage. The gray shaded regions indicate the halo masses not directly probed in our analysis, as in Figure 3.6.

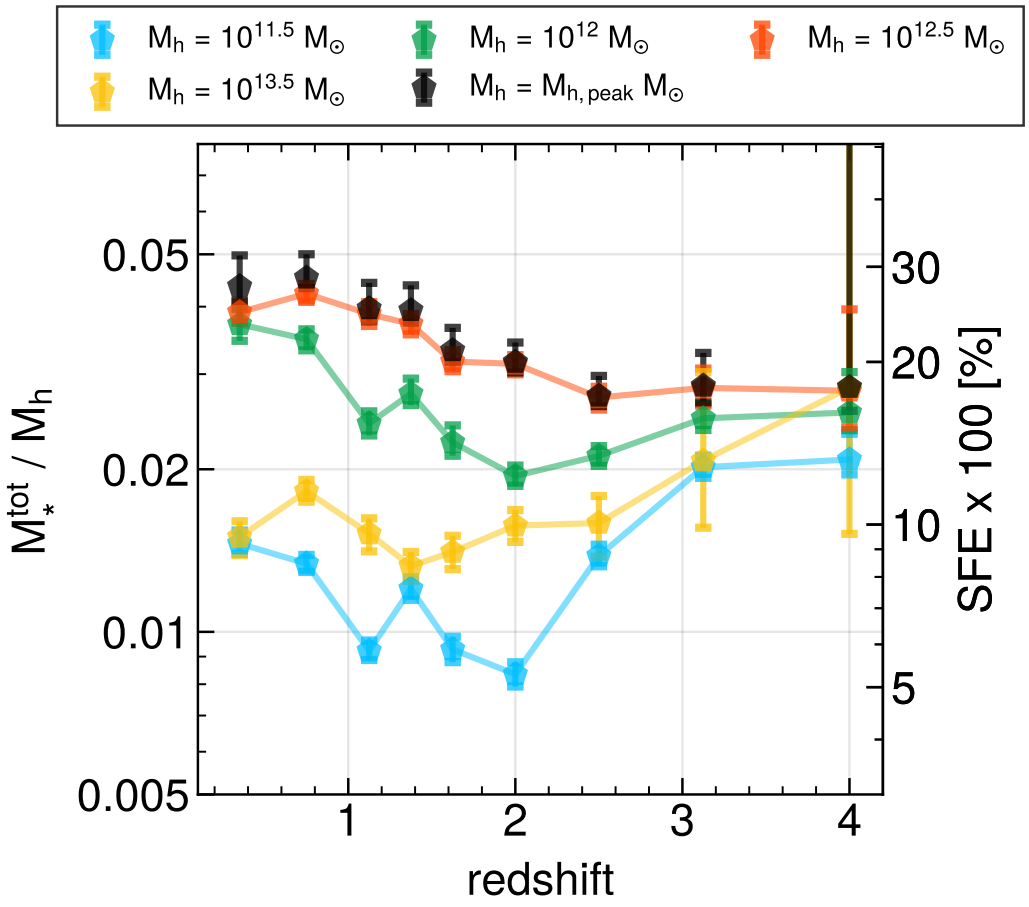

Figure 3.10: Redshift evolution of the total stellar mass-to-halo mass ratio, M_*^{tot}/M_h in halos of fixed mass. Different halo masses are indicated in the legend, and the SFE ($= (M_*^{\text{mean cen}}/M_h)/f_b$) in percentage is shown on the right y-axis.

3.5.7 Redshift evolution of SHMR along the halo merger trees

So far in this paper, halo masses or SHMR (total or otherwise) have been inferred from HOD fitting of large samples of galaxies in fixed z-bins. We have then made comparisons of the evolution of various galaxy and halo properties with redshift either at fixed galaxy or halo mass. However, galaxies and halos are both evolving in mass with redshift and so by comparing at fixed mass we are not comparing the 'same' halos or galaxies at each redshift. Ideally we would like to be tracing the halo mass along the merger trees of the same halos. Using our fitted SHMR, we can retrieve the stellar masses of halos along their merger histories from a dark-matter-only simulation. For this analysis, we retrieved the publicly available data from the dark-matter-only TNG300-1-Dark simulation (Pillepich et al. 2018). Starting from $z \sim 0.2$ (snapshot number 84), the lowest redshift in our galaxy sample, we trace the sub-halo mass history along the main progenitor branch for the 2000 most massive halos. This results in final halo masses at $z \sim 0.2$ ranging from $\sim 10^{12}$ $M_\odot$ to $\sim 10^{16}$ $M_\odot$ and their halo mass histories.

We divide the final halo masses in bins of 0.2 dex and within each bin, trace the mean of the halo mass histories. We convert these mean halo mass histories into M_*^{tot}/M_h histories using the M_*^{tot}/M_h at fixed halo masses shown in Figure 3.9. By calculating M_*^{tot}/M_h for each mean progenitor halo mass along the merger trees at medians of our z-bins, we trace the evolution of M_*^{tot}/M_h as the halo masses grow along the merger trees. The resulting M_*^{tot}/M_h along the merger trees are shown in the left panel of Figure 3.11. Similarly, we convert the $M_*^{\text{mean cen}}/M_h$ at fixed halo masses shown in Fig-

ure 3.6 into $M_*^{\rm mean\,cen}/M_h$ along the merger trees and show the results in the right panel of Figure 3.11. In Figure 3.11, we indicate the mean halo masses with the sizes of the circles on the legend at the top. The final mean log halo mass, $\log_{10}$ (M_h [$M_\odot$]) and the mean central galaxy stellar masses, $\log_{10}$ ($<M_{*cen}>$ [$M_\odot$]) at $z = 0.35$ (median of the lowest z-bin) are highlighted in different shades of red, as shown in the legend at the bottom, to guide the eye to trace their evolution.

Perhaps the most visually striking aspect of Figure 3.11 is the change in the hierarchy of SFE in halos with redshift. The highest mass halos have the highest SFE at $z \gtrsim 3$ and the order gets completely flipped by $z = 0.35$, where the lowest mass halos have the highest SFE. For instance, the highest average mass of halos ($M_h = 10^{13.8}$ $M_\odot$) at $z = 0.35$ shown in the darkest shade of red, accumulates most of its stellar mass early on at $z \gtrsim 3$, both in terms of the total stellar mass and that of only centrals, with SFE reaching ∼18% and ∼12%, respectively. Below $z \sim 3$, this halo is found to be continually declining in SFE, getting completely quenched by $z = 0.35$.

This is in stark contrast to the lowest mean halo mass ($M_h = 10^{12.27}$ $M_\odot$) considered at $z = 0.35$ which started building its stellar content much more recently at $z \lesssim 2$, with its SFE peaking (∼28% for the total stellar mass buildup in the left panel) only at $z \sim 0.35$. This is a clear sign of *downsizing* (Cowie et al. 1996) manifesting in the decline of the halo mass scale over time at which the SFE peaks - the highest mass halos assemble their stellar mass first followed by lesser and lesser massive halos.

Comparing the evolution of $M_*^{\rm mean\,cen}/M_h$ (right panel) with M_*^{tot}/M_h (left panel), the increasing contribution of satellites with decreasing redshift is apparent and unsurprising as demonstrated by the overall increase in satellite

fraction toward low redshift as seen in Figure D1. It also shows how the stellar mass growth of the most massive halos at low redshift is being driven almost entirely by satellite galaxies, with the opposite true of the lowest mass halos.

3.6 Discussion

The evolution of SHMR has imprints of various physical processes integrated over halos' lifetimes such as cold gas flows towards the centers of halos giving rise to star formation, stellar/dark matter halo mass growth due to mergers, and on the flip side, the feedback effects that heat up or eject the gas out of galaxies quenching star formation. In light of our SHMR results in §3.5, here we try to build an intuition of the galaxy-halo connection and in turn its connection to galaxy formation.

3.6.1 The most star-forming halo mass - $\mathbf{M}_h^{peak}$

We shall first look at the redshift evolution of the halo mass with peak SFE, M_h^{peak} as shown in Figure 3.7 (black line). This peak mass where the central galaxies peak in SFE equally applies to the total SHMR where the peak is dominated by the centrals.

Downsizing in $\mathbf{M}_h^{peak}$

Interestingly, M_h^{peak} drops by 0.11 dex from z $\sim$ 1.4 to z $\sim$ 0.35, but remains consistent with no evolution beyond $z = 1.5$, as described in §3.5.4. Previous M_h^{peak} results in the literature (shown in Figure 3.7) had seen *downsizing* trends unanimously stronger than our ours ($\sim$0.26 dex as opposed to 0.11 dex in our case) at $z < 1.5$ and much stronger than ours in some studies

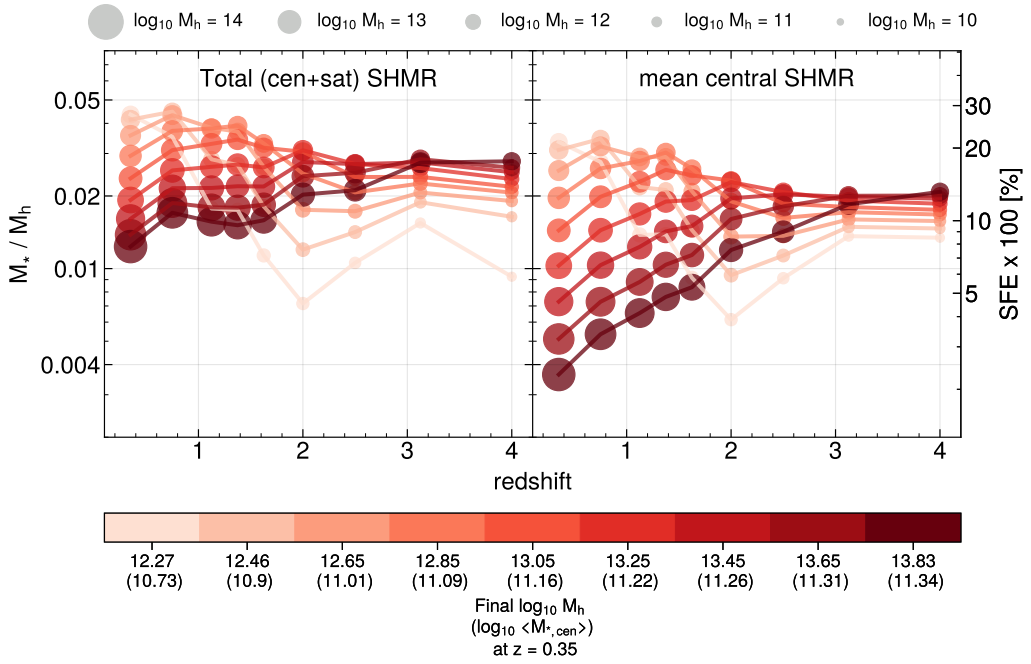

Figure 3.11: Left: Evolution of the total (centrals+satellites) stellar-halo mass ratio along the main progenitor branches (most massive branch) of z = 0.35 (median of the lowest z-bin in our analysis) dark matter halos in the TNG300-Dark-1 simulation (Pillepich et al. 2018). The total stellar mass in a given TNG300-Dark-1 halo mass (average in a 0.2 dex bin), depicted by the sizes of the red circles, is calculated using the fitted HOD model parameters at each redshift as described in §3.5.5. The shading of the circles and the lines connecting them reflect the final halo mass at $z = 0.35$ (as shown in the legend), where lighter shades of red depict lower final halo masses. Right: Evolution of the mean central stellar-halo mass ratio along the main progenitor branches of $z = 0.35$ dark matter halos in the TNG300-Dark-1 simulation as in the left panel. Again, this is evaluated at the best fit HOD parameters. As in the left panel, the colors of the circles represent the final halo masses (at $z = 0.35$), and their sizes capture the current halo masses.

like Shuntov et al. (2022)) at $z > 1.5$. However, since these earlier results had been heavily based on a single field of COSMOS, it was not clear if the *downsizing* trends in M_h^{peak} were commonplace in the Universe. The use of two large fields in our work (UDS in conjunction with COSMOS) greatly mitigated the effect of cosmic variance on M_h^{peak} evolution, and in doing so we found a much weaker *downsizing* signal throughout the redshifts probed.

Even though more widely separated and large fields are needed to constrain the M_h^{peak} out high redshift with better confidence, its *downsizing* trend might not be as strong or as universal as deduced from the COSMOS-based data. Peak integrated star formation occurs in a narrow channel of M_h^{peak} $\sim 10^{12.2}$ $M_\odot$ $-$ $10^{12.4}$ $M_\odot$ out to $z = 4.5$, indicating tight coupling between halo mass assembly and galaxy stellar mass assembly, and that quenching strongly depends on halo mass. This is also supported by the small scatter in stellar mass at a fixed halo mass, σ_{logM_*} ($\sim$ 0.25 dex), again out to $z = 4.5$.

Implications on the strength of AGN feedback

The M_h^{peak} observed at any redshift results from the star formation history that preceded in a way to result in the most stellar mass per halo mass (M_*/M_h). Therefore, however minimal, the change in the M_h^{peak} holds clues to the strength of star formation regulating AGN feedback (Silk & Rees 1998; Benson et al. 2003) especially at the high-mass end. AGN feedback supposedly acts on the high mass halos ($\gtrsim 10^{12} M_\odot$) at least in the redshift range considered in this study. There are two broad classes of AGNs in the literature: i- *'jet-mode'* AGN: they expel their energy primarily through mechanical jets as they are radiatively inefficient, have low optical/ultraviolet

(UV)/X-ray luminosities and low accretion rates with a thick toroidal disk. They generally host more massive black holes and are found more in redder, more massive quiescent galaxies and richer environments; ii- *'radiative-mode'* AGN: they are quite the opposite as their radiatively efficient thin disks allow them to expel their energy primarily through radiation instead of jets and are therefore much more luminous in optical/ultraviolet (UV)/X-ray wavelengths (Cielo et al. 2018; Heckman & Best 2014).

There is accumulating observational evidence that *'jet-mode'* AGN is responsible for much of the quenching in massive galaxies. For instance, Kondapally et al. (2023) analyzed a sample of $\sim$ 1000 radio-excess AGNs within $0.5 < z < 2.5$ from the radio catalogs of LoTSS Deep Fields DR1 (Sabater et al. 2021; Tasse et al. 2021) and found that low-excitation (jet-mode) AGNs dominate the feedback across the full redshift range considered. On the contrary, *'radiative-mode'* AGNs are actually found to coexist with high levels of star formation. Florez et al. (2020) compared a sample of X-ray luminous AGNs with the galaxies that did not host an X-ray luminous AGN within $0.5 < z < 3$. They found X-ray luminous AGNs to be hosting much higher (by a factor of $\sim 3-10$) star-formation rates across the full range of redshifts and stellar mass considered. Florez et al. (2021) find these results to be in general agreement with the simulations SAG (Cora et al. 2018) and IllustriTNG (Pillepich et al. 2018; Weinberger et al. 2017), while they see the opposite in their comparison to SIMBA (Davé et al. 2019). i.e. in SIMBA, high X-ray luminous AGNs are in fact found to have much lower levels of star formation and many are even quenched.

Therefore, for the following, we assume that it is the 'jet-mode' AGN feedback which is the dominant suppressor of star formation in massive galaxies

and simply call it AGN feedback. Moving from high redshift to low redshift, if the AGN feedback had been weaker recently, higher mass halos become the most efficient in star formation, shifting M_h^{peak} toward higher masses. On the other hand, stronger AGN feedback in the recent past of the observed redshift manifests in a lower M_h^{peak} at the observed redshift than otherwise.

Given the decrease in M_h^{peak} with time, albeit minimal, at least in our case, it is likely that the AGN feedback progresses by quenching the massive galaxies in decreasing order of stellar mass. This is also supported by the hierarchy of quenching as seen in Figure 3.11, where the most massive galaxies quench first followed by lesser and lesser massive galaxies. By z $\sim$ 1, most of the massive galaxies in massive halos have quenched, whereas the lower mass halos are now peaking in SFE, thereby shifting the M_h^{peak} to its lowest value at $z \sim 0.35$.

3.6.2 Stellar-Halo mass growth

Figure 3.11 beautifully captures the *downsizing* trend in stellar-halo mass growth, i.e. they grow in decreasing order of mass. Group and cluster scales halos ($M_h \gtrsim 10^{13.5}$) at $z \sim 0.3$ hosting quenched galaxy population accumulated the bulk of the stellar mass early on, at $z \gtrsim 2$. Below $z \lesssim 2$, they are found with residual levels of star formation. On the other hand, the less massive halos ($M_h \lesssim 10^{12.7}$ $M_\odot$) at $z \sim 0.3$ only started to build up their stellar mass at $z \lesssim 2$, reaching peak levels of star formation at $z \sim 0.3$. Now, we will examine the stellar-halo mass growth in these two regimes more closely.

Low mass regime

Throughout the redshift range considered in this study ($0.2 < z < 4.5$), on the lower side of the peak stellar-halo mass ratio (either of the M_h^{peak} or the peak in M_*^{tot}/M_h), stellar mass increases with increasing halo mass. As time passes, halos are expected to grow via merging with much smaller halos with lower stellar mass (Leauthaud et al. 2012; Legrand et al. 2019). Therefore, with an increase in halo mass, there will be a mild increase in stellar mass, but the M_*^{tot}/M_h should decline. Mergers alone cannot explain the steep rise in M_*^{tot}/M_h at $z \lesssim 2$; secular 'in-situ' star formation has to be invoked. At $z \gtrsim 2$, the low mass slope becomes shallower indicating lower levels of 'in-situ' star formation which is not surprising as star-forming efficiencies of low mass halos only pick up at $z \lesssim 2$, as seen through Figure 3.11.

High mass regime

As visible in Figure 3.9, the central galaxies dominate stellar mass within halos below the peak at $M_h \sim 10^{12}$ $M_\odot$. Beyond this mass, in group-scale halos, the central contribution starts to decline and is taken over by the satellites. At first glance, this might indicate that the central galaxies suffer from a decline in SFE because the star-forming ability of the massive halos ($M_h \gtrsim 10^{12}$ $M_\odot$) now gets distributed between centrals and satellites. However, upon close inspection, this does not seem to be the case, at least at $z \lesssim 1.5$. In the "satellite dominated" regime (beyond $M_h \sim 10^{12}$ $M_\odot$), even if all the stellar mass in satellites is somehow transferred to centrals, i.e. the solid M_*^{tot}/M_h curve becomes the $M_{*,\,cen}/M_h$ curve, the peak SFE remains at $M_h \sim 10^{12}$ $M_\odot$, i.e there is no second peak on the high mass end. This means that centrals in group-scale halos do not exceed the SFE obtained at

$M_h \sim 10^{12}$ $M_\odot$, *not* because the satellites take away their share of SFE, but rather that beyond $M_h \sim 10^{12}$ $M_\odot$, there is a global decline in SFE of halos, again at least out to $z \sim 1.5$. Whatever quenching mechanism is at play in group-scale halos impacts both the central and satellites. This is in general agreement with Leauthaud et al. (2012) and Shuntov et al. (2022).

Interestingly at z > 1.5, M_*^{tot}/M_h seems to rise again beyond $M_h \sim 10^{13}$ $M_\odot$ but it is hard to ascertain if it indeed overtakes the peak M_*^{tot}/M_h ratio obtained at $M_h \sim 10^{12}$ $M_\odot$. This is because the upper limit on the halo mass probed by our analysis does not extend to high enough halo masses disallowing us to confidently conclude if M_*^{tot}/M_h indeed overtakes the peak of $M_h \sim 10^{12}$ $M_\odot$. Even if we extrapolate our models beyond the maximum halo mass probed at each redshift, the highest physically possible halo mass also decreases with redshift, thus leaving a very faint possibility of M_*^{tot}/M_h having a second peak at higher halo masses $M_h \gtrsim 10^{13}$ $M_\odot$.

At $z > 1.5$, regardless if there is a second peak in M_h or not, there is an upturn in M_*^{tot}/M_h beyond $M_h \sim 10^{13}$ $M_\odot$ due to satellites. The presence of this upturn and the lack thereof at $z < 1.5$ possibly conveys an important message: At $z > 1.5$, satellite accretion onto the massive halos is more prevalent which is plausible as the Universe was indeed more merger-ridden back then (Governato et al. 1999; Wetzel et al. 2009; Stewart et al. 2009). However, at $z < 1.5$, many of these satellites might have merged onto central galaxies, hence the suppression in the satellite contribution to M_*^{tot}/M_h for high halo masses.

3.7 Summary

We present a comprehensive analysis of the stellar-to-halo mass relationship (SHMR) of central and satellite galaxies since $z = 4.5$. To derive the SHMRs, we fit the halo occupation distributions (HOD) models using measurements of galaxy clustering and abundance in two large legacy UDS and COSMOS fields, covering a total effective area of ~ 1.61 deg^2. This is the first time that two photometric datasets (FENIKS v1 (UDS; Zaidi et al. (2024)) and UVISTA DR3 ultra-deep stripes (COSMOS; Muzzin et al. (2013)) of such area coverage and depth have been homogeneously integrated to trace stellar mass assembly within halos.

We also introduce a novel HOD fitting framework that allows the high constraining power of low-z data, due to more than ~ 3 dex of dynamical range in stellar mass where we are complete, to communicate with high-z ($z \gtrsim 2.5$), where we are complete only for $\lesssim 1$ dex. Data in all z-bins are jointly fitted where HOD parameters between adjacent z-bins are connected via continuity (smoothing) priors. This is motivated by the simple physical fact that galaxy clustering and abundance measurements between adjacent z-bins evolve from one to another, instead of being independent of each other. Therefore, HOD parameters between neighboring z-bins should not see abrupt changes. This allowed dramatically reducing the uncertainties in SHMR at $z > 2.5$, giving a much clearer picture of SHMR evolution out to $z = 4.5$. Our findings are as follows:

• We show that the halo mass with peak integrated star-forming efficiency (SFE), M_h^{peak} evolves very little: it stays between $\sim 10^{12.2} M_\odot$ and $\sim 10^{12.4} M_\odot$. At $z < 1.5$, we observe weak *downsizing* by 0.11 dex as opposed to ~ 0.26

dex in the COSMOS-based studies (see §3.5.4). We attribute this difference to the presence of over-dense SMFs of COSMOS due to many reported over-dense structure (see §3.5.4). Beyond $z > 1.5$, we see insignificant evolution in M_h^{peak} as opposed to steep *downsizing* found in other COSMOS-based studies like Shuntov et al. (2022), albeit with large error-bars.

• The total stellar mass locked up in halos less massive than $M_h \sim 10^{12} M_\odot$ is dominated by central galaxies, whereas satellites dominate the stellar mass budget beyond $M_h \sim 10^{13} M_\odot$ in group and cluster scale halos (see §3.5.5).

• The conditional stellar mass function (SMF) of central galaxies depends strongly on halo mass. It peaks at increasingly larger stellar masses for increasing halo masses, and tails off on either end. Satellites' conditional SMFs, however, depend much more weakly on halo mass; satellites are equally scattered across $M_h = 10^{11.5} - 10^{13.5} M_\odot$ at $M_* \lesssim 10^{10}$ $M_\odot$. Overall, the conditional satellite SMFs have a classic SMF shape with a power law at low masses, 'knee' at intermediate masses, beyond which it declines sharply (see §**??**).

• The fraction of mass in satellites, f_{sat} generally increases with decreasing redshift and stellar mass thresholds. f_{sat} never exceeds $\sim 30\%$ attained within the lowest z-bin and stellar mass threshold considered (see §**??**).

• We show in Figure 3.11 the evolution in SFE of halos, for the first time, as they grow along their progenitor merger trees. Star formation occurs hierarchically: most massive halos and galaxies observed within the lowest z-bin ($z \sim 0.35$) accumulate their stellar mass first followed by progressively lighter systems. This hierarchy applies to both central and satellite galaxies (see §3.5.7).

3.8 Acknowledgments

KZ acknowledges the John F. Burlingamme Graduate Fellowship in the Physics & Astronomy Department at Tufts University for support while conducting majority of this work. KZ and DM acknowledge support from the National Science Foundation under grant AST-2009442. Additionally, CP and JAD acknowledge support from the National Science Foundation under grant AST-2009632. K.G. acknowledges support from Australian Research Council Laureate Fellowship FL180100060.

The authors acknowledge the Tufts University High Performance Compute Cluster (https://it.tufts.edu/high-performance-computing) which was utilized for the research reported in this paper.

`Astropy` (Astropy Collaboration et al. 2022), `numpy` (Harris et al. 2020), `halomod` (Murray et al. 2021), `CORRFUNC` (Sinha & Garrison 2020; Sinha & Garrison 2019), `pycorr` (https://github.com/cosmodesi/pycorr)

A Measured Stellar Mass Functions

In addition to the comparison between the SMFs of UDS (FENIKS v1 catalog; Zaidi et al. (2024)) and COSMOS (ultra-deep stripes of UVISTA DR3; Muzzin et al. (2013)) fields in §3.5.4 at $z < 1.5$, we present the raw SMFs for both the individual fields and their combination in Figure A1 The joint use of two, instead of the usual one, homogeneously prepared photometric catalogs covering large sky areas is one of the novelties of this paper, key for mitigating errors due to cosmic variance, σ_{CV}. We highlight the contribution of σ_{CV} to the total error budget for the SMFs in Figure A1 through thick

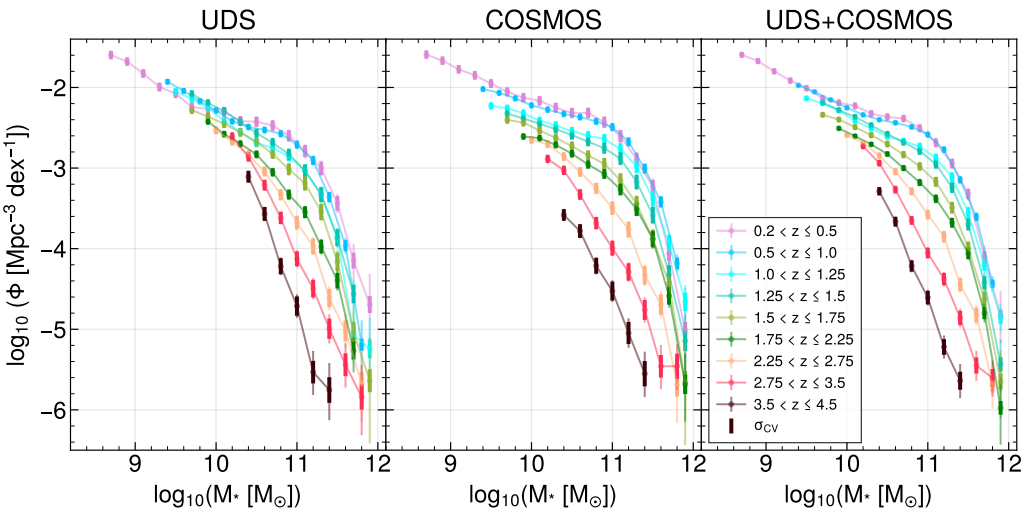

Figure A1: Measured SMFs from the individual UDS and COSMOS fields and their combination (UDS+COSMOS). The thinner error bars indicate the total errors as described in §3.3.3, whereas the contributions of cosmic variance (indicated via thicker bar) decrease due to combining the two fields in the right-most panel.

bars while the total error is shown via thinner bars. The decrease in σ_{CV} can be easily seen through the diminishing extent of these thick bars in the third panel as compared to their extent in the first two panels showing the SMFs of the individual fields.

In Tables A1 - A9, we also present the combined SMFs (UDS+COSMOS), including the percentage errors due to the Poisson ($\sigma_{Poisson}$), cosmic variance (σ_{CV}), and the fixed (3%) uncertainty in the survey area (σ_{Area}).

Table A1: $0.2 < z < 0.5$

$\log_{10}(M_* [M_\odot])_{bin}$	N_{gal}	$\log_{10}(\Phi$ [Mpc^{-3} dex^{-1}])	$\sigma_{Poisson}$	σ_{CV}	σ_{Area}
8.6 - 8.8	5130	$-1.5951^{+0.0392}_{-0.0392}$	$^{+1.41\%}_{-1.41\%}$	8.4%	3%
8.8 - 9.0	4290	$-1.6728^{+0.0396}_{-0.0396}$	$^{+1.54\%}_{-1.54\%}$	8.48%	3%
9.0 - 9.2	3207	$-1.7991^{+0.0403}_{-0.0403}$	$^{+1.78\%}_{-1.78\%}$	8.59%	3%
9.2 - 9.4	2459	$-1.9145^{+0.041}_{-0.041}$	$^{+2.04\%}_{-2.04\%}$	8.71%	3%
9.4 - 9.6	1982	$-2.0081^{+0.0419}_{-0.0419}$	$^{+2.27\%}_{-2.27\%}$	8.88%	3%
9.6 - 9.8	1476	$-2.1362^{+0.0429}_{-0.0429}$	$^{+2.64\%}_{-2.64\%}$	9.04%	3%
9.8 - 10.0	1305	$-2.1896^{+0.0441}_{-0.0441}$	$^{+2.81\%}_{-2.81\%}$	9.29%	3%
10.0 - 10.2	1193	$-2.2286^{+0.0454}_{-0.0454}$	$^{+2.94\%}_{-2.94\%}$	9.59%	3%
10.2 - 10.4	956	$-2.3248^{+0.047}_{-0.0471}$	$^{+3.29\%}_{-3.29\%}$	9.88%	3%
10.4 - 10.6	869	$-2.3662^{+0.0483}_{-0.0483}$	$^{+3.45\%}_{-3.45\%}$	10.14%	3%
10.6 - 10.8	832	$-2.3851^{+0.0494}_{-0.0495}$	$^{+3.53\%}_{-3.53\%}$	10.4%	3%
10.8 - 11.0	631	$-2.5052^{+0.0535}_{-0.0536}$	$^{+4.06\%}_{-4.06\%}$	11.25%	3%
11.0 - 11.2	417	$-2.6851^{+0.0596}_{-0.0596}$	$^{+5.01\%}_{-5.02\%}$	12.41%	3%
11.2 - 11.4	230	$-2.9435^{+0.0672}_{-0.0672}$	$^{+6.8\%}_{-6.82\%}$	13.57%	3%
11.4 - 11.6	96	$-3.323^{+0.0801}_{-0.0803}$	$^{+10.69\%}_{-10.74\%}$	14.74%	3%
11.6 - 11.8	16	$-4.1011^{+0.1388}_{-0.1419}$	$^{+27.57\%}_{-28.38\%}$	15.9%	3%
11.8 - 12.0	3	$-4.8281^{+0.3038}_{-0.3479}$	$^{+67.78\%}_{-78.21\%}$	17.06%	3%

Table A2: $0.5 < z < 1.0$

$\log_{10}(M_* [M_\odot])_{bin}$	N_{gal}	$\log_{10}(\Phi$ [Mpc^{-3} dex^{-1}])	$\sigma_{Poisson}$	σ_{CV}	σ_{Area}
9.3 - 9.5	10173	$-1.9732^{+0.0276}_{-0.0276}$	$^{+1.0\%}_{-1.0\%}$	5.52%	3%
9.5 - 9.7	8434	$-2.0546^{+0.028}_{-0.028}$	$^{+1.09\%}_{-1.09\%}$	5.59%	3%
9.7 - 9.9	6716	$-2.1535^{+0.0284}_{-0.0284}$	$^{+1.23\%}_{-1.23\%}$	5.68%	3%
9.9 - 10.1	5351	$-2.2522^{+0.0293}_{-0.0293}$	$^{+1.38\%}_{-1.38\%}$	5.89%	3%
10.1 - 10.3	4372	$-2.3399^{+0.0302}_{-0.0302}$	$^{+1.52\%}_{-1.52\%}$	6.09%	3%
10.3 - 10.5	3805	$-2.4003^{+0.0312}_{-0.0312}$	$^{+1.63\%}_{-1.63\%}$	6.33%	3%
10.5 - 10.7	3488	$-2.438^{+0.0323}_{-0.0323}$	$^{+1.71\%}_{-1.71\%}$	6.59%	3%
10.7 - 10.9	3102	$-2.489^{+0.0338}_{-0.0338}$	$^{+1.81\%}_{-1.81\%}$	6.96%	3%
10.9 - 11.1	2490	$-2.5844^{+0.0381}_{-0.0381}$	$^{+2.02\%}_{-2.02\%}$	7.98%	3%
11.1 - 11.3	1616	$-2.7722^{+0.0426}_{-0.0426}$	$^{+2.52\%}_{-2.52\%}$	9.0%	3%
11.3 - 11.5	684	$-3.1456^{+0.0485}_{-0.0485}$	$^{+3.89\%}_{-3.9\%}$	10.02%	3%
11.5 - 11.7	237	$-3.6059^{+0.0576}_{-0.0576}$	$^{+6.7\%}_{-6.71\%}$	11.04%	3%
11.7 - 11.9	35	$-4.4365^{+0.0956}_{-0.0965}$	$^{+18.16\%}_{-18.41\%}$	12.06%	3%

Table A3: $1.0 < z < 1.25$

$\log_{10}(M_* [M_\odot])_{bin}$	N_{gal}	$\log_{10}(\Phi$ [Mpc^{-3} dex^{-1}])	$\sigma_{Poisson}$	σ_{CV}	σ_{Area}
9.4 - 9.6	5246	$-2.1343^{+0.0345}_{-0.0345}$	$^{+1.39\%}_{-1.39\%}$	7.22%	3%
9.6 - 9.8	4531	$-2.1979^{+0.0349}_{-0.0349}$	$^{+1.5\%}_{-1.5\%}$	7.31%	3%
9.8 - 10.0	3717	$-2.2839^{+0.0359}_{-0.0359}$	$^{+1.65\%}_{-1.65\%}$	7.53%	3%
10.0 - 10.2	2753	$-2.4143^{+0.0373}_{-0.0373}$	$^{+1.92\%}_{-1.92\%}$	7.82%	3%
10.2 - 10.4	2218	$-2.5082^{+0.0388}_{-0.0388}$	$^{+2.15\%}_{-2.15\%}$	8.13%	3%
10.4 - 10.6	1765	$-2.6074^{+0.041}_{-0.041}$	$^{+2.41\%}_{-2.41\%}$	8.63%	3%
10.6 - 10.8	1541	$-2.6663^{+0.0432}_{-0.0432}$	$^{+2.58\%}_{-2.58\%}$	9.13%	3%
10.8 - 11.0	1326	$-2.7316^{+0.0486}_{-0.0486}$	$^{+2.78\%}_{-2.78\%}$	10.41%	3%
11.0 - 11.2	953	$-2.875^{+0.0561}_{-0.0561}$	$^{+3.29\%}_{-3.29\%}$	12.12%	3%
11.2 - 11.4	550	$-3.1138^{+0.0643}_{-0.0643}$	$^{+4.35\%}_{-4.36\%}$	13.83%	3%
11.4 - 11.6	200	$-3.5531^{+0.0757}_{-0.0757}$	$^{+7.31\%}_{-7.33\%}$	15.53%	3%
11.6 - 11.8	45	$-4.2009^{+0.1027}_{-0.1032}$	$^{+15.9\%}_{-16.07\%}$	17.24%	3%
11.8 - 12.0	10	$-4.8541^{+0.1752}_{-0.1816}$	$^{+35.5\%}_{-37.17\%}$	18.95%	3%

Table A4: $1.25 < z < 1.5$

$\log_{10}(M_* [M_\odot])_{bin}$	N_{gal}	$\log_{10}(\Phi$ [Mpc^{-3} dex^{-1}])	$\sigma_{Poisson}$	σ_{CV}	σ_{Area}
9.6 - 9.8	5318	$-2.1876^{+0.035}_{-0.035}$	$^{+1.38\%}_{-1.38\%}$	7.35%	3%
9.8 - 10.0	4287	$-2.2812^{+0.036}_{-0.036}$	$^{+1.54\%}_{-1.54\%}$	7.58%	3%
10.0 - 10.2	3530	$-2.3655^{+0.0374}_{-0.0374}$	$^{+1.7\%}_{-1.7\%}$	7.88%	3%
10.2 - 10.4	2777	$-2.4697^{+0.0389}_{-0.0389}$	$^{+1.92\%}_{-1.92\%}$	8.22%	3%
10.4 - 10.6	2160	$-2.5789^{+0.0417}_{-0.0417}$	$^{+2.17\%}_{-2.17\%}$	8.85%	3%
10.6 - 10.8	1717	$-2.6785^{+0.0444}_{-0.0444}$	$^{+2.44\%}_{-2.44\%}$	9.47%	3%
10.8 - 11.0	1333	$-2.7885^{+0.0508}_{-0.0508}$	$^{+2.78\%}_{-2.78\%}$	10.96%	3%
11.0 - 11.2	918	$-2.9505^{+0.0594}_{-0.0594}$	$^{+3.35\%}_{-3.36\%}$	12.91%	3%
11.2 - 11.4	468	$-3.2431^{+0.069}_{-0.069}$	$^{+4.73\%}_{-4.73\%}$	14.86%	3%
11.4 - 11.6	164	$-3.6985^{+0.0821}_{-0.0821}$	$^{+8.1\%}_{-8.12\%}$	16.81%	3%
11.6 - 11.8	31	$-4.422^{+0.1178}_{-0.1187}$	$^{+19.37\%}_{-19.67\%}$	18.76%	3%
11.8 - 12.0	3	$-5.4362^{+0.3081}_{-0.3516}$	$^{+67.78\%}_{-78.21\%}$	20.71%	3%

Table A5: $1.5 < z < 1.75$

$\log_{10}(M_* [M_\odot])_{bin}$	N_{gal}	$\log_{10}(\Phi$ [Mpc^{-3} dex^{-1}])	$\sigma_{Poisson}$	σ_{CV}	σ_{Area}
9.6 - 9.8	4075	$-2.3388^{+0.0362}_{-0.0362}$	$^{+1.58\%}_{-1.58\%}$	7.6%	3%
9.8 - 10.0	3554	$-2.3983^{+0.0373}_{-0.0373}$	$^{+1.69\%}_{-1.69\%}$	7.86%	3%
10.0 - 10.2	2865	$-2.4918^{+0.0387}_{-0.0387}$	$^{+1.89\%}_{-1.89\%}$	8.18%	3%
10.2 - 10.4	2279	$-2.5912^{+0.0404}_{-0.0404}$	$^{+2.12\%}_{-2.12\%}$	8.56%	3%
10.4 - 10.6	1769	$-2.7012^{+0.0439}_{-0.0439}$	$^{+2.41\%}_{-2.41\%}$	9.34%	3%
10.6 - 10.8	1307	$-2.8327^{+0.0474}_{-0.0474}$	$^{+2.8\%}_{-2.8\%}$	10.11%	3%
10.8 - 11.0	933	$-2.9791^{+0.0549}_{-0.0549}$	$^{+3.33\%}_{-3.33\%}$	11.83%	3%
11.0 - 11.2	600	$-3.1708^{+0.065}_{-0.065}$	$^{+4.16\%}_{-4.17\%}$	14.06%	3%
11.2 - 11.4	287	$-3.4911^{+0.0766}_{-0.0766}$	$^{+6.07\%}_{-6.08\%}$	16.28%	3%
11.4 - 11.6	90	$-3.9947^{+0.0945}_{-0.0947}$	$^{+11.05\%}_{-11.11\%}$	18.51%	3%
11.6 - 11.8	21	$-4.6268^{+0.1378}_{-0.1396}$	$^{+23.83\%}_{-24.37\%}$	20.74%	3%
11.8 - 12.0	2	$-5.6479^{+0.3779}_{-0.4597}$	$^{+83.88\%}_{-103.28\%}$	22.96%	3%

Table A6: $1.75 < z < 2.25$

$\log_{10}(M_* [M_\odot])_{bin}$	N_{gal}	$\log_{10}(\Phi$ [Mpc^{-3} dex^{-1}])	$\sigma_{Poisson}$	σ_{CV}	σ_{Area}
9.8 - 10.0	5870	$-2.5075^{+0.0305}_{-0.0305}$	$^{+1.31\%}_{-1.31\%}$	6.22%	3%
10.0 - 10.2	4695	$-2.6045^{+0.0316}_{-0.0316}$	$^{+1.47\%}_{-1.47\%}$	6.47%	3%
10.2 - 10.4	3720	$-2.7056^{+0.033}_{-0.033}$	$^{+1.65\%}_{-1.65\%}$	6.79%	3%
10.4 - 10.6	2821	$-2.8257^{+0.0363}_{-0.0363}$	$^{+1.9\%}_{-1.9\%}$	7.56%	3%
10.6 - 10.8	1887	$-3.0003^{+0.0397}_{-0.0397}$	$^{+2.33\%}_{-2.33\%}$	8.33%	3%
10.8 - 11.0	1248	$-3.1799^{+0.0464}_{-0.0465}$	$^{+2.87\%}_{-2.87\%}$	9.86%	3%
11.0 - 11.2	785	$-3.3812^{+0.0552}_{-0.0552}$	$^{+3.63\%}_{-3.63\%}$	11.8%	3%
11.2 - 11.4	396	$-3.6784^{+0.065}_{-0.065}$	$^{+5.15\%}_{-5.15\%}$	13.74%	3%
11.4 - 11.6	164	$-4.0613^{+0.0777}_{-0.0778}$	$^{+8.1\%}_{-8.12\%}$	15.68%	3%
11.6 - 11.8	30	$-4.799^{+0.1155}_{-0.1166}$	$^{+19.71\%}_{-20.02\%}$	17.62%	3%
11.8 - 12.0	2	$-5.9751^{+0.3743}_{-0.4567}$	$^{+83.88\%}_{-103.28\%}$	19.56%	3%

Table A7: $2.25 < z < 2.75$

$\log_{10}(M_* [M_\odot])_{bin}$	N_{gal}	$\log_{10}(\Phi$ [Mpc^{-3} dex^{-1}])	$\sigma_{Poisson}$	σ_{CV}	σ_{Area}
9.9 - 10.1	4968	$-2.5889^{+0.0359}_{-0.0359}$	$^{+1.43\%}_{-1.43\%}$	7.56%	3%
10.1 - 10.3	3953	$-2.6882^{+0.0371}_{-0.0371}$	$^{+1.6\%}_{-1.6\%}$	7.84%	3%
10.3 - 10.5	2749	$-2.8459^{+0.0407}_{-0.0407}$	$^{+1.93\%}_{-1.93\%}$	8.67%	3%
10.5 - 10.7	1739	$-3.0448^{+0.0457}_{-0.0457}$	$^{+2.43\%}_{-2.43\%}$	9.8%	3%
10.7 - 10.9	993	$-3.2881^{+0.052}_{-0.052}$	$^{+3.22\%}_{-3.22\%}$	11.13%	3%
10.9 - 11.1	511	$-3.5767^{+0.0636}_{-0.0636}$	$^{+4.52\%}_{-4.52\%}$	13.6%	3%
11.1 - 11.3	271	$-3.8521^{+0.076}_{-0.076}$	$^{+6.25\%}_{-6.26\%}$	16.06%	3%
11.3 - 11.5	69	$-4.4462^{+0.0984}_{-0.0986}$	$^{+12.7\%}_{-12.79\%}$	18.53%	3%
11.5 - 11.7	29	$-4.8227^{+0.1268}_{-0.1278}$	$^{+20.06\%}_{-20.4\%}$	21.0%	3%
11.7 - 11.9	4	$-5.683^{+0.2726}_{-0.2999}$	$^{+58.14\%}_{-64.87\%}$	23.46%	3%

Table A8: $2.75 < z < 3.5$

$\log_{10}(M_* [M_\odot])_{bin}$	N_{gal}	$\log_{10}(\Phi$ [Mpc^{-3} dex^{-1}])	$\sigma_{Poisson}$	σ_{CV}	σ_{Area}
10.1 - 10.3	5261	$-2.7287^{+0.0387}_{-0.0387}$	$^{+1.39\%}_{-1.39\%}$	8.28%	3%
10.3 - 10.5	3246	$-2.9384^{+0.0432}_{-0.0432}$	$^{+1.77\%}_{-1.77\%}$	9.31%	3%
10.5 - 10.7	1521	$-3.2676^{+0.0497}_{-0.0497}$	$^{+2.6\%}_{-2.6\%}$	10.74%	3%
10.7 - 10.9	631	$-3.6497^{+0.0579}_{-0.058}$	$^{+4.06\%}_{-4.06\%}$	12.35%	3%
10.9 - 11.1	249	$-4.0536^{+0.0722}_{-0.0723}$	$^{+6.53\%}_{-6.54\%}$	15.0%	3%
11.1 - 11.3	118	$-4.3779^{+0.0882}_{-0.0883}$	$^{+9.6\%}_{-9.64\%}$	17.65%	3%
11.3 - 11.5	39	$-4.8587^{+0.1162}_{-0.1168}$	$^{+17.15\%}_{-17.36\%}$	20.31%	3%
11.5 - 11.7	10	$-5.4498^{+0.1841}_{-0.1902}$	$^{+35.5\%}_{-37.17\%}$	22.96%	3%
11.7 - 11.9	7	$-5.6047^{+0.2178}_{-0.2286}$	$^{+43.02\%}_{-45.9\%}$	25.61%	3%

Table A9: $3.5 < z < 4.5$

$\log_{10}(M_* [M_\odot])_{bin}$	N_{gal}	$\log_{10}(\Phi$ [Mpc^{-3} dex^{-1}])	$\sigma_{Poisson}$	σ_{CV}	σ_{Area}
10.3 - 10.5	1795	$-3.2876^{+0.0539}_{-0.0539}$	$^{+2.39\%}_{-2.39\%}$	11.81%	3%
10.5 - 10.7	753	$-3.6649^{+0.0638}_{-0.0638}$	$^{+3.71\%}_{-3.71\%}$	13.9%	3%
10.7 - 10.9	212	$-4.2153^{+0.0778}_{-0.0778}$	$^{+7.09\%}_{-7.11\%}$	16.16%	3%
10.9 - 11.1	86	$-4.6072^{+0.0982}_{-0.0983}$	$^{+11.32\%}_{-11.38\%}$	19.34%	3%
11.1 - 11.3	21	$-5.2195^{+0.143}_{-0.1447}$	$^{+23.83\%}_{-24.37\%}$	22.52%	3%
11.3 - 11.5	8	$-5.6386^{+0.207}_{-0.2156}$	$^{+40.03\%}_{-42.38\%}$	25.7%	3%

B "Discrete-z" versus "Smooth-z"

B1 Example HOD posteriors for "smooth-z" versus "discrete-z" models

We showcase the improvement in fitting due to the implementation of "smooth-z" model over "discrete-z" model, specially at high-z, through their corner plots showing HOD posteriors in Figures B1−B4.

B2 Comparison of the mean SHMR for centrals between "smooth-z" and "discrete-z" models

In Figure B5, we compare the $M_{*,\mathrm{mean\,cen}}/M_h$ resulting from the "smooth-z" model fits in blue with the "discrete-z" model fits in orange, highlighting the gain in precision, which generally becomes more noticeable with increasing redshift and/or increasing halo mass, by employing the "smooth-z" model. The disagreement between the two results at the high mass end in some of the high redshift bins could be attributed to the following running theme of this paper. The constraining power of a singular redshift bin weakens with redshift due to stellar mass limit (due to completeness) of the sample utilized. This can also be seen through the narrowing range in the halo masses probed with redshift as indirectly indicated via the enlarging gray shaded regions showing unprobed halo mass regimes. As the "discrete-z" fits rely on singular redshift bins at a time, as opposed to the "smooth-z" fits which compensates for the shrinking dynamical range in mass with redshift

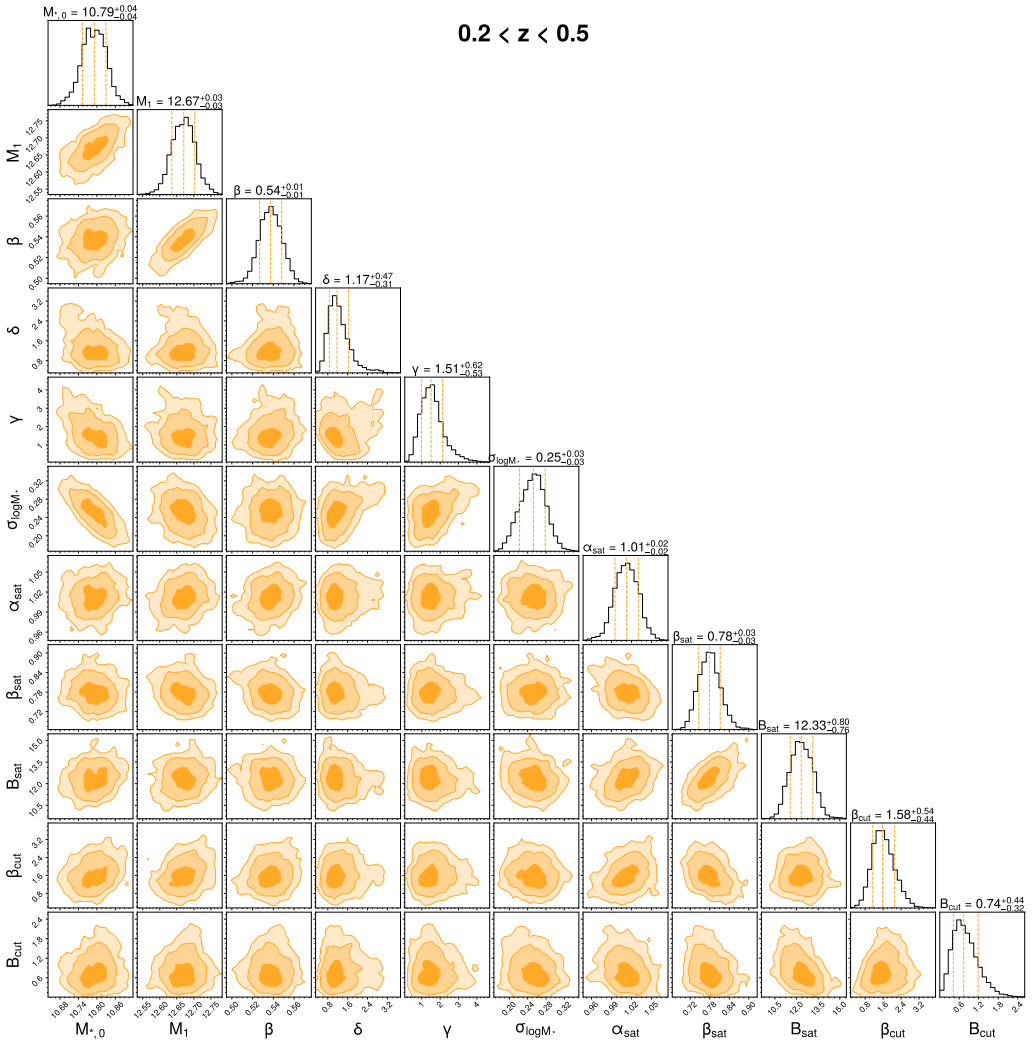

Figure B1: "smooth-z" model corner plot showing HOD posteriors at 0.2 $<$ z $<$ 0.5, where we jointly fitted data in all the z-bins out to z = 4.5 as described in §3.4.8). The median values and the 1σ uncertainty are shown on top of the 1-D projected marginalized distributions along the diagonal of the figure. The 2-D projected marginalized distributions of pairs of parameters are shown with contours with lightening shades of yellow going from 1σ (darkest) to 2σ to 3σ (lightest).

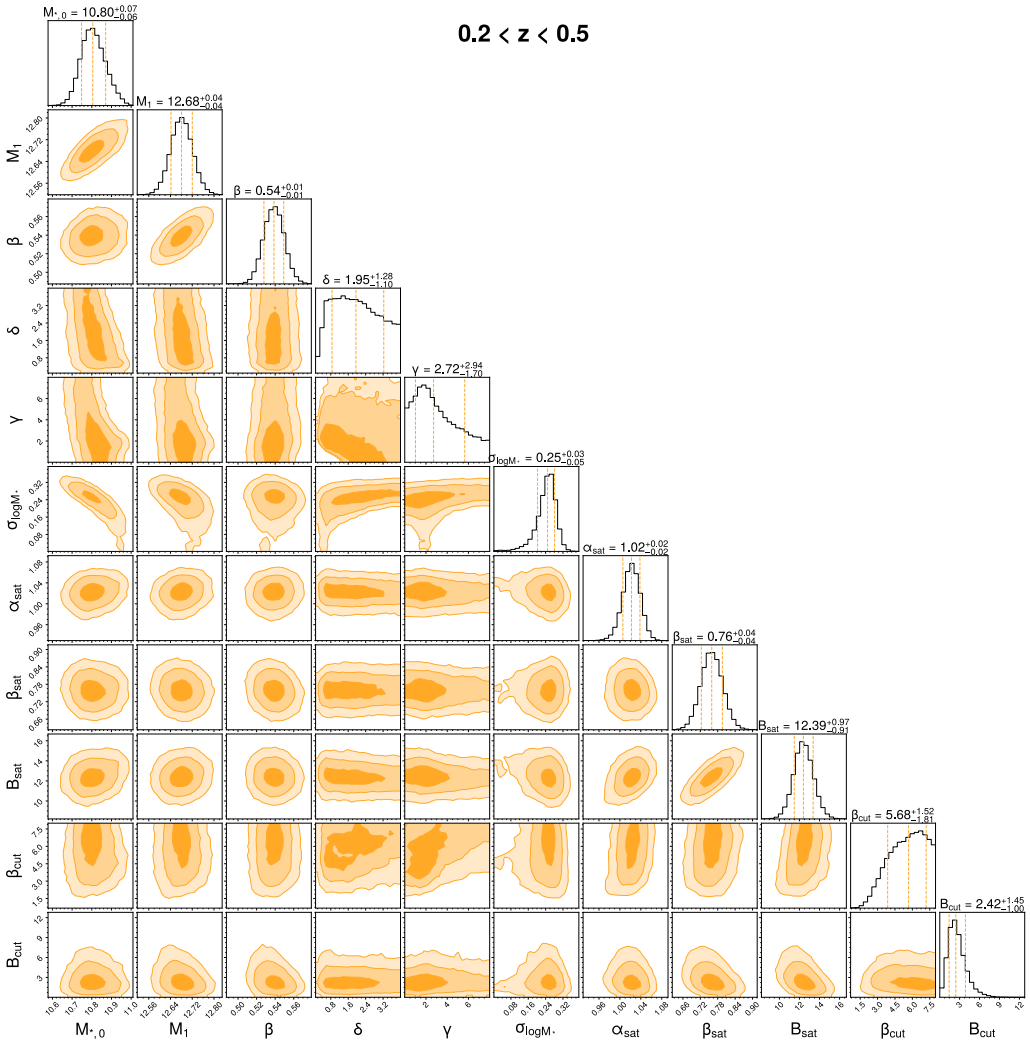

Figure B2: "discrete-z" model corner plot showing HOD posteriors at $0.2 < z < 0.5$ plotted as in Figure B1.

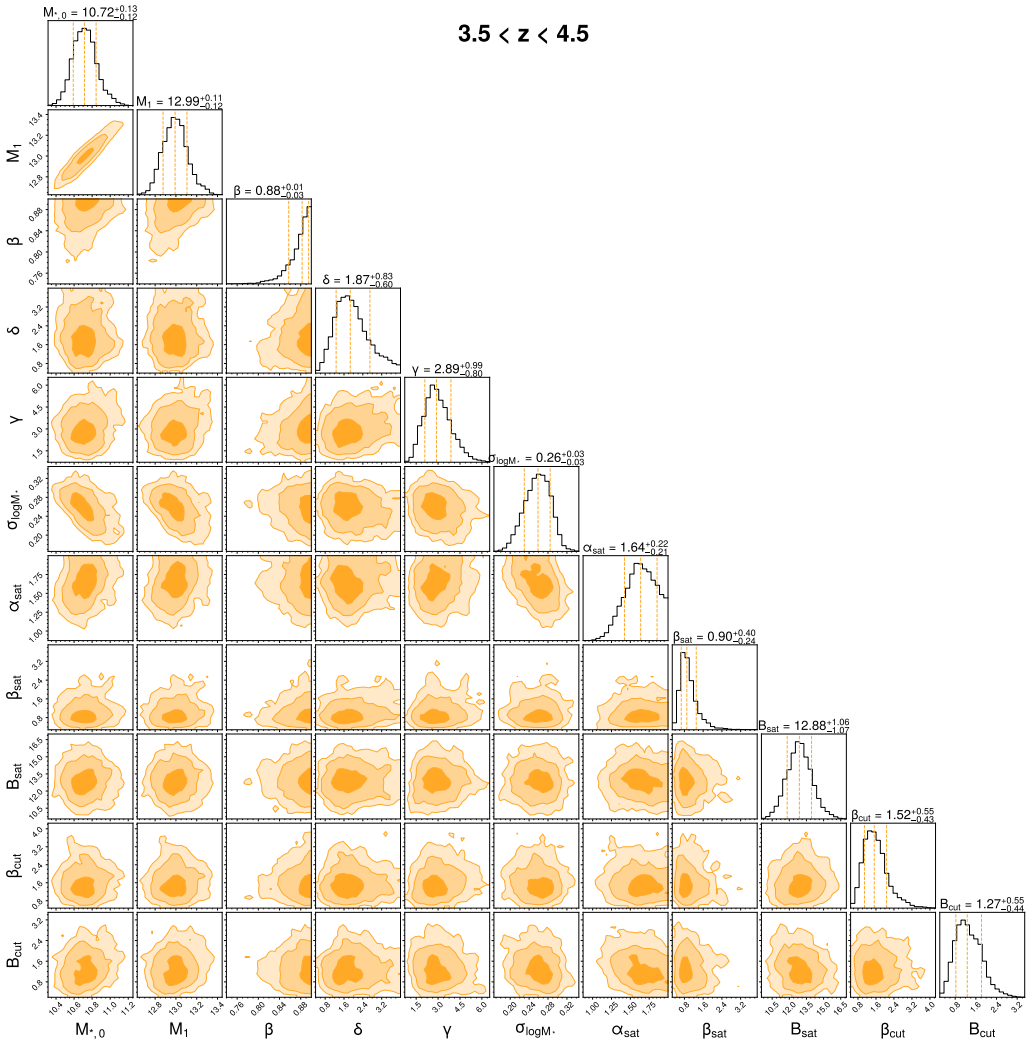

Figure B3: "smooth-z" model corner plot showing HOD posteriors at $3.5 < z < 4.5$ plotted as in Figure B1.

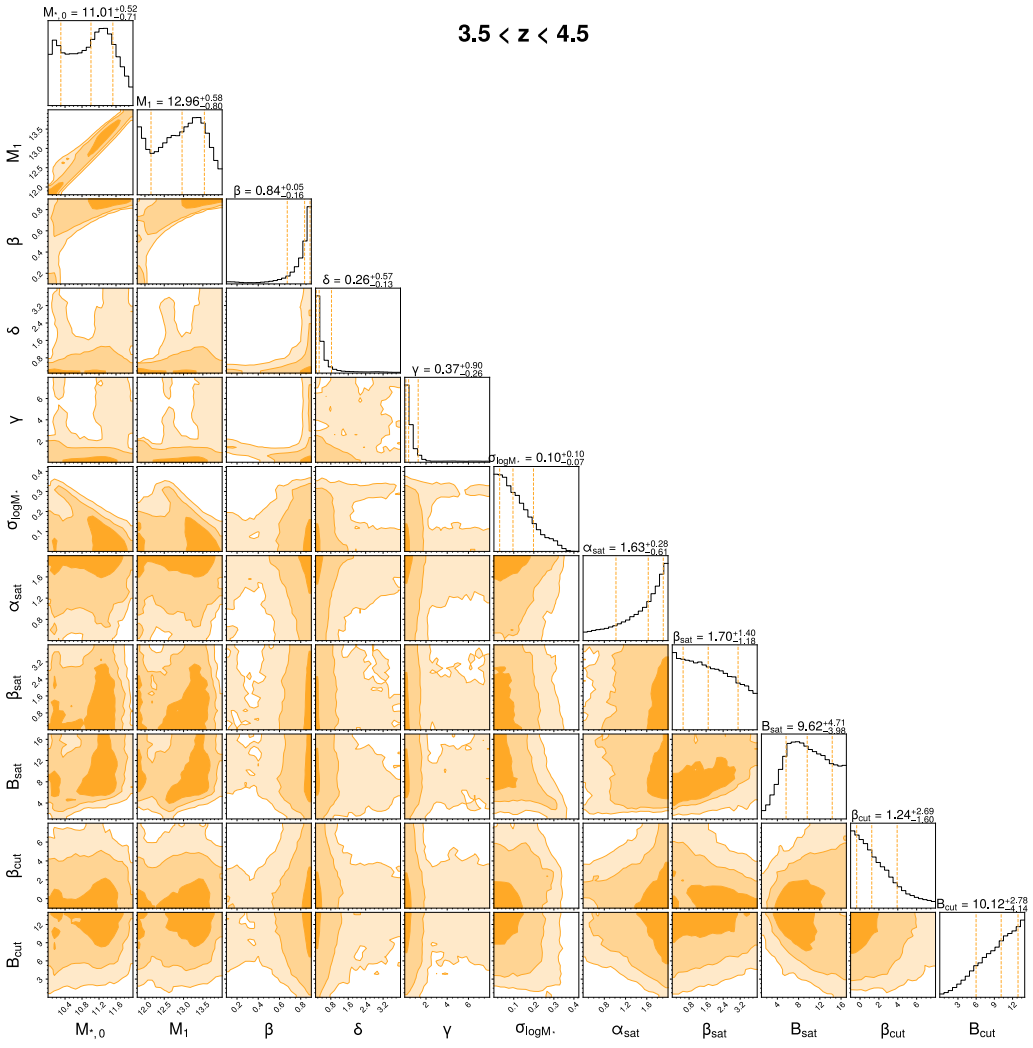

Figure B4: "discrete-z" model corner plot showing HOD posteriors at $3.5 < z < 4.5$ plotted as in Figure B1.

via jointly fitting all redshift bins, the disagreement in the $M_{*,\mathrm{mean\,cen}}/M_h$ at high redshift from the two approaches is not surprising.

C Redshift evolution of the SMFs for halos with fixed masses

We derive the conditional (on halo mass) SMFs for our best-fit models by adjusting the integration limits in Equation 3.19. The SMFs of 0.5 dex wide halo mass bins from $10^{11.5} - 10^{13.5}$ $M_\odot$ are shown in Figure C1, with the legend at the top indicating the four halo mass bins. The contributions from centrals and satellites are shown separately through dashed and solid curves, respectively, and the shaded regions represent the 1σ uncertainties. The central galaxies see a tight correlation with all halo mass bins out to z =4.5; their SMF peak at a certain stellar mass which increases with halo mass, and falls off abruptly on either end. This tight correlation of centrals with halo mass is not enforced by HOD modeling, as we fitted for the the scatter in stellar mass at fixed halo mass, σ_{logM_*} which is linked to the scatter in M_h at fixed M_*, σ_{logM_h} relevant here. As described in L11, since the SHMR behaves like a power law at low-mass, $\sigma_{logM_h} = \beta \times \sigma_{logM_*}$, where β controls the low-mass slope. At high-mass, σ_{logM_h} rises at a rate set by δ and γ. In our fits, σ_{logM_*} was allowed to reach 0.5 dex, i.e. $\sigma_{logM_*} \in [0.001, 0.5]$, but it never surpassed ~ 0.3 dex in any z-bin, indicating that the tight correlation of centrals with M_h we observe was not enforced via HOD modeling.

On the other hand, satellites have huge scatter; there is almost no preference for the halo mass for satellites less massive than $M_* \sim 10^{10} M_\odot$. This highlights how simple abundance matching by rank ordering halos and galax-

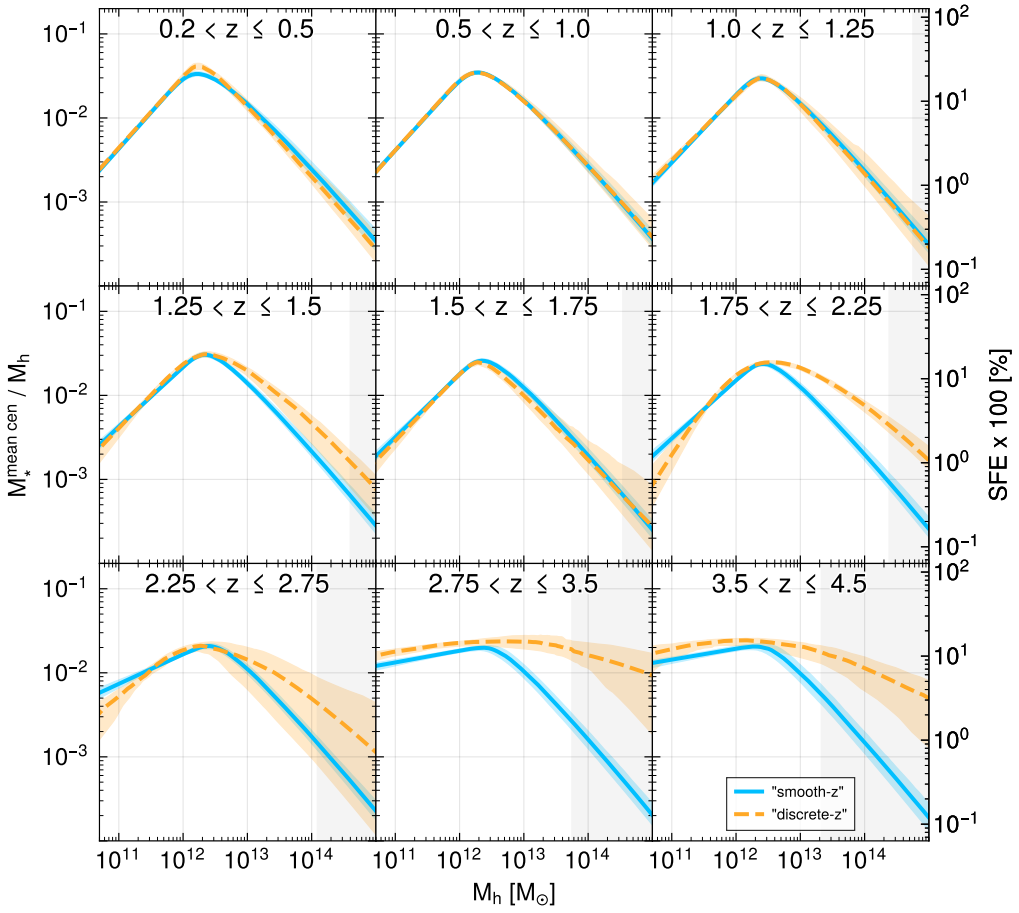

Figure B5: Mean stellar-to-halo mass ratio of central galaxies, $M_{*,\mathrm{mean\,cen}}/M_h$ as a function of halo mass for all 9 z-bins based on "smooth-z" fits in blue compared with the "discrete-z" results in orange. The SFE ($= (M_*^{\mathrm{mean\,cen}}/M_h)/f_b$) in percentages is indicated on the right-most ordinates. The gray shaded regions represent halo masses not directly probed in our analysis, calculated based on "smooth-z" fits as described in §3.5.3.

ies with mass is not enough. Clustering provides additional information which can then be used to constrain the contribution of number densities to central and satellite galaxies individually. Satellites more massive than $M_* \gtrsim 10^{10} M_\odot$ are naturally hosted according to their mass; more massive satellites reside in more massive halos hierarchically.

D Redshift evolution of the satellite fractions

As shown in §3.5.5, satellites start to contribute noticeably to the stellar mass budget of the halos beyond $M_h \sim 10^{12}$ $M_\odot$. Their contributions continue to rise and end up dominating the stellar mass budget beyond $M_h \sim 10^{13}$ $M_\odot$ in the group and cluster-scale halos. As satellites are acquired through mergers, analyzing the redshift evolution of satellite fraction, f_{sat} (calculated using Equation 3.10) can shed light on the evolution of the environment of these halos.

In Figure D1, we show the stellar mass threshold dependence of f_{sat} in bins of redshift (left panel) as well as the redshift dependence of f_{sat} for different stellar mass thresholds (right panel). Within each z-bin at, f_{sat} only weakly grows as the stellar mass threshold is lowered to include more and more galaxies. For instance, in the $1.0 < z < 1.25$ bin (shown in cyan), f_{sat} only evolves from $\sim 19\%$ to 23%. Indeed at any redshift, lowering the stellar mass threshold does not increase the f_{sat} by more than $\sim 7\%$ within the stellar mass ranges considered. Furthermore, even within the lowest z-bin with the lowest stellar mass threshold ($M_* > 10^{8.7} M_\odot$) considered here, where f_{sat} is maximized as expected, it only reaches $\sim 30\%$.

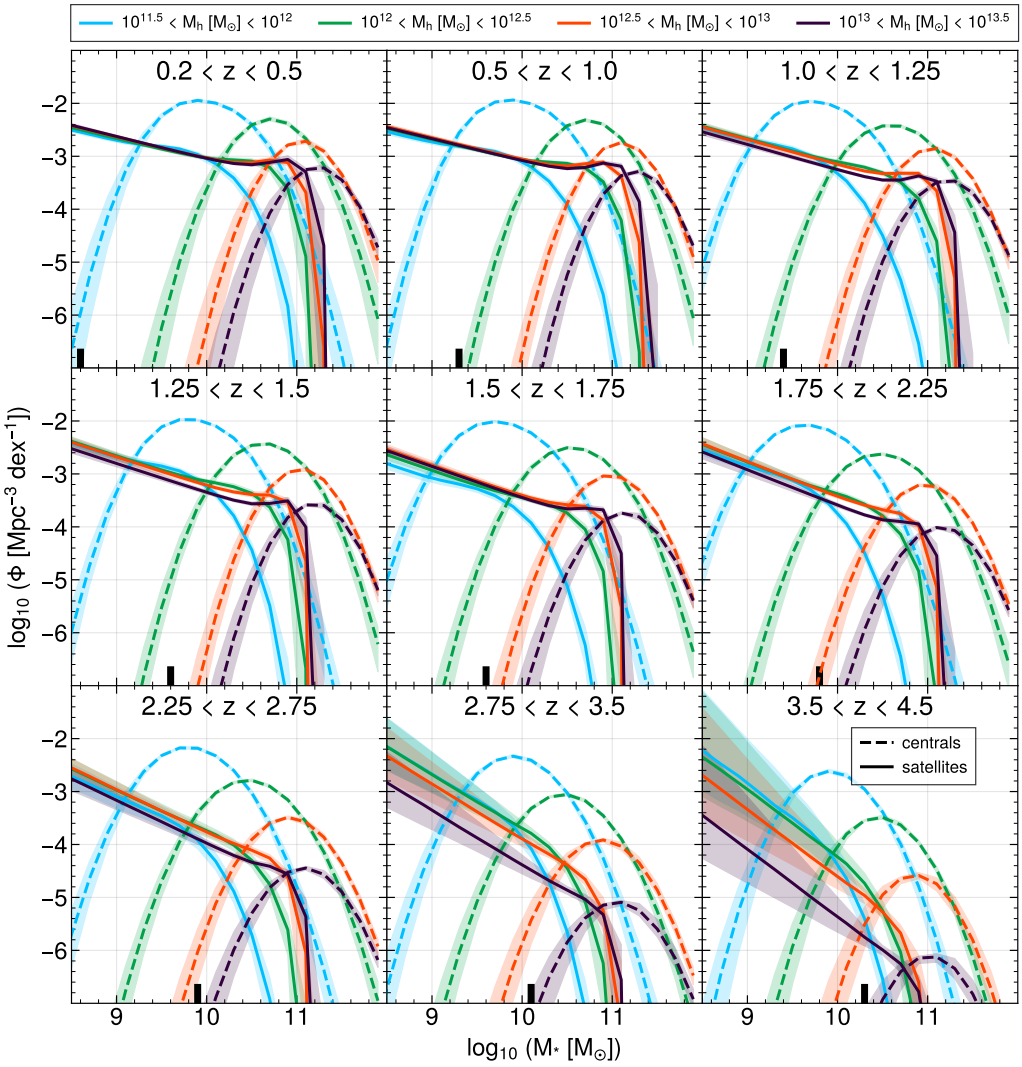

Figure C1: SMFs for centrals (dashed curves) and satellites (solid curves) within 0.5 dex wide bins of halos masses highlighted in different colors and indicated in the legend. The lowest stellar mass threshold considered within each z-bin, due to the stellar mass completeness (as described in §3.2.2 and quoted in Table 3.1) is indicated with a black marker at the bottom of each panel.

The plateauing of f_{sat} with decreasing stellar mass threshold, visible at $z < 1$, can be attributed to the much higher abundance of the low-mass halos and the sharp decline at the high-mass end of the HMF. Therefore, as the stellar mass threshold is lowered to include more and more low-mass galaxies, they are more likely to be allocated as central galaxies in highly abundant low-mass halos than as satellites in rare high-mass halos. Even though the occupation number of satellites increases as a power law as halo mass increases, due to the shape of the HMF, there are just too few high-mass halos available to host many low-mass satellites. Therefore, low-mass galaxies are more often hosted as centrals in low-mass halos (Shuntov et al. 2022).

While the dependence of f_{sat} on stellar mass threshold is relatively weak, it depends much more strongly on redshift such that it increases sharply as the Universe ages as seen by the vertical offsets between different z-bins again in the left panel of Figure D1. This is shown more clearly in the right panel of Figure D1, where f_{sat} sees an overall increasing trend with decreasing redshift, and plateaus at $z \lesssim 0.75$. In general, at each redshift, the lower the stellar mass threshold, the higher its f_{sat}, unsurprisingly.

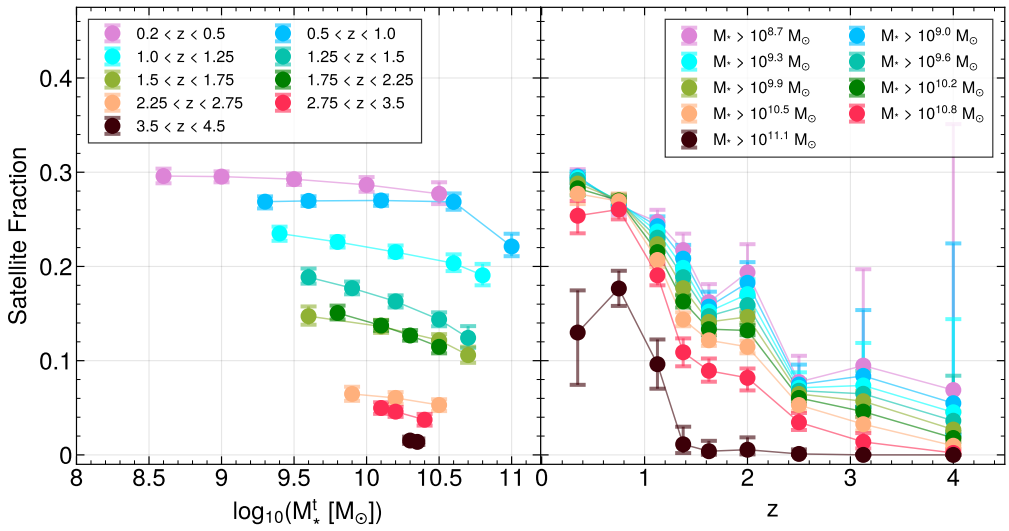

Figure D1: Left: Satellite fractions within each z-bin (as indicated in the legend) as a function of the stellar mass threshold, M_*^t of the sub-samples. Right: Satellite fractions of increasing stellar mass thresholds (as shown in the legend) as a function of redshift.

CHAPTER 4

SUMMARY AND FUTURE DIRECTIONS

This dissertation covers the entire journey from raw astronomical datasets to the laying down of the evolution of galaxy-halo connection underpinned by the SHMR, out to the first billionth year of cosmic history. In chapter 2, photometry from multiple bands spread over multiple surveys were combined into a robust catalog of more than 200, 000 galaxies out to z=8 in the UDS field, providing photometric redshifts and stellar population properties like stellar mass, etc. The catalog (Zaidi et al. 2024) was made publicly available[1]. In chapter 3, we combined this catalog with a similarly sized UVISTA COSMOS DR3 catalog (Muzzin et al. 2013) to expand the total effective area to be ~ 1.61 deg^2 to probe the cosmic SHMR out to $z \sim 5$, through the measurements of galaxy clustering and abundance. We introduced a novel Bayesian HOD fitting technique, which by jointly fitting clustering and abundance (SMF) data across redshifts and stellar mass, was able to dramatically plummet the previously reported much larger uncertainties (e.g. Shuntov et al. (2022)) on the SHMR at high-z ($z > 2$). For instance, we could constrain the $\log_{10}(\mathrm{M}_h^{peak}\ [\mathrm{M}_\odot])$ by a factor of up to ~ 6.5 better at high-z when compared to a similar analysis done by Shuntov et al. (2022) but without using our new HOD fitting approach.

With the use of the new approach, we confidently show that the integrated SFE efficiencies of dark matter halos remain low ($\lesssim 20\%$) out to $z \sim 5$, and M_h^{peak} barely evolves with time, staying within $\sim 10^{12.2} - 10^{12.4}$ $\mathrm{M}_\odot$ again out to $z \sim 5$, contrary to some of the previous studies contemplating a rising M_h^{peak} with redshift. Moreover, we highlight the importance of mitigating

[1]https://www.zaidikumail.com/feniks-uds-catalogs

cosmic variance by incorporating two fields instead of one typical in such analyses. Previous SHMR results were either based purely on, or dominated by, the data in the COSMOS field. Utilizing the UDS field in addition to COSMOS, we showed that the M_h^{peak} remains systematically lower, up to $\sim$ 0.15 dex at z $<$ 1.5, if only COSMOS is used. This study marks a concrete understanding of the evolution of SHMR out to $z \sim 5$.

Current generation of spectroscopic surveys like PFS are targeting stellar mass complete samples out to $z \sim 1.7$. Beyond such redshifts, PFS will only conduct intergalactic medium (IGM) tomography at $z > 2$, and capture active galactic nuclei (AGN)/Quasars out to $z > 4$. We are still years away from stellar mass complete spectroscopic surveys at $2 < z < 5$, let alone out to *JWST*'s dominance at $z \sim$ 8-10. Therefore, any clustering and abundance measurements out to $z \sim$ 8-10 need to rely on *JWST*. However, a single JWST pointing−or even a few−does not cover a sufficiently large area for the statistical studies of the nature described in chapter 3 extracting cosmic SHMR. Additionally, combining data from multiple surveys with heterogeneously constructed photometric catalogs introduces biases in the derivation of stellar population properties, convoluting the resulting SHMR.

Therefore, currently we are utilizing the ASTRODEEP- *JWST* catalog (Merlin et al. 2024), which homogeneously integrates data from several *JWST* programs−CEERS, PRIMER-UDS, PRIMER-COSMOS, JADES-GN, JADES-GS, and NGDEEP. I am also supplementing it with the 29-band (HST+JWST) CANUCS-Technicolor (Hubble Frontier Fields) catalog prepared in homogeneity with ASTRODEEP, totaling an effective area of $\sim$ 720 arcmin^2, a herculean effort as a single JWST's NIRCam pointing covers only 9.7 arcmin^2. While this area is not sufficient to probe the SHMR at very low redshifts,

it is large enough to put meaningful constraints on the SHMR at high redshift, due to how the differential volume captured by a given survey area scales with redshift. Not only that, but the use of multiple widely separated fields probe different lines of sights, greatly mitigating the effects of cosmic variance in the measurements of galaxy clustering/abundance, crucial in robustly constraining the SHMR with the HOD framework. This endeavor will answer some of the outstanding question in high-z galaxy formation today, such as, how does the SFE of halos look at the earliest epochs? Are they high enough to explain the observed overabundance of massive galaxies at cosmic dawn ($z \gtrsim 6$) (e.g. Labbé et al. (2023))?

So far, I have been focusing on the fundamental relationship of the galaxy-halo connection−galaxy stellar mass and halo mass relationship, the SHMR. However, even though halo mass guides most of the relationship between galaxies and halos, differences in halo assembly history, e.g. in terms of halo formation time, halo concentration, etc. might lead to second-order effects on galaxy assembly. There might be subtle changes in galaxy clustering, due to the halo properties beyond mass. This is termed in the literature as *galaxy assembly bias* (Zentner et al. 2014). There have been many observational efforts to detect this *galaxy assembly bias* signal (e.g. Miyatake et al. 2016; Zentner et al. 2019; Vakili & Hahn 2019), however with contradictory conclusions regarding its support in data.

The lack of clear success in probing assembly bias is not helped by the fact that most of the efforts mentioned above have been limited to spectroscopic surveys like SDSS DR7 (Abazajian et al. 2009). Even though, they exquisitely chart the 3D distributions of galaxies in the Universe, they only do so at low redshift ($z \lesssim 0.7$). *Galaxy assembly bias* depends on the *assembly*

histories of halos, and therefore, should change with time. Therefore, such efforts need to extend beyond low redshift which will be possible with the PFS (Greene et al. 2022). The comprehensive dependence of the galaxy-halo connection on all halo properties, beyond just mass, will be the focus of my future research as a postdoctoral appointee at the Argonne National Lab.

REFERENCES

Abazajian K. N., et al., 2009, ApJS, 182, 543

Aihara H., et al., 2019, PASJ, 71, 114

Aihara H., et al., 2022, PASJ, 74, 247

Antwi-Danso J., et al., 2023, ApJ, 943, 166

Ashby M. L. N., et al., 2018, ApJS, 237, 39

Astropy Collaboration et al., 2022, ApJ, 935, 167

Bahcall N. A., Soneira R. M., 1983, ApJ, 270, 20

Bardeen J. M., Bond J. R., Kaiser N., Szalay A. S., 1986, ApJ, 304, 15

Behroozi P. S., Conroy C., Wechsler R. H., 2010, ApJ, 717, 379

Behroozi P. S., Wechsler R. H., Conroy C., 2013, ApJ, 770, 57

Behroozi P., Wechsler R. H., Hearin A. P., Conroy C., 2019, MNRAS, 488, 3143

Benson A. J., Bower R. G., Frenk C. S., Lacey C. G., Baugh C. M., Cole S., 2003, ApJ, 599, 38

Berlind A. A., et al., 2003, ApJ, 593, 1

Bertin E., Arnouts S., 1996, A&AS, 117, 393

Blake C., Collister A., Lahav O., 2008, MNRAS, 385, 1257

Blanton M. R., et al., 2017, AJ, 154, 28

Bower R. G., Benson A. J., Malbon R., Helly J. C., Frenk C. S., Baugh C. M., Cole S., Lacey C. G., 2006, MNRAS, 370, 645

Bradshaw E. J., et al., 2013, MNRAS, 433, 194

Brammer G., 2021, eazy-py, doi:10.5281/zenodo.5012704, https://github.com/gbrammer/eazy-py

Brammer G. B., van Dokkum P. G., Coppi P., 2008, ApJ, 686, 1503

Brammer G. B., et al., 2012, ApJS, 200, 13

Bruzual G., Charlot S., 2003, MNRAS, 344, 1000

Byler N., Dalcanton J. J., Conroy C., Johnson B. D., 2017, ApJ, 840, 44

Calzetti D., Armus L., Bohlin R. C., Kinney A. L., Koornneef J., Storchi-Bergmann T., 2000, ApJ, 533, 682

Casali M., et al., 2007, A&A, 467, 777

Chabrier G., 2003, PASP, 115, 763

Childress M. J., et al., 2017, MNRAS, 472, 273

Chworowsky K., et al., 2023, arXiv e-prints, p. arXiv:2311.14804

Cielo S., Bieri R., Volonteri M., Wagner A. Y., Dubois Y., 2018, MNRAS, 477, 1336

Coil A. L., et al., 2011, ApJ, 741, 8

Cole S., Lacey C. G., Baugh C. M., Frenk C. S., 2000, MNRAS, 319, 168

Colless M., et al., 2001, MNRAS, 328, 1039

Conroy C., Gunn J. E., 2010, ApJ, 712, 833

Conroy C., Wechsler R. H., 2009, ApJ, 696, 620

Conroy C., Gunn J. E., White M., 2009, ApJ, 699, 486

Cool R. J., et al., 2013, ApJ, 767, 118

Cora S. A., et al., 2018, MNRAS, 479, 2

Coupon J., et al., 2012, A&A, 542, A5

Coupon J., et al., 2015, MNRAS, 449, 1352

Cowie L. L., Songaila A., Hu E. M., Cohen J. G., 1996, AJ, 112, 839

Cowley W. I., et al., 2018, ApJ, 853, 69

Croton D. J., et al., 2006, MNRAS, 365, 11

Darvish B., Sobral D., Mobasher B., Scoville N. Z., Best P., Sales L. V., Smail I., 2014, ApJ, 796, 51

Davé R., Anglés-Alcázar D., Narayanan D., Li Q., Rafieferantsoa M. H., Appleby S., 2019, MNRAS, 486, 2827

Davidzon I., et al., 2017, A&A, 605, A70

De Lucia G., Blaizot J., 2007, MNRAS, 375, 2

Eikenberry S., et al., 2008, in McLean I. S., Casali M. M., eds, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 7014, Ground-based and Airborne Instrumentation for Astronomy II. p. 70140V, doi:10.1117/12.788326

Eikenberry S., et al., 2012, in McLean I. S., Ramsay S. K., Takami H., eds, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 8446, Ground-based and Airborne Instrumentation for Astronomy IV. p. 84460I, doi:10.1117/12.925679

Eke V. R., Cole S., Frenk C. S., Navarro J. F., 1996, MNRAS, 281, 703

Esdaile J., et al., 2021, AJ, 162, 225

Fall S. M., Efstathiou G., 1980, MNRAS, 193, 189

Ferland G. J., et al., 2017, Rev. Mex. Astron. Astrofis., 53, 385

Florez J., et al., 2020, MNRAS, 497, 3273

Florez J., et al., 2021, MNRAS, 508, 762

Fontana A., et al., 2014, A&A, 570, A11

Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, PASP, 125, 306

Forrest B., et al., 2020, ApJ, 903, 47

Furusawa H., Kosugi G., Akiyama M., Takata T., Sekiguchi K., Furusawa J., 2008, in Kodama T., Yamada T., Aoki K., eds, Astronomical Society of the Pacific Conference Series Vol. 399, Panoramic Views of Galaxy Formation and Evolution. p. 131

Garcet O., et al., 2007, A&A, 474, 473

Gehrels N., 1986, ApJ, 303, 336

Gomez P. L., et al., 2012, in American Astronomical Society Meeting Abstracts #219. p. 413.07

Goodman J., Weare J., 2010, Communications in Applied Mathematics and Computational Science, 5, 65

Governato F., Gardner J. P., Stadel J., Quinn T., Lake G., 1999, AJ, 117, 1651

Greene J., Bezanson R., Ouchi M., Silverman J., the PFS Galaxy Evolution Working Group 2022, arXiv e-prints, p. arXiv:2206.14908

Groth E. J., Peebles P. J. E., 1977, ApJ, 217, 385

Guo Q., et al., 2011, MNRAS, 413, 101

Hambly N. C., et al., 2008, MNRAS, 384, 637

Harris C. R., et al., 2020, Nature, 585, 357

Hatfield P. W., Lindsay S. N., Jarvis M. J., Häußler B., Vaccari M., Verma A., 2016, MNRAS, 459, 2618

Heckman T. M., Best P. N., 2014, ARA&A, 52, 589

Henriques B. M. B., White S. D. M., Thomas P. A., Angulo R., Guo Q., Lemson G., Springel V., Overzier R., 2015, MNRAS, 451, 2663

Henriques B. M. B., Yates R. M., Fu J., Guo Q., Kauffmann G., Srisawat C., Thomas P. A., White S. D. M., 2020, MNRAS, 491, 5795

Hewett P. C., Warren S. J., Leggett S. K., Hodgkin S. T., 2006, MNRAS, 367, 454

Hodgkin S. T., Irwin M. J., Hewett P. C., Warren S. J., 2009, MNRAS, 394, 675

Hopkins P. F., Quataert E., Murray N., 2012, MNRAS, 421, 3522

Hopkins P. F., et al., 2023, MNRAS, 519, 3154

Infante L., 1994, A&A, 282, 353

Ishikawa S., Kashikawa N., Toshikawa J., Tanaka M., Hamana T., Niino Y., Ichikawa K., Uchiyama H., 2017, ApJ, 841, 8

Iyer K., Gawiser E., 2017, ApJ, 838, 127

Iyer K. G., Gawiser E., Faber S. M., Ferguson H. C., Kartaltepe J., Koekemoer A. M., Pacifici C., Somerville R. S., 2019, ApJ, 879, 116

Jarvis M. J., et al., 2013, MNRAS, 428, 1281

Jones D. H., et al., 2009, MNRAS, 399, 683

Juneau S., et al., 2005, ApJ, 619, L135

Kaiser N., 1984, ApJ, 284, L9

Kelson D. D., et al., 2014, ApJ, 783, 110

Klypin A. A., Trujillo-Gomez S., Primack J., 2011, ApJ, 740, 102

Kondapally R., et al., 2023, MNRAS, 523, 5292

Kravtsov A. V., Berlind A. A., Wechsler R. H., Klypin A. A., Gottlöber S., Allgood B., Primack J. R., 2004, ApJ, 609, 35

Kriek M., van Dokkum P. G., Labbé I., Franx M., Illingworth G. D., Marchesini D., Quadri R. F., 2009, ApJ, 700, 221

Kron R. G., 1980, ApJS, 43, 305

Labbé I., et al., 2003, AJ, 125, 1107

Labbe I. F., Bouwens R., Illingworth G., Franx M., 2006, in American Astronomical Society Meeting Abstracts. p. 132.02

Labbé I., et al., 2023, Nature, 616, 266

Lacy M., Petric A. O., Sajina A., Canalizo G., Storrie-Lombardi L. J., Armus L., Fadda D., Marleau F. R., 2007, AJ, 133, 186

Lacy M., et al., 2013, ApJS, 208, 24

Lacy M., et al., 2021, MNRAS, 501, 892

Laigle C., et al., 2016, The Astrophysical Journal Supplement Series, 224, 24

Landy S. D., Szalay A. S., 1993, ApJ, 412, 64

Lawrence A., et al., 2007, MNRAS, 379, 1599

Le Fèvre O., et al., 2013, A&A, 559, A14

Leauthaud A., Tinker J., Behroozi P. S., Busha M. T., Wechsler R. H., 2011, ApJ, 738, 45

Leauthaud A., et al., 2012, ApJ, 744, 159

Legrand L., et al., 2019, MNRAS, 486, 5468

Leja J., et al., 2019, ApJ, 877, 140

Limber D. N., 1954, ApJ, 119, 655

Lower S., Narayanan D., Leja J., Johnson B. D., Conroy C., Davé R., 2020, ApJ, 904, 33

Mandelbaum R., Seljak U., Kauffmann G., Hirata C. M., Brinkmann J., 2006, MNRAS, 368, 715

Marchesini D., et al., 2010, ApJ, 725, 1277

Marinoni C., Hudson M. J., 2002, ApJ, 569, 101

Marsan Z. C., et al., 2022, ApJ, 924, 25

Masters D. C., et al., 2019, ApJ, 877, 81

Mauduit J. C., et al., 2012, PASP, 124, 714

McCracken H. J., et al., 2012, A&A, 544, A156

McCracken H. J., et al., 2015, MNRAS, 449, 901

McLeod D. J., McLure R. J., Dunlop J. S., Cullen F., Carnall A. C., Duncan K., 2021, MNRAS, 503, 4413

McLure R. J., et al., 2013, MNRAS, 428, 1088

McLure R. J., et al., 2018, MNRAS, 479, 25

Merlin E., et al., 2024, A&A, 691, A240

Miyatake H., More S., Takada M., Spergel D. N., Mandelbaum R., Rykoff E. S., Rozo E., 2016, Phys. Rev. Lett., 116, 041301

Mo H. J., White S. D. M., 1996, MNRAS, 282, 347

Momcheva I. G., et al., 2016, ApJS, 225, 27

Moster B. P., Somerville R. S., Maulbetsch C., van den Bosch F. C., Macciò A. V., Naab T., Oser L., 2010, ApJ, 710, 903

Moster B. P., Somerville R. S., Newman J. A., Rix H.-W., 2011, ApJ, 731, 113

Moster B. P., Naab T., White S. D. M., 2018, MNRAS, 477, 1822

Moustakas J., et al., 2013, ApJ, 767, 50

Mowla L., et al., 2022, ApJ, 937, L35

Murray S. G., Diemer B., Chen Z., Neuhold A. G., Schnapp M. A., Peruzzi T., Blevins D., Engelman T., 2021, Astronomy and Computing, 36, 100487

Muzzin A., et al., 2013, ApJS, 206, 8

Newman J. A., Moster B. P., 2014, QUICKCV: Cosmic variance calculator, Astrophysics Source Code Library, record ascl:1402.012

Oke J. B., Gunn J. E., 1983, ApJ, 266, 713

Ono Y., et al., 2010, MNRAS, 402, 1580

Ouchi M., et al., 2008, ApJS, 176, 301

Papovich C., et al., 2016, ApJS, 224, 28

Peebles P. J. E., 1973, ApJ, 185, 413

Penzias A. A., Wilson R. W., 1965, ApJ, 142, 419

Persson S. E., et al., 2013, PASP, 125, 654

Pillepich A., et al., 2018, MNRAS, 473, 4077

Planck Collaboration et al., 2020, A&A, 641, A6

Pozzetti L., et al., 2010, A&A, 523, A13

Quadri R. F., Williams R. J., 2010, ApJ, 725, 794

Quadri R., et al., 2007, AJ, 134, 1103

Quadri R. F., Williams R. J., Lee K.-S., Franx M., van Dokkum P., Brammer G. B., 2008, ApJ, 685, L1

Riebe K., et al., 2013, Astronomische Nachrichten, 334, 691

Roche N., Eales S. A., 1999, MNRAS, 307, 703

Sabater J., et al., 2021, A&A, 648, A2

Sargsyan L. A., Weedman D. W., 2009, ApJ, 701, 1398

Sawicki M., et al., 2019, MNRAS, 489, 5202

Scharré L., Sorini D., Davé R., 2024, MNRAS, 534, 361

Schaye J., et al., 2015, MNRAS, 446, 521

Schlafly E. F., Finkbeiner D. P., 2011, ApJ, 737, 103

Schmidt M., 1968, ApJ, 151, 393

Schreiber C., et al., 2018, A&A, 618, A85

Scodeggio M., et al., 2018, A&A, 609, A84

Scoville N., et al., 2007a, ApJS, 172, 1

Scoville N., et al., 2007b, ApJS, 172, 150

Shipley H. V., et al., 2018, ApJS, 235, 14

Shuntov M., et al., 2022, A&A, 664, A61

Silk J., Rees M. J., 1998, A&A, 331, L1

Sinha M., Garrison L., 2019, in Majumdar A., Arora R., eds, Software Challenges to Exascale Computing. Springer Singapore, Singapore, pp 3–20, https://doi.org/10.1007/978-981-13-7729-7_1

Sinha M., Garrison L. H., 2020, MNRAS, 491, 3022

Skelton R. E., et al., 2014, ApJS, 214, 24

Somerville R. S., Primack J. R., 1999, MNRAS, 310, 1087

Springel V., et al., 2005, Nature, 435, 629

Springel V., Frenk C. S., White S. D. M., 2006, Nature, 440, 1137

Springel V., et al., 2018, MNRAS, 475, 676

Stalin C. S., Petitjean P., Srianand R., Fox A. J., Coppolani F., Schwope A., 2010, MNRAS, 401, 294

Steidel C. C., et al., 2014, ApJ, 795, 165

Stevens A. R. H., Croton D. J., Mutch S. J., 2016, MNRAS, 461, 859

Stevens A. R. H., Sinha M., Rohl A., Sammons M. W., Hadzhiyska B., Hernández-Aguayo C., Hernquist L., 2023, arXiv e-prints, p. arXiv:2312.04137

Stevens A. R. H., Sinha M., Rohl A., Sammons M. W., Hadzhiyska B., Hernández-Aguayo C., Hernquist L., 2024, Publ. Astron. Soc. Australia, 41, e053

Stewart K. R., Bullock J. S., Barton E. J., Wechsler R. H., 2009, ApJ, 702, 1005

Straatman C. M. S., et al., 2016, ApJ, 830, 51

Tasse C., et al., 2021, A&A, 648, A1

Tinker J. L., Robertson B. E., Kravtsov A. V., Klypin A., Warren M. S., Yepes G., Gottlöber S., 2010, ApJ, 724, 878

Tinker J. L., Leauthaud A., Bundy K., George M. R., Behroozi P., Massey R., Rhodes J., Wechsler R. H., 2013, ApJ, 778, 93

Totsuji H., Kihara T., 1969, PASJ, 21, 221

Vakili M., Hahn C., 2019, ApJ, 872, 115

Vogelsberger M., et al., 2014, MNRAS, 444, 1518

Wake D. A., et al., 2011, ApJ, 728, 46

Wall J. V., Jenkins C. R., 2012, Practical Statistics for Astronomers

Weaver J. R., et al., 2022, ApJS, 258, 11

Weaver J. R., et al., 2023, A&A, 677, A184

Wechsler R. H., Tinker J. L., 2018, ARA&A, 56, 435

Weinberger R., et al., 2017, MNRAS, 465, 3291

Wetzel A. R., Cohn J. D., White M., 2009, MNRAS, 395, 1376

Wetzel A., et al., 2023, ApJS, 265, 44

Whitaker K. E., et al., 2011, ApJ, 735, 86

White S. D. M., Rees M. J., 1978, MNRAS, 183, 341

Williams R. J., Quadri R. F., Franx M., van Dokkum P., Labbé I., 2009, ApJ, 691, 1879

Wuyts S., 2007, in Metcalfe N., Shanks T., eds, Astronomical Society of the Pacific Conference Series Vol. 379, Cosmic Frontiers. p. 356

Yamada T., et al., 2005, ApJ, 634, 861

York D. G., et al., 2000, AJ, 120, 1579

Zaidi K., et al., 2024, ApJ, 969, 84

Zehavi I., et al., 2004, ApJ, 608, 16

Zehavi I., et al., 2005, ApJ, 630, 1

Zehavi I., et al., 2011, ApJ, 736, 59

Zentner A. R., Hearin A. P., van den Bosch F. C., 2014, MNRAS, 443, 3044

Zentner A. R., Hearin A., van den Bosch F. C., Lange J. U., Villarreal A. S., 2019, MNRAS, 485, 1196

Zheng Z., et al., 2005, ApJ, 633, 791

Zheng Z., Coil A. L., Zehavi I., 2007, ApJ, 667, 760

Zu Y., Mandelbaum R., 2015, MNRAS, 454, 1161

Zu Y., Mandelbaum R., 2016, MNRAS, 457, 4360

van Dokkum P. G., et al., 2009, PASP, 121, 2