Galaxy Formation and Evolution from Deep Optical and Infrared Surveys

Nathan James Adams

New College
University of Oxford

A thesis submitted for the degree of Doctor of Philosophy

Trinity 2021

Abstract

In this thesis, I present results from population studies conducted on galaxies across most of cosmic time. Using the unique depth and area combination of modern, ground-based surveys allows for strong constraints to be placed on the number density of the most massive and luminous galaxies between redshifts of $0.1 < z < 5.2$, while simultaneously allowing for contextualisation of results with the fainter/lower-mass population. The dataset compiles together $\approx 2,000,000$ objects over 6 square degrees of the sky and I use these to tackle two issues facing extragalactic research.

First, I measure the stellar mass function of galaxies from $0.1 < z < 2.0$, covering the peak of global star formation ($z \sim 1.5$–2.0). A sample of 320,000 galaxies is used to map the exponential cut off in galaxy number density at high masses. The survey area used allows for variance attributed to large scale structure in the Universe to be significantly reduced compared to similar studies in the past and provides evidence this exponential cut off in galaxy number density shifts to greater masses with time, matching predictions from some hydrodynamical simulations such as EAGLE. I suggest this is driven by an increasing population of massive, passive galaxies produced by mergers.

Second, I measure and examine the ultraviolet luminosity function of galaxies in the redshift range $3.5 < z < 5.2$. With high precision over the luminosity range in which the dominance of ultraviolet emission transitions from galaxies to active galactic nuclei (AGN), I find that both the Schechter function and double power law models can be used to describe the galaxy population. However, they predict drastically different number counts for ultra-luminous galaxies ($M_{UV} < -23$). Alongside, I find the number density of faint AGN ($M_{UV} > -24$) are high, generating a steep power law slope $-2.1 < \alpha_{AGN} < -1.7$. I propose that the large rise in space density of AGN over this time period can be attributed to a global increase in the number density of active black holes. Measurements of the galaxy population are found to match models which suggest that a steep bright-end slope is driven by dust obscuration and that mass-dependent quenching from AGN feedback is not yet significant at this time.

Galaxy Formation and Evolution from Deep Optical and Infrared Surveys

Nathan James Adams

New College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2021

Acknowledgements

I pass on my deepest gratitude to my supervisors Matt Jarvis and Rebecca Bowler. Their insight and encouragement has been invaluable and their patience most appreciated. This is extended to the rest of the research group which I have worked with over the past four years: Anastasia, Imogen, Ian, Aprajita, Peter, James, Leah, Madalina, Nijin, Catherine, Josie and Zahra. You all made me feel a part of something so much more. I thank Ashling, Leanne, Jonathan and all of the administrative and IT staff in the department for their constant support. I also thank the people abroad with whom I've worked over the years: Sarah White, Natasha Maddox, Eliab Malefahlo and many others with whom we've produced some great science. The ICRAR SU2 group in Australia (Luke Davies, Aaron Robotham, Sabine Bellstedt, Simon Driver, Luca Cortese, Claudia Lagos and their army of students) hosted me for six months and I am grateful for the unique opportunity they offered to me.

I also thank my family (Wendy, Lee, Michael, Katie, Jazz and the grandparents) for their continued support in my mission to avoid getting a 'real' job for as long as possible! To Eleanor, I thank you for fixing my many... many misplaced hyphens and for keeping me sane during the coronavirus pandemic. To the other DPhil students in the Astrophysics and AOPP sub-departments who started alongside me... congrats everyone, we made it!

Abstract

In this thesis, I present results from population studies conducted on galaxies across most of cosmic time. Using the unique depth and area combination of modern, ground-based surveys allows for strong constraints to be placed on the number density of the most massive and luminous galaxies between redshifts of $0.1 < z < 5.2$, while simultaneously allowing for contextualisation of results with the fainter/lower-mass population. The dataset compiles together $\approx 2,000,000$ objects over 6 square degrees of the sky and I use these to tackle two issues facing extragalactic research.

First, I measure the stellar mass function of galaxies from $0.1 < z < 2.0$, covering the peak of global star formation ($z \sim 1.5$–2.0). A sample of 320,000 galaxies is used to map the exponential cut off in galaxy number density at high masses. The survey area used allows for variance attributed to large scale structure in the Universe to be significantly reduced compared to similar studies in the past and provides evidence this exponential cut off in galaxy number density shifts to greater masses with time, matching predictions from some hydrodynamical simulations such as EAGLE. I suggest this is driven by an increasing population of massive, passive galaxies produced by mergers.

Second, I measure and examine the ultraviolet luminosity function of galaxies in the redshift range $3.5 < z < 5.2$. With high precision over the luminosity range in which the dominance of ultraviolet emission transitions from galaxies to active galactic nuclei (AGN), I find that both the Schechter function and double power law models can be used to describe the galaxy population. However, they predict drastically different number counts for ultra-luminous galaxies ($M_{UV} < -23$). Alongside, I find the number density of faint AGN ($M_{UV} > -24$) are high, generating a steep power law slope $-2.1 < \alpha_{AGN} < -1.7$. I propose that the large rise in space density of AGN over this time period can be attributed to a global increase in the number density of active black holes. Measurements of the galaxy population are found to match models which suggest that a steep bright-end slope is driven by dust obscuration and that mass-dependent quenching from AGN feedback is not yet significant at this time.

Contents

List of Figures

As for me, I am tormented with an everlasting itch for things remote. I love to sail forbidden seas, and land on barbarous coasts.

— Herman Melville's *Moby-Dick*

1 Introduction

Contents

1.1 The Big Picture

Over the past 50 years, the field of extragalactic astrophysics has been in a near-constant state of revolution. During this period, we have witnessed increasing observational evidence for dark matter (e.g. Rubin & Ford, 1970) and the discovery of galaxies located at such distances that we observe them as they were when the Universe was less than 10 per cent of its current age (e.g. Warren et al., 1987a,b).

Advances in telescope technology and survey strategies have granted access to vast volumes of the Universe, allowing for a census of galaxies to be taken at different epochs and ultimately leading to a shift from studying individual cases to handling upwards of millions of extragalactic sources. This has enabled astronomers to map global trends in the galaxy population (such as star formation rates, stellar masses and luminosity) across much of cosmic time. This DPhil thesis details the results from studies that I have conducted on galaxy populations that span over 90 per cent of the history of the Universe. I begin this work by providing an introduction to the key concepts that drive this research, before discussing new observational results derived from upwards of 2 million luminous sources found in 6 square degrees of the sky.

1.1.1 The Timeline of the Universe

Spanning around 13.7 billion years of evolution, the Universe is a violent and turbulent machine that has led to the formation of structure across a variety of spatial scales. These range from individual hydrogen atoms through to grains of dust, stars, solar systems, galaxies and clusters. Yet despite this complexity, the history of the Universe can be broken down into several key phases (See Figure.1.1). The exact timings where the Universe transitions between these phases is dependent on the fine tuning of the cosmological model being used, but they can be broadly described as follows:

1. **Pre-CMB Era:** During the first $\sim 380,000$ years after the Big Bang, the Universe is in a dense, hot phase and consists of an ionised plasma of unbound particles, such as protons and electrons. The ability for these free particles to interact with light leads to a Universe that is opaque and greatly limits the information available to researchers about this time. As the Universe expands and cools to an average temperature of $\sim$3000K, these particles bind together to form the first atoms, mostly hydrogen and helium, in a process commonly referred to as Recombination or Nucleosynthesis. After Recombination, the Universe quickly transitions to a phase of transparency,

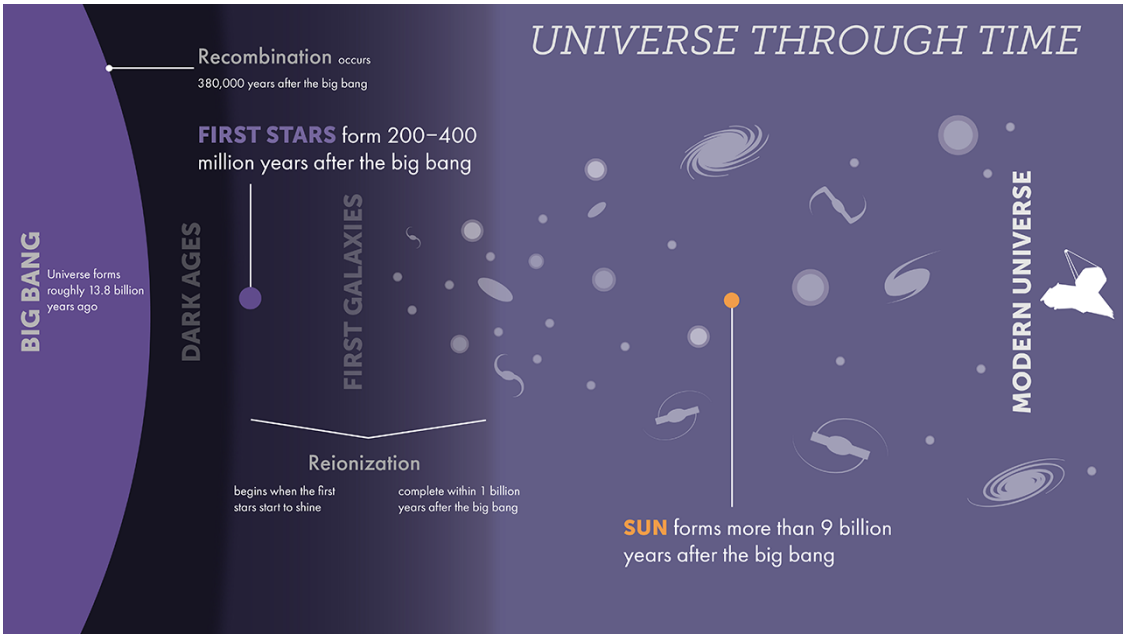


Figure 1.1: A cartoon diagram showing a broad timeline of the Universe. Time flows from left to right and some of the key periods of the Universe's history are labelled. Source: First Stars: Timeline of the Universe, STSci on behalf of NASA.

emitting thermal radiation which is observed today as the Cosmic Microwave Background (CMB; McKellar, 1941; Penzias & Wilson, 1965). Today, the CMB is measured as a background of low energy light that follows a black body with a temperature of $T = 2.73$K (Planck Collaboration et al., 2020).

2. **The Cosmic Dark Ages:** Following the emission of the CMB, random over-densities undergo a process of accumulating additional mass as a consequence of gravitational attraction. This in turn forms the first halos of dark and baryonic matter which become the seeds for future galaxies, galaxy clusters and the wider cosmic web. This era ends when clouds of gas become dense enough to initiate nuclear fusion, forming the first generation of stars.

3. **The Era of Reionisation:** The first generation of stars creates a new source for emitting energetic photons of light into the Universe. Photons with an energy greater than 13.6 electron-volts (eV) are capable of breaking apart hydrogen atoms into the constituent proton and electron, a process called ionisation. With high surface temperatures, stars in infant galaxies begin a

process of ionising the surrounding medium in a process known as Reionisation (since the Universe began in an ionised state). The result is a neutral Universe which is embedded with growing bubbles of ionised matter that surround young galaxies (see Wise, 2019, for a recent review on the subject). The ionisation process can be mapped with multiple observational signatures, such as the polarisation of the CMB (Planck Collaboration et al., 2016), 21cm hydrogen emission (e.g. Morales & Wyithe, 2010; Pritchard & Loeb, 2012) and Lyα emission (e.g. Dijkstra, 2014), and is thought to be complete before the Universe is 1Gyr in age (redshift: $z > 6$). However, debate has persisted over the exact duration of the ionisation process, as well as the contribution from exotic objects, such as Active Galactic Nuclei (e.g. Madau & Haardt, 2015).

4. **Cosmic Noon:** Once the Universe is ionised, galaxies continue to accumulate mass through accretion from the surrounding intergalactic medium (IGM) and from merging events. The peak in the global star formation rate is widely measured to have occurred when the Universe was between 3Gyr–4Gyr in age (e.g. Madau & Dickinson, 2014), and is commonly referred to as Cosmic Noon.

5. **The Low Redshift Universe:** Beyond Cosmic Noon, the Universe becomes increasingly passive due to a number of factors. For example, dark energy limits the distance scales in which gravity can induce mergers of galaxies and galaxy clusters. Moreover, intracluster dynamics, such as ram-pressure stripping, remove reservoirs of gas from galaxies, and heating induced by supernovae and AGN activity suppresses star formation rates (a process commonly referred to as 'feedback'). The endless expansion of the Universe combined with a gradual decrease in star formation will ultimately result in the 'heat-death' of the Universe or a 'Big Freeze', where galaxies will mostly consist of various types of dead star (e.g. white dwarfs, neutron stars and black holes). The fall in global star formation rates is much slower than the

initial rise to Cosmic Noon. It is currently ~ 10 Gyr since Cosmic Noon and the global star formation rate is still measured to be around 10 per cent of the peak (Madau & Dickinson, 2014).

1.1.2 The Cosmological Model

The current leading theory for the evolution of the Universe is commonly called 'ΛCDM'. The name refers to the use of Dark Energy (Λ, which provides the energy that drives the expansion of the Universe) and cold dark matter (CDM, which drives the collapse of matter onto halos that ultimately form galaxies). The theory relies on a number of core physical constants. For the purpose of this thesis, I assume the values of:

$$\Omega_\Lambda = 0.7, \Omega_m = 0.3, H_0 = 70\text{km/s/Mpc}^3 \tag{1.1}$$

where the parameter Ω indicates the fraction of the energy density contained in the form of matter (both baryonic and dark) or in dark energy (Λ). When $\Omega_\Lambda + \Omega_m = 1$ the Universe is geometrically flat and infinite. The Hubble Constant (H_0) is the present rate of expansion of the Universe.

Since the Universe is expanding, there is a connection between the measured recessional velocity (v) of an object and its distance from an observer (d) through the simple relation $v = H_0 d$. On cosmological scales, such recessional velocities are large enough that the Doppler effect stretches observed wavelengths (λ) of light with the following relation:

$$\lambda_{\text{observed}} = (1 + z)\lambda_{\text{rest}} \tag{1.2}$$

where the parameter z is commonly referred to as 'redshift' and is connected to a scale factor (a) that changes with time:

$$\frac{a(t_{\text{observed}})}{a(t_{\text{emitted}})} = 1 + z. \tag{1.3}$$

The scale factor (a) translates a length scale between two epochs of the Universe.

Since the speed of light (c) is a fixed constant in the Universe, a distance from an observer can be converted into a look-back time (time that has passed since emission). Ultimately, this enables an individual to make observations of the Universe at different epochs by simply examining objects at various distances. However, in a reference frame that expands with time, distance is not such a simple concept, particularly if the so-called Hubble Constant H_0 is not, in fact, constant.

A time-dependant Hubble Constant (or Hubble Parameter H) can be defined as:

$$H(t)^2 = \left(\frac{\dot{a}(t)}{a(t)}\right)^2 = \frac{8\pi G}{3}\rho(t) - \frac{kc^2}{a(t)^2} + \frac{\Lambda c^2}{3} \tag{1.4}$$

where G is the gravitational constant, ρ is the energy density of the Universe, k is the curvature of the Universe (0 if flat) and Λ is the cosmological constant derived from Einstein's equations. The rate of change of H is thus:

$$\frac{dH}{dt} = \frac{d}{dt}\left(\frac{\dot{a}(t)}{a(t)}\right) = -\left(\frac{\dot{a}(t)}{a(t)}\right)^2 - \left(\frac{\ddot{a}(t)}{a(t)}\right) = -\left(\frac{\dot{a}(t)}{a(t)}\right)^2 \left(1 + \left(\frac{\ddot{a}(t)a(t)}{\dot{a}(t)^2}\right)\right) = -H^2(1+q) \tag{1.5}$$

where q is denoted as the 'deceleration' parameter.

Since $k = 0$, Λ is theorised to be constant and the energy density in a given volume ρ will decrease in an ever-expanding Universe. The Hubble parameter decreases with time, but ultimately converges to a value $\frac{\Lambda c^2}{3}$ or $\sim 57\text{km/s/Mpc}^3$ as $t \to \infty$.

An expanding Universe allows a number of distances to be defined, replacing the traditional rest-frame distance in various mathematical formulae. Examples include the distance used to convert an apparent luminosity to an absolute luminosity and the distance used to calculate the proper size of an object based on its angular size. Under simple circumstances, these relations would be defined as follows:

$$S = \frac{L}{4\pi d_L^2} \tag{1.6}$$

$$d_A = \frac{x}{\sin\theta} \tag{1.7}$$

where S is the measured flux of an object of luminosity L at a distance d_L and θ is the angular size of an object of proper size x at a distance of d_A. On small length

scales, where the expansion of the Universe has a negligible impact, both of these distances are the same. However, at cosmic length scales these distances (called the Luminosity Distance and the Angular Diameter Distance, respectively) are found to have dependencies on the changing scale factor $a(t)$, which in turn can be related to factors of $(1 + z)$. These two distances are thus defined under ΛCDM as:

$$d_L = a_0 r_0 (1 + z) \tag{1.8}$$

$$d_A = \frac{a_0 r_0}{(1 + z)} \tag{1.9}$$

where the factor $a_0 r_0$ correspond to a 'proper' distance in the case of a static Universe. This adds a redshift dependence to the distances defined in Equation 1.8 and Equation 1.9, leading to:

$$S = \frac{L}{4\pi a_0^2 r_0^2 (1 + z)^2}, \tag{1.10}$$

$$\frac{a_0 r_0}{(1 + z)} = \frac{x}{\sin\theta}. \tag{1.11}$$

This causes the flux S to decrease at a faster rate with distance. Similarly, the redshift dependence on angular size causes objects with redshifts greater than $z \sim 1.5$ to increase in angular size because space was more compressed and the galaxies initially much closer to the observer at emission than they are at reception.

Key to making observations of galaxies is the surface brightness of the source (I), the amount of flux (S) per solid angle (ω) subtended on the sky. The additional redshift terms in the formulae for luminosity distance and angular diameter distance have a significant affect on the measured surface brightness of galaxies, resulting in the following:

$$I_{\text{observed}} = \frac{S}{\omega} = \frac{L}{4\pi d_L^2} \frac{d_A^2}{A} = \frac{L}{4\pi A} \frac{1}{(1 + z)^4} = I_{\text{emitted}} \frac{1}{(1 + z)^4}, \tag{1.12}$$

where A is the surface area of the sky the object subtends and I_{emitted} is the surface brightness of the object if it was at rest / there was no expansion in the Universe. This additional factor of $(1 + z)^{-4}$ causes galaxies of greater distances to become increasingly difficult to detect as a consequence of cosmology.

The galaxy population is dynamic and ever-changing. When discussing how this population evolves in an expanding Universe, a key unit of measurement is the co-moving volume (V_C). This volume is defined so that the number density of a galaxy population with no time dependence on its physical properties remains constant in an expanding Universe. The use of this volume essentially removes the impact of an expanding Universe when considering the number of galaxies per unit volume, enabling for evolution caused by physical processes affecting galaxies to be observed in a manner that is not dominated by the cosmological model. Under ΛCDM, the co-moving volume is defined as:

$$dV_c = \frac{c}{H_0} \frac{(1+z)^2 d_A^2}{\sqrt{\Omega_M (1+z)^3} + \Omega_k (1+z)^2 + \Omega_\Lambda} d\omega dz. \tag{1.13}$$

Subsequently, a redshift interval (dz) and solid angle $d\omega$ of the sky can be converted into a co-moving volume.

1.2 The Formation and Evolution of Galaxies

Arising from the ΛCDM model is much of the physics governing the formation and subsequent evolution of galaxies. Galaxies are gravitationally bound structures that consist of both baryonic (stars, dust, gas etc.) and dark matter. The interactions that these constituent components have with gravity, light, magnetism and pressure determine how and when galaxies form, and how they became the population that we observe today. In this section, I will introduce the current understanding of various components of galaxy evolution.

1.2.1 The First Galaxies

Galaxies are believed to form under a process known as hierarchical growth. In this model, small dark matter halos, generated by random perturbations in the very early Universe, assimilate baryonic matter and merge with each other through gravitational attraction (e.g. White & Rees, 1978; Peebles, 1982; Blumenthal et al., 1984; White & Frenk, 1991). The capability for baryonic matter to cool by radiating

light, causes it to sink to the centre of these halos. It is here that the proto-galaxy is formed, consisting of clouds of mostly hydrogen and helium gas that cool via thermal radiation and compress under gravity.

The random initial distribution of over-densities that form dark matter halos is supported by the non-uniform nature of the CMB, which is measured to exhibit temperature fluctuations (Planck Collaboration et al., 2020). Furthermore, without dark matter halos, simulations have found galaxies would form too slow in order to generate the population that we observe today. Additional observational evidence for dark matter halos stem from observations of the rotational dynamics of local galaxies (Rubin & Ford, 1970), where mass requirements for the observed velocities of the outer parts of galaxies were much larger than could be accounted for based on knowledge of stars, gas and dust distributions. The orbital dynamics of galaxies within clusters (Zwicky, 1933; Smith, 1936; Zwicky, 1937), as well as gravitational lensing, also imply the presence of excess mass than cannot be accounted for by baryonic matter (Walsh et al., 1979; Soucail et al., 1988).

As baryonic gas cools and sinks to the centre of these halos, compression as a result of gravitational attraction will increase the pressure within the gas until nuclear fusion is initiated and a star is formed. This first generation of stars is commonly referred to as Population III and they are characterised by their lack of metal content (Kashlinsky & Rees, 1983; McDowell, 1986). A gas seeded with random density fluctuations will form a population of stars with a mixture of masses. The distribution in the number density of these different mass stars is commonly referred to as an Initial Mass Function (IMF: Salpeter, 1955; Carr et al., 1984; Kroupa, 2001; Chabrier, 2003). The IMF of Population III stars has not been measured because such metal-poor stars will have existed in only the earliest times, further than present-day observations have managed to achieve. However, some authors report evidence from studying newer stellar generations and metal poor stars that the IMF of Population III stars must be 'top-heavy', producing large numbers of ultra-massive stars (100's to 1000's of times the mass of our Sun, $M_{\odot}$) compared to later generations (e.g. Schwarzschild & Spitzer, 1953; Larson, 1998;

Sokasian et al., 2004; Marks et al., 2012). The presence of a significant population of ultra-massive stars can explain why observations of high-redshift galaxies have struggled to detect strong signatures of a distinct set of Population III stars, since more massive stars have shorter lifespans and thus be limited to existing at very high redshifts (see Bromm, 2013, for a review on the subject).

A population of very high mass stars would also generate a population of intermediate mass black holes as they reach the end of their lives (Madau & Rees, 2001). These black holes would have masses of tens to hundreds of solar masses and have the potential to form the seeds for the central supermassive black holes we see in modern galaxies today ($M > 10^6 M_\odot$). These black holes are a key component of a galaxy and their evolution is fundamentally tied to the evolution of their host galaxy (see e.g. Section 1.2.3). However, observations of galaxies at $z \sim 7$ indicate that a population of black holes with masses of around $M \sim 10^9 M_\odot$ are already in place, when the Universe is only a few per cent of its current age (e.g. Bañados et al., 2018; Wang et al., 2019; Yang et al., 2020; Wang et al., 2021). Such observations challenge theories of black hole seed masses, requiring those theories that produce low initial masses to invoke the need for high and almost constant levels of efficient accretion in order to explain observations.

There is a natural limit imposed on the growth rates of black holes, the Eddington Limit. This limit is driven by the balance between the forces of radiation pressure from an accretion disk heated by viscous stresses and gravitational attraction. It is defined as:

$$L_{\text{Edd}} = \frac{4\pi G M_{BH} m_p c}{\sigma_T} \tag{1.14}$$

where the Eddington Limit (L_{Edd}) is dependent on the gravitational constant G, mass of the black hole M_{BH}, the mass of the proton m_p and the Thompson cross section σ_T. An accretion disk generating sufficient radiation pressure will prevent new gas from being accreted on to the supermassive black hole, leading to this Eddington limit. Super-Eddington accretion (accretion faster than allowed by the Eddington limit) may be possible as a consequence of non-spherical accretion by

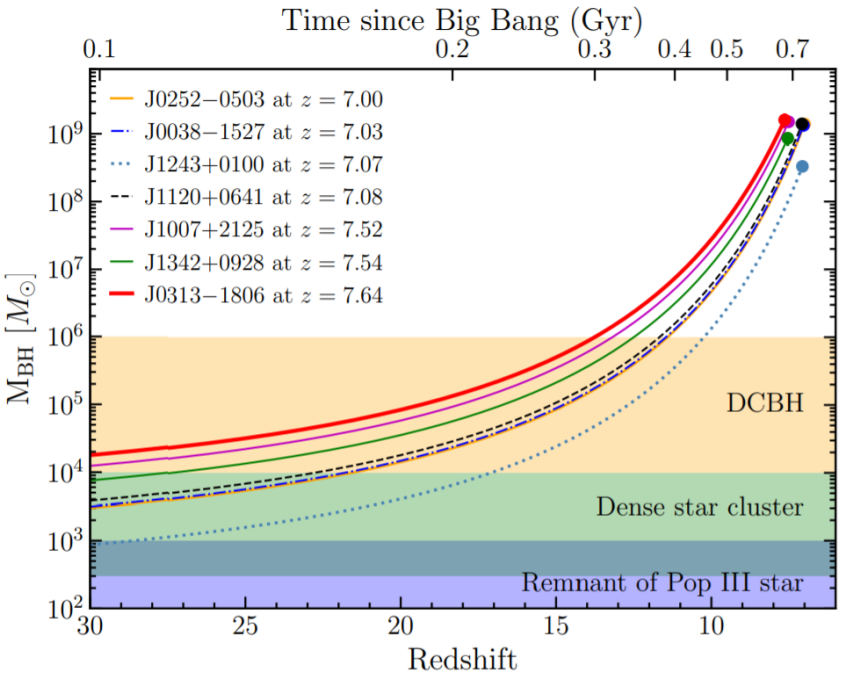


Figure 1.2: A compilation of supermassive black hole mass measurements at $z > 7$. The lines trace the predicted growth of each black hole if it accreted consistently at 10 per cent of the Eddington Limit. Shaded regions indicate different black hole formation scenarios such as Populations III stars and direct collapse of entire clouds (DCBH). Credit: Wang et al. (2021)

black-holes (Abramowicz et al., 1988). This is because conservation of angular momentum causes clouds of infalling gas to increase in rotational velocities and form disk-like structures known as accretion disks. However, pressure resulting from thermal radiation from the infalling gas would still be a significant limiting factor in black hole growth rates. Consequently, if Population III remnants are the seeds for these black holes, they must grow at both high and consistent rates (greater than 1 per cent of the Eddington Limit) in order to form the black hole masses observed (see Fig. 1.2).

Strong UV emission from Population III stars could also be capable of disassociating molecular hydrogen gas H_2, which is more efficient at cooling than atomic hydrogen H. This would reduce the capability of small overdensities to cool at high rates and form other stars, leading the gas cloud to instead collapse as a singular entity into a black hole (known as a direct collapse black hole, DCBH) with the potential for more extreme masses at earlier times (up to $M \sim 10^4$–$10^6 M_\odot$). Under

such a formation mechanism, black holes would not need to have such high and consistent growth rates in order to reproduce observations (e.g. Haehnelt & Rees, 1993; Umemura et al., 1993; Begelman et al., 2006).

1.2.2 Forming New Stars in Galaxies

A key component of the evolution of galaxies is the star formation history, the rate at which a galaxy forms stars versus time. After their initial formation, gravitational attraction provides galaxies with a pathway to accumulate new mass and generate new populations of stars. The rate at which galaxies accumulate this new mass and form new stars is not trivial and it depends on the environment in which the galaxy and its dark matter halo resides.

As observations of galaxies have advanced, a strong correlation has been revealed between the star formation rate of a galaxy and its total stellar mass (the total amount of mass of a galaxy in the form of stars). Galaxies with larger stellar masses tend to have greater star formation rates than those with lower stellar masses (Brinchmann et al., 2004; Noeske et al., 2007; Elbaz et al., 2007; Daddi et al., 2007; Magdis et al., 2010; Schreiber et al., 2015). The relation is consistent up to masses of around $10^{11} M_{\odot}$, at which point some observational studies have shown it plateaus (Schreiber et al., 2015; Tomczak et al., 2016). The result is that galaxies of different stellar masses exhibit very similar specific star formation rates (SSFR), defined as the star formation rate divided by the total stellar mass. This correlation of stellar mass and star formation rate is termed the star forming main sequence.

The formation of new stars in a galaxy requires a sustained reservoir of cold gas. Thus, in order to increase the stellar mass of a galaxy, there must be mechanisms in place for galaxies to accumulate new stores of cold gas. Measurements of the star forming main sequence show that there has been a slow decline in the star formation rates of galaxies across all masses since the Era of Reionisation. In particular, the most massive galaxies have been measured to experience the greatest decline (See Fig. 1.3). This evolution in star formation rates generates the peak in the global star formation rate in the universe at $z \sim 2$ and suggests that mass-dependent

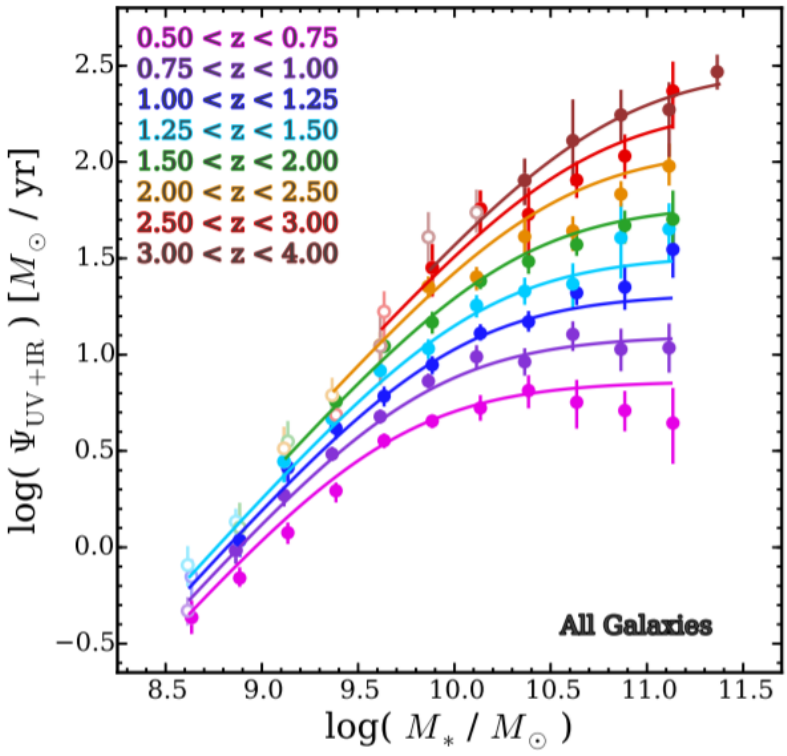


Figure 1.3: The star forming main sequence as measured by Tomczak et al. (2016), showing star formation rate (Ψ) against stellar mass (M_*). It is clear here that galaxies were more star forming in the earlier universe (higher z) and that high-mass galaxies have experienced the largest decrease in their star formation rates with time.

mechanisms enabling the accumulation of new gas and the formation of new stars are needed. Below, I describe some physical processes which can enable galaxies to increase their total stellar mass .

Accretion from the Cosmic Web

Galaxies are not isolated systems, but reside within a larger cosmic structure known as the 'Cosmic Web'. This Cosmic Web consists of high-density nodes connected to each other by filaments of matter and voids occupying the remaining space. These nodes and filaments are made up of both dark and baryonic matter. Galaxies can reside within these filaments, and galaxy clusters are found to occupy the nodes of this web-like structure. Subsequently, galaxies can accumulate new material from these structures, replacing cold gas reservoirs at rates of 10's of solar masses a year (e.g. Kereš et al., 2005; Dekel et al., 2009; Kleiner et al., 2017; Martin et al., 2019).

Without such accretion of new fuel, galaxies would run out of cold gas within a few Gyr based on measurements of gas content and star formation rates (Bigiel et al., 2008; Leroy et al., 2008; Rahman et al., 2012; Ho et al., 2019).

Direct evidence for the infall of new material onto galaxies is difficult to measure due to the low density of the gas involved. Some experiments have used luminous background sources in order to search for absorption from intervening gas surrounding lower redshift galaxies (e.g. Tumlinson et al., 2013). They have found that substantial quantities of gas do exist beyond the main disks of galaxies. Simulations have shed further light on this by showing that accretion from filaments is a dominant form of of gas replenishment in galaxies of all sizes in the early Universe and remains significant for galaxies of low masses today (e.g. Kereš et al., 2005, 2009; Brooks et al., 2009). This process is thought to enable low-mass galaxies to maintain relatively high levels of star formation for longer periods of time.

Merger events

A consequence of gravitational attraction, is that galaxies and dark matter halos attract each other. On small length scales, where the expansion rate of the Universe is insufficient to push two dark matter halos apart, the two galaxies can merge with each other, leading to the assimilation of two smaller galaxies/dark matter halos into a single larger galactic system. The result is a new galaxy with a mass of the two original galaxies combined. The merging process of galaxies can take more than hundreds of millions of years and the increased turbulence of gas can lead to a temporary elevation in star formation rate (e.g. Duc et al., 1997; Conselice et al., 2003; Bridge et al., 2007). The original stellar populations are mostly unaffected by the merger. This is because the combination of the small sizes of stars, large distances between them and high velocities during the merging process lead to highly improbable cases of direct collision between individual stars. However, mergers have been observed to eject a small portion of their stellar populations during the process (Puglisi et al., 2021).

Merger events are commonly categorised as two types: major mergers and minor mergers (Bournaud et al., 2005; Lotz et al., 2010). Major mergers occur when two galaxies of approximately the same mass merge with each other, while minor mergers are the case where galaxies with large mass differences (a factor of 10 or greater) merge with each other. Major mergers can lead to large changes in the morphology of galaxies (e.g. Toomre, 1977; Negroponte & White, 1983; Hopkins et al., 2009; Ferreras et al., 2009). Minor mergers typically consist of the accretion of satellite galaxies onto larger galaxies. They are thought to be the reason that galaxies follow a size-mass relation (where more massive galaxies have larger effective radii), indicating that satellite galaxies merge by coalescing with the outer edges of the larger system (Naab et al., 2009; McLure et al., 2013).

Galaxy mergers can also be defined as 'wet' or 'dry' mergers, based on how gas rich the two parent galaxies are at the time of merging (e.g. Lin et al., 2008, 2010). So-called 'wet mergers' generate spikes in star formation rates during the merging process. Dry mergers occur when two gas-poor galaxies (typically very massive galaxies) merge, which leads to no change in the overall star formation rate and a simple summation of the two stellar populations (Khochfar & Silk, 2009).

Galaxy mergers are thus mechanisms that lead to large increases in stellar mass over shorter periods of time compared to cold gas accretion. However the process leads to a net decrease in the total number of galaxies. Simulations have shown that the most massive galaxies contain the highest fractions of stars which were formed outside of the primary progenitor galaxy (e.g. Rodriguez-Gomez et al., 2015; Qu et al., 2017). In other words, more stars were formed in a secondary galaxy before merging.

1.2.3 Galaxy Quenching and Feedback Mechanisms.

As well as physical processes that encourage the growth of galaxies and the formation of new stars, there are processes which can hinder the accumulation of matter and formation of stars. Galaxies with substantially low star formation rates compared to the wider population are often called 'passive' or 'quenched'. They are characterised by their lack of massive, blue, short-lived stars, leading to redder optical colours

than their star-forming counterparts. Quenched galaxies can therefore be selected using rest-frame colours (e.g. the NUV-r-J technique; Williams et al., 2009; Ilbert et al., 2010) or by selecting galaxies with measured star formation rates which fall significantly below the star-forming main sequence.

Passive galaxies are found to be more prevalent in high-density environments, such as at the centre of galaxy clusters, which may indicate that environment plays an important role in how a galaxy quenches (e.g. Kauffmann et al., 2004; Bernardi et al., 2006). Processes which introduce heating into reservoirs of baryonic gas are also key to reducing star formation rates, since the formation of stars is dependent on the cooling of gas through thermal emission. In this section, I will introduce a selection of key physical processes which can hinder star formation and the accretion of new baryonic matter onto galaxies. These processes include physics resulting from interactions external to the galaxy (known as environmental quenching) and internal effects (known as feedback).

Environmental Quenching

Nodes within the Cosmic Web are dense environments, where upwards of hundreds of galaxies can group together in large, gravitationally bound systems known galaxy clusters. Outside of clusters, galaxies can accumulate smaller satellite companions either through the hierarchical merging of dark matter halos or the collapses of baryonic matter within a sub-halo (Mandelbaum et al., 2006; Shi et al., 2020). Gravitational attraction between galaxies generate complex orbits that cause galaxies to travel through dense filamentary/intracluster media or have close passes and mergers with their neighbours. Each of these interactions has consequences on the matter contained within a galaxy.

A key theory for the quenching of galaxies in dense environments is ram pressure stripping. This process occurs when a galaxy travels through hot intracluster gas. Direct collision of cold gas reservoirs with this medium leads to an effective pressure that can strip cold gas from a galaxy (e.g. Gunn & Gott, 1972; Kapferer et al., 2009; Ebeling et al., 2014). This pressure is dependent on the densities of the gas

Figure 1.4: An example of a jellyfish galaxy ESO 137-001, a member of the Norma Cluster. Observations with the Hubble Space Telescope (blue, white, yellow), MUSE (purple) and ALMA (red) show a long tail of matter trailing behind it, believed to be a consequence of ram pressure stripping. Credit: ALMA / ESO / NAOJ / NRAO / Jáchym et al. (2019)

as well as the relative speed between the galaxy and the medium it is travelling through. Observational evidence of ram pressure stripping is focused on galaxies in clustered environments, particularly so-called 'jellyfish galaxies' (see Fig.1.4) which have long tails of matter trailing behind them and evidence of bow-shocks on their leading edges. There has also been propositions that dwarf galaxies can be stripped by the filamentary medium under a similar process (Benítez-Llambay et al., 2013). Ram pressure stripping ultimately leads to an increasingly gas-poor galaxy and a reduction in star formation rate.

A second process is named Galaxy Harassment and involves the close passing of two galaxies. Since galaxies are 3-dimensional structures, when two galaxies pass each other with close proximity, the force of attraction from gravity varies across the disk of each galaxy. This varying force across each galaxy distorts their morphology, generating long tidal tails of stars and gas (Farouki & Shapiro, 1981; Mihos et al., 2005; Wetzel & White, 2010; Bahé et al., 2019). As with ram pressure stripping, this physical effect leads to the stripping of material from the

galaxies. The large numbers of stars present in these tidal tails make observations of this process more simple than in the case of ram pressure stripping, and it generates galaxies with more extended or complex morphologies. The velocities of these tidal disruptions can exceed the escape velocity of the galaxy, leading to the loss of stars and gas to the intergalactic/intracluster medium and reductions in total mass and star formation rate.

Supernova Feedback

Another form of quenching occurs when stars reach the end of their lives, those with masses greater than eight solar masses will enter a short violent phase known as a supernova. Triggering mechanisms include runaway nuclear reactions of heavy elements, accretion between binary pairs of stars and the collapse of stellar cores in the most massive stars. It is these massive stars undergoing core collapse that release the most energy, in the form of both radiation and winds of matter, into their surroundings (see Janka, 2012, for a recent review of core-collapse supernova). The timescale of these events is very short, of order tens of days verses the millions to billions of years that stars can live for. The mass expelled by supernovae can be both dense and of high velocity, forming a strong wind which can impart energy on the surrounding intergalactic medium. In small dwarf galaxies, the velocities of these winds can exceed the escape velocity of the galaxy, leading to an ejection of baryonic material over time (e.g. Larson, 1974; Dekel & Silk, 1986; Efstathiou, 2000). In galaxies with slightly higher masses, the particles making up these winds can collide with the atoms/molecules that make up the cold gas in the interstellar medium, inducing shocks and leading to a net increase in the temperature of the gas. For more massive galaxies, the effects of supernovae are more localised due to the much stronger gravitational potential that these galaxies possess.

Supernova feedback thus disproportionately affects the low-mass population of galaxies. It is postulated to create gas-poor ultra-faint dwarfs (UFDs; e.g. Simon, 2019) and limit the rates at which more intermediate-mass galaxies can form new

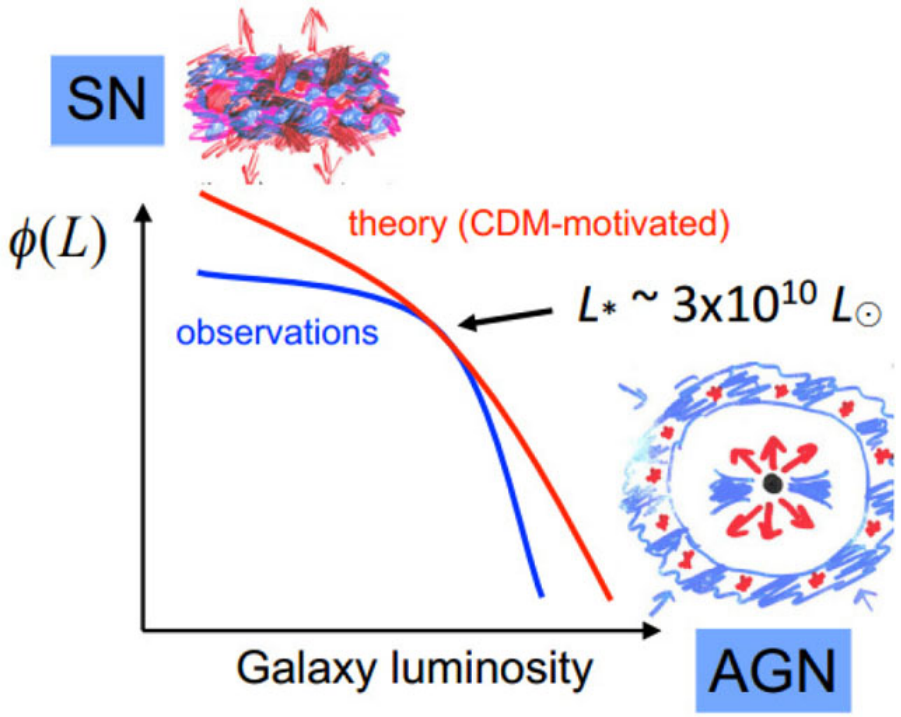


Figure 1.5: A cartoon diagram showing how feedback effects lead to a deficit of low-luminosity/mass systems through supernovae feedback and high-luminosity/mass systems through AGN feedback. These diagrams are commonly referred to as Luminosity Functions, they show the space density (Φ) of galaxies as a function of intrinsic luminosity (L). The red line indicates the number of galaxies that would be expected at different luminosities based on a simple ΛCDM simulation, while observations provide results more akin to the blue line. Luminosity Functions are discussed in greater detail in Section 1.3. Source: Silk & Mamon (2012)

stars. The energy introduced by supernova feedback has been found to solve 'over-cooling' problems found in simulations, where such low-mass galaxies were found to have much greater star formation rates than observed (e.g. Rees & Ostriker, 1977; White & Rees, 1978; Cole et al., 2001). Supernovae are also essential mechanisms for distributing metals within a galaxy. The high speeds of supernova ejecta quickly pollutes the surrounding gas with high-mass elements, which leads to a greater metal content in the next generation of stars and provides the material required to form planetary systems (e.g. Gnedin, 1998; Scannapieco et al., 2006).

AGN Feedback

The quenching of a galaxy can also be linked to the evolution of supermassive black holes. It is now widely accepted that the majority of the massive galaxy population

have a central supermassive black hole (Kormendy & Richstone, 1995; Magorrian et al., 1998). Such black holes can undergo phases of activity when material (be it a star or gas) passes by with insufficient speed to escape. Due to conservation of angular momentum, when material falls onto a black hole the increased speeds cause a randomly shaped cloud to condense into a more disk-like structure called 'accretion disks'. With high velocities and large internal velocity gradients, viscous stresses (friction) within an accretion disk can lead to the conversion of gravitational potential energy to thermal energy and the emission of light (see Abramowicz & Fragile, 2013, for a review on black hole accretion disks). A black hole undergoing this process is often referred to as an Active Galactic Nuclei (AGN). The temperature and surface area of such disks can be significant enough for AGN to have a luminosity comparable to, or greater than, the total population of stars contained within a galaxy.

AGN are capable of generating two primary sources of feedback which inhibit star formation in galaxies, one which is mechanical in nature and one which is based on radiation (e.g. Cole et al., 2002; Benson et al., 2003; Begelman, 2004; Bower et al., 2006). The mechanical mechanism is a consequence of bipolar jets of material which are produced by AGN. The physics behind the creation of these jets is currently poorly understood, but theories of jet production generally call upon strong twisted magnetic fields generated by the black hole and the accretion disk (e.g. Lu, 1991; Rawlings & Saunders, 1991; Natarajan & Pringle, 1998; Caproni et al., 2004; Narayan & Quataert, 2005) in order to generate the energies required to match observations of the velocities of these jets as well as produce the bipolar morphology (see Blandford et al., 2019, for a review on AGN jets). These jets have been measured to have extreme velocities that enable them to travel multiple effective radii of the host galaxy. When the jets impact the medium immediately surrounding the galaxy, shocks are generated which leads to a rise in temperature and a slowing of the accretion of new baryonic matter into the host galaxy. The impacts of these jets are observed in radio and x-ray imaging, which reveals the extent of these shocks and increased IGM temperatures (See Fig. 1.5 for an example of such an observation).

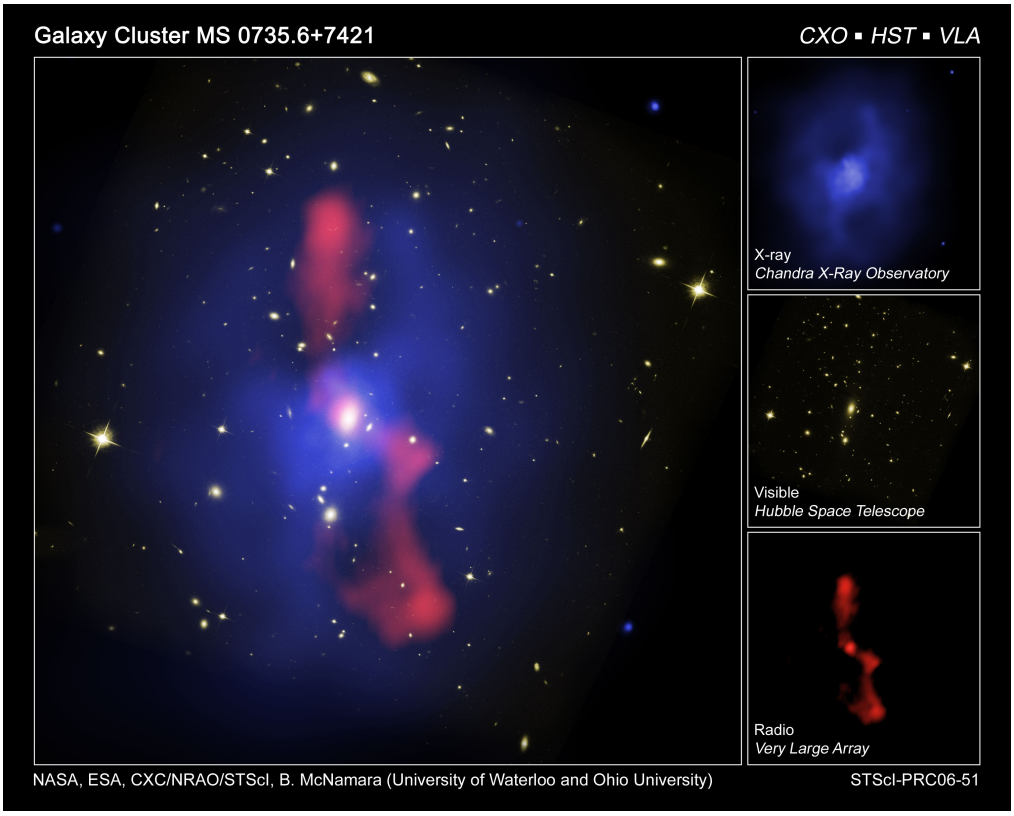


Figure 1.6: A composite image showing the galaxy cluster MS 0725.6+7421. This cluster features a massive central galaxy with ongoing AGN activity. This is evidenced by large radio jets (radio: red) which have carved out regions of the hot intracluster medium surrounding the galaxy (xray: blue). Credit location within the image.

Radiation-based feedback is the result of the extreme luminosities that can be generated by AGN accretion disks. Radiation pressure can generate outflows of gas and dust from the central regions of galaxies as well as increase the net temperatures of gas reservoirs in the galaxy. It requires the central black hole to be accreting at efficient rates, greater than around 1 per cent of the Eddington Limit. Observational evidence of this mechanism can be obtained through spectroscopic studies and surveys (e.g. The Sloan Digital Sky Survey contains of order 1 million AGN spectra; Hall et al., 2002; Albareti et al., 2017) targeting the excess blueshifting of broad absorption lines (BAL). These are driven by high-velocity outflowing gas (> 1000 km/s) in galaxies known to contain AGN (e.g. Weymann et al., 1991; Di Matteo et al., 2005; Tremonti et al., 2007; Arav et al., 2008; Moe et al., 2009; Dunn et al., 2010; Alexander et al., 2010; Hopkins & Elvis, 2010). Simulations have also

found relative success in implementing this mechanism and replicating observations of the global galaxy population (e.g. Moll et al., 2007; Gabor & Bournaud, 2014; Koudmani et al., 2019; Richings et al., 2021).

There are additional plausible mechanisms through which AGN can impart energy on their surroundings. One additional example is magnetohydrodynamical (MHD) winds, where magnetic forces drive outflows from accretion disks on top of winds generated by thermal radiation pressure (e.g. Konigl & Kartje, 1994; Ouyed & Pudritz, 1997; Balbus & Hawley, 1998; Krasnopolsky et al., 1999; Proga, 2003). If the cooling time (energy loss rate) of these thermal and MHD winds is longer than the flowing time (time taken to traverse through the galactic bulge and disk), then shocks can be generated in the IGM of the host galaxy. These absorb energy from the AGN driven wind, slowing it down, and generates an accelerating forward shock which expels gas from the system (Dyson & Williams, 1997; King, 2003; Zubovas & King, 2012).

Since the Eddington Limit scales with black hole mass, the greatest AGN feedback effects are generated by the most massive black holes. There is an observed trend for galaxies of larger stellar mass to have more massive supermassive black holes (e.g. Reines & Volonteri, 2015), and so the strength of AGN feedback increases with galaxy mass (King & Pounds, 2015). This results in AGN feedback being a key limiting factor in how massive galaxies can be in terms of baryonic mass and how many stars they form. Observations of galaxies reveal there to be a drastic cut-off in the numbers of galaxies with stellar masses greater than around $M > 10^{11} M_{\odot}$ (e.g. Fontana et al., 2004; Pérez-González et al., 2008; Marchesini et al., 2009; Pozzetti et al., 2010; Ilbert et al., 2013; Muzzin et al., 2013; Davidzon et al., 2017a; Wright et al., 2018). The relatively few galaxies that do exist above this characteristic mass have been found to be very red in colour and highly quenched (e.g. Baldry et al., 2012; Ilbert et al., 2013; Davidzon et al., 2017a; McLeod et al., 2020).

1.2.4 The Emission of Light from Galaxies

To understand the physical processes that govern the formation and evolution of galaxies, an understanding of the processes that lead to photon emission is required in order to correctly interpret observations. Galaxies consist of multiple different light-emitting sources, from cold hydrogen gas to hot accretion disks around black holes, which generate photons of energies ranging from gamma rays through to radio waves. The work conducted as a part of this thesis focuses on light within the wavelength range 3000-45000Å , otherwise known as the near-ultraviolet, optical and near-infrared regimes. In this section, I provide an introduction to the contributions provided by the constituent components of galaxies provide to the total spectral energy distribution (SED, the wavelength-dependent energy output) of a galaxy and explore common methods used to observe them.

For most galaxies, stars dominate the rest-frame ultraviolet and optical emission. The light emitted by a star is dependent on its surface temperature and can vary from warm brown dwarfs (with a temperature of a few hundred Kelvin) to massive O and B type stars that have a surface temperature of tens of thousands of Kelvin. The shape of an individual star's SED largely depends on thermal black body emission, with the exception of small-scale features caused by elements absorbing and emitting light in the stellar atmosphere. A star's photon emission can thus be estimated to follow Planck's law (Equation 1.15) and Wein's displacement law (Equation 1.16):

$$B_\lambda(T) = \frac{2hc^2}{\lambda^5}\frac{1}{e^{hc/\lambda k_B T} - 1}, \tag{1.15}$$

$$\lambda_{\text{peak}} T = 2.9 \times 10^{-3}\text{mK}, \tag{1.16}$$

where $B_\lambda(T)$ is flux per unit of wavelength and steradian. h is Planck's constant, k_B is Boltzmann's constant, T is the mean temperature and λ_{peak} is the wavelength corresponding to peak emission (Planck & Masius, 1914).

Hotter stars are consequently 'bluer' in their emission, while colder stars are 'redder'. Stars with high surface temperatures are more massive and have shorter lifespans. Consequently, the presence of significant ultraviolet and 'blue' optical

emission is an indicator of a young stellar population. This is because a galaxy with an older stellar population would have had the 'bluer' population of stars advance beyond their main sequence lives. They would then take the form of a stellar remnant (white dwarf, neutron star or black hole), which are not as luminous in ultraviolet/blue-optical light. Using an initial mass function, the SED generated by the sum of all stars can be used to estimate the star formation rate of a galaxy through a detailed examination of the balance of stellar emission from various masses/temperatures/'colours' of star (e.g. Chabrier, 2003; Bruzual & Charlot, 2003). In addition, the same principle can be used to estimate the total mass in the form of stars (stellar mass $M_{\odot}$).

While the sum of many black bodies describes the light that is emitted externally from the stars, this light has to pass through the interstellar medium (ISM) of the galaxy and the intergalactic medium (IGM) between galaxies before it can be observed by a telescope. These are made up of gas and dust, which are capable of absorbing and reprocessing stellar light. These factors are required to be understood in order to uncover the properties of the stellar population in a galaxy. In the following subsections, I introduce three primary features that are imprinted on the total SED by these processes.

Lyman and Balmer Breaks

The first feature is commonly known as the Lyman Break. It is a feature that is imprinted on a galaxy SED by cold hydrogen gas located within the galaxy and in the intergalactic medium. Hydrogen atoms in the ground state ($n = 1$) can be ionised by absorbing light of ultraviolet wavelengths shorter than 912Å. In addition, between 912Å and 1216Å, there are a series of electron energy levels that enable hydrogen atoms to absorb light of specific wavelengths in this regime, exciting the hydrogen atom to a higher state ($n > 1$). Ionised and excited hydrogen will over time recombine/relax back into the ground state, causing the emission of an ultraviolet photon corresponding to the same energy. However, the direction in which this photon is radiated is random, with a very low probability that the

emitted photon has the same directional vector as the photon initially absorbed. This absorption and remission from hydrogen gas in intervening gas consequently leads to a net reduction in the amount of ultraviolet light radiated by a target source as it travels to an observer, a process known as 'scattering'. This results in a deep absorption feature in the observed SED of galaxies in rest-frame wavelengths less than 1216Å , named the Lyman Break. A mathematical treatment for estimating the amount of absorption caused by the Lyman Break was derived in Madau (1995). It shows that there is a strong redshift dependence on the depth of the absorption feature, with upwards of 50 per cent of the far-ultraviolet flux emitted by a source scattered by intervening hydrogen gas at $z > 3.5$. In Figure.1.7 I show how the Lyman Break leads to a significant drop in the blue-end luminosity of the galaxy SED at high redshifts, causing galaxies to 'drop out' of the images taken with the bluest filters. This deep absorption feature can thus be used to estimate redshifts of sources using photometric images (Guhathakurta et al., 1990; Steidel & Hamilton, 1992; Steidel et al., 1996), this is discussed in more detail in Section 1.3.4.

The second feature is known as the 'Balmer Break' and it follows the same process as the Lyman Break, but is applied in the case that hot hydrogen gas collisionally sustains an excitation level of $n = 2$. Such a hot gas primarily occurs in the uppermost atmosphere of stars and their coronae. Although hotter stars can sustain larger quantities of excited hydrogen, a stellar atmosphere that is too hot (e.g. O and B-class stars) can totally ionise the gas instead. The excited gas absorbs light of wavelengths shorter than 3646Å and recombination re-emits the photons in random directions, resulting in the Balmer Break. Since the stars of highest temperature do not exhibit a Balmer Break, because they fully ionise their surrounding hydrogen, an observation of a strong Balmer Break in a galaxy SED is indicative of a moderately aged stellar population $\geq$ 1Gyr (e.g. Poggianti & Barbaro, 1997; de Barros et al., 2014; López-Corredoira et al., 2017). This break can have increased depth and extend redwards to 4000Å if the stellar population has a deficiency in hot, blue stars and a high metal content in their atmospheres,

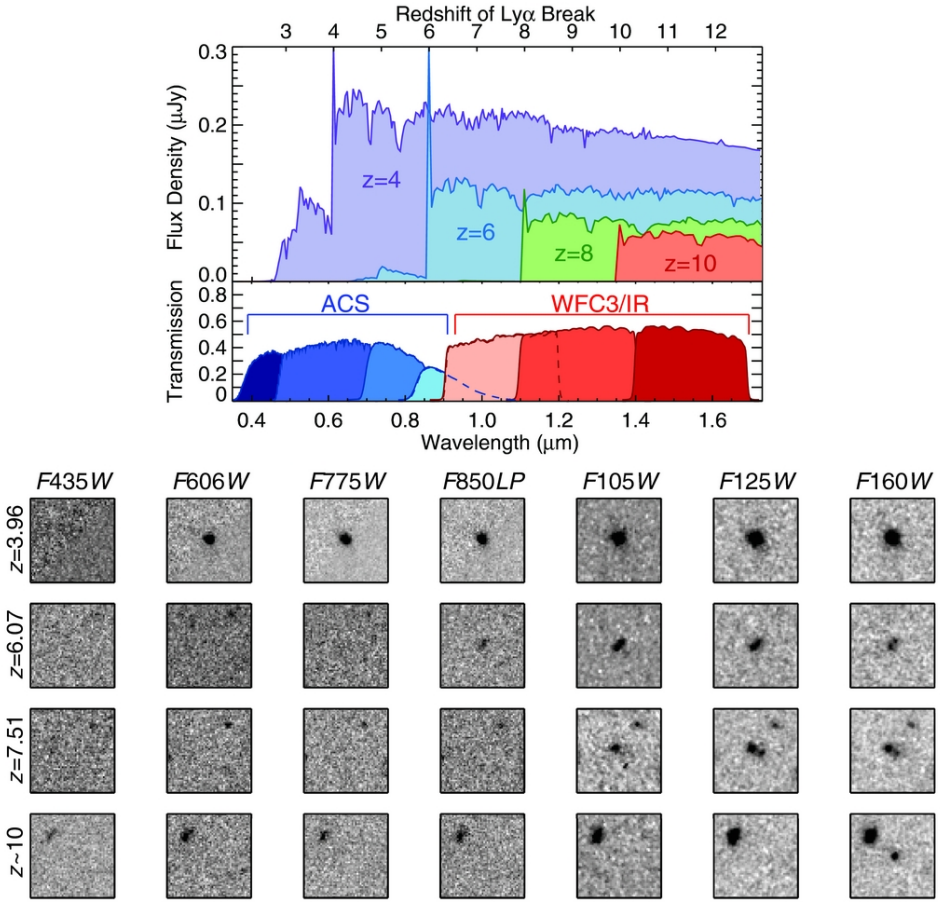


Figure 1.7: Top: The impact of the Lyman Break on a galaxy SED. For wavelengths lower than 1216Å in the rest-frame, absorption from hydrogen gas in the IGM causes a significant deficit in the observed flux from an extragalactic source. Bottom: The observational impact of the Lyman Break on a photometric survey using the Hubble Space Telescope. Images from blue (left) to red (right) of galaxies at different redshifts show that absorption from the Lyman Break can cause galaxies to become undetectable in the bluer bands. Source: Finkelstein (2016)

generating absorption lines. One key contributor is Calcium, with the Ca-H and Ca-K lines located at 3934Å and 3969Å (Bruzual A., 1983; Poggianti & Barbaro, 1997)

Dust Extinction

A third feature in a galaxy SED is a consequence of absorption by micron-sized dust grains within the interstellar medium (ISM). Dust is capable of absorbing ultraviolet and optical light, which in turn causes a rise in temperature in the dust grain. In equilibrium, dust in the ISM has a typical temperature of a few tens to a few hundreds of Kelvin. As a result of this, the dust favourably emits in the far-infrared

through black body radiation. Consequently, ultraviolet and optical light absorbed by dust is reprocessed to longer wavelengths, resulting in dusty galaxies having a more 'red' colour than their stellar population would otherwise create. The effect of dust on UV/blue light can be modelled through the Calzetti Law (Calzetti et al., 2000), which assumes that the dust acts as a foreground screen between an observer and some point-like stars, or by adopting an empirical dust-extinction model e.g. using observations of the Milky Way or the Large/Small Magellanic Clouds (e.g. Witt et al., 1984; Prevot et al., 1984; Clayton & Martin, 1985; Gordon et al., 2003). The effects of dust can be implemented using the following formula:

$$f_{\rm intrinsic}(\lambda) = f_{\rm observed}(\lambda) \times 10^{0.4\kappa^e(\lambda)E(B-V)}. \tag{1.17}$$

Here, the difference between the intrinsic stellar continuum to the one observed is related to an 'extinction law' (κ^e) and an 'extinction strength' ($E(B-V)$). The extinction strength corresponds to the change in the colour between classical broadband photometric filters B and V ($(B-V)_{\rm observed} - (B-V)_{\rm intrinsic}$). The Calzetti extinction law is defined as:

$$\begin{aligned}\kappa^e(\lambda) &= 2.659(-1.857 + 1.040/\lambda) + 4.04 \ (0.63\mu{\rm m} \leq \lambda \leq 2.20\mu{\rm m}) \\ &= 2.659(-2.156 + 1.509/\lambda - 0.198/\lambda^2 + 0.011/\lambda^3) + 4.04 \ (0.12\mu{\rm m} \leq \lambda < 0.63\mu{\rm m})\end{aligned} \tag{1.18}$$

For the purpose of the work presented in this thesis, I use the Calzetti law because it is derived from local starburst galaxies, which are thought to be more analogous to higher-redshift sources, which generally have higher star formation rates.

The role of dust attenuation in the ultraviolet emission of a galaxy is a key consideration that must be taken into account when making inferences from galaxy SEDs. The presence of dust in galaxies reduces the net output of ultraviolet emission and is found to contribute towards the drastic fall in the number density of ultra-luminous galaxies (e.g. Gonzalez-Perez et al., 2013; Ma et al., 2019). Ultraviolet emission thus needs to be contextualised with dust emission in the infrared in order to successfully discern between obscured activity (be it star formation or AGN emission) or a genuine lack of activity in the source. Starburst galaxies have been

found to display a relation between the ultraviolet spectral slope (how blue the object is) and the far-infrared excess (the ratio of UV emission and far-infrared emission) which can be used to estimate the dust content of the galaxy. As these starburst galaxies exhibit star formation levels which are similar to those at higher redshifts, this relation has previously been applied to obtain dust-corrected star formation rates in the early Universe (Meurer et al., 1999; Bouwens et al., 2014).

Active Galactic Nuclei

Alongside stars and dust, AGN accretion disks also emit light thermally, however the temperature of the disk is not uniform and is higher towards lower radii. Accretion disks thus display a much broader spectrum than would be expected from a black body of a single temperature. The subsequent 'Big Blue Bump' (BBB) is a broad spectral feature that inflates the total ultraviolet and blue-optical emission of a galaxy. Examining this feature can provide insight into the physics of AGN. For example, observations have found that the peak of the BBB lies deep in the UV at around 100Å (Marchese et al., 2012). This indicates that temperatures within the innermost parts of AGN disks will be greater than 10^5 K (Bonning et al., 2007). Alongside this, upwards of 50 per cent of the total bolometric luminosity of an AGN is contained within the feature (Koratkar & Blaes, 1999). This indicates that most of the gravitational potential energy of matter accretion onto a black hole is emitted in these wavelengths.

Alongside continuum emission from the BBB, AGN are a source of broad emission lines in galaxy spectra (see Gaskell, 2009, for a review on the subject). These features can contain sufficient flux to impact measurements made in broadband photometric filters (e.g. Richards et al., 2006; Brown et al., 2019). Such emission lines (e.g. CIV, MgII, Hα) have the potential to cause the perceived normalisation and slope of the continuum emission of a source to be incorrect if they are ignored in any analysis. In addition, large dust reservoirs within the torus structures that surround many AGN (see Ramos Almeida & Ricci, 2017) provide a mechanism to absorb and reprocess ultraviolet and optical emission from AGN into thermal infrared emission.

This can result in a significant infrared 'bump' (wavelengths from a few microns to a few hundred microns) comprising many tens of percent of the total infrared power of a galaxy hosting an AGN (e.g. Fritz et al., 2006; Zhuang et al., 2018).

Emission from AGN can fall foul of the same physics described for the stellar emission. This means it can exhibit a Lyman Break as well as generate a peak of infrared emission as a consequence of reprocessing from dust in the ISM. A galaxy that has significant AGN contributions to its overall SED can complicate the aforementioned methodologies for modelling the stellar population and determining the total stellar mass and star formation rate. This is because AGN exhibit a variety of strengths, with some outshining their host galaxy, and levels of dust obscuration may be different in the central parts of a galaxy compared to regions of active star formation. If contributions from AGN to the galaxy SED are not taken into consideration, physical parameter estimations are likely to be incorrect (e.g. Salvato et al., 2009).

1.3 Luminosity and Mass Functions

Each mechanism detailed in Section 1.2 leaves an impact on how the population of galaxies in the Universe grow and form new stars. Key to understanding the magnitude of these, alongside what time dependence they may have, are observational measurements known as Luminosity Functions and Mass Functions. These map the time evolution of the space density of galaxies (Φ) as a function of intrinsic luminosity (L) or mass (M), enabling for changes in the population of galaxies to be mapped. However, the expanding Universe complicates these measurements as a proper volume at one time is not directly comparable with a proper volume at another time. This is where the co-moving volume (defined in Equation 1.13) is very useful, since its use removes the effects of Universe expansion, allowing trends that are the result of physics directly affecting galaxies to be revealed. In this thesis, I measure the Galaxy Stellar Mass Function (GSMF) between $0.1 < z < 2.0$ and the Ultraviolet Luminosity Function (UV LF) between $3.5 < z < 5.2$, using new data sets to tackle outstanding questions regarding the

growth of the general galaxy population. The properties of these Luminosity/Mass Functions are detailed further throughout this section.

Luminosity and mass functions are key measurements for assessing how galaxies evolve. This is because the combined effects of processes that both encourage and hinder galaxy growth are reflected in the time evolution of these functions. Cosmological simulations use them as a benchmark for examining success in replicating observations. Local ($z = 0$) luminosity and mass functions often form the basis for tuning each physical process involved and trialling alternative theories for galaxy evolution. Consequently, accurate observations are required in order to draw the correct conclusions from theoretical work.

In both mass and luminosity functions, a steep decline is observed in the number of galaxies of mass/luminosity above a characteristic value. In its most basic form, these functions are modelled as a power law slope to describe the faint/low-mass population that transitions into an exponential fall-off at some value of luminosity/mass. Such a form is named the Schechter function (Schechter, 1976) and is defined as:

$$\Phi(L) = \frac{\Phi^*}{L^*}(\frac{L}{L^*})^{\alpha} \exp(-\frac{L}{L^*}), \tag{1.19}$$

where Φ is the luminosity (or mass) dependent number density, Φ^* is a normalisation for the profile, L^* is the characteristic luminosity (or mass) and α is a power law slope used in describing the low-luminosity/mass population. The Schechter function has been extensively used over the years, however improvements in the accuracy and precision of observations have shown deviations from such a functional form for both mass- and luminosity-dependent number densities. I discuss these, along with other nuances about each measurement, in the following subsections.

1.3.1 Ultraviolet Luminosity Function of Galaxies

The Ultraviolet Luminosity Function (UV LF) measures the number densities of galaxies based on their rest-frame ultraviolet emission, typically wavelengths around 1500Å . At high redshifts ($3 < z < 6$) the rest-frame ultraviolet emission of galaxies

is redshifted into the optical regime of a distant observer. Ultraviolet emission is particularly important when discussing the high-redshift/early Universe as it drives the reionisation process and is an indicator for both star formation and AGN activity, key processes in how galaxies have evolved. Integrating under the UV LF provides the total UV power output of all galaxies. It can be used determine which types of galaxies made the greatest contribution towards the reionisation process (e.g. was it the more numerous but fainter galaxies or the rarer ultra-luminous sources?).

Different survey strategies are more capable of measuring the UV LF in different luminosity regimes. Due to the observed steep decline in the number density of the most luminous objects, surveys with large areas are required to probe enough volume to find these rare objects. In order to be time efficient, these surveys consequently sacrifice the image depth required to successfully detect the fainter galaxies. Surveys using the Hubble Space Telescope and gravitational lenses (using galaxy clusters as a natural magnifying glass to search for fainter more distant sources) have thus provided the strongest constraints on the faint-end of the UV LF at high redshifts (e.g. Atek et al., 2015; Bouwens et al., 2017; Atek et al., 2018), while wide-field, ground-based studies have been used to establish the number density of the rare and luminous sources (e.g. Bowler et al., 2015; Ono et al., 2018; Bowler et al., 2020).

Results in the past decade have found a slight excess in the numbers of ultra-luminous galaxies than would be described by an exponential cut-off to high luminosities. This has led to the proposition that the UV LF of galaxies follows a double power law (DPL) rather than a Schechter function. This functional form replaces the exponential cut-off with a bright-end power law slope β and follows:

$$\Phi(L) = \frac{\Phi^*}{L^*} \frac{1}{(L/L^*)^{\alpha} + (L/L^*)^{\beta}}. \tag{1.20}$$

The use of the DPL functional form is not entirely new to luminosity functions and has previously been used to model the AGN luminosity function (discussed further in the next subsection). Most studies treat AGN and galaxies as two separate populations and subsequently fit individual functional forms to each. The

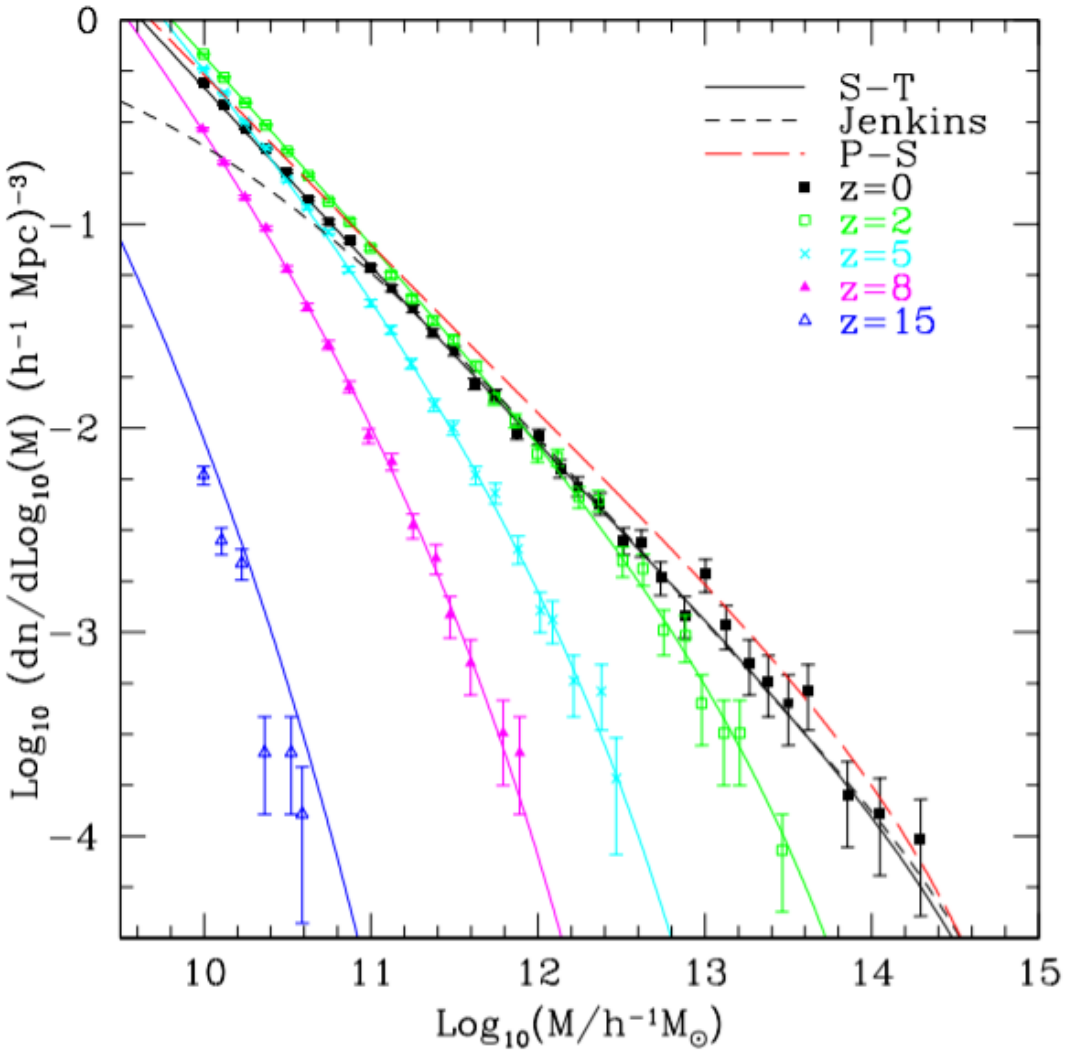


Figure 1.8: The time evolution of the Halo Mass Function (HMF) as found in the simulations conducted by Reed et al. (2003). Of note is the stabilisation of the low mass end at lower redshift and the continued strong evolution of the massive end ($z < 5$).

total ultraviolet luminosity function is thus the sum of the galaxy and AGN luminosity functions.

The underlying physics governing the evolution of these galaxies can be revealed by examining the shape and time evolution of the UV LF. The work presented in this thesis covers the bright end of the UV LF ($L > L^*$), but in this introduction, I provide a brief summary of the current understanding of the UV LF.

For the galaxy population, the faint-end slope has been found to flatten as time advances forward, with the value of α found to be approximately -2.0 at

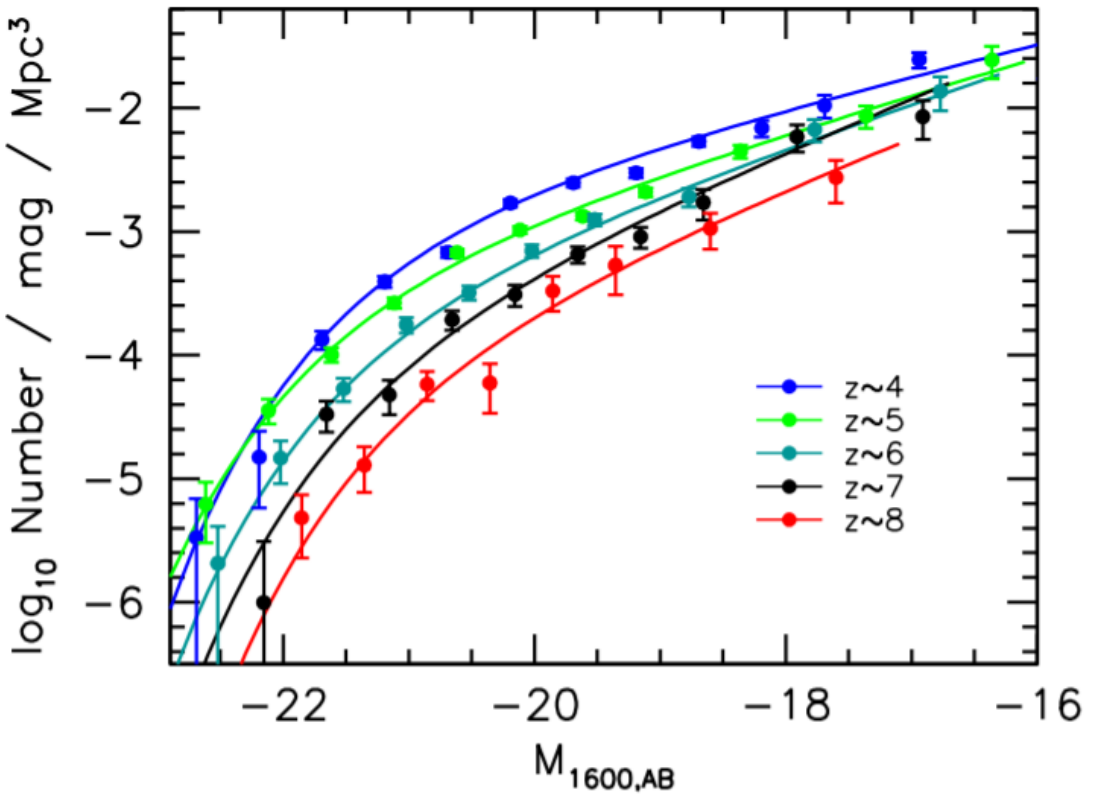


Figure 1.9: The time evolution of the UV LF of galaxies as measured in Bouwens et al. (2015). The plot shows how the galaxy population has built up from $z \sim 8$ through to $z \sim 4$ (when the Universe was about 10 per cent of its current age).

$z \sim 7$ (e.g. Bouwens et al., 2015; Finkelstein et al., 2015; Ono et al., 2018; Bowler et al., 2015) and -1.25 in the local Universe (e.g. Wyder et al., 2005; Hammer et al., 2012). The cause for such a change is attributed to evolution in the low-mass component of the halo mass function (HMF; mass function of dark matter halos) which is strongly evolving from $z \sim 7$ through to $z \sim 2$ (e.g. Reed et al., 2003; Trenti et al., 2010, see Fig.1.8) and shocks induced by supernova explosions which heat the reservoirs of gas used to form new stars.

Under the assumption that the evolution in the galaxy UV LF follows the HMF, there would be an expected growth in the number of the most luminous systems which will shift the value of L^* brightwards leading up to $z \sim 2$ (peak global star formation). However the degree of such evolution is found to be much smaller than would be expected under ΛCDM (e.g. Croton et al., 2006).

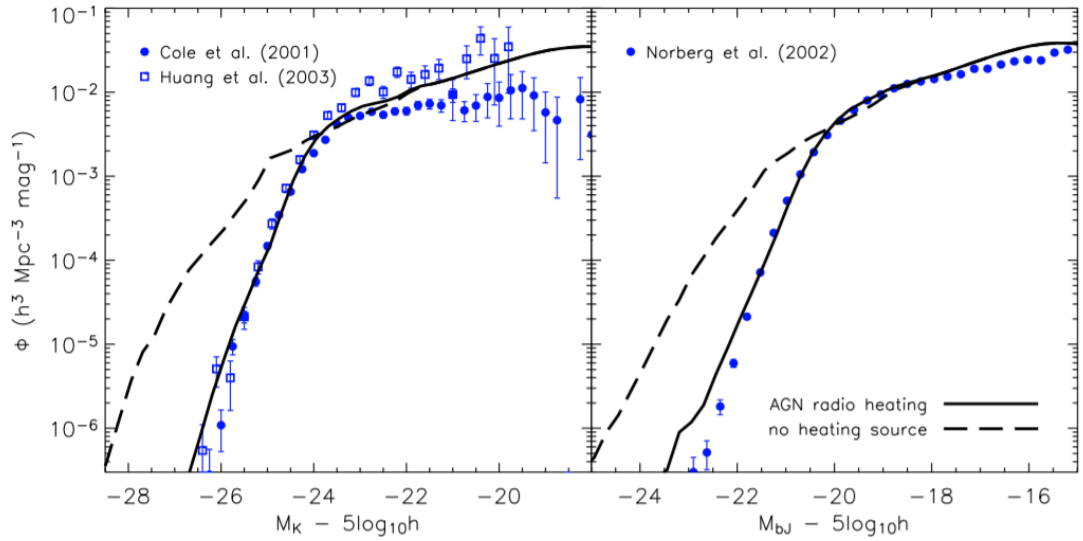


Figure 1.10: An example of the impact that a model for AGN feedback has when applied to a simulation. Shown are the K (left) and B (right) band luminosity functions derived from a simulation and compared to observations in the local Universe. Without AGN (dashed line) the numbers of ultra-luminous sources are drastically overestimated. Credit: Croton et al. (2006)

These findings have led to the development of theories of AGN Feedback (Section 1.2.3), where high-intensity radiation, winds and jets arising from highly accreting supermassive black holes causes a reduction in star formation by heating or expelling gas within/around galaxies.

AGN are thus postulated to be a significant contributor towards the exponential decrease in the number density of bright galaxies. Presented in Figure.1.10 is an example of a simulation that shows the impact of heating from AGN activity on the measured luminosity function of galaxies. Consequently, any time dependence in the characteristic luminosity (M^*) and the steepness of the bright-end slope (β) could provide insight into the time evolution of any physical processes that limit the number densities of highly luminous galaxies. Therefore, in order to successfully understand the evolution of these luminous systems, we require a two-way discussion between observers and simulators, examining the impact that postulated physics has on predicted luminosity functions and comparing to observational results. One of the primary goals explored in Chapter 3 and 4 is to map the bright-end slope of the galaxy population at $z \sim 4$ and $z \sim 5$ in order to determine if they are better

described by a Schechter or DPL functional form and explore the consequences for this on theories of AGN feedback.

1.3.2 Ultraviolet Luminosity Function of AGN

Constraining the number density of galaxies exhibiting ongoing AGN activity through the AGN luminosity function is essential to understanding the impact of AGN on their host galaxies. This is because it directly measures how common they are as well as the distribution of their power outputs. Within the redshift range $4 < z < 6$, shortly after the Era of Reionisation, the AGN population has been observed to experience a drastic evolution, with an increasing number density of over a factor of 10 in less than 500Myr and their number density is found to steadily increase to $z \sim 2$ before turning over and decreasing towards $z \sim 0$ (Richards et al., 2006; Croom et al., 2009). The precise number count of AGN at different stages of the Universe is required alongside the models for feedback in order to successfully determine the total impact of AGN on the galaxy population and how it varies with time.

A key source of uncertainty in AGN occurrence rates are the number of AGN which are faint in the ultraviolet. These can be AGN which are low in power or have significant levels of obscuration due to high dust content in galaxies. Constraining the numbers of these AGN is challenging because AGN and galaxies have comparable number densities in the luminosity range $-24 < M_{\rm UV} < -23$. This leads to high risks of contamination between the two types of sources, making it difficult to measure the galaxy luminosity function brighter than $M_{\rm UV} \simeq -23$, where AGN greatly outnumber galaxies, and the number densities of AGN fainter than $M_{\rm UV} \simeq -23$, where galaxies outnumber AGN. In Chapters 3 and 4, I present an attempt to simultaneously model the AGN and galaxy luminosity functions in order overcome these issues and measure the number densities of both ultra-luminous galaxies as well as faint AGN sources.

1.3.3 Galaxy Stellar Mass Function

Stellar mass is the mass of a galaxy in the form of stars and does not include mass from gas, dust, dark matter etc. A stellar mass function can thus map the changes in the stellar populations of galaxies with time. An integration under the stellar mass function provides the total stellar mass density of the Universe at different redshifts, the time evolution of this provides a route to calculating the global star formation rate. Akin to examining which galaxies are emitting the most ionising photons, the integral of the stellar mass function can show which kinds of galaxies hold the majority of the stars in the Universe.

As with the luminosity function, galaxies which have greater stellar mass tend to be bright but are much rarer beyond a characteristic mass. The stellar mass function thus presents the same observational challenges as the luminosity function, where different survey strategies are more effective at probing different mass and redshift ranges. The cause of this steep drop in galaxy number density at high mass is again attributed to AGN feedback and its effects on inhibiting star formation. Stellar mass can be accumulated in galaxies through multiple mechanisms, such as through the star formation and merger events. The time evolution in the stellar mass function thus captures the summed effects of all of these processes that stimulate or hinder star formation. As a consequence of this, they are a key benchmark for simulations to replicate. It is common to see observations of the $z \sim 0$ stellar mass function used in calibration of simulations and significant discussion of the stellar mass function at other times. It is through this co-analysis of luminosity/mass functions, halo mass functions and simulations that we obtain the present day theories of AGN feedback and supernova feedback that respectively influence the high and low mass population of galaxies.

As observations of the stellar mass function have advanced, it has become apparent that the classical Schechter function is inadequate to explain the shape of the stellar mass function. The source of this is attributed to the natural bi-modality in the types of galaxies that exist in the Universe, namely star-forming and passive galaxies. The galaxy population can be split into star-forming and

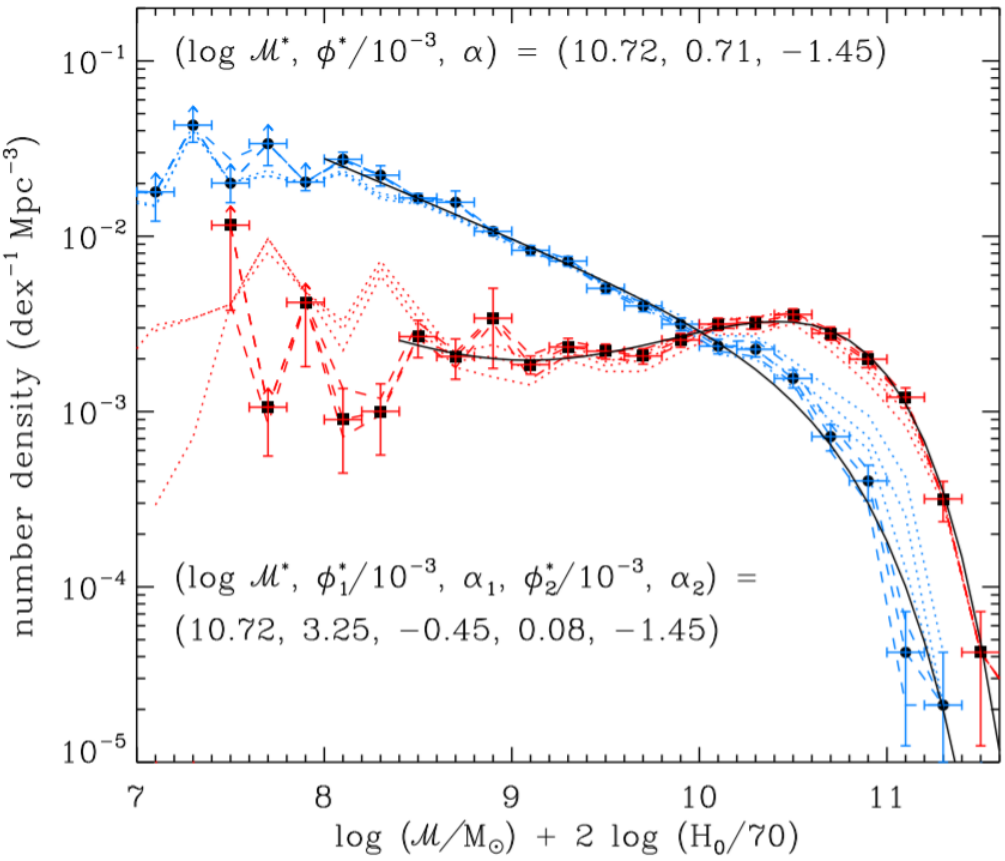


Figure 1.11: The stellar mass function of star forming (blue) and passive (red) galaxies at $z \sim 0$ as measured in Baldry et al. (2012). This shows how the passive population of galaxies adds an additional bump to the high-mass component of the total stellar mass function (sum of the two populations) such as the results shown in Fig.1.12.

passive categories either through the use of colour-colour cuts or the modelling of their spectral energy distributions (SEDs). Both methodologies find that an additional bump in the shape of the stellar mass function is driven by the population of massive and passive galaxies (see Fig.1.11). To improve the modelling of the stellar mass function, a double Schechter function is used:

$$\Phi(M)\delta M = \left[\Phi_1 \left(\frac{M}{M^*}\right)^{\alpha_1} + \Phi_2 \left(\frac{M}{M^*}\right)^{\alpha_2}\right] \exp\left(-\frac{M}{M^*}\right) \frac{\delta M}{M^*}, \tag{1.21}$$

where M^* is the characteristic mass, Φ_1/α_1 are the normalisation and low-mass slope of one Schechter function and Φ_2/α_2 are the normalisation and low-mass slope of a second Schechter function. Commonly, the same M^* value is used for both components. An example of the application of this functional form to measurements

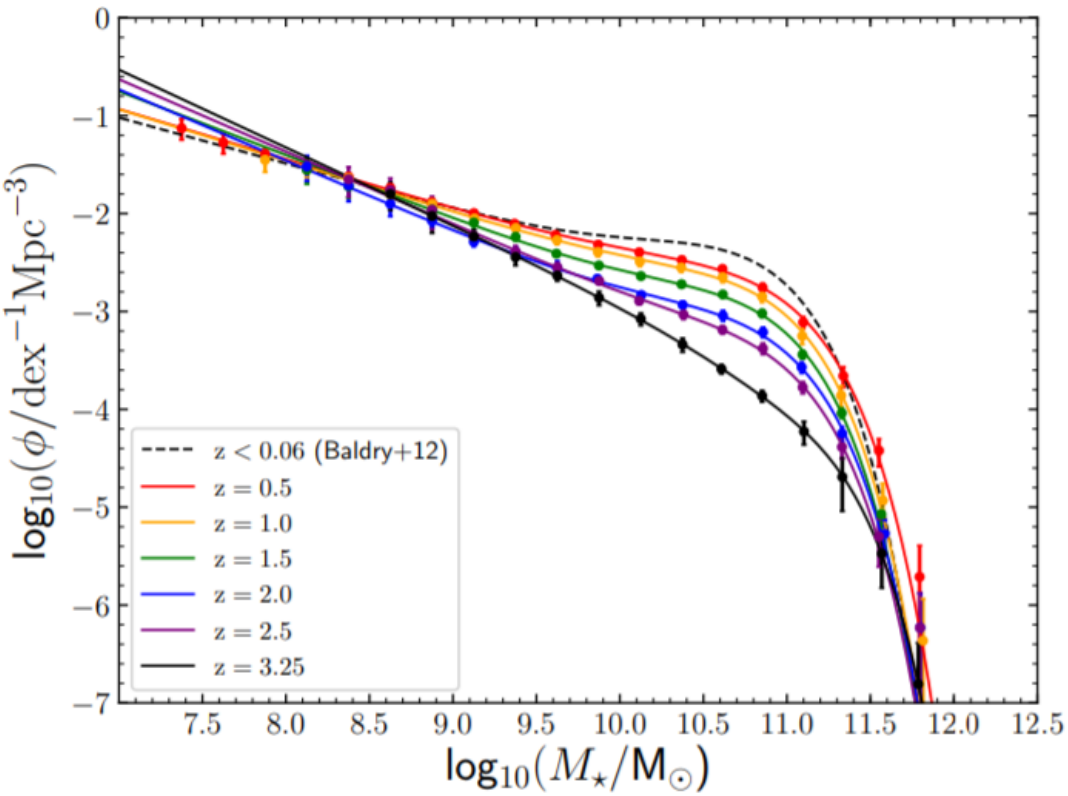


Figure 1.12: An example of recent observations of the galaxy stellar mass function across a wide redshift range. Here the massive population is shown to undergo strong evolution across the mid to late aged universe $z < 3.25$ and a very steep decline is observed for galaxies above a mass of around $\log_{10}(M^*/M_\odot) > 10.5$. Source: McLeod et al. (2020)

of the stellar mass function is shown in Figure.1.12.

A significant problem in previous attempts to measure the galaxy stellar mass function is the broad disagreement between studies on the value of M^* and how it changes with time. At redshifts $0.1 < z < 2.0$, the values of M^* have varied between $10.5 < \log_{10}(M^*/M_\odot) < 11.0$ across a number of studies. A key issue plaguing measurements of M^* in mass functions is a bias known as Eddington Bias. This effect is the combination of the very steep fall in the number density of ultra-massive galaxies and observational errors on the stellar mass. These errors can cause some of the much more numerous (many orders of magnitude) but lower mass galaxies to scatter up and inflate the numbers of very rare but high mass galaxies. Between different studies of the stellar mass function, there has been no consensus on how to handle this issue (e.g. Grazian et al., 2015; Davidzon et al.,

2017a). Some have attempted to model Eddington bias by folding in the errors in stellar mass in their fitting process (e.g. Leja et al., 2020; Thorne et al., 2020), while other studies have attempted to measure the distribution of mass uncertainties and convolved it with the Mass Function (Ilbert et al., 2013; Davidzon et al., 2017a; McLeod et al., 2020). The variations in methodology consequently lead to variations in the conclusions drawn by each of their studies.

The number density of high-mass ($M \geq M^*$) galaxies is very sensitive to AGN feedback and the rate of which these galaxies undergo merger events. It is thus highly important to obtain reliable observational constraints in this regime in order provide the basis with which to advance hydrodynamical simulations and robustly test theories on feedback processes. This forms the primary goal of the work conducted in Chapter 2, where I measure the stellar mass function in nine time steps from $0.1 < z < 2.0$ and vary the methodology used in order to explore the potential bias' that can arise.

1.3.4 Photometric Redshifts

Measuring the time evolution of the luminosity function or stellar mass function of galaxies requires a large sample of galaxies with measured redshifts. While the use of spectroscopy allows highly precise redshift measurements for extragalactic sources to be made, through atomic emission lines, the process is time consuming. This is because the high spectral resolution of spectrographs spread the light from a galaxy over a large number of pixels, meaning longer exposure lengths are required to obtain good signal to noise. Spectrographs are also limited in the number of simultaneous observations that can be taken, with the most modern multi-object spetrographs able to observe around 1000 galaxies simultaneously. For the purpose of measuring luminosity and mass functions at different epochs, a small sacrifice in precision can be afforded in order to increase sample sizes and improve efficiency. Photometric surveys, while low in spectral resolution, can provide enough information in order to constrain the redshift of millions of galaxies.

Redshift estimations using photometry can be conducted using two primary methods: colour-colour cuts and SED modelling. With colour-colour cuts, we can use as few as three photometric bands in order to select samples of galaxies with certain redshifts and physical properties for study. However, the method relies on a compromise of sample completeness and sample purity. This is because galaxies vary in the age of their stellar populations, dust content and redshift, which can cause an overlap in the colours of different populations. As a result, care is required to ensure the methodology takes into account the effect of sample completeness and sources of contamination. This often requires simulations to be conducted which insert fake sources into images in order to test the selection procedure.

In the case that more than just a few photometric bands are available, the whole optical/NIR SED of each galaxy can be modelled in order to estimate redshifts and other basic properties such as stellar mass and star formation rate. This procedure reduces the risks of contamination, due to degeneracy in colours, that are experienced with colour-colour cuts. Template SEDs of galaxies combine together our knowledge and assumptions regarding the initial mass function, star formation history, stellar metallicity, SEDs of individual stars, dust extinction and IGM absorption/Lyman Breaks. The star formation history of a galaxy is the time-dependent star formation rate, and alongside the age of the galaxy observed and the initial mass function, it provides the population of stars present at the time the observed SED was emitted.

The quality of data presented by photometric surveys (of order a dozen bands) is currently incapable of distinguishing between complex star formation histories, so most studies use simplified forms, such as exponential decays or lognormal distributions. However, the real star formation histories of galaxies are likely to be more complex and stochastic due to events like mergers. These caveats with the star formation histories are important when making measurements of the physical properties of galaxies (e.g. stellar mass or current star formation rate), although the redshifts are constrained by strong features such as the Lyman Break, which are negligibly affected by these simplifications in the template sets.

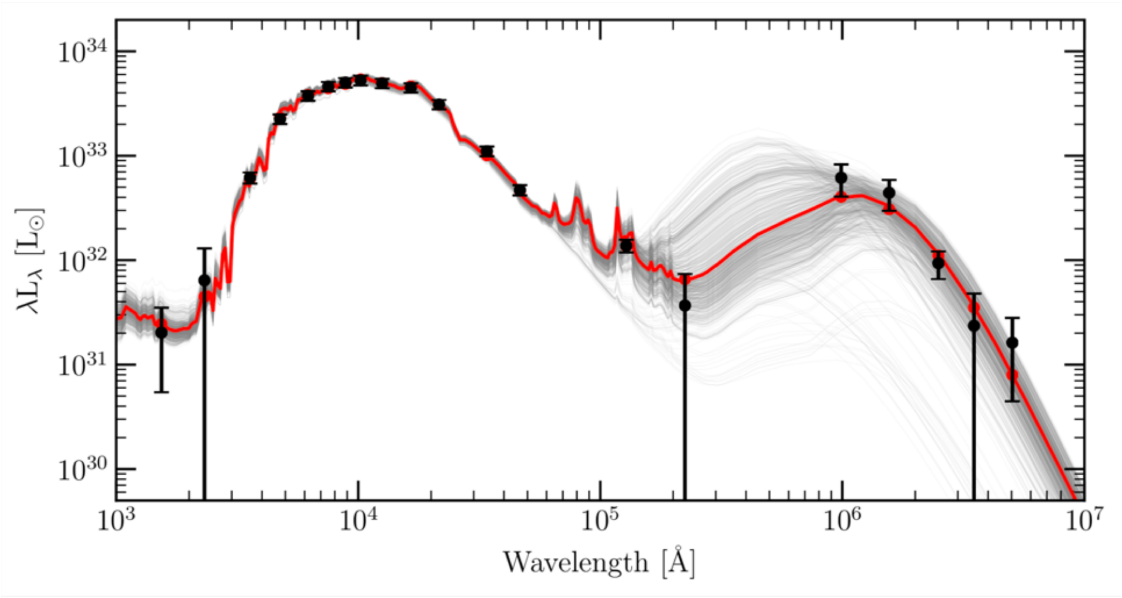


Figure 1.13: Example of an SED modelling procedure conducted with the ProSpect algorithm on a low-redshift galaxy. Photometric data points are shown in solid black while the best-fitting galaxy template is shown as the red line. Examples of other galaxy templates are shown in faint black. This example clearly shows the 'blue bump' induced by the stellar population at low wavelengths and a 'dust bump' at longer wavelengths which is generated by warm dust reprocessing stellar light. Image credit: Sabine Bellstedt, Aaron Robotham & Simon Driver.

In addition to measuring the redshifts and physical properties of galaxies, the template fitting process can be used to conduct object classification. This is conducted by comparing the quality of fits (e.g. using minimisation of χ^2) for SED templates of different types of objects. These can include singular stars or galaxies with significant AGN contributions. These classifications can be used to further refine sample selection and remove potential contaminants such as foreground Milky Way stars. This ultimately ensures that measurements of stellar mass functions and luminosity functions contain pure and complete samples of galaxies, minimising the risk of subsequent systematic errors and misinterpretation of the results. In this thesis, redshift and stellar mass estimations are conducted on $\sim$ 2 million sources in order to tackle the outlined science objectives with the stellar mass and ultraviolet luminosity functions. The full procedure is outlined along with the data used in Section 2.2.

1.4 Thesis Outline

With the recent advances the depths and wavelength coverage achieved in square degree scale extragalactic fields, there is scope to build upon studies that have previously measured the Luminosity/Mass function of galaxies and solve outstanding issues in extragalactic astrophysics that are derived from those results. This thesis presents the work that I have conducted on the now-complete VIDEO and UltraVISTA survey programmes. I have used these surveys, alongside complementary optical and mid-infrared data, to measure the stellar mass function between $0.1 < z < 2.0$ and the ultraviolet luminosity function between $3.5 < z < 5.2$ in order to further understand the growth in number densities of the most massive and luminous sources.

This thesis is structured around the three journal publications that have been developed over the course of my work. In Chapter 2, I begin by using my catalogues to measure the galaxy stellar mass function in the redshift range of $0.1 < z < 2.0$ in nine redshift bins. The mass functions are measured in multiple ways by varying the choice of source extraction software and the photometric bands used in the SED fitting process. The subsequent fits to the stellar mass function show that the characteristic mass has grown between $0.5 < z < 1.5$ and that the dominant source of this growth is in the large rise in the number density of massive and passive galaxies. This is found to closely match results from the hydrodynamical simulation EAGLE. In Chapter 3, I modify the catalogues in order to effectively select luminous sources at $z \sim 4$ to measure the combined AGN and galaxy ultraviolet luminosity function between $3.5 < z < 4.5$. The use of SED modelling across optical and near-infrared wavelengths enables a highly efficient selection of sources in this wavelength range, with a completeness of ~ 90 per cent and the unique depth-area combination of surveys used enable galaxy number densities to be constrained across 5 orders of magnitude. This enables me to map the transition from AGN-dominated ultraviolet sources to galaxy-dominated ultraviolet sources and shows that Schechter and DPL based fits to the galaxy UV LF predict drastically different transitions. In Chapter 4, this work is expanded closer to the Era of Reionisation ($4.5 < z < 5.2$)

and comparisons drawn to the results in Chapter 3. It is found that the AGN population evolves significantly within this time frame, increasing in number density by around 0.8 dex from $z \sim 4.8$ to $z \sim 3.9$. Combining these findings with other AGN-focused work shows a rapid onset in AGN activity in the first 500MYr after the end of the reionisation epoch at $z \sim 6$.

-He who can no longer pause to wonder and stand rapt in awe, is as good as dead; his eyes are closed

-If a cluttered desk is a sign of a cluttered mind, of what, then, is an empty desk?

— Albert Einstein

2

Deep Field Photometry and the Stellar Mass Function over 10 Billion Years

Contents

2.1 Introduction

2.1.1 The evolution of the galaxy stellar mass function

The galaxy stellar mass function (GSMF) is a measurement of the cumulative effects of physical processes that enhance or hinder star formation within galaxies. These processes include merger events, internal feedback mechanisms (both supernova and active galactic nuclei (AGN) driven) and environmental effects. Understanding the balance between these processes is key to understanding how galaxies have been assembled over cosmic time. Measurements of the local GSMF reveal a steep cut-off in the number of high-mass galaxies above a characteristic mass $\log_{10}(M^*/M_\odot) \sim 10.7$ (e.g. Baldry et al., 2012). In addition to this, the population of very high-mass galaxies becomes increasingly quenched with time (e.g. Davidzon et al., 2017a; McLeod et al., 2020). Many theories have been proposed to explain why there is significant suppression in the star formation (often called quenching) of galaxies above this stellar mass, examples include but are not limited to: starvation/strangulation (Larson et al., 1980; Kawata & Mulchaey, 2008; McCarthy et al., 2008; Feldmann et al., 2011; Bahé et al., 2013; Feldmann & Mayer, 2015), virial shock heating (Birnboim & Dekel, 2003; Dekel & Birnboim, 2006; Cattaneo et al., 2006) and AGN feedback (Binney & Tabor, 1995; Di Matteo et al., 2005; Springel et al., 2005; Bower et al., 2006; Croton et al., 2006; Cattaneo et al., 2009; Fabian, 2012; Bongiorno et al., 2016; Beckmann et al., 2017). To adequately test these theories and increase understanding of how these processes influence galaxy evolution, simulations are required. Key milestones in testing theories such as these include being able to accurately reproduce the observed galaxy population through Luminosity/Mass Functions (see Somerville & Davé, 2015; Vogelsberger et al., 2020, for recent reviews). The evolving shape of the GSMF is dependent on all forms of stellar mass growth, including growth via merging events in addition to

internal star formation (e.g. Rodriguez-Gomez et al., 2015; Qu et al., 2017; O'Leary et al., 2020). Consequently, to increase understanding of both quenching and merger rates, stellar mass functions have become a key benchmark for simulations in the past few years (Henriques et al., 2013; Schaye et al., 2015; Pillepich et al., 2018; Lagos et al., 2018) and hence, require observations to be reliable in order to tune and test these simulations.

The advance in deep extra-galactic surveys over the past decade has led to a better measurement of the GSMF at high-redshift. Using predominantly photometric redshifts, several studies have now mapped out the evolution of the GSMF from $z = 0$–3 (e.g. Fontana et al., 2004; Pérez-González et al., 2008; Marchesini et al., 2009; Pozzetti et al., 2010; Ilbert et al., 2013; Muzzin et al., 2013; Davidzon et al., 2017a; Wright et al., 2018) and even up to $z \simeq 7$ (e.g. Verma et al., 2007; McLure et al., 2009; Stark et al., 2009; Grazian et al., 2015; Song et al., 2016a; Thorne et al., 2020). Despite the increased cosmological volume probed by these studies, there is no consensus on the exact form of the GSMF over this epoch. For example Wright et al. (2018), McLeod et al. (2020) and Thorne et al. (2020) find an approximately constant characteristic mass with redshift, but offset from each other by up to $M^* \sim 0.5\,\mathrm{dex}$. Alongside this, other studies of high mass systems suggest merger events are required to explain the observed change in the size-mass relation of these objects (e.g. McLure et al., 2013). The rate of mergers required would consequently influence the evolution of the shape of the stellar mass function at the high mass end. To solve these issues, the precise number densities of galaxies with masses greater than M^* are required. Uncertainties in both the stellar mass and number density of these objects can have a much larger impact on the measured shape and characteristic mass of the derived GSMF than the more numerous low-mass sources. This is due to the need to quantify a steep exponential fall off, with relatively few galaxies, that suffer from higher levels of cosmic variance (e.g. Moster et al., 2011). In this work I exploit deep, wide-area optical and IR imaging to measure the GSMF over $> 5\,\mathrm{deg}^2$, leveraging the large area to understand, in particular, the evolution in the number density of the most massive galaxies between $0.1 < z < 2.0$.

2.1.2 Measuring accurate photometric redshifts and stellar masses

In the coming decade, the Vera Rubin Observatory (Ivezić et al., 2019) and *Euclid* (Racca et al., 2016) survey programmes will be conducting broadband observations to further improve both the area and depths achieved within popular multi-wavelength fields such as COSMOS and XMM-LSS, among others. In parallel to these photometric surveys, a number of spectroscopic campaigns are also set to begin operations in the next couple of years. MOONRISE (Cirasuolo et al., 2012) and WAVES (Driver et al., 2019) are examples of multi-object spectrograph (MOS) surveys and both have an immediate need for high quality photometric redshift and physical parameter estimates in order to plan survey operations and develop final target catalogues ahead of commissioning. Consequently, there is demand for a maintained compilation of broadband photometry and photometric redshifts based on the improved optical and NIR data that has been obtained in recent years in order to best prepare for these upcoming projects (past examples including: Laigle et al., 2016; Alarcon et al., 2020).

Moreover, with next-generation telescopes and survey programmes comes next-generation software and analysis pipelines. A workhorse in photometric source extraction for over 20 years has been `SExtractor` (Bertin & Arnouts, 1996), but in recent years there has been a renewed effort in tackling the problem of obtaining accurate flux measurements from images with robust uncertainties and improved handling of source blending. One product of this effort has been `ProFound` (Robotham et al., 2018), which potentially provides improved photometry, particularly for extended sources, over a variety of wavelengths due to the non-parametric apertures used (e.g. Davies et al., 2018; Hale et al., 2019; Bellstedt et al., 2020). In this study I produce multiple catalogues of broadband photometry and using these different source extraction methods. The use of two source extraction tools, as well as varying the use of *Spitzer*/IRAC data, allows for any potential bias' in the GSMF due to these different effects to be explored.

This paper is presented as follows: In Section 2.2 I describe the data used in producing my photometric catalogues. In Section 2.3 I describe the methods used in source extraction, fitting for photometric redshifts and derivation of basic physical properties of galaxies. In Section 2.4 I present the measured GSMF in the redshift range $0.1 < z < 2.0$ and in Section 2.5 I present the results of modelling the GSMF, and compare the models with the results from previous studies. In section 2.6 I explore the time evolution of the GSMF and how my observations compare to results from simulations. I finally present my conclusions in Section 2.7. Throughout this paper I use the AB magnitude system (Oke, 1974; Oke & Gunn, 1983) and assume ΛCDM cosmology with $H_0 = 70\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, $\Omega_\mathrm{M} = 0.3$ and $\Omega_\Lambda = 0.7$.

2.2 Data

The following sub-sections describe the data in the two fields used in the construction of my photometric catalogues and subsequent measurement of the GSMF.

2.2.1 COSMOS

The COSMOS field (Scoville et al., 2007) is one of the most widely studied multi-wavelength fields in extra-galactic astrophysics, with data spanning the X-ray through to the radio domains over $\sim 2\,\mathrm{deg}^2$ of the sky centred on the J2000 coordinates of RA = 150.12 deg (10:00:28.6) DEC = +2.21 deg (+02:12:21.0).

The imaging data over this field that are used in the construction of my catalogues is derived from four telescopes. The bluest coverage is from the Canada-France-Hawaii Telescope Legacy Survey (CFHTLS; Cuillandre et al., 2012) which has an ultra-deep pointing in the central square degree of the field, and I restrict my analysis to this area for this reason. From this survey I make use of the u^*-band. For the optical coverage I take data from the ultra-deep component of the HyperSuprimeCam Strategic Survey Programme DR1 (HSC; Aihara et al., 2018b,a). Near-infrared imaging is sourced from the UltraVISTA survey (McCracken et al., 2012). I make use of the fourth data release of UltraVISTA (DR4) which has a tiered observing strategy, leading to a striped pattern of near-infrared coverage

across the field. The 'deep' component is approximately 1 magnitude shallower than the 'ultra-deep' region. 3.6 and 4.5 micron photometry is obtained from the *Spitzer* Extended Deep Survey (SEDS; Ashby et al., 2013), *Spitzer* Matching Survey of the UltraVISTA ultra-deep Stripes survey (SMUVS; Ashby et al., 2018) and *Spitzer* Large Area Survey with Hyper-Suprime-Cam (SPLASH; Capak et al., 2012). The 5σ detection limits for the 2 (2.8) arcsecond photometry in each optical/NIR (*Spitzer*) bands are outlined in Table 2.1. This table also highlights the impact of the tiered structure of the UltraVISTA survey.

2.2.2 XMM-LSS

With a much greater area (at $\simeq 4.5$ deg^2), the XMM-Large Scale Structure field is one of 3 deep fields that make up the Vista Infrared Deep Extra-galactic Observations (VIDEO) survey (Jarvis et al., 2013). Located at RA = 35.5 deg (02:22:00.0) DEC = -4.8 deg (-04:48:00.0), my study focuses on regions of the field with high quality HSC observations. I mask parts of the edge of the field due to the lack of overlapping coverage between HSC and VISTA (see Bowler et al., 2020, for more information on overlapping coverage from the different surveys and variable depths). The total area of the field that is used after considering the overlap of telescope footprints is 4.23 deg^2 (giving a total of 5.23 deg^2 when combined with COSMOS).

I use the same photometric bands as those used in the COSMOS field. Unlike in the COSMOS field, the optical coverage in XMM-LSS is not uniform while the near-infrared is more uniform. u-band imaging over the full XMM-LSS field was obtained from the CFHTLS Wide survey in addition to the CFHTLS-D1 field which covers a 1 deg^2 patch where u^* observations are 1 magnitude deeper than the rest of the field (Cuillandre et al., 2012). HSC SSP covers a different 1.5 deg^2 region of the field (centred on UKIDSS UDS Lawrence et al., 2007) which has 'ultra-deep' coverage. Near-infrared photometry is derived from the final data release of VIDEO (Jarvis et al., 2013) and *Spitzer* data is sourced from the *Spitzer* Extragalactic Representative Volume Survey (SERVS; Mauduit et al., 2012). The depths of the images in each broadband filter are also outlined in Table 2.1.

Table 2.1: Summary of the 5σ detection depths within the COSMOS and XMM-LSS fields. Depths are calculated in 2" diameter circular apertures (2.8" for IRAC bands due to *Spitzer*'s larger point spread function) that were placed on empty regions of the image. Values are grouped into Deep (D) and Ultra-Deep (UD) sub-regions. The XMM-LSS field has a single ultra-deep pointing from HSC centred on the UDS field and so a third of XMM-LSS has this deeper optical coverage. Similarly the u^* coverage has a single square degree of ultra-deep coverage in a separate part of XMM-LSS as part of the CFHT Legacy Survey observing programme.

Filter	COS-UD	COS-D	XMM-UD	XMM-D	Origin
u^*	27.2	–	25.9	$25.9 - 26.9$	CFHT
g	27.1	–	27.0	26.5	HSC
r	26.7	–	26.5	25.9	HSC
i	26.6	–	26.4	25.5	HSC
z	25.9	–	25.7	24.7	HSC
y	25.5	–	25.0	24.1	HSC
Y	26.3	25.2	25.4	–	VISTA
J	26.1	25.0	24.9	–	VISTA
H	25.7	24.6	24.4	–	VISTA
K_s	25.4	25.0	24.0	–	VISTA
3.6	24.8	–	24.3	–	Spitzer
4.5	24.8	–	24.0	–	Spitzer

2.3 Photometric catalogues and derived products

For my study I define and produce four separate stellar mass estimates for all galaxies identified in the near-infrared. The various stellar mass estimates are determined by using two source extraction algorithms and including/excluding the *Spitzer*/IRAC 1 and 2 bands. The goal here is to examine the impact of these two variables in the methodology, such as how different measurements of total flux translates to differences in stellar-mass estimation and how the use of *Spitzer*/IRAC bands affects photometric redshift performance and the final distributions in stellar mass.

I refer to these catalogues as 'SExtractor/SE' and 'ProFound/PF' when using them in a lengthened/shortened format. When the two *Spitzer*/IRAC bands are included in any analysis, the suffix '+IRAC' is added to the catalogue name.

2.3.1 SExtractor photometry

Source finding is performed in SExtractor (Bertin & Arnouts, 1996) with the K_s-band image used for object selection. The fiducial photometry was derived using 2

arcsecond diameter circular apertures placed at the location of sources found by SExtractor. As *Spitzer* has a larger point spread function, for the IRAC bands I use 2.8 arcsecond diameter circular apertures (Bowler et al., 2020). The optical and NIR photometry is aperture corrected through a point-spread-function (PSF) generated in each band by PSFEx (Bertin, 2011) using cut-outs of point-sources. This is calculated separately for the COSMOS field and over each of the three VISTA/VIDEO tiles in XMM-LSS. For *Spitzer* photometry, I use an aperture correction derived in the *Spitzer* handbook[1]. Alongside this I also measure the `MAG_AUTO` parameter from SExtractor to estimate the total flux for each object. The `MAG_AUTO` measurement uses an elliptical aperture which is scaled to encompass the total flux of an object in the detection image following the prescription developed in Kron (1980). Due to the significantly larger PSF of the two *Spitzer* bands, carrying over the same aperture from the K_s band would lead to an underestimation of the total flux. To solve this problem, I derive a correction to translate aperture flux (f) to total flux (F). This correction follows the average value of $F_{\nu,IRAC} = f_{\nu,IRAC} \times \frac{F_{\nu,K_s}}{f_{\nu,K_s}}$ for objects in bins of K_s magnitude and redshift used in the later calculation of the GSMF. The use of this correction is sanity checked against the methodology I use with the ProFound catalogues and is found to agree well for luminous objects. Errors on the SExtractor photometry are calculated based on local depth maps generated by inserting apertures in empty locations of the field (the same method as applied in Bowler et al., 2020).

2.3.2 Profound photometry

In addition to the SExtractor photometry, I produce ProFound (Robotham et al., 2018) catalogues selected on a weighted stack of the VISTA-Y, J, H and K_s bands[2]. ProFound generates two photometric measurements for each object, a 'total' flux and a 'colour' flux. ProFound operates by iteratively dilating the aperture encompassing each galaxy in each image until it meets the local background. This

[1]The Spitzer handbook can be found at: `https://irsa.ipac.caltech.edu/data/SPITZER/docs/spitzermission/missionoverview/spitzertelescopehandbook/`

[2]I note that this does not affect my results as I subsequently apply a conservative limiting magnitude cut based on the K_s-band

results in a morphologically derived aperture (one that follows the isophote of the source rather than assuming any shape e.g. elliptical) around each identified object for each photometric band. For my purposes I only make use of the total flux measurements and the associated errors from the ProFound output. ProFound flux errors are calculated by summing in quadrature: 1) the standard deviation of the distribution that results from subtracting the sky model from local unoccupied pixels; 2) the error on the sky model itself around the target source, calculated through the root mean square (RMS) method and 3) the Poisson noise resulting from the raw counts in the image. Because this dilation process is performed for each photometric band, the larger *Spitzer*/IRAC PSF is taken into account and thus no alterations are required to obtain total fluxes in these bands.

To validate both the ProFound and SExtractor photometry. I compare my measured total photometry against measurements made in other catalogues which have targeted these fields. These comparison catalogues are the COSMOS2015 catalogue for the COSMOS field (Laigle et al., 2016) and the Subaru/XMM-Newton Deep Field (SXDF) catalogue (Mehta et al., 2018). I find that the vast majority of my photometric bands are consistent ($\ll 0.1$ mag median offsets) with the measurements from these reference catalogues (see Fig.2.1). The methodology I use to determine total SExtractor *Spitzer* fluxes for bright, resolved sources is also found to be consistent with these catalogues. The greatest difference in photometry is found when running ProFound on *Spitzer* imagery in XMM-LSS. For bright sources ($m < 20$), the photometry is consistent with my SExtractor catalogue and the reference catalogue. However, for fainter objects the measured flux in ProFound is found to be smaller than with SExtractor (up to 0.2 mags offset at $m \sim 24$). Similarly, colour distributions match those of the reference catalogues, with the exception of the aforementioned ProFound IRAC offsets in XMM-LSS.

ProFound photometry is only used to calculate a second set of stellar masses for each galaxy in my sample. This is to examine what differences, if any, are found to the GSMF measured when alternate software is used in source extraction. Because this IRAC offset has arisen with no change to the methodology between two fields,

I assume no truth to the catalogues and proceed with this offset present in the XMM-LSS IRAC bands. The IRAC bands do have a significant impact on stellar mass calculation, with the study by Elsner et al. (2008) showing that without any constraints, template fits tend to over estimate infrared fluxes. Including *Spitzer* bands was found to reduce stellar mass by up to 0.2dex for galaxies at $z < 2.5$. One can thus expect that fainter sources in the ProFound catalogues will have slightly lower masses than in the SExtractor case when *Spitzer* bands are used.

2.3.3 Photometric Redshifts

With the reliability of the photometry confirmed, I proceeded to conduct object classification and estimate redshifts, stellar mass and star formation rates for all sources within the COSMOS and XMM-LSS fields. The process uses the publicly available code `LePhare` in order to conduct these estimations (Arnouts et al., 1999; Ilbert et al., 2006). `LePhare` functions by minimising the χ^2 of a series of template galaxy SEDs against the measured photometry of each object. The templates used fall into three primary categories: 1) Galaxies; 2) Galaxies with AGN; 3) Single Stars. The best-fitting template within each category is provided as an output alongside a probability density function (PDF) for the redshift. The galaxy templates are the same as those used to construct the popular COSMOS2015 catalogue (Laigle et al., 2016). These consist of 43 templates of synthetic galaxies that cover a wide variety of galaxy morphologies (e.g. E, S0, Sa, Sb, Sc, Sd, Sdm on the Hubble Tuning Fork). Each template covers a wavelength range of 0.1μm to 1000μm and is generated by the stellar population synthesis code GRASIL (Silva et al., 1998). Due to issues simulating dust reprocessing, the work by (Polletta et al., 2007) further develops these templates in the regime where emission transitions from being stellar dominated to dust dominated (5-12μm) by incorporating observations from the *Infrared Space Observatory* and *Spitzer*. Since the work undertaken in this study does use not data extending redder than 5μm, it is not anticipated that errors modelling dust emission in the templates will significantly affect the template fitting process. One other possible source of uncertainty in the infrared regime is

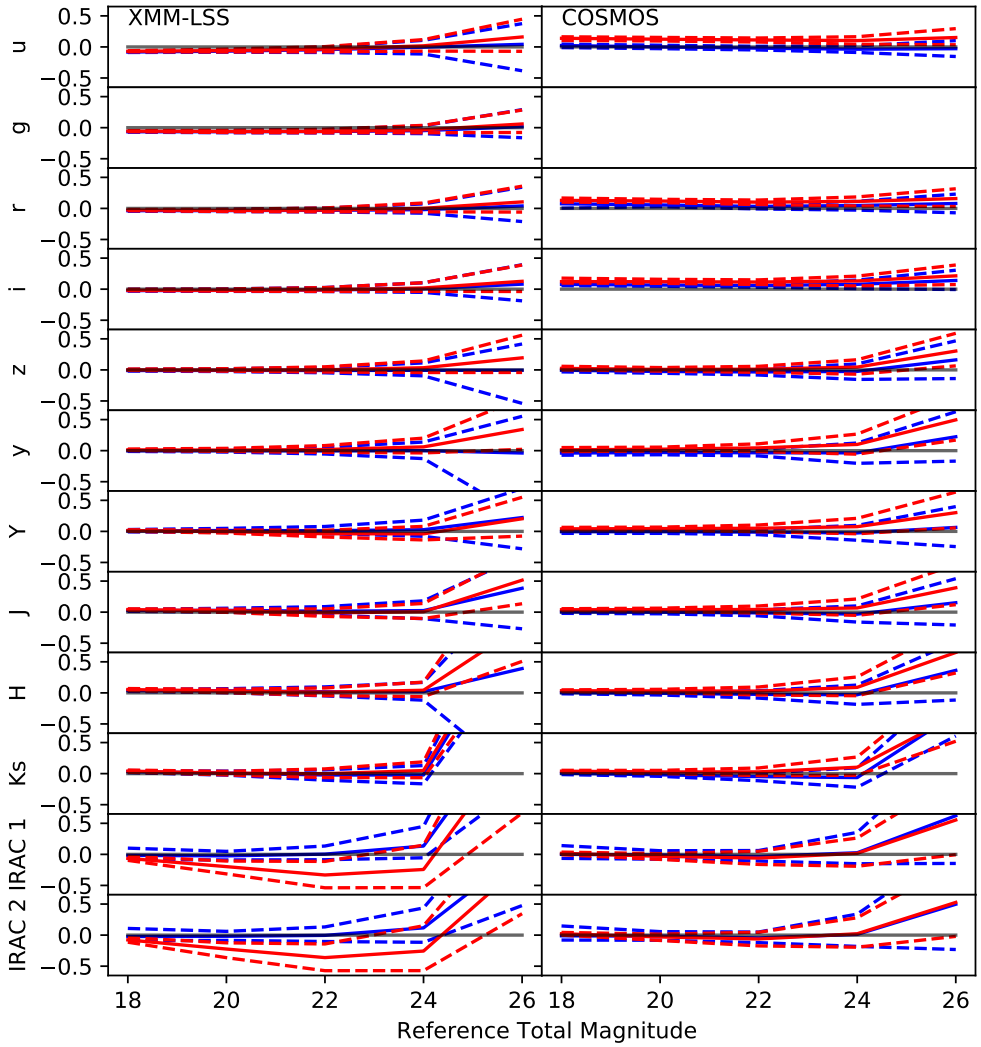


Figure 2.1: Direct comparison of the photometry derived from SExtractor (blue) and ProFound (red) against a reference catalogue from each extragalctic field, Mehta et al. (2018) for XMM-LSS and Laigle et al. (2016) for COSMOS. The y-axis is the difference between the reference catalogue and the catalogue produced as part of this thesis ($m_{\rm ref} - m_{\rm SE/ProF}$). The x-axis is the total magnitude of the source as listed in the reference catalogue. The solid line is the median offset and the dashed lines indicate the inter-quartile range of all sources cross-matched within 1 arcsecond. The upturn displayed in the red bands is a consequence of the K_s-band selection.

the contribution of pulsing asymptotic giant branch stars, which are not as well understood in stellar population synthesis yet have been shown to be influential in calculating stellar masses and ages (Maraston, 2005; Maraston et al., 2009).

Each of the templates used in this methodology are modified for dust attenuation according to the Calzetti et al. (2000) extinction law with E(B-V) = 0, 0.05, 0.1, 0.15, 0.2, 0.3, 0.6, 1.0, 1.5. Each SED is also recalculated with the effects of redshift and the Lyman Break in steps of 0.04 from $z = 0$–6 and steps of 0.1 from $z = 6$–9. The Lyman Break strength follows the treatment derived in Madau (1995) The resultant probability space is then interpolated in order to find the optimal redshift solution. A similar process is applied to galaxies with contributions from an AGN. Here the templates used are obtained from Salvato et al. (2009), which generate templates with varying fractional contributions from an AGN to the total SED of the source. In order to identify foreground stars, stellar templates from Hamuy et al. (1992, 1994); Bohlin et al. (1995); Pickles (1998); Chabrier et al. (2000) were also fit. Photometric errors are set to a minimum of 5 per cent in flux during the template fitting process, this is to alleviate the consequences of using templates that probe the colour space discretely while the real galaxy population is continuous.

Zero-point calibration with spectroscopic samples

The two fields have been targeted by a large number of spectroscopic campaigns which can be used to calibrate my methods and examine photometric redshift accuracy. I make use of the spectroscopic catalogue compiled by the HSC team[3]. Included are spectra from the VIMOS VLT Deep Survey (VVDS; LeFèvre et al., 2013), z-COSMOS (Lilly et al., 2009), Sloan Digital Sky Survey (SDSS-DR12; Alam et al., 2015), 3D-HST (Skelton et al., 2014; Momcheva et al., 2016), Primus (Coil et al., 2011; Cool et al., 2013) and the Fiber-Multi Object Spectrograph (FMOS; Silverman et al., 2015). From these I select only those with high quality flags (>95 per cent confidence) to ensure secure redshifts are being used. Together these provide a spectroscopic sample of 22,409 in COSMOS and 35,125 in XMM-LSS.

[3]The source of the spectroscopic catalogue compiled by the HSC team can be found at: `https://hsc-release.mtk.nao.ac.jp/doc/index.php/dr1_specz/`

Table 2.2: The list of zero-point offsets applied to the 2 arcsecond aperture photometry after optimising the template fitting process against a large spectroscopic sample. Offsets listed are in magnitudes and are calculated separately for each field.

Band	XMM Offset	COS Offset
u^*	0.138	0.136
g	-0.062	-0.066
r	-0.096	-0.078
i	-0.011	-0.026
z	-0.048	-0.037
y	-0.064	-0.054
Y	0.078	-0.006
J	0.037	0.020
H	0.077	0.084
Ks	0.025	0.024
IRAC1	0.006	0.005
IRAC2	-0.048	0.013

I use these spectroscopic redshifts to examine the accuracy of my photometric redshift estimates. In addition to this, I also make use of LePhare functionality to make iterative adjustments to the zero-points of each photometric filter in order to optimise the results against the spectroscopic sample. Minor shifts in the zero-points can occur as a result of inaccurate filter transmission functions, through biases within the choice of SED templates and from the calibration of the images. The inclusion of a very large sample of spectroscopically confirmed objects from a variety of different surveys minimises the risk of introducing an additional bias through calibration on a non-representative sample. For each catalogue, I run LePhare once on the spectroscopic sample to obtain the zero-point corrections, these offsets are applied and the entire field is then run. I show the results of this process in Table.2.2 and the majority are small compared to the errors ($\ll 0.1$ mags). The large offset in the bluest band (u^*) is a result commonly found when conducting this process on similar photometric data sets, with work conducted by Laigle et al. (2016) and Mehta et al. (2018) finding offsets in 'B', 'u^*' and 'NUV' photometric bands of ± 0.2 magnitudes. The likely cause of this are imperfect stellar populations as the extremes of the wavelength ranged covered by the models ($\sim 0.1\mu$m) is approached. The minimal offsets found in the *Spitzer* bands provides reassuring

Table 2.3: Summary of the photometric redshift statistics. Displayed are the outlier rates and Normalised Median Absolute Deviation (NMAD) of the catalogues when compared to a large spectroscopic sample. The cut of $M_i < 22$ corresponds to the brightest 50 per cent of the spectroscopic sample I compare to in COSMOS and the brightest 60 per cent in XMM-LSS.

	Outlier Rate (%) $M_i < 22$	NMAD $M_i < 22$	Outlier Rate (%) $M_i < 26$	NMAD $M_i < 26$
COS	3.3	0.027	4.5	0.029
COS+IRAC	3.3	0.027	4.5	0.029
XMM	3.2	0.028	5.3	0.029
XMM+IRAC	3.4	0.030	6.2	0.031

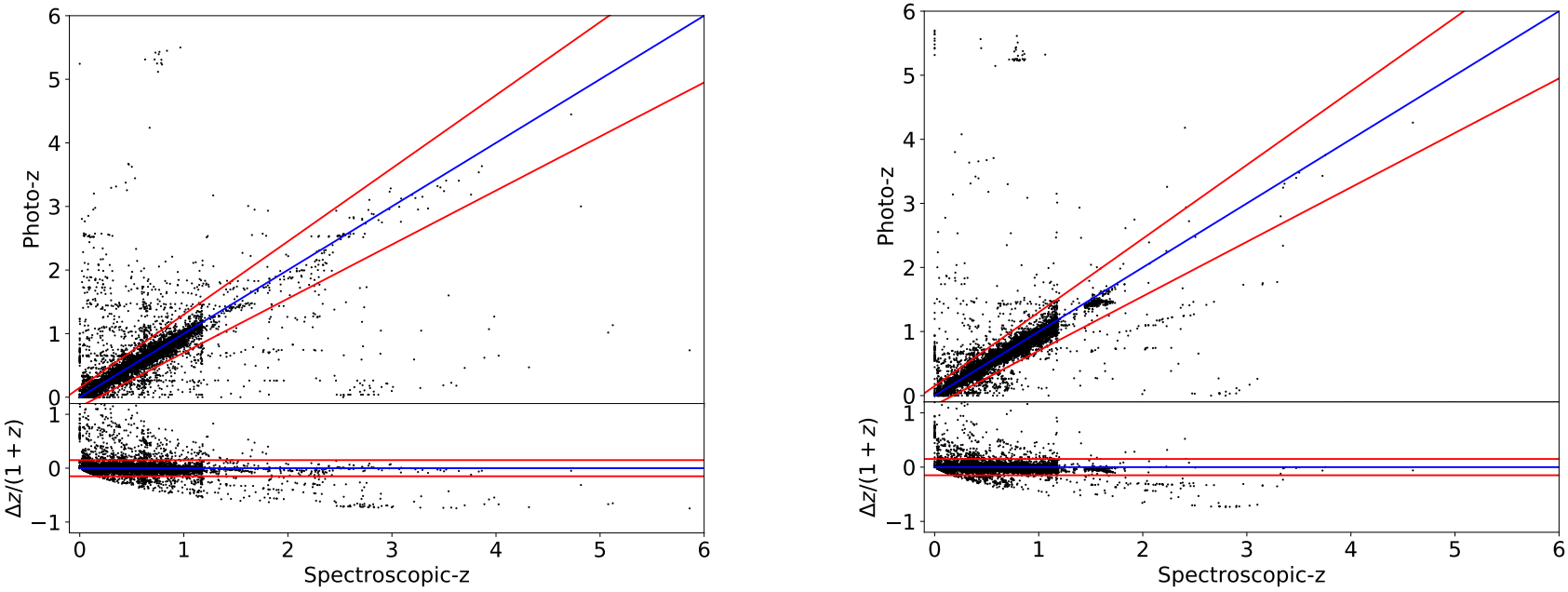
evidence that the template set functions well up to $\sim$5μm.

2.3.4 Photometric redshift accuracy

The quality of the photometric redshift catalogues can be described with two numerical values. The outlier rate, defined as percentage of objects satisfying $\mathrm{abs}(z_{spec} - z_{phot})/(z_{spec} + 1) > 0.15$, and the Normalised Mean Absolute Deviation Hoaglin et al. (NMAD; 1983), defined as $1.48 \times \mathrm{median}[|z_{spec} - z_{phot}|/(1 + z_{spec})]$. These two values quantify a) the rate at which the photometric redshift method produces a redshift value that is in significant tension to the measured spectroscopic redshift and b) the spread around $z_{spec} - z_{phot}$ in a manner that is resistant to influence from the relatively small number of extreme outliers. Comparing the two fields, COSMOS has greater depth and uniformity while XMM-LSS is shallower and wider, and has around 1 magnitude variability in its optical coverage . It is therefore expected that XMM-LSS would produce photometric redshifts of lower quality. The results of this comparison are displayed in Table 2.3. For each field, I produce two sets of photometric redshift estimations one with and one without the inclusion of the *Spitzer* IRAC 3.6 and 4.5 μm bands. The addition of the *Spitzer* IRAC bands to COSMOS makes minimal difference in the quality of the photometric redshifts. However, redshifts are found to decrease in quality for the faintest objects in the XMM-LSS catalogue due to the relatively shallow coverage provided by *Spitzer*. Results are displayed in Fig. 2.2 and Fig. 2.3.

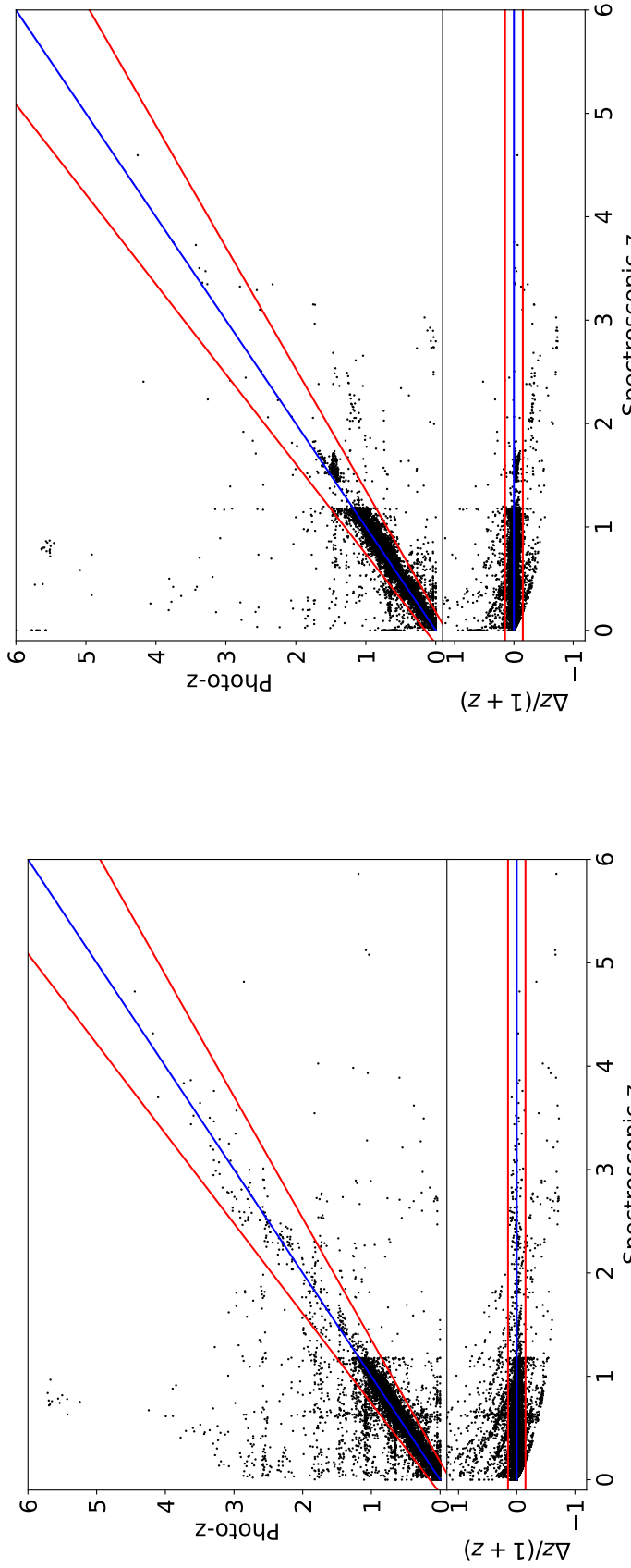
2.3.5 Stellar mass determination

With photometric redshifts and object classification determined for each source, I proceed to measure the stellar mass. This is performed by fixing the redshift to the best-fit value (template with minimum χ^2) and rerunning LePhare using the total flux measurements, rather than aperture fluxes. For SExtractor I use fluxes from the `MAG_AUTO` parameter and for ProFound this is `magt`. Compared to the aperture fluxes, the total fluxes are essential to making accurate measurements of the normalisation of the SED for resolved objects and hence the total luminosity



(a) XMM-LSS **(b)** COSMOS

Figure 2.2: A comparison of the photometric redshifts derived in this study, without the use of the *Spitzer* bands, to a compilation of spectroscopic redshifts. On the left I show the results for XMM-LSS and on the right, the results for COSMOS. The blue line shows the one-to-one correlation in the ideal case, and the upper and lower red lines show the 15 per cent margin in $1 + z$ that defines a significant outlier. The secondary-plot in each figure shows the same figure with the y-axis instead scaling with $\Delta z = (z_{\rm spec} - z_{\rm phot})$. .



(a) XMM-LSS (+IRAC) (b) COSMOS (+IRAC)

Figure 2.3: A comparison of the photometric redshifts derived in this study, including the two *Spitzer* bands, to a compilation of spectroscopic redshifts. On the left I show the results for XMM-LSS and on the right, the results for COSMOS. The blue line shows the one-to-one correlation in the ideal case, and the upper and lower red lines show the 15 per cent margin in $1 + z$ that defines a significant outlier. The secondary-plot in each figure shows the same figure with the y-axis instead scaling with $\Delta z = (z_{\rm spec} - z_{\rm phot})$..

of the galaxy and stellar mass. In the case that an object has a spectroscopic redshift from one of the surveys described in Section 2.3.3, this value is used instead of the photometric redshift.

For the SED fitting and hence, stellar mass determination, I use a selection of SED templates from Bruzual & Charlot (2003). These include templates with constant star formation rate, exponentially decaying star formation rates with timescales $\tau = [0.1, 0.3, 1, 2, 3, 5, 10, 15, 30]$ Gyr, 58 ages for the stellar population (from 0.01 to 13.5 Gyr) and two metallicities ($Z_{\odot}$ and 20 per cent of $Z_{\odot}$). No additional evolutionary corrections are applied to the template sets because the set already includes 58 time stamps for each star formation history. Similarly, since I am directly fitting redshifted spectra to the photometry, k-corrections are not required to obtain rest-frame luminosities because these can be obtained from the original template.

2.4 Measuring the GSMF

I select objects for use in measuring the GSMF based on a number of criteria to maintain purity and completeness.

1. The source exists in both SExtractor and ProFound derived catalogues. The majority of sources that fail this are either artefacts on the edge of manual masking or are a consequence of the initial ProFound selection on a VISTA stack verses just the Ks-band. Such sources would also removed by the K_s magnitude cuts detailed below.

2. The source has a 2 arcsecond aperture magnitude following the condition $K_s < 24.5$ in COSMOS and $K_s < 23.2$ in XMM-LSS. This corresponds to a SNR cut of 8σ and is employed to minimise the potential for contamination while enabling M^* to be well constrained up to $z \sim 2$.

3. The source has a best fit SED that is a galaxy or QSO with a redshift between 0.1 and 2.0 ($\chi^2_{\rm Gal/QSO} < \chi^2_{\rm Star}$). In the case a source has a spectroscopic redshift, that value is used in place of the photometric redshift.

4. I apply an upper limit on the quality of the photometric redshift of $\chi^2 < 250$ (removing the worst 0.5 per cent of objects).

This results in a sample of $\sim 320,000$ galaxies in the combined COSMOS and XMM-LSS fields used in measuring the GSMF. Of order 97 per cent of the sample have a quality of fit corresponding to $\chi^2 < 50$. Sources can experience large increases in χ^2 if they are influenced by an image artefact, causing the measurement in a single photometric band to be incorrect. Additionally, if a significant emission line is unmodelled/missing from the template, a similar spike in flux within a photometric band may occur. Source blending can also lead to issues with the aperture photometry in crowded regions, especially when considering the larger PSF of *Spitzer* images. The $\chi^2 < 250$ cut compromises between being inclusive of objects affected by minor photometric issues while preventing objects which are clearly junk from entering the sample. It is inevitable that some degree of contamination in the sample will occur, but this is expected to be small ($\lesssim 1$ per cent).

I present the galaxy number counts in each redshift bin in Tables 2.5 and 2.6. The inclusion of *Spitzer* bands leads to a significant reduction in the number of highly massive galaxies at redshifts $0.1 < z < 2$. This is the result of *Spitzer* bands breaking degeneracies between stellar templates and certain combinations of galaxy template, leading to a number of these massive objects being reclassified as stars. The use of ProFound photometry over SExtractor leads to minimal difference to the general population of objects, cases will be discussed in Section 2.5.1.

2.4.1 The $1/V_{\rm max}$ method

I first compute the GSMFs using the 1/Vmax methodology (Schmidt, 1968; Rowan-Robinson, 1968). I determined the $V_{\rm max}$ for each galaxy by redshifting the best-fitting template (from the 2 arcsecond aperture photometry) until the object no longer meets my Ks-band magnitude limit.

The GSMF is then determined using:

$$\Phi(M)d\log(M) = \frac{1}{\Delta M}\sum_{i}^{N}\frac{1}{V_{\max,i}}, \tag{2.1}$$

$$\delta\Phi(M) = \frac{1}{\Delta M}\sqrt{\sum_{i}^{N}\left(\frac{1}{V_{\max,i}}\right)^{2}}, \tag{2.2}$$

where M is stellar mass, ΔM is the bin width which I set to 0.25 dex and $V_{\max,i}$ is the maximum volume for which galaxy i could have been successfully detected.

2.4.2 Stellar mass completeness

Towards lower stellar masses, galaxies become intrinsically less luminous. This ultimately leads to a regime where the detection limits of the data are reached and galaxy number counts begin to fall as they are lost to noise. As the science goals of this study focus on the massive end of the GSMF, I adopt a conservative approach while still probing a significant mass range. In my model fitting procedures I elect to only use bins of redshift and mass where over 95 per cent of the galaxy sample are brighter than the 8σ 2 arcsecond aperture detection limit of the respective field. Due to the shallower coverage in XMM-LSS, the criteria for completeness is approximately 0.5-0.75 dex higher in stellar mass than in COSMOS.

I apply a simple correction to the survey areas based on the fraction of the field that is occupied by other sources or masked by foreground stars. For COSMOS this is 15 per cent and XMM it is 7 per cent. This corrects the GSMF for the probability that sources are highly blended ($>$ 50 per cent overlap) with other sources or significantly effected by bright foreground stars.

2.4.3 Cosmic Variance

My measurements of the GSMF are based on data that only probe a limited volume of the Universe. As a result, they are susceptible to biases that are a consequence of large-scale fluctuations in density in the galaxy distribution. This is commonly referred to as cosmic variance (σ^2_{cv}). As I am measuring the GSMF across a wide range of mass and redshift, there is no single quantitative value that can be used to describe this effect. In order to model the effects of cosmic variance on my

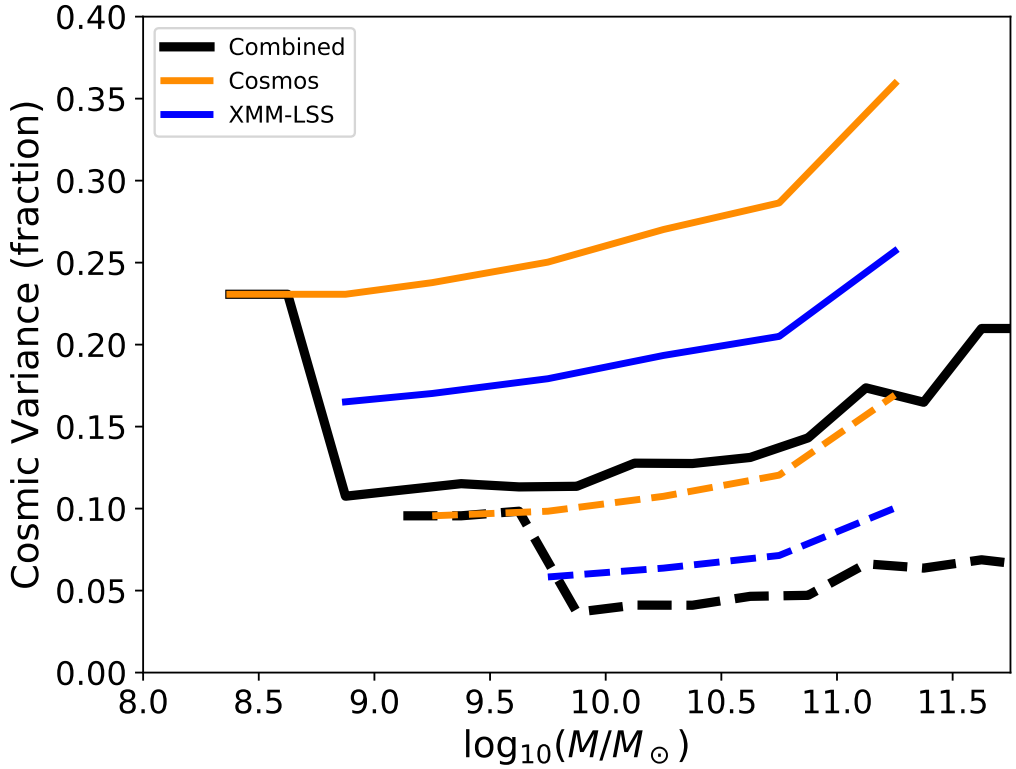


Figure 2.4: The estimated cosmic variance for two example redshift bins. The solid lines show $0.2 < z < 0.3$, the dashed lines show $0.75 < z < 1.0$. Here I display the cosmic variance for each field if treated independently and the result of combining the fields using the cosmic variance calculator from Moster et al. (2011) and my measured number counts. COSMOS is shown in gold, XMM-LSS in blue and the combination of the two fields in black. Where XMM-LSS becomes incomplete the cosmic variance value for the combined case is just the COSMOS cosmic variance.

measurements, I use the treatment from Moster et al. (2011), which provides an estimate of the cosmic variance as a function of both stellar mass and redshift. My dataset consists of two fields with differing area and dimensions, thus allowing us to mitigate some of the effects of cosmic variance. Where both the XMM-LSS and COSMOS fields are used in measuring the GSMF I calculate the cosmic variance for each field independently ($\sigma^2_{cv,i}$) and combine the values together with a co-moving volume weighting (Equation 7 in Moster et al., 2011). The output is a percentage error on the GSMF, and so to combine the fields this value is converted back into variance (σ^2_{cv}) using my galaxy number counts. When data from the XMM-LSS field drops out of consideration due to its shallower depth, cosmic variance is determined

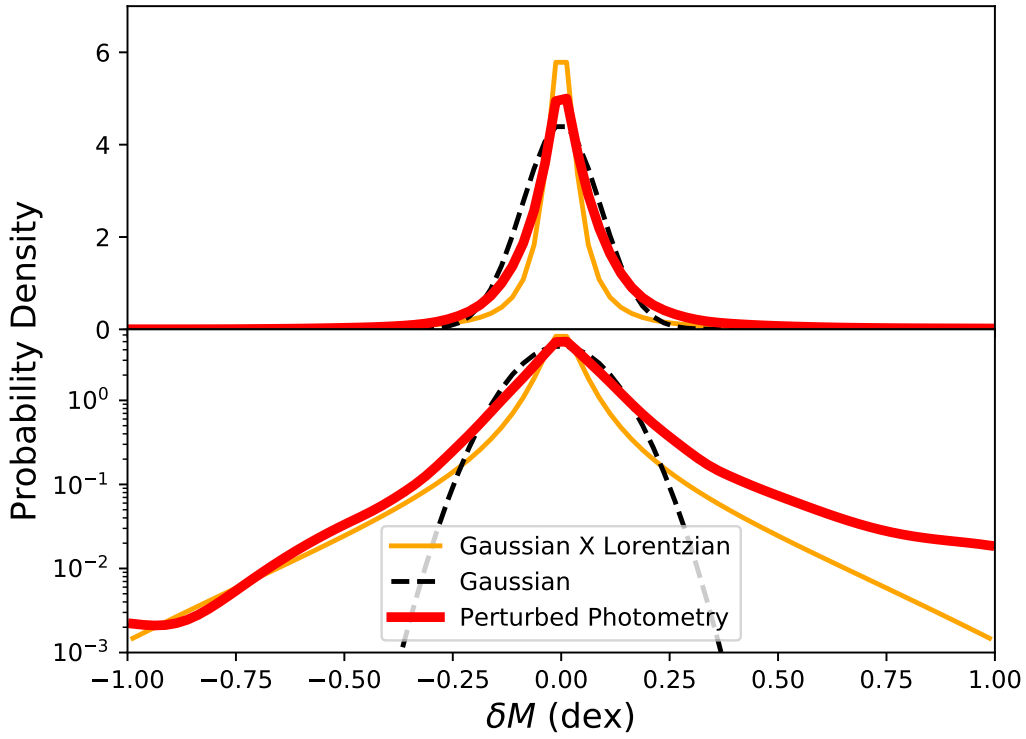


Figure 2.5: The shapes of the three Eddington bias corrections implemented as a part of my study. I show the probability of a certain mass scatter ($\delta M = M_{\rm original} - M_{\rm perturbed}$) for all objects meeting the selection and completeness criteria
as a result of photometric errors in the SED fitting process. The lower plot shows identical data to the upper plot, however I have logged the y-axis to reveal the low-amplitude wings of the distribution. The non-parametric distribution derived directly from my data is shown as the thick red line. The black dashed line is the best fit Gaussian to this data and the orange line is for a Gaussian multiplied by a Lorentzian.

from the area of the COSMOS field alone. Fig. 2.4 shows the results of my cosmic variance calculations for two redshift bins ($0.2 < z < 0.3$ and $0.75 < z < 1.0$), highlighting how the increased area from including the XMM-LSS field, coupled with combining two widely separated fields, results in a factor of ~ 2 decrease in the cosmic variance uncertainty.

2.4.4 Eddington bias

The steep drop in the GSMF beyond the knee can lead to a bias in the derived number densities of the most massive galaxies due to Eddington bias (Eddington, 1913).

All galaxies in the sample have an uncertainty in the derived stellar mass, however as low-mass galaxies are significantly more numerous this leads to more galaxies scattering to higher stellar mass than the number that scatter to lower masses. To account for this effect and hence, determine an estimate of the intrinsic GSMF, I require an estimate of the uncertainty in the stellar masses derived for my sample. With this distribution I can then deconvolve (or in reality fit a convolved double Schechter form) to my observations to determine the intrinsic GSMF. To measure the uncertainty in the stellar masses derived in this study, I repeat both the photometric redshift measurement and the SED fitting process after perturbing the photometry of my sources according to the photometric errors in each band. This process is repeated multiple times to produce the distributions shown in Fig. 2.5. Based on this, I examine three possible methodologies to uncover the intrinsic stellar mass function from this observed distribution in my analysis described in Section 2.4.5.

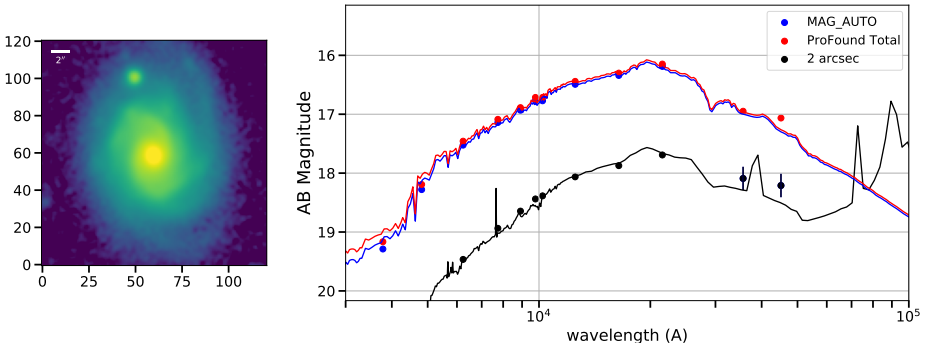
2.4.5 The measured galaxy stellar mass functions

The GSMFs, as measured from the samples produced from my four catalogues, are presented in Fig. 2.9. They probe stellar masses from $7.75 < \log_{10}(M_{\odot}) < 11.75$ and are split into nine redshift bins between $0.1 < z < 2.0$. I provide an example of the SED fit to a galaxy from each of these redshift bins in Figures. 2.6, 2.7 and 2.8.

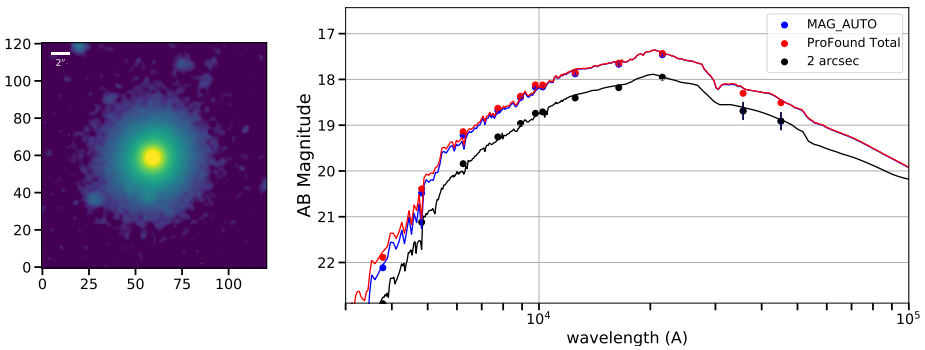
To each GSMF I fit a double Schechter function (Schechter, 1976, Eqn. 2.3) using a Markov-Chain Monte Carlo (MCMC) implemented in `emcee` (Foreman-Mackey et al., 2013). In this redshift range, past studies have found the double Schecher functional form better describes the GSMF due to the underlying bimodality in the galaxy population (e.g. Strateva et al., 2001; Driver et al., 2006; Baldry et al., 2012; Ilbert et al., 2013; Davidzon et al., 2017a) and I clearly see an upturn at the low mass end of my GSMFs.

$$\Phi(M)\delta M = \left[\Phi_1 \left(\frac{M}{M^*}\right)^{\alpha_1} + \Phi_2 \left(\frac{M}{M^*}\right)^{\alpha_2}\right] \exp\left(-\frac{M}{M^*}\right) \frac{\delta M}{M^*}. \tag{2.3}$$

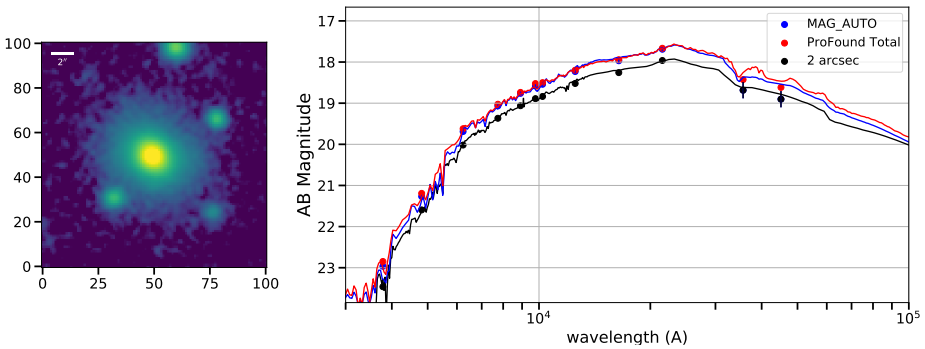
A series of priors are applied to prevent parameters from flipping due to the symmetry of the double Schechter functional form shown in Equation 2.3. The



(a) $0.1 < z < 0.2$

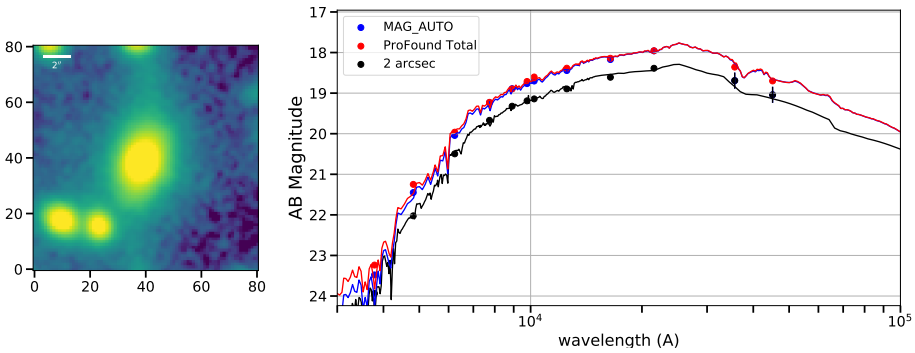


(b) $0.2 < z < 0.3$

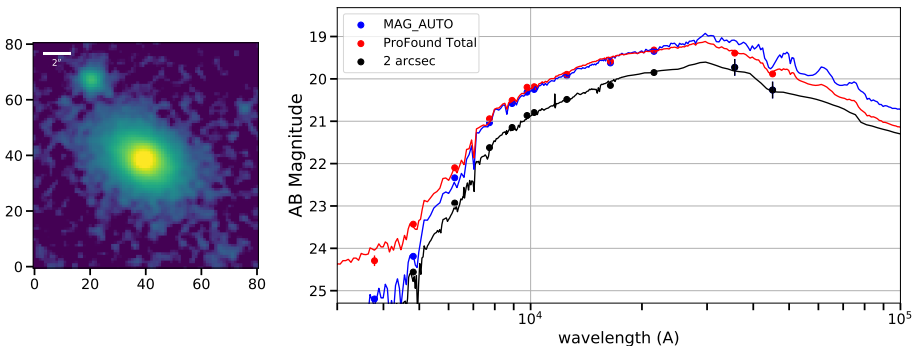


(c) $0.3 < z < 0.5$

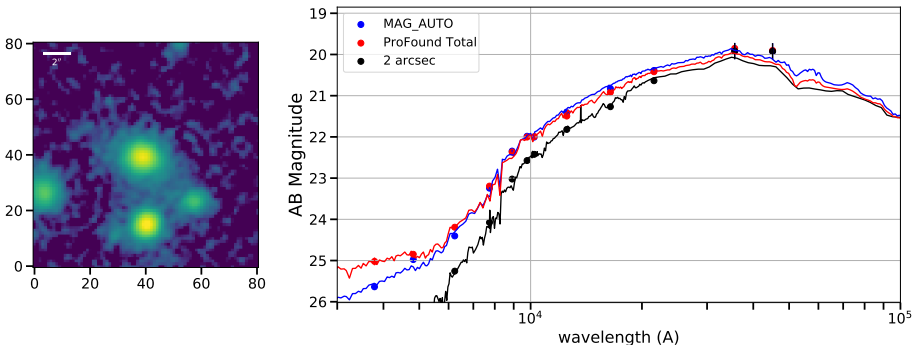
Figure 2.6: Example galaxies from the lowest three redshift bins alongside their measured photometry and SED fits. The left figure shows a cut out of the galaxy from the K_s images, axis labels are in pixels and a scale bar showing 2 arcseconds is added for reference. The right plots show the measured photometry in 2 arcsecond apertures (black), SExtractor `MAG_AUTO` (blue) and ProFound total photometry (red) as well as the best fitting SED for each set of photometry. Each galaxy is taken from one of the redshift bins used in measuring the stellar mass function and has a mass approximately that of the characteristic mass ($10.5 < \log_{10}(M/M_{\odot}) < 11.0$).



(a) $0.5 < z < 0.75$

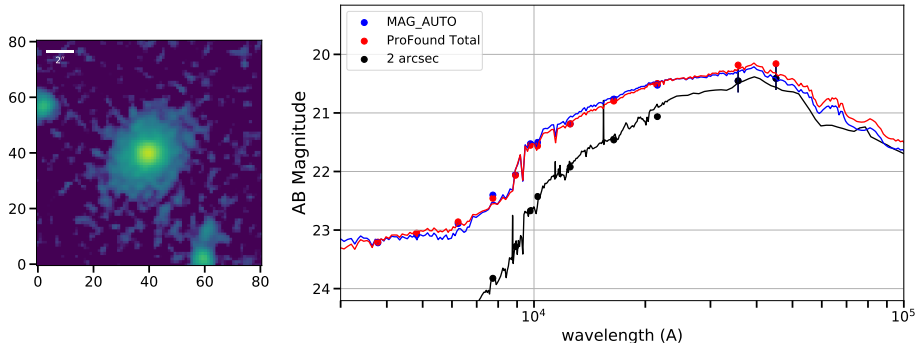


(b) $0.75 < z < 1.0$

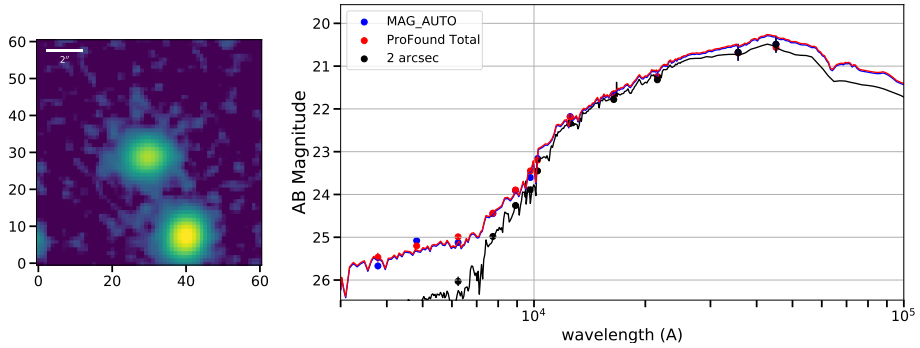


(c) $1.0 < z < 1.25$

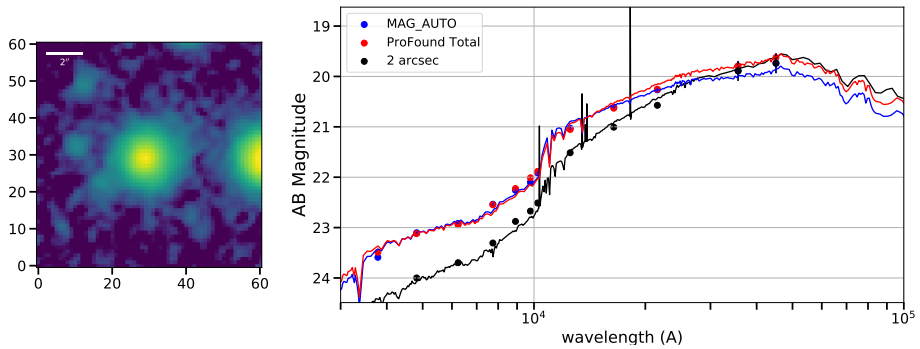
Figure 2.7: Example galaxies from the central three redshift bins alongside their measured photometry and SED fits. The left figure shows a cut out of the galaxy from the K_s images, axis labels are in pixels and a scale bar showing 2 arcseconds is added for reference. The right plots show the measured photometry in 2 arcsecond apertures (black), SExtractor `MAG_AUTO` (blue) and ProFound total photometry (red) as well as the best fitting SED for each set of photometry. Each galaxy is taken from one of the redshift bins used in measuring the stellar mass function and has a mass approximately that of the characteristic mass ($10.5 < \log_{10}(M/M_{\odot}) < 11.0$).



(a) $1.25 < z < 1.5$



(b) $1.5 < z < 1.75$



(c) $1.75 < z < 2.0$

Figure 2.8: Example galaxies from the highest three redshift bins alongside their measured photometry and SED fits. The left figure shows a cut out of the galaxy from the K_s images, axis labels are in pixels and a scale bar showing 2 arcseconds is added for reference. The right plots show the measured photometry in 2 arcsecond apertures (black), SExtractor `MAG_AUTO` (blue) and ProFound total photometry (red) as well as the best fitting SED for each set of photometry. Each galaxy is taken from one of the redshift bins used in measuring the stellar mass function and has a mass approximately that of the characteristic mass ($10.5 < \log_{10}(M/M_{\odot}) < 11.0$).

normalisation (Φ_1 & Φ_2) follows the condition $\Phi_1 > \Phi_2$, the low-mass slopes (α_1 & α_2) are limited to being between [-1.8,1.5] and [-3.0,-0.9] respectively for the two components. To compare against past studies, I make the same assumption that each Schechter component has a single, shared value of M^* in the range $10 < \log_{10}(M^*/\mathrm{M}_\odot) < 12$. I only fit to data points where the bins in stellar mass are greater than 95 percent complete. The MCMC is set up with 500 walkers that burn in for 100,000 steps before conducting a further 20,000 steps for use in mapping the posterior distributions. Each walker is distributed uniformly in the parameter space and limited by the above priors. I perform the fitting procedure four times, once on the observed GSMF and three times with the double Schechter function modified with the convolution with one of three Eddington bias methods that I describe below.

First, I modify the fit to be a convolution of the double Schechter function with a Gaussian distribution, where the standard deviation (σ_M) of the distribution is calculated by fitting a Gaussian to the measured scatter in masses shown in Fig. 2.5. This is a commonly used method in studies of the GSMF (e.g. Wright et al., 2018). In the second method I extend this model by multiplying the Gaussian with a Lorentzian in the same manner as described in Ilbert et al. (2013); Davidzon et al. (2017a). This adds extended wings to the function, which more adequately reproduces the distribution, however it continues with the assumption that the mass scatter is symmetric. Any asymmetry is important to account for, as it means that there is a greater probability of scattering to lower masses than towards higher masses (due to the photometric redshift uncertainty), and this would impact on the measured GSMF. Therefore, in the third case, I do not fit any parametric distribution to the mass scatter, instead I use the smoothed histogram of the scatter convolved directly with the double Schechter function. This should improve upon the use of the analytic forms because it captures the observed asymmetry in the mass scatter (δM).

Visual inspection of the distribution in mass scatter derived in Section 2.4.4 reveals there to be minimal dependence on redshift. The two lowest redshift bins are slightly broader as a consequence of photometric redshift uncertainty but these bins are much more complete with spectroscopic redshifts (30 per cent). As a

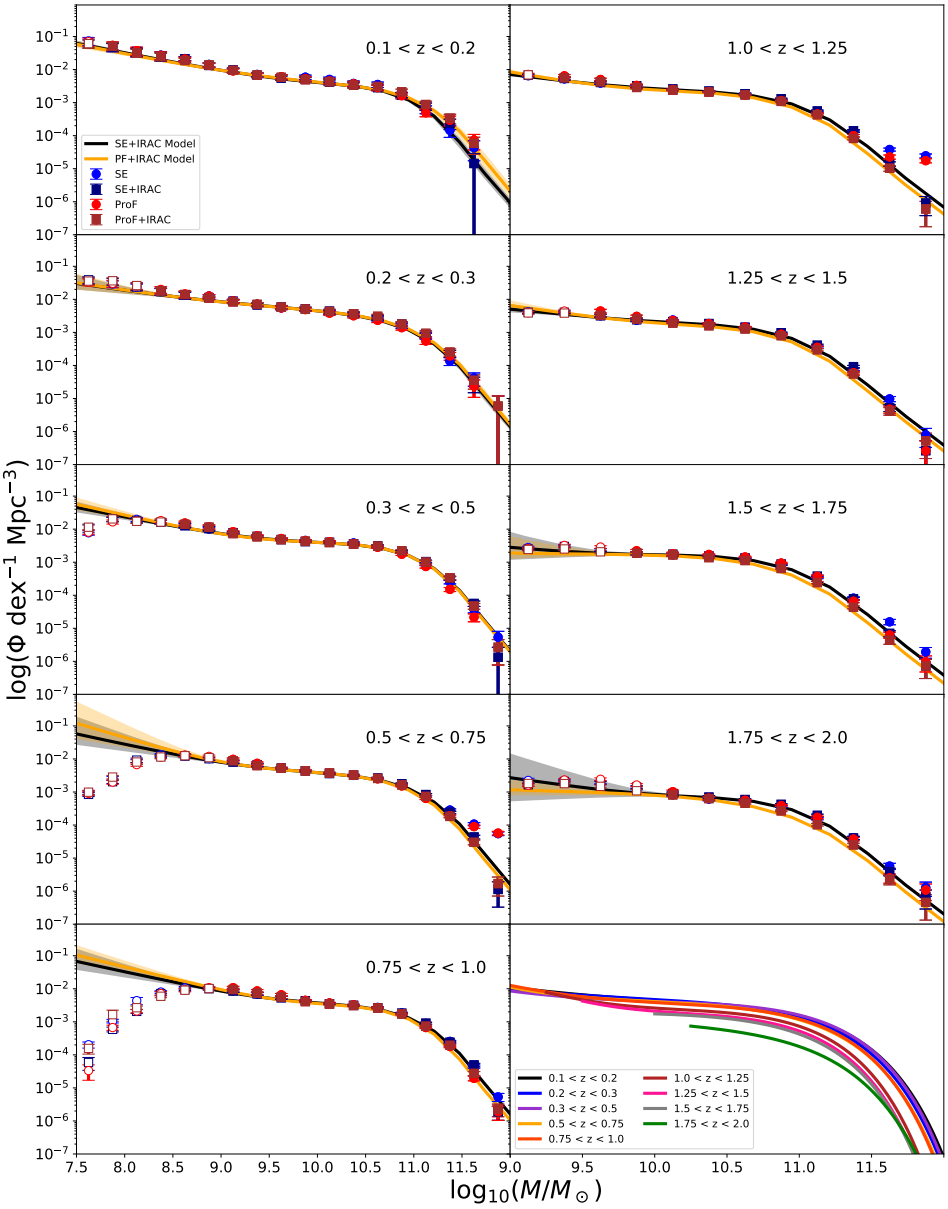


Figure 2.9: The GSMF from my analysis of the COSMOS and XMM-LSS data. In blue and red are the results from SExtractor and ProFound, while navy and brown data points denote the points measured after including *Spitzer*/IRAC photometry. Unfilled data points indicate those with less than 95 per cent completeness and are not used in any fitting procedure. Displayed data points are based on raw observations and do not correct for Eddington bias (see Tables 2.5 and 2.6). The black line shows the median of the MCMC results for the SExtractor+IRAC data when fit with a double Schechter function convolved with my Eddington correction. The grey shaded region shows the region contained within 1σ of the model fit and is based on 10,000 random samples of the final posterior. The gold line and shaded region follows the same process with the ProFound+IRAC data. The high redshift bins in the right column have the x-axis truncated to higher masses to focus on the complete regime. In the final panel the Eddington corrected ProFound models are shown simultaneously and cut where data is incomplete.

result the real scatter within these bins is likely much smaller. Using the method which convolves a Gaussian distribution with the measured GSMF, I find values with $0.08 < \sigma_M < 0.10$ across all redshifts. For the second method, which uses a Gaussian distribution multiplied with a Lorentzian, I find $0.43 < \sigma_M < 0.58$. Since minimal evolution was found, I remove the redshift dependence on the Lorentzian component that was introduced in Ilbert et al. (2013) to minimise total fit parameters. The resultant formula for the extended wings is thus $L(x) = \frac{0.08}{2\pi} \frac{1}{\frac{0.08}{2\pi}^2 + x^2}$ which is equivalent to fixing the redshift to 1.0 in the original formula from Ilbert et al. (2013). The range of $0.43 < \sigma_M < 0.58$ for the Gaussian component is in agreement with the findings of Ilbert et al. (2013) and slightly larger than the value found by Davidzon et al. (2017a). I note that the σ_M values for the two methodologies are not directly comparable due to the different functional forms used between them. In addition, the studies conducted by Grazian et al. (2015) and Davidzon et al. (2017a) find evidence for redshift and stellar mass dependence on the measured mass scatter when approaching the completeness limited regime. The probable explanation for the lack of such observation in my data is the conservative SNR cuts that have been implemented. Following previous studies, I fix the σ_M values in my final fits to 0.09 in the Gaussian case and 0.5 for the Gaussian $\times$ Lorentzian case. The shapes of these distributions are presented in Fig. 2.5. I discuss the impacts of this correction on the measured GSMF in Section 2.5.2.

My preferred method for correcting for Eddington bias is to use the histogram presented in Fig. 2.5 directly in the convolution. This method directly uses the results of the perturbed catalogues and captures the subsequent asymmetry found in the distribution. Such an asymmetry has previously been described in recent studies exploring the Eddington bias such as Grazian et al. (2015). The best-fitting double Schechter function fit parameters using this Eddington bias correction are presented in Table 2.4. Corner plots showing the posterior probability distribution for the parameters in this model are shown in Fig. 2.10, 2.11, 2.12, 4.6, 2.14, 2.15, 2.16, 2.17, 2.18. For completeness, I also present my results without the application

of the Eddington bias correction in Tables 2.5, 2.6, 2.7, 2.8 alongside the results obtained using the various parametric fits to the mass scatter.

My double Schechter model is only fit up to stellar masses of $10^{11.75} M_{\odot}$. While there are a small number of objects with stellar masses greater than this limit in my sample, these are increasingly likely to be subject to forms of contamination such as AGN activity, source blending, misclassification of stars or artefacts. In the high redshift regime I am also unable to constrain the low-mass Schechter component. If I instead fit with a single Schechter function for $z > 1.25$, I find the shift in the fit parameters to be minimal ($< 1\sigma$) compared to results obtained with a double Schechter function. For consistency I proceed with the double Schechter functional form for all redshift bins.

I note that visual inspection of the GSMFs measured reveal a rather significant change between $0.75 < z < 1.0$ and $1.0 < z < 1.25$ (see the last panel of Figure.2.9). This evolution is present with both photometric methods and regardless of the inclusion of the *Spitzer* data. Inspection of the redshift distributions reveal no significant features, such as peaks or troughs, within these redshift bins. The total shift in the best-fit Schechter parameters of these two mass functions amounts to around 2 sigma, so the combination of statistical errors/cosmic variance or an unknown systematic could be the driver of such a change.

Table 2.4: The best-fitting double Schechter function parameters derived from my observed GSMF when corrected for Eddington bias through the direct implementation of the mass scatter over fitting an analytic form. Priors are set to ensure that α_1 & $\log_{10}(\Phi_1)$ refer to the high mass component and α_2 & $\log_{10}(\Phi_2)$ refer to the low mass component.

Redshift SExtractor+IRAC	$\log_{10}(M^*/M_\odot)$	α_1	$\log_{10}(\Phi_1)$ [mag^{-1}Mpc^{-3}]	α_2	$\log_{10}(\Phi_2)$ [mag^{-1}Mpc^{-3}]
0.1-0.2	$10.71^{+0.11}_{-0.11}$	$-0.47^{+0.44}_{-0.36}$	$-2.61^{+0.10}_{-0.12}$	$-1.60^{+0.09}_{-0.12}$	$-3.34^{+0.23}_{-0.36}$
0.2-0.3	$10.89^{+0.08}_{-0.13}$	$-1.06^{+0.59}_{-0.17}$	$-2.76^{+0.12}_{-0.10}$	$-1.57^{+0.24}_{-0.65}$	$-4.12^{+1.03}_{-3.28}$
0.3-0.5	$10.83^{+0.05}_{-0.05}$	$-0.64^{+0.22}_{-0.17}$	$-2.60^{+0.05}_{-0.05}$	$-1.61^{+0.14}_{-0.18}$	$-3.62^{+0.30}_{-0.44}$
0.5-0.75	$10.86^{+0.05}_{-0.05}$	$-0.80^{+0.30}_{-0.18}$	$-2.71^{+0.05}_{-0.06}$	$-1.69^{+0.27}_{-0.47}$	$-3.78^{+0.56}_{-1.06}$
0.75-1.0	$10.83^{+0.05}_{-0.05}$	$-0.67^{+0.28}_{-0.20}$	$-2.68^{+0.04}_{-0.05}$	$-1.67^{+0.19}_{-0.32}$	$-3.63^{+0.38}_{-0.68}$
1.0-1.25	$10.67^{+0.05}_{-0.05}$	$-0.19^{+0.35}_{-0.30}$	$-2.75^{+0.05}_{-0.07}$	$-1.58^{+0.18}_{-0.28}$	$-3.39^{+0.25}_{-0.44}$
1.25-1.5	$10.60^{+0.05}_{-0.05}$	$0.00^{+0.37}_{-0.42}$	$-2.87^{+0.09}_{-0.10}$	$-1.44^{+0.18}_{-0.47}$	$-3.29^{+0.20}_{-0.60}$
1.5-1.75	$10.66^{+0.05}_{-0.06}$	$-0.45^{+0.40}_{-0.23}$	$-2.84^{+0.04}_{-0.09}$	$-1.79^{+0.59}_{-0.78}$	$-4.39^{+1.01}_{-2.39}$
1.75-2.0	$10.66^{+0.07}_{-0.08}$	$-0.25^{+0.54}_{-0.42}$	$-3.19^{+0.04}_{-0.10}$	$-1.94^{+0.60}_{-0.70}$	$-4.36^{+0.69}_{-2.88}$
ProFound+IRAC					
0.1-0.2	$10.88^{+0.16}_{-0.15}$	$-0.74^{+0.50}_{-0.40}$	$-2.71^{+0.15}_{-0.20}$	$-1.62^{+0.12}_{-0.23}$	$-3.54^{+0.35}_{-0.91}$
0.2-0.3	$10.90^{+0.09}_{-0.13}$	$-1.04^{+0.56}_{-0.18}$	$-2.75^{+0.12}_{-0.11}$	$-1.60^{+0.26}_{-0.64}$	$-4.13^{+1.02}_{-3.05}$
0.3-0.5	$10.83^{+0.05}_{-0.05}$	$-0.67^{+0.21}_{-0.16}$	$-2.58^{+0.04}_{-0.05}$	$-1.70^{+0.17}_{-0.22}$	$-3.79^{+0.38}_{-0.52}$
0.5-0.75	$10.81^{+0.04}_{-0.05}$	$-0.86^{+0.25}_{-0.13}$	$-2.67^{+0.04}_{-0.05}$	$-1.94^{+0.39}_{-0.54}$	$-4.21^{+0.78}_{-1.09}$
0.75-1.0	$10.75^{+0.04}_{-0.04}$	$-0.54^{+0.22}_{-0.18}$	$-2.62^{+0.03}_{-0.04}$	$-1.76^{+0.17}_{-0.24}$	$-3.64^{+0.30}_{-0.45}$
1.0-1.25	$10.63^{+0.05}_{-0.05}$	$-0.29^{+0.31}_{-0.24}$	$-2.74^{+0.03}_{-0.05}$	$-1.81^{+0.22}_{-0.31}$	$-3.64^{+0.32}_{-0.50}$
1.25-1.5	$10.57^{+0.05}_{-0.05}$	$-0.08^{+0.39}_{-0.34}$	$-2.89^{+0.06}_{-0.10}$	$-1.65^{+0.24}_{-0.43}$	$-3.46^{+0.26}_{-0.55}$
1.5-1.75	$10.66^{+0.04}_{-0.06}$	$-0.70^{+0.32}_{-0.13}$	$-2.91^{+0.04}_{-0.05}$	$-1.86^{+0.62}_{-0.75}$	$-5.05^{+1.35}_{-3.19}$
1.75-2.0	$10.76^{+0.05}_{-0.05}$	$-0.98^{+0.19}_{-0.13}$	$-3.38^{+0.06}_{-0.07}$	$-1.86^{+0.64}_{-0.74}$	$-6.64^{+2.07}_{-2.29}$

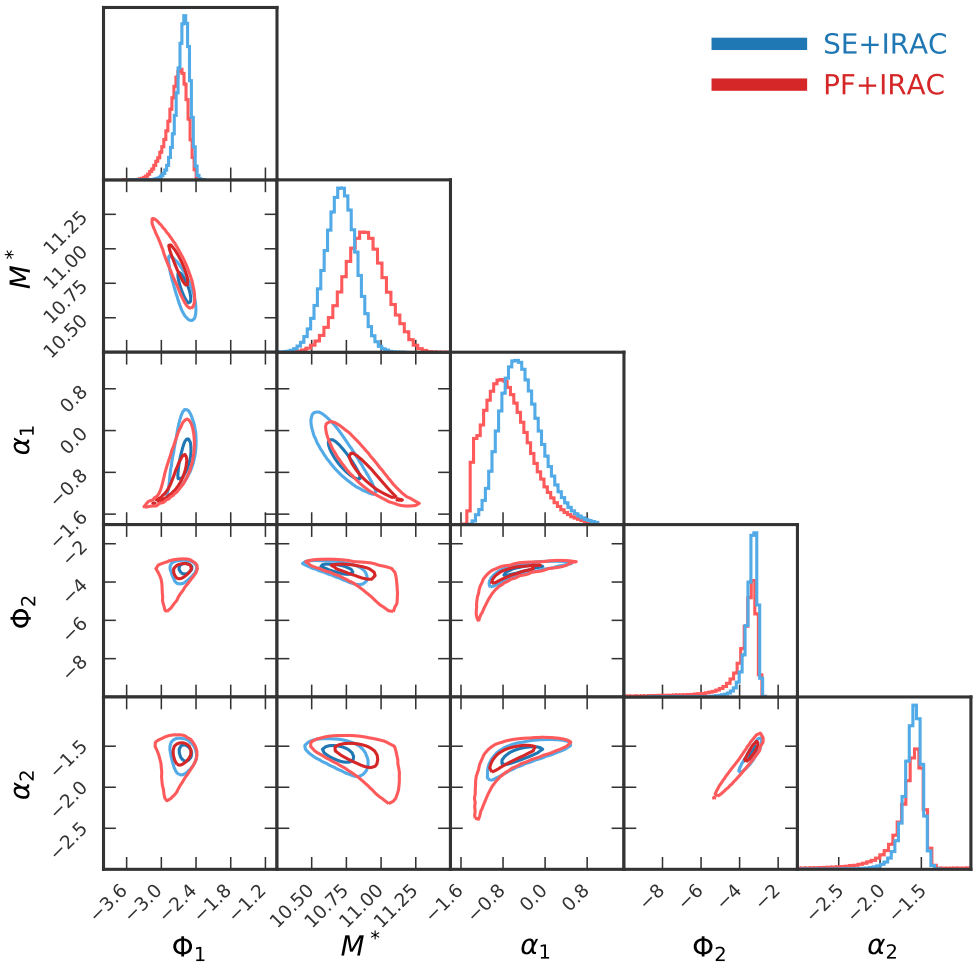


Figure 2.10: The posterior distributions generated by the MCMC fitting procedure for the $0.1 < z < 0.2$ redshift bin. Contours shown are the 2-dimensional 1σ and 2σ levels for the final 20,000 steps undertaken by each of the 500 walkers. Results from SExtractor derived results are shown in blue while ProFound derived results are shown in red. Contours are shown for the runs that include the use of *Spitzer* photometry and correct for Eddington bias through the mass scatter measured by perturbing the photometry in the SED fitting process.

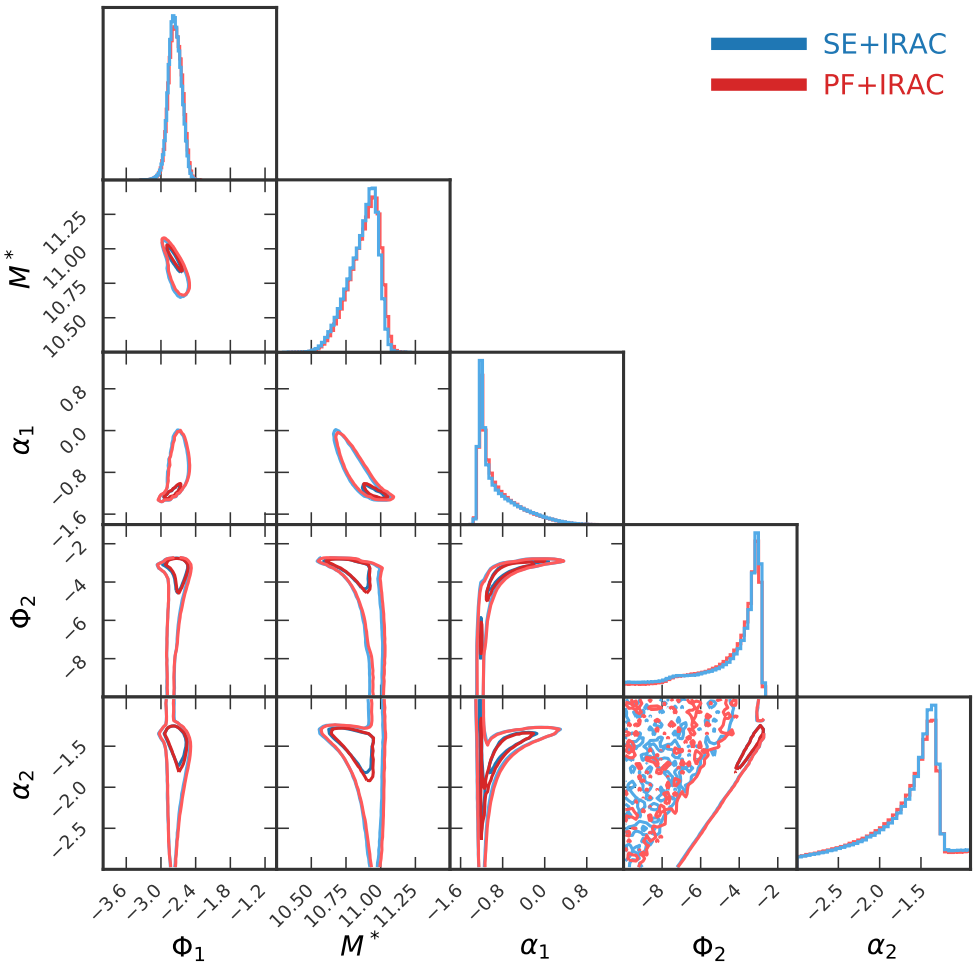


Figure 2.11: The posterior distributions generated by the MCMC fitting procedure for the $0.2 < z < 0.3$ redshift bin. Contours shown are the 2-dimensional 1σ and 2σ levels for the final 20,000 steps undertaken by each of the 500 walkers. Results from SExtractor derived results are shown in blue while ProFound derived results are shown in red. Contours are shown for the runs that include the use of *Spitzer* photometry and correct for Eddington bias through the mass scatter measured by perturbing the photometry in the SED fitting process.

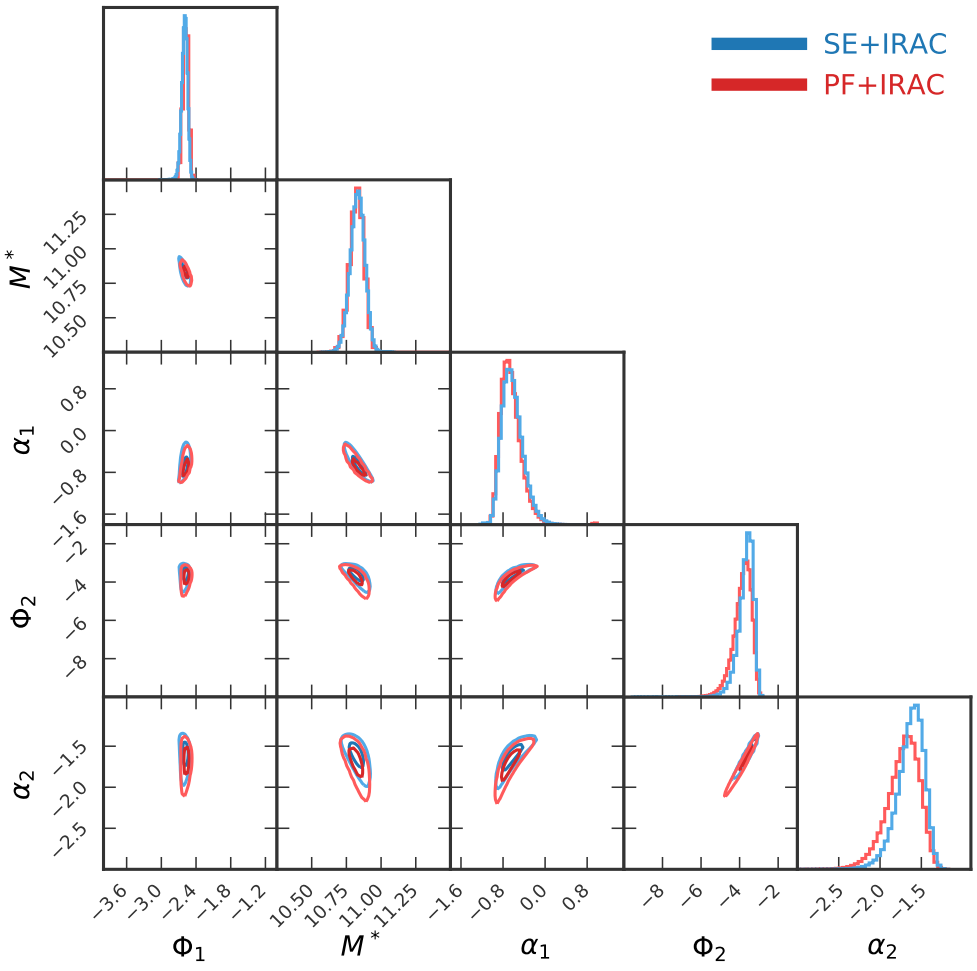


Figure 2.12: The posterior distributions generated by the MCMC fitting procedure for the $0.3 < z < 0.5$ redshift bin. Contours shown are the 2-dimensional 1σ and 2σ levels for the final 20,000 steps undertaken by each of the 500 walkers. Results from SExtractor derived results are shown in blue while ProFound derived results are shown in red. Contours are shown for the runs that include the use of *Spitzer* photometry and correct for Eddington bias through the mass scatter measured by perturbing the photometry in the SED fitting process.

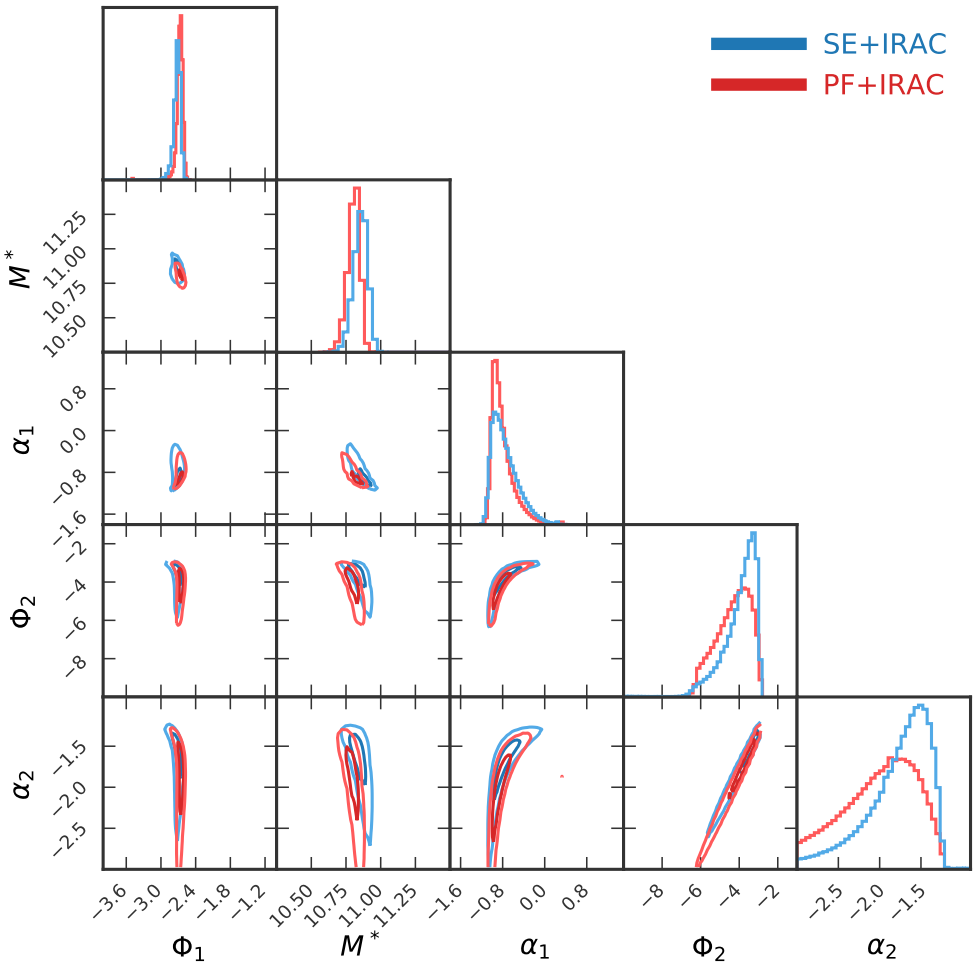


Figure 2.13: The posterior distributions generated by the MCMC fitting procedure for the $0.5 < z < 0.75$ redshift bin. Contours shown are the 2-dimensional 1σ and 2σ levels for the final 20,000 steps undertaken by each of the 500 walkers. Results from SExtractor derived results are shown in blue while ProFound derived results are shown in red. Contours are shown for the runs that include the use of *Spitzer* photometry and correct for Eddington bias through the mass scatter measured by perturbing the photometry in the SED fitting process.

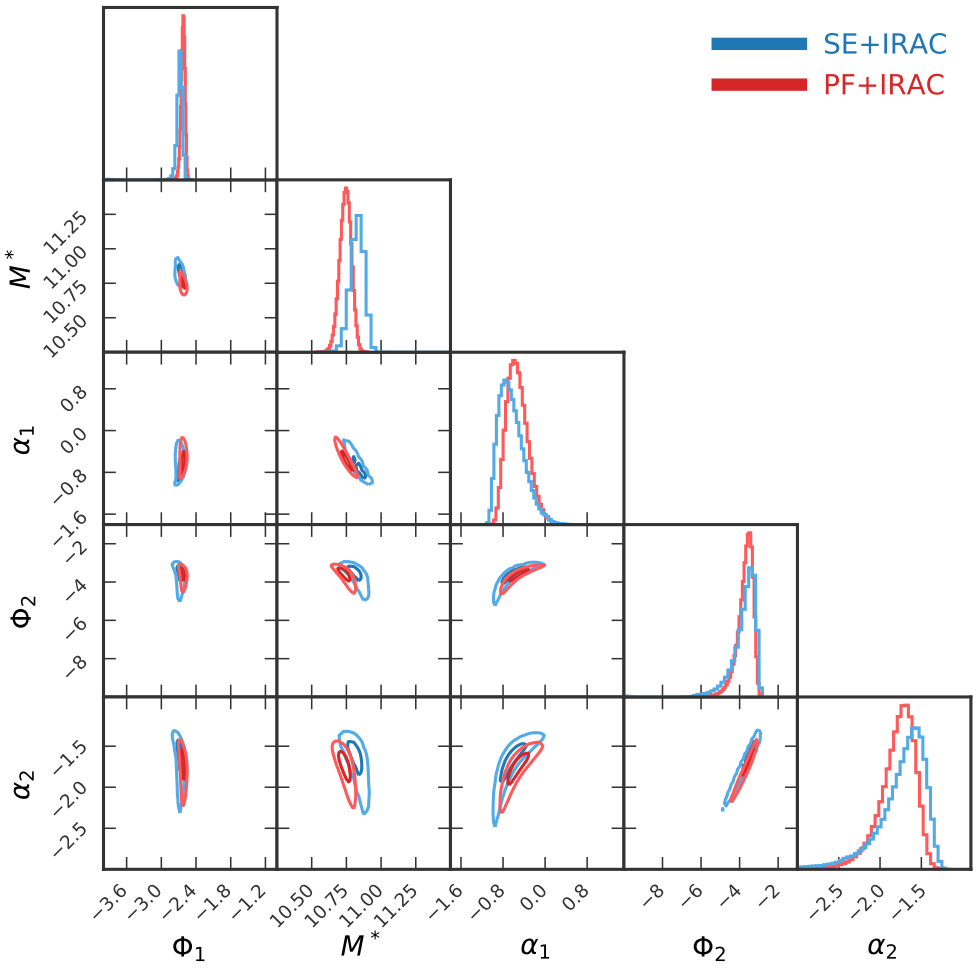


Figure 2.14: The posterior distributions generated by the MCMC fitting procedure for the $0.75 < z < 1.0$ redshift bin. Contours shown are the 2-dimensional 1σ and 2σ levels for the final 20,000 steps undertaken by each of the 500 walkers. Results from SExtractor derived results are shown in blue while ProFound derived results are shown in red. Contours are shown for the runs that include the use of *Spitzer* photometry and correct for Eddington bias through the mass scatter measured by perturbing the photometry in the SED fitting process.

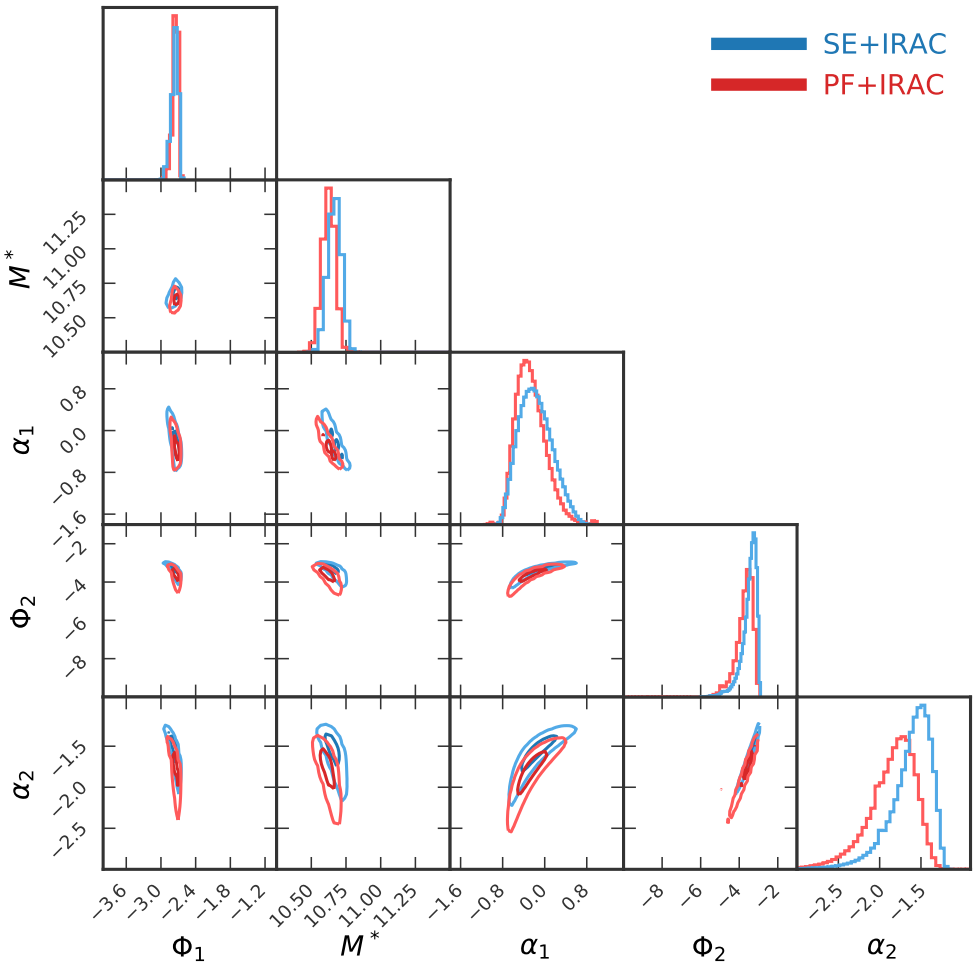


Figure 2.15: The posterior distributions generated by the MCMC fitting procedure for the $1.0 < z < 1.25$ redshift bin. Contours shown are the 2-dimensional 1σ and 2σ levels for the final 20,000 steps undertaken by each of the 500 walkers. Results from SExtractor derived results are shown in blue while ProFound derived results are shown in red. Contours are shown for the runs that include the use of *Spitzer* photometry and correct for Eddington bias through the mass scatter measured by perturbing the photometry in the SED fitting process.

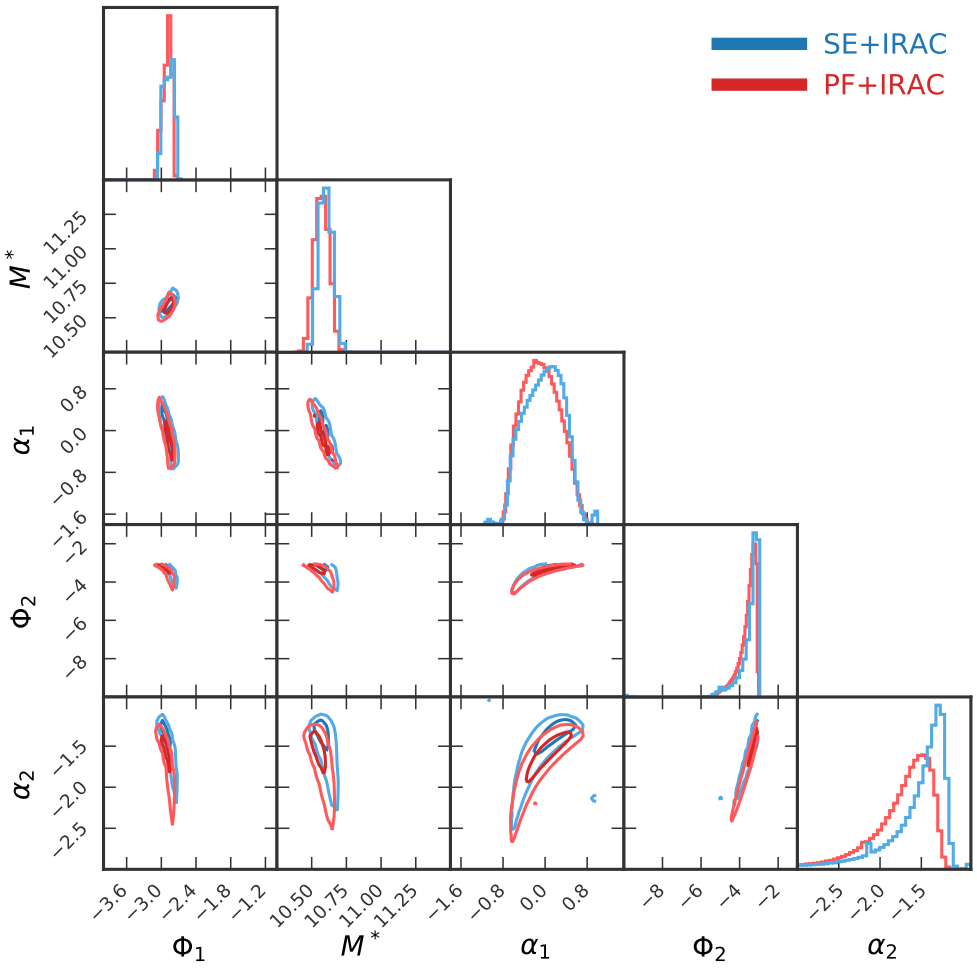


Figure 2.16: The posterior distributions generated by the MCMC fitting procedure for the $1.25 < z < 1.5$ redshift bin. Contours shown are the 2-dimensional 1σ and 2σ levels for the final 20,000 steps undertaken by each of the 500 walkers. Results from SExtractor derived results are shown in blue while ProFound derived results are shown in red. Contours are shown for the runs that include the use of *Spitzer* photometry and correct for Eddington bias through the mass scatter measured by perturbing the photometry in the SED fitting process.

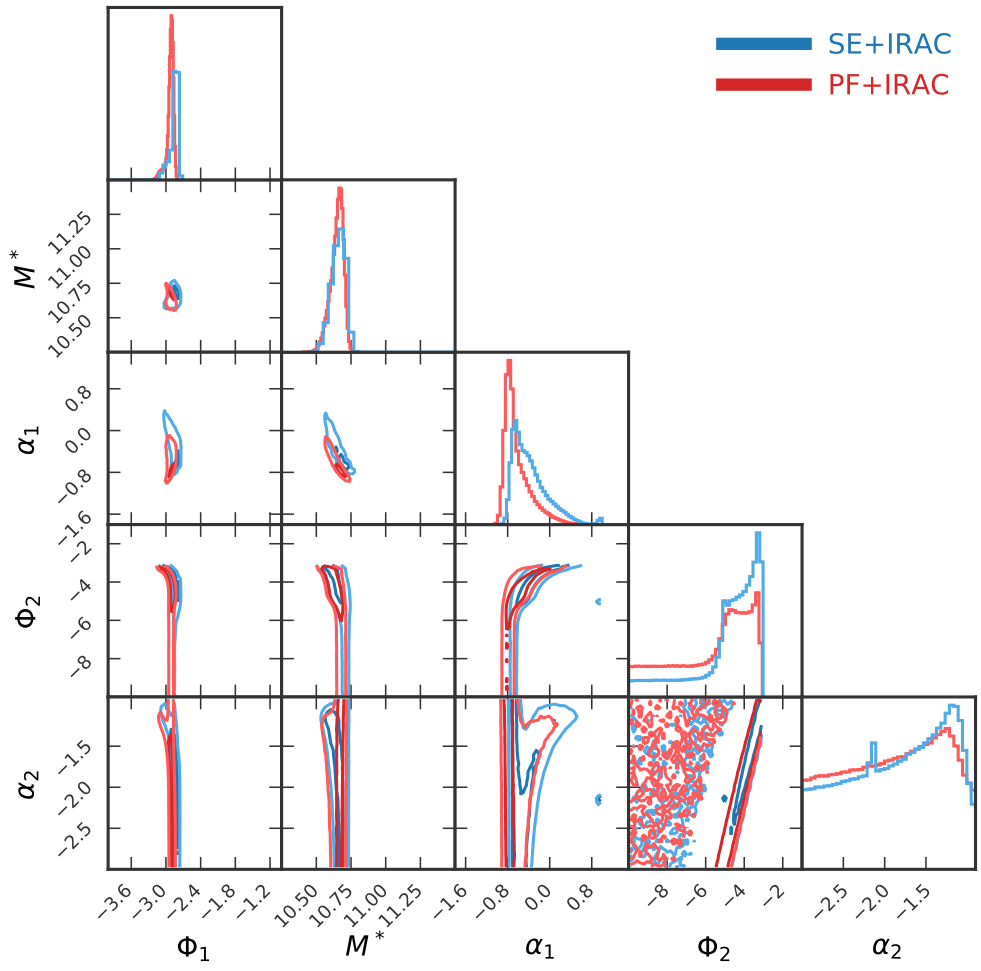


Figure 2.17: The posterior distributions generated by the MCMC fitting procedure for the $1.5 < z < 1.75$ redshift bin. Contours shown are the 2-dimensional 1σ and 2σ levels for the final 20,000 steps undertaken by each of the 500 walkers. Results from SExtractor derived results are shown in blue while ProFound derived results are shown in red. Contours are shown for the runs that include the use of *Spitzer* photometry and correct for Eddington bias through the mass scatter measured by perturbing the photometry in the SED fitting process.

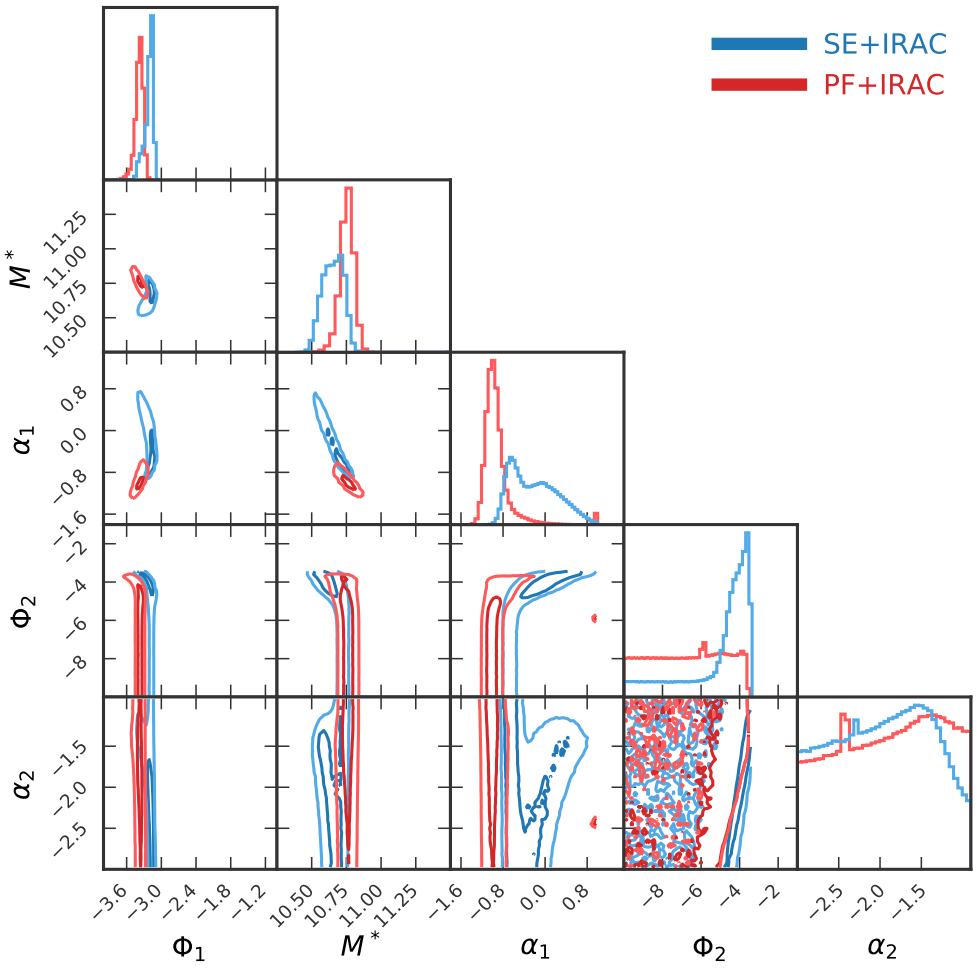


Figure 2.18: The posterior distributions generated by the MCMC fitting procedure for the $1.75 < z < 2.0$ redshift bin. Contours shown are the 2-dimensional 1σ and 2σ levels for the final 20,000 steps undertaken by each of the 500 walkers. Results from SExtractor derived results are shown in blue while ProFound derived results are shown in red. Contours are shown for the runs that include the use of *Spitzer* photometry and correct for Eddington bias through the mass scatter measured by perturbing the photometry in the SED fitting process.

Table 2.5: The median and standard deviations of the MCMC samples for the double Schechter function when using SExtractor based total photometry. Priors are set to ensure that α_1 & $\log_{10}(\Phi_1)$ refer to the high mass component and α_2 & $\log_{10}(\Phi_2)$ refer to the low mass component. Alongside are the number of galaxies from each of the two fields that contribute to the GSMF as shown in Fig. 2.9. These mass functions are based on the raw observations and are not corrected for Eddington bias.

Redshift SExtractor	$\log_{10}(M^*/M_\odot)$	α_1	$\log_{10}(\Phi_1)$ [mag^{-1}Mpc^{-3}]	α_2	$\log_{10}(\Phi_2)$ [mag^{-1}Mpc^{-3}]	COSMOS	XMM
0.1-0.2	$10.74^{+0.12}_{-0.12}$	$-0.72^{+0.37}_{-0.31}$	$-2.61^{+0.12}_{-0.15}$	$-1.63^{+0.12}_{-0.18}$	$-3.48^{+0.31}_{-0.59}$	3099	4269
0.2-0.3	$10.95^{+0.06}_{-0.12}$	$-1.24^{+0.46}_{-0.07}$	$-2.89^{+0.11}_{-0.10}$	$-1.61^{+0.29}_{-0.69}$	$-5.31^{+2.02}_{-3.11}$	5423	7011
0.3-0.5	$10.90^{+0.04}_{-0.04}$	$-0.94^{+0.15}_{-0.10}$	$-2.69^{+0.05}_{-0.05}$	$-1.85^{+0.25}_{-0.34}$	$-4.30^{+0.62}_{-0.86}$	15988	21292
0.5-0.75	$11.11^{+0.03}_{-0.03}$	$-1.23^{+0.04}_{-0.03}$	$-2.97^{+0.04}_{-0.05}$	$-1.81^{+0.61}_{-0.77}$	$-6.94^{+2.07}_{-2.04}$	26864	30713
0.75-1.0	$10.96^{+0.03}_{-0.03}$	$-1.01^{+0.13}_{-0.07}$	$-2.78^{+0.03}_{-0.04}$	$-2.13^{+0.39}_{-0.49}$	$-4.77^{+0.82}_{-1.01}$	33588	35190
1.0-1.25	$10.96^{+0.04}_{-0.04}$	$-1.09^{+0.9}_{-0.09}$	$-2.99^{+0.05}_{-0.06}$	$-2.41^{+0.64}_{-0.41}$	$-5.60^{+0.97}_{-0.83}$	25752	21395
1.25-1.5	$10.87^{+0.02}_{-0.02}$	$-1.08^{+0.05}_{-0.04}$	$-2.99^{+0.03}_{-0.04}$	$-1.81^{+0.64}_{-0.74}$	$-6.94^{+2.03}_{-2.08}$	18876	21575
1.5-1.75	$10.85^{+0.03}_{-0.03}$	$-0.94^{+0.07}_{-0.06}$	$-2.99^{+0.04}_{-0.05}$	$-1.81^{+0.66}_{-0.74}$	$-7.23^{+2.15}_{-1.89}$	13907	14130
1.75-2.0	$10.91^{+0.04}_{-0.04}$	$-1.01^{+0.13}_{-0.10}$	$-3.43^{+0.05}_{-0.07}$	$-1.87^{+0.67}_{-0.70}$	$-6.94^{+1.82}_{-2.08}$	9570	3778
SExtractor+IRAC							
0.1-0.2	$10.83^{+0.10}_{-0.11}$	$-0.72^{+0.38}_{-0.31}$	$-2.73^{+0.11}_{-0.13}$	$-1.60^{+0.06}_{-0.10}$	$-3.51^{+0.29}_{-0.55}$	3031	4264
0.2-0.3	$10.98^{+0.05}_{-0.08}$	$-1.20^{+0.27}_{-0.06}$	$-2.85^{+0.09}_{-0.09}$	$-1.69^{+0.37}_{-0.70}$	$-5.64^{+2.12}_{-2.77}$	5674	7239
0.3-0.5	$10.94^{+0.04}_{-0.04}$	$-0.88^{+0.15}_{-0.11}$	$-2.69^{+0.05}_{-0.05}$	$-1.76^{+0.07}_{-0.09}$	$-4.09^{+0.48}_{-0.63}$	15568	21246
0.5-0.75	$10.98^{+0.03}_{-0.03}$	$-1.09^{+0.12}_{-0.07}$	$-2.83^{+0.05}_{-0.05}$	$-2.02^{+0.55}_{-0.59}$	$-5.24^{+1.33}_{-1.61}$	25900	30209
0.75-1.0	$10.96^{+0.03}_{-0.03}$	$-1.01^{+0.13}_{-0.07}$	$-2.78^{+0.04}_{-0.04}$	$-2.10^{+0.38}_{-0.48}$	$-4.71^{+0.81}_{-1.00}$	31657	33791
1.0-1.25	$10.85^{+0.03}_{-0.03}$	$-0.84^{+0.17}_{-0.09}$	$-2.83^{+0.03}_{-0.04}$	$-2.12^{+0.39}_{-0.48}$	$-4.52^{+0.71}_{-0.76}$	25792	22288
1.25-1.5	$10.81^{+0.03}_{-0.03}$	$-0.82^{+0.28}_{-0.13}$	$-2.92^{+0.04}_{-0.08}$	$-2.07^{+0.62}_{-0.58}$	$-4.67^{+1.04}_{-0.90}$	18482	21133
1.5-1.75	$10.82^{+0.02}_{-0.02}$	$-0.85^{+0.09}_{-0.06}$	$-2.95^{+0.03}_{-0.04}$	$-1.86^{+0.68}_{-0.71}$	$-6.74^{+1.87}_{-2.22}$	13747	14020
1.75-2.0	$10.83^{+0.04}_{-0.04}$	$-0.84^{+0.20}_{-0.12}$	$-3.29^{+0.04}_{-0.06}$	$-1.92^{+0.70}_{-0.68}$	$-6.44^{+1.78}_{-2.42}$	9221	4476

Table 2.6: The median and standard deviations of the MCMC samples for the double Schechter function when using ProFound based total photometry. Priors are set to ensure that α_1 & $\log_{10}(\Phi_1)$ refer to the high mass component and α_2 & $\log_{10}(\Phi_2)$ refer to the low mass component. Alongside are the number of galaxies from each of the two fields that contribute to the GSMF as shown in Fig. 2.9. These mass functions are based on the raw observations and are not corrected for Eddington bias.

Redshift ProFound	$\log_{10}(M^*/M_\odot)$	α_1	$\log_{10}(\Phi_1)$ [mag^{-1}Mpc^{-3}]	α_2	$\log_{10}(\Phi_2)$ [mag^{-1}Mpc^{-3}]	COSMOS	XMM
0.1-0.2	$10.88^{+0.23}_{-0.19}$	$-1.06^{+0.42}_{-0.31}$	$-2.81^{+0.21}_{-0.30}$	$-1.76^{+0.20}_{-0.39}$	$-4.06^{+0.66}_{-2.29}$	3176	4062
0.2-0.3	$11.01^{+0.05}_{-0.07}$	$-1.30^{+0.11}_{-0.05}$	$-2.99^{+0.09}_{-0.09}$	$-1.73^{+0.49}_{-0.72}$	$-6.47^{+2.52}_{-2.34}$	5384	6886
0.3-0.5	$10.83^{+0.04}_{-0.04}$	$-0.90^{+0.16}_{-0.11}$	$-2.66^{+0.05}_{-0.05}$	$-1.88^{+0.22}_{-0.29}$	$-4.20^{+0.53}_{-0.68}$	16073	21130
0.5-0.75	$11.01^{+0.04}_{-0.04}$	$-1.22^{+0.07}_{-0.04}$	$-2.92^{+0.06}_{-0.06}$	$-2.14^{+0.79}_{-0.62}$	$-6.23^{+1.50}_{-2.20}$	27059	38937
0.75-1.0	$10.86^{+0.03}_{-0.03}$	$-0.94^{+0.13}_{-0.08}$	$-2.72^{+0.03}_{-0.04}$	$-2.11^{+0.26}_{-0.37}$	$-4.45^{+0.57}_{-0.67}$	33820	33885
1.0-1.25	$10.85^{+0.03}_{-0.04}$	$-0.92^{+0.15}_{-0.09}$	$-2.89^{+0.04}_{-0.04}$	$-2.35^{+0.37}_{-0.38}$	$-4.70^{+0.57}_{-0.65}$	25894	28041
1.25-1.5	$10.77^{+0.04}_{-0.04}$	$-0.82^{+0.24}_{-0.12}$	$-2.91^{+0.04}_{-0.05}$	$-2.30^{+0.46}_{-0.47}$	$-4.54^{+0.71}_{-0.62}$	18959	14280
1.5-1.75	$10.78^{+0.03}_{-0.03}$	$-0.84^{+0.11}_{-0.07}$	$-2.89^{+0.03}_{-0.04}$	$-1.90^{+0.70}_{-0.71}$	$-6.46^{+1.86}_{-2.41}$	13962	15248
1.75-2.0	$10.82^{+0.05}_{-0.06}$	$-0.83^{+0.44}_{-0.16}$	$-3.33^{+0.05}_{-0.08}$	$-2.06^{+0.72}_{-0.60}$	$-5.49^{+1.33}_{-2.84}$	9595	4140
ProFound+IRAC							
0.1-0.2	$10.99^{+0.14}_{-0.14}$	$-0.96^{+0.42}_{-0.31}$	$-2.83^{+0.16}_{-0.19}$	$-1.66^{+0.16}_{-0.33}$	$-3.81^{+0.49}_{-1.52}$	2941	4120
0.2-0.3	$10.99^{+0.06}_{-0.08}$	$-1.19^{+0.28}_{-0.07}$	$-2.84^{+0.10}_{-0.09}$	$-1.71^{+0.38}_{-0.70}$	$-5.45^{+1.98}_{-2.73}$	5600	7093
0.3-0.5	$10.93^{+0.04}_{-0.04}$	$-0.89^{+0.15}_{-0.10}$	$-2.68^{+0.05}_{-0.05}$	$-1.89^{+0.15}_{-0.10}$	$-4.32^{+0.50}_{-0.70}$	15556	20820
0.5-0.75	$10.92^{+0.03}_{-0.03}$	$-1.07^{+0.09}_{-0.07}$	$-2.78^{+0.04}_{-0.05}$	$-2.29^{+0.58}_{-0.48}$	$-5.38^{+1.08}_{-0.96}$	26014	38403
0.75-1.0	$10.89^{+0.03}_{-0.03}$	$-0.94^{+0.12}_{-0.08}$	$-2.74^{+0.03}_{-0.03}$	$-2.17^{+0.30}_{-0.38}$	$-4.60^{+0.58}_{-0.75}$	31848	33188
1.0-1.25	$10.79^{+0.03}_{-0.03}$	$-0.81^{+0.15}_{-0.10}$	$-2.84^{+0.03}_{-0.04}$	$-2.30^{+0.34}_{-0.40}$	$-4.62^{+0.59}_{-0.59}$	25860	27097
1.25-1.5	$10.77^{+0.04}_{-0.04}$	$-0.80^{+0.26}_{-0.12}$	$-2.96^{+0.04}_{-0.06}$	$-2.21^{+0.44}_{-0.51}$	$-4.66^{+0.77}_{-0.59}$	18539	12523
1.5-1.75	$10.79^{+0.03}_{-0.03}$	$-0.96^{+0.10}_{-0.07}$	$-3.04^{+0.04}_{-0.05}$	$-1.88^{+0.69}_{-0.68}$	$-6.77^{+1.77}_{-2.20}$	13785	11226
1.75-2.0	$10.88^{+0.04}_{-0.04}$	$-1.17^{+0.14}_{-0.10}$	$-3.52^{+0.06}_{-0.08}$	$-1.86^{+0.65}_{-0.69}$	$-6.89^{+1.86}_{-2.11}$	9246	2898

Table 2.7: The median and standard deviations of the MCMC samples for the double Schechter function when corrected for Eddington bias by convolving with a Gaussian distribution using a width of $\sigma_M = 0.09$. Priors are set to ensure that α_1 & $\log_{10}(\Phi_1)$ refer to the high mass component and α_2 & $\log_{10}(\Phi_2)$ refer to the low mass component.

Redshift SExtractor+IRAC	$\log_{10}(M^*/M_\odot)$	α_1	$\log_{10}(\Phi_1)$ [mag^{-1}Mpc^{-3}]	α_2	$\log_{10}(\Phi_2)$ [mag^{-1}Mpc^{-3}]
0.1-0.2	$10.78^{+0.10}_{-0.11}$	$-0.63^{+0.40}_{-0.33}$	$-2.68^{+0.11}_{-0.13}$	$-1.61^{+0.10}_{-0.14}$	$-3.44^{+0.26}_{-0.47}$
0.2-0.3	$10.95^{+0.06}_{-0.10}$	$-1.17^{+0.39}_{-0.08}$	$-2.82^{+0.10}_{-0.09}$	$-1.65^{+0.32}_{-0.69}$	$-5.07^{+1.19}_{-3.05}$
0.3-0.5	$10.90^{+0.04}_{-0.05}$	$-0.79^{+0.18}_{-0.13}$	$-2.66^{+0.05}_{-0.05}$	$-1.69^{+0.17}_{-0.22}$	$-3.86^{+0.39}_{-0.55}$
0.5-0.75	$10.93^{+0.04}_{-0.04}$	$-1.01^{+0.18}_{-0.09}$	$-2.78^{+0.05}_{-0.06}$	$-1.94^{+0.44}_{-0.58}$	$-4.60^{+1.03}_{-1.42}$
0.75-1.0	$10.91^{+0.03}_{-0.04}$	$-0.92^{+0.18}_{-0.10}$	$-2.74^{+0.04}_{-0.04}$	$-1.93^{+0.31}_{-0.46}$	$-4.28^{+0.67}_{-0.97}$
1.0-1.25	$10.79^{+0.03}_{-0.04}$	$-0.68^{+0.24}_{-0.15}$	$-2.79^{+0.03}_{-0.04}$	$-1.90^{+0.31}_{-0.45}$	$-4.02^{+0.52}_{-0.76}$
1.25-1.5	$10.74^{+0.04}_{-0.05}$	$-0.60^{+0.39}_{-0.20}$	$-2.88^{+0.04}_{-0.11}$	$-1.79^{+0.43}_{-0.67}$	$-3.94^{+0.63}_{-1.06}$
1.5-1.75	$10.77^{+0.02}_{-0.03}$	$-0.78^{+0.11}_{-0.07}$	$-2.91^{+0.03}_{-0.04}$	$-1.85^{+0.66}_{-0.76}$	$-6.34^{+1.95}_{-2.48}$
1.75-2.0	$10.78^{+0.04}_{-0.05}$	$-0.74^{+0.32}_{-0.14}$	$-3.25^{+0.04}_{-0.06}$	$-1.91^{+0.67}_{-0.71}$	$-5.91^{+1.67}_{-2.77}$
ProFound+IRAC					
0.1-0.2	$10.94^{+0.15}_{-0.14}$	$-0.88^{+0.45}_{-0.35}$	$-2.79^{+0.15}_{-0.20}$	$-1.64^{+0.14}_{-0.28}$	$-3.69^{+0.42}_{-1.27}$
0.2-0.3	$10.96^{+0.06}_{-0.10}$	$-1.15^{+0.39}_{-0.09}$	$-2.81^{+0.11}_{-0.10}$	$-1.67^{+0.33}_{-0.69}$	$-4.94^{+1.66}_{-2.93}$
0.3-0.5	$10.89^{+0.04}_{-0.04}$	$-0.81^{+0.16}_{-0.12}$	$-2.64^{+0.04}_{-0.05}$	$-1.80^{+0.20}_{-0.25}$	$-4.08^{+0.47}_{-0.62}$
0.5-0.75	$10.88^{+0.03}_{-0.03}$	$-1.01^{+0.13}_{-0.07}$	$-2.74^{+0.04}_{-0.05}$	$-2.21^{+0.50}_{-0.50}$	$-4.97^{+1.02}_{-1.01}$
0.75-1.0	$10.84^{+0.03}_{-0.03}$	$-0.83^{+0.15}_{-0.11}$	$-2.70^{+0.03}_{-0.04}$	$-1.99^{+0.24}_{-0.34}$	$-4.21^{+0.49}_{-0.66}$
1.0-1.25	$10.74^{+0.03}_{-0.04}$	$-0.69^{+0.19}_{-0.13}$	$-2.80^{+0.03}_{-0.03}$	$-2.14^{+0.33}_{-0.41}$	$-4.28^{+0.51}_{-0.65}$
1.25-1.5	$10.71^{+0.03}_{-0.04}$	$0.66^{+0.28}_{-0.15}$	$-2.92^{+0.04}_{-0.05}$	$-2.08^{+0.45}_{-0.54}$	$-4.19^{+0.61}_{-0.75}$
1.5-1.75	$10.74^{+0.03}_{-0.03}$	$-0.90^{+0.10}_{-0.07}$	$-2.99^{+0.04}_{-0.04}$	$-1.84^{+0.65}_{-0.76}$	$-6.71^{+2.03}_{-2.23}$
1.75-2.0	$10.83^{+0.04}_{-0.04}$	$-1.11^{+0.13}_{-0.11}$	$-3.47^{+0.06}_{-0.07}$	$-1.82^{+0.63}_{-0.76}$	$-6.96^{+2.05}_{-2.07}$

Table 2.8: The median and standard deviations of the MCMC samples for the double Schechter function when corrected for Eddington bias through the scattering of photometry in the photo-z process, leading to a convolution with a Gaussian × Lorentzian functional form with a strength of $\sigma_M = 0.50$. Priors are set to ensure that α_1 & $\log_{10}(\Phi_1)$ refer to the high mass component and α_2 & $\log_{10}(\Phi_2)$ refer to the low mass component.

Redshift SExtractor+IRAC	$\log_{10}(M^*/M_\odot)$	α_1	$\log_{10}(\Phi_1)$ [mag^{-1}Mpc^{-3}]	α_2	$\log_{10}(\Phi_2)$ [mag^{-1}Mpc^{-3}]
0.1-0.2	$10.73^{+0.11}_{-0.11}$	$-0.53^{+0.42}_{-0.35}$	$-2.64^{+0.10}_{-0.13}$	$-1.60^{+0.10}_{-0.13}$	$-3.38^{+0.24}_{-0.40}$
0.2-0.3	$10.91^{+0.07}_{-0.12}$	$-1.13^{+0.51}_{-0.11}$	$-2.79^{+0.11}_{-0.10}$	$-1.61^{+0.28}_{-0.66}$	$-4.56^{+1.39}_{-3.21}$
0.3-0.5	$10.85^{+0.05}_{-0.05}$	$-0.70^{+0.21}_{-0.16}$	$-2.62^{+0.05}_{-0.05}$	$-1.64^{+0.15}_{-0.20}$	$-3.70^{+0.34}_{-0.49}$
0.5-0.75	$10.88^{+0.04}_{-0.05}$	$-0.88^{+0.29}_{-0.15}$	$-2.74^{+0.05}_{-0.06}$	$-1.76^{+0.31}_{-0.51}$	$3.98^{+0.68}_{-1.18}$
0.75-1.0	$10.85^{+0.04}_{-0.05}$	$-0.75^{+0.25}_{-0.18}$	$-2.70^{+0.04}_{-0.04}$	$-1.72^{+0.22}_{-0.37}$	$-3.76^{+0.44}_{-0.78}$
1.0-1.25	$10.70^{+0.05}_{-0.05}$	$-0.33^{+0.29}_{-0.54}$	$-2.77^{+0.04}_{-0.07}$	$-1.62^{+0.20}_{-0.33}$	$-3.49^{+0.29}_{-0.54}$
1.25-1.5	$10.64^{+0.05}_{-0.05}$	$-0.16^{+0.25}_{-0.77}$	$-2.88^{+0.08}_{-0.11}$	$-1.48^{+0.21}_{-0.56}$	$-3.38^{+0.25}_{-0.77}$
1.5-1.75	$10.69^{+0.04}_{-0.05}$	$-0.58^{+0.36}_{-0.16}$	$-2.86^{+0.04}_{-0.07}$	$-1.78^{+0.59}_{-0.79}$	$-4.68^{+1.20}_{-2.97}$
1.75-2.0	$10.69^{+0.06}_{-0.08}$	$-0.40^{+0.56}_{-0.33}$	$-3.22^{+0.04}_{-0.09}$	$-1.95^{+0.62}_{-0.68}$	$-4.57^{+0.82}_{-3.04}$
ProFound+IRAC					
0.1-0.2	$10.90^{+0.15}_{-0.15}$	$-0.80^{+0.48}_{-0.38}$	$-2.74^{+0.15}_{-0.20}$	$-1.63^{+0.13}_{-0.25}$	$-3.60^{+0.37}_{-1.06}$
0.2-0.3	$10.91^{+0.08}_{-0.12}$	$-1.08^{+0.52}_{-0.15}$	$-2.78^{+0.12}_{-0.11}$	$-1.62^{+0.28}_{-0.66}$	$-4.37^{1.21}_{-3.05}$
0.3-0.5	$10.85^{+0.05}_{-0.05}$	$-0.73^{+0.20}_{-0.15}$	$-2.60^{+0.05}_{-0.05}$	$-1.73^{+0.18}_{-0.23}$	$-3.89^{+0.41}_{-0.56}$
0.5-0.75	$10.83^{+0.04}_{-0.05}$	$-0.91^{+0.22}_{-0.10}$	$-2.69^{+0.04}_{-0.05}$	$-2.02^{+0.43}_{-0.54}$	$-4.39^{+0.87}_{-1.10}$
0.75-1.0	$10.77^{+0.04}_{-0.04}$	$-0.63^{+0.21}_{-0.17}$	$-2.65^{+0.03}_{-0.04}$	$-1.80^{+0.19}_{-0.26}$	$-3.75^{+0.34}_{-0.52}$
1.0-1.25	$10.65^{+0.04}_{-0.05}$	$-0.39^{+0.30}_{-0.56}$	$-2.76^{+0.03}_{-0.05}$	$-1.86^{+0.24}_{-0.33}$	$-3.74^{+0.35}_{-0.56}$
1.25-1.5	$10.60^{+0.05}_{-0.05}$	$-0.23^{+0.31}_{-0.19}$	$-2.90^{+0.05}_{-0.10}$	$-1.70^{+0.27}_{-0.47}$	$-3.56^{+0.31}_{-0.63}$
1.5-1.75	$10.68^{+0.03}_{-0.05}$	$-0.76^{+0.27}_{-0.11}$	$-2.94^{+0.04}_{-0.05}$	$-1.87^{+0.64}_{-0.73}$	$-5.44^{+1.58}_{-3.00}$
1.75-2.0	$10.78^{+0.05}_{-0.05}$	$-1.02^{+0.18}_{-0.13}$	$-3.41^{+0.06}_{-0.07}$	$-1.85^{+0.64}_{-0.73}$	$-6.65^{+2.02}_{-2.26}$

2.5 Results

2.5.1 Changes in the GSMF with varying methodology

Within each redshift bin, I measure a total of four stellar mass functions. First I compare the results with and without the inclusion of *Spitzer*/IRAC [3.6] and [4.5] data, and second I compare the GSMF derived from SExtractor based photometry in comparison to that derived with Profound. While the low mass component of the GSMF is consistent between these different methods, I find some differences in the results at stellar masses greater than M^*.

The impact of including *Spitzer*/IRAC photometry on the measured GSMF

The most immediately apparent feature is the offset between the GSMFs that include or exclude *Spitzer*/IRAC data in the redshift bins of $0.5 < z < 0.75$ and $1.0 < z < 1.25$. This is due to two effects. Firstly, many high-mass objects that lie in the $0.5 < z < 0.75$ bin have a different redshift solution ($z \sim 0.1$) when *Spitzer*/IRAC is included. They consequently have smaller masses in this lower redshift bin and their influence is negated by the high number densities within this mass-redshift range. The cause of the different redshift solutions is the uncertainty of the SED slope redder than the K_s band, with a red slope favouring the higher redshift solutions and a blue slope favouring low redshift solutions. Secondly, there is a likely contamination from stars in the $0.5 < z < 0.75$ and $1.0 < z < 1.25$ bins. Many high mass objects have smooth red optical slopes that turn over around the H or K_s bands. With a limited wavelength range, some black body-like spectra and red galaxies are indistinguishable. Introducing the *Spitzer*/IRAC bands vastly increases the χ^2 of the galaxy models and reduces the χ^2 of the stellar models for a significant number of these high mass objects. Consequently these objects no longer meet my selection criteria (either the χ^2 increases above 250 or the classification switches to a star) when using the values associated with the inclusion of the mid-infrared bands. Both of these cases are examples of degeneracies between template sets as a result of the use of a limited number of broadband filters. Thus

I find that the inclusion of the [3.6] and [4.5] bands makes a significant difference to the derived number density of the most massive galaxies in my data. Therefore, I focus on the GSMFs measured with the inclusion of the *Spitzer*/IRAC when discussing redshift evolution and comparisons to simulations.

The impact of varying the choice of source extraction software

While the global mass distributions between SExtractor and ProFound based catalogues are broadly very similar, there are individual cases where mass estimates can vary widely between objects. Issues can typically be put down to artefacts affecting the photometry, proximity to bright sources and disagreement between the two source extraction measurements in the *Spitzer* bands. Instances of significant differences in mass estimations tend to occur in high redshift and/or low mass objects that fall within my incomplete regime and so do not affect my final results.

At low redshifts, galaxies become increasingly resolved and so any systematic variation between the SExtractor and ProFound photometry would be expected to be more apparent. In my measurements I find any such variation between the two to be very small, with the low-mass components ($M < M^*$) between $0.1 < z < 1.5$ being highly consistent between the two source extraction methods. I find that at the lowest redshift ($0.1 < z < 0.2$) the ProFound based GSMF produces more galaxies of very high mass compared to the SExtractor derived measurements. This is then reversed at higher redshifts $z > 0.5$, where the ProFound GSMF produces fewer galaxies of very high mass. However, these differences are of relatively low significance (of order 1σ) and demonstrate that over the redshift and mass ranges probed in this study, the choice of source extraction software makes no significant difference to GSMF results.

2.5.2 The intrinsic GSMF corrected for Eddington bias

To recover the intrinsic GSMF from my observations, which is affected by the uncertainties in the estimate of stellar mass, I trial three different methods described in Section 2.4.4. In the first method, I assume that the scatter induced by

uncertainties on stellar mass are described by a Gaussian distribution of $\sigma_M = 0.09$. I find that this imposes minimal changes to the shape of the high-mass end of the MF, corresponding to a shift in M^* of order 0.05 dex lower when compared to the observed GSMF. With the second method, where I convolve a Gaussian $\times$ Lorentzian combination with $\sigma_M = 0.5$, I find the shift between the observed and intrinsic parameters to be more significant with M^* shifting to lower masses by ~ 0.1 dex compared to the observed GSMF. For completeness, I display the MCMC results for these two methods in Tables 2.7 and 2.8 respectively. Each of these methods make a fundamental assumption in that the scatter in the derivation of the stellar mass is symmetric in logarithmic stellar mass space. I find this assumption to work best when redshifts are confident e.g. if I conduct the photometry scatter procedure on just the mass calculations and assume the photometric redshift is correct or the objects have spectroscopic redshifts. However, when the uncertainty on the photometric redshifts is included, the measured distribution is found to be broader and more asymmetric. This results in the Gaussian $\times$ Lorentizan method underestimating the amount of scattering in certain regimes (small up-scatter and most of the down-scattering), even with the extended wings of the function. It is for this reason that I elect to instead use the measured distribution of mass scatter to convolve with the intrinsic double Schechter function (see Fig. 2.19 for an example of the impact).

The GSMF, when corrected with this distribution in mass scatter, undergoes a similar shift in the Schechter parameters to that of the Gaussian $\times$ Lorentzian. The shift in M^* is 0.12 dex compared to the observed GSMF and all the fit parameters lie within 1σ of the results found with the Gaussian $\times$ Lorentzian method. While the broad wings of the distribution of mass scatter are very small in probability beyond shifts of 0.3 dex (see Fig. 2.5), the nature of the GSMF spanning many orders of magnitude in number density around the knee requires these wings to be modelled in order to capture the impact of a small number of objects scattering into the less populated, high-mass bins. This highlights that the intrinsic GSMF is very sensitive to the strength of the Eddington bias correction and could be a key source

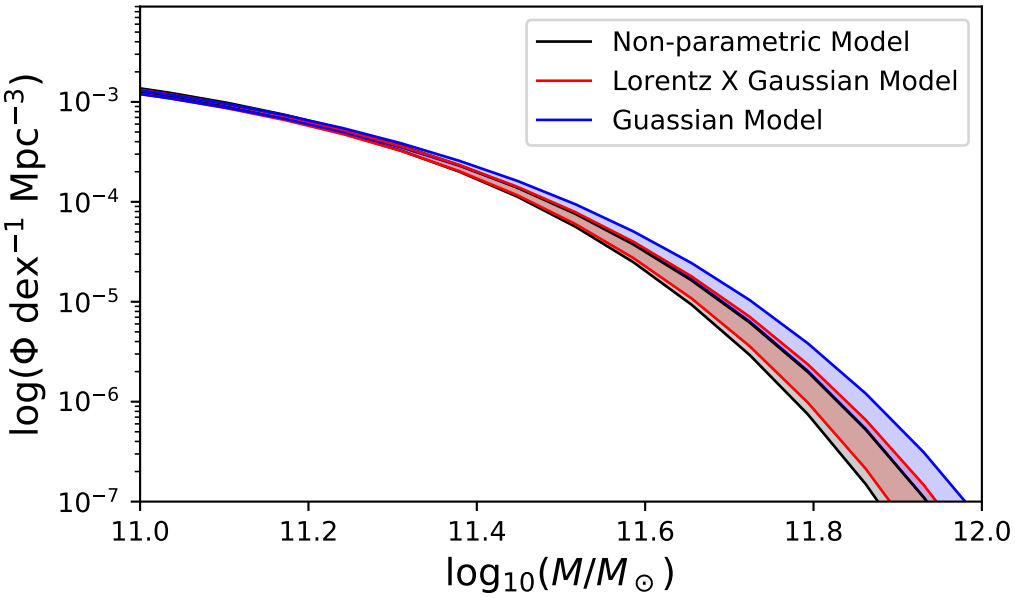


Figure 2.19: The impact that differing models of the Eddington Bias have on the measurement of the intrinsic GSMF at high masses. Shown is the intrinsic GSMF at $0.5 < z < 0.75$ as measured with SExtractor photometry when recovered using the three methods applied in this study. In blue I show the results of using a simple Gaussian to model mass errors, in red I expand the model to include Lorentzian wings and in black I show the results when using a non-parametric approach. Shaded regions indicate the 1 sigma uncertainty and are derived from 10,000 random samples of the posterior probability distribution. The edges of the shaded regions are made more bold to assist in readability.

of inconsistency between results of observational studies. The best-fit parameters for the double Schechter function when using my preferred non-parametric Eddington correction are shown in Table 2.4.

The use of high completeness spectroscopic surveys will enable for mass errors due to photometric redshift uncertainty to be drastically reduced. Surveys such as GAMA and SDSS have enabled for this to be applied to the very low redshift Universe ($z < 0.1$: Baldry et al., 2012; Weigel et al., 2016; Wright et al., 2017). Other spectroscopic surveys have made narrower but deeper observations in order to constrain the high mass end to higher redshifts (z<1.0: Pozzetti et al., 2010; Moustakas et al., 2013; Leauthaud et al., 2016). While such studies have been more limited to the brighter and more massive objects (e.g. $\log_{10}(M/M_\odot) > 10.5$), it is

these very objects that are most at risk at having their number counts inflated by Eddington Bias. With new surveys and instruments using multi-object spectrographs coming online in the coming years (e.g. DEVILS and WAVES), such studies will soon be capable of measuring the GSMF to lower masses and higher redshifts (Davies et al., 2018; Driver et al., 2019)

2.5.3 Comparisons to previous studies

To put this study into greater context I compare against a number of past studies. The studies selected for these comparisons are Davidzon et al. (2017a) which utilised only the COSMOS field in their study with the Laigle et al. (2016) catalogue, Wright et al. (2018) which uses a combination of the GAMA (Driver et al., 2011), COSMOS (Davies et al., 2015; Andrews et al., 2017) and 3D-HST surveys (Brammer et al., 2012; Skelton et al., 2014; Momcheva et al., 2016) and the recent study conducted by McLeod et al. (2020) that uses components of the COSMOS & XMM-LSS fields as well as *Hubble* CANDELS fields. The Davidzon et al. (2017a) sample consists of entirely photometric redshifts derived from of order 30 photometric bands (including narrowband filters) which extends to cover the near-ultraviolet (0.23μm) to *Spitzer* IRAC channel 4 (8μm), the derivation of these photometric redshifts is described in (Laigle et al., 2016). The Wright et al. (2018) sample uses redshifts from mixed sources, with some galaxies having full spectroscopic data, a fraction with grism spectra and photometric redshifts for those without spectra using up to 22 photometric filters through the inclusion of some narrowband filters and Hubble data. The sample from McLeod et al. (2020) uses entirely photometric redshifts derived from their own treatment (22 filters, mostly broadband filters with overlapping wavelength coverage from multiple telescopes) and from public CANDELS catalogues (Barro et al., 2019). Since each study utilises different data sets, the quality of redshifts, mass estimations and the impact of Eddington bias will be different in each case. The results of the the double Schechter fits conducted in these studies are presented alongside my own in Fig. 2.20 and all were calculated with the same cosmological model.

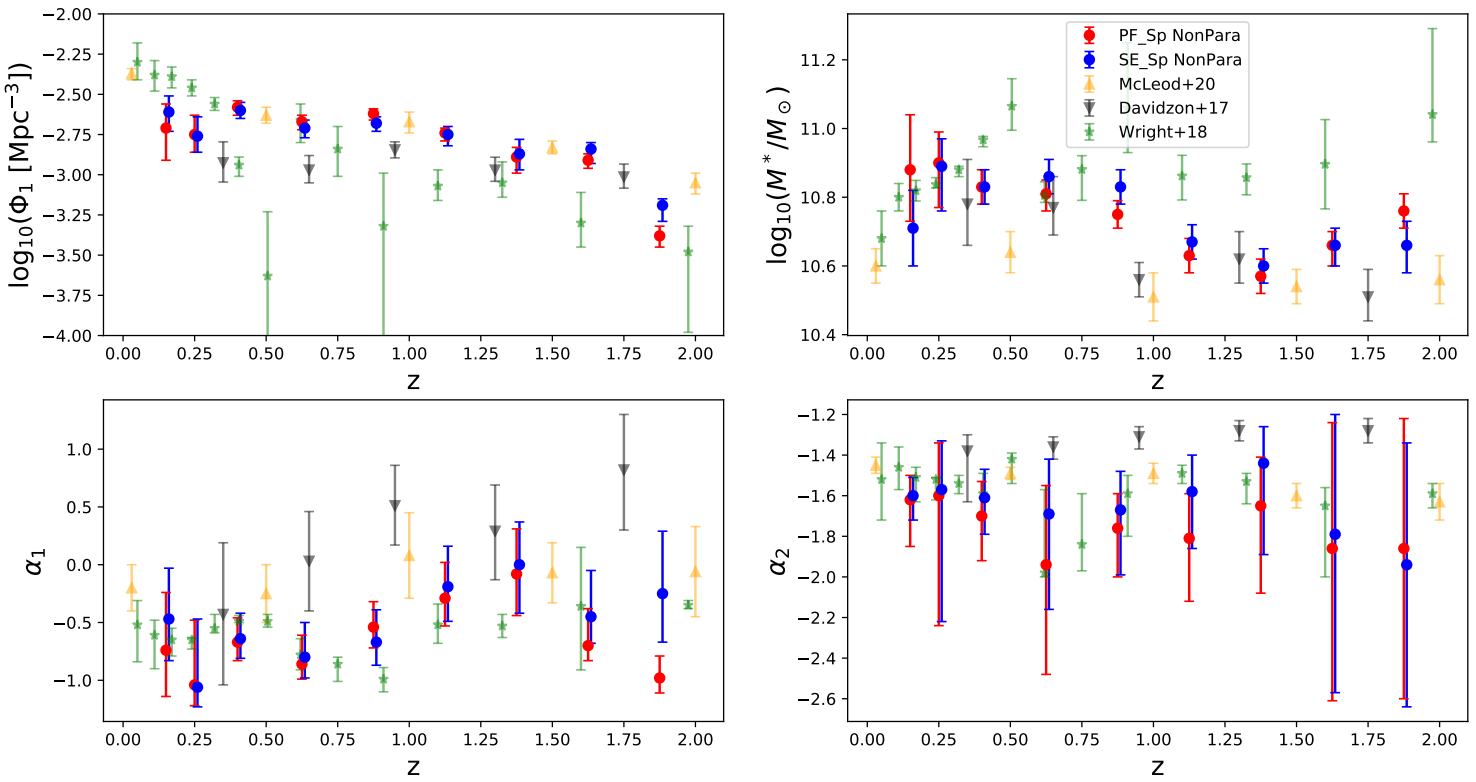


Figure 2.20: The time evolution of the double Schechter parameters as measured with photometry derived by SExtractor (blue) and ProFound (red) and including the use of *Spitzer*/IRAC photometry in both redshift and mass calculations. These data points are for the fits that utilise the non-parametric correction for Eddington bias. Since the parameter Φ_2 is largely unconstrained for a significant portion of the redshift range, I do not show those results in this figure. Alongside my results I display the results of Davidzon et al. (2017a); Wright et al. (2018); McLeod et al. (2020) in black, gold and green respectively. The lowest redshift data point from McLeod et al. (2020) is a modified value from Baldry et al. (2012)

DRAFT Printed on October 15, 2021

Examining the Schechter parameters individually, the strongest agreement is that of the high mass normalisation (Φ_1), where there is very close agreement between my results and those from McLeod et al. (2020). In contrast, the slope of the high mass component α_1 is found to agree more with the results from Wright et al. (2018), however this is subject to degeneracy based effects from my poorer capability to constrain the low-mass slope α_2. Inspection of the corner plots show that the more negative values of α_1 are driven by the steeper slopes found in my broad uncertainties of the low-mass slope (see e.g. Fig. 2.16).

The Eddington bias corrections implemented in previous work all use different functional forms. Wright et al. (2018) and McLeod et al. (2020) use corrections that are independent of mass and redshift as in my study, whereas Davidzon et al. (2017a) implement a minor redshift dependence in the Lorentzian terms. Furthermore, Wright et al. (2018) use a Gaussian of width $\sigma_M = 0.1$ dex while Ilbert et al. (2013); Davidzon et al. (2017a) use the Gaussian $\times$ Lorentzian distribution of widths 0.5 dex and 0.35 dex respectively. McLeod et al. (2020) uses a log-normal distribution of width $\sigma_M = 0.15$ dex to correct for the Eddington bias, leading to larger changes in their fit parameters, with M^* changing by up to 0.2 dex compared to the raw observations. Other examples of recent studies include Thorne et al. (2020) and Leja et al. (2020), which fold their mass uncertainties into their model fitting procedures. They respectively find a constant $M^* = 10.78$ and mildly evolving values between 10.8 and 10.9. The work conducted in Leja et al. (2020) is an example of a study that included redshift evolution terms within a model fitting procedure that did not bin by redshift. Such a methodology can result in a smoothing of the evolution of the model parameters, which has its pros and cons depending on the timescales examined and assumptions made (e.g. fixing the slope of α_1). For the propose of this study, I have elected to assume no particular evolutionary form for the GSMF.

These past studies, combined with my work trialling different functional forms to account for Eddington bias, suggest that up to 0.15 dex in variation of the characteristic mass M^* can be attributed to implementing different methodologies. While significant enough to cause changes of a few σ, it is not enough to explain the

~ 0.5 dex discrepancy among M^* values found between these studies, indicating that there are further systematics at play between the data sets.

2.6 Discussion

2.6.1 Time evolution of the high-mass component of the GSMF

Including *Spitzer*/IRAC photometry is found to be important in obtaining accurate galaxy classification and photometric redshifts, I therefore discuss the redshift evolution in the context of the GSMFs that make use of these data. I show the time evolution for the best-fit Schechter parameters $(\Phi_1, M^*, \alpha_1, \alpha_2)$, corrected for Eddington bias via my non-parametric method, in Fig. 2.20. Examining these best-fit Schechter parameters reveals evolution to be initially driven by the normalisation (Φ_1) of the double Schechter components, which increases with time. This evolution is found to be much stronger at higher redshifts, with a change of over 0.5 dex between $z = 2$ through to $z = 1.25$. For $z < 1.25$, the normalisation Φ_1 stabilises and the evolution of the mass function becomes dominated by evolution in M^* (there is also a possible flattening of α_1, but this is degenerate with α_2 which is poorly constrained at higher redshift).

Under the assumption that M^* is constant for $0 < z < 2$, the best fitting value of $\log_{10}(M^*/M_\odot) = 10.73 \pm 0.04$ is found for both SExtractor and ProFound derived photometry. However, I find that a constant M^* model gives a poor quality of fit, with $\chi^2_{\rm red} = 3.4\,(3.4)$ for SExtractor (ProFound). The observed rise in the value of M^* at $z < 1.25$ is present in all versions of the intrinsic GSMF that I produce with the differing Eddington corrections. I therefore fit a simple linear model to the evolution of M^* of the form $\log_{10}(M^*/M_\odot) = Az + B$, where A and B are free parameters. This fit is conducted with a simple minimisation of χ^2 method. Introducing this simple time dependence into M^* produces significantly better fits ($\chi^2_{\rm red} = 1.4\,(2.4)$ for SExtractor (ProFound)) with the evolution described by,

$$\frac{\Delta \log_{10}(M^*/M_\odot)}{\Delta z} = \begin{cases} -0.164 \pm 0.047 & \text{SExtractor} \\ -0.113 \pm 0.054 & \text{ProFound} \end{cases} \tag{2.4}$$

a 2–3σ disagreement with the evolution described by a constant M^*.

However, the observed evolution in M^* could also be attributed to a steeper evolution over a smaller redshift range, rather over the entire redshift range probed (see top-right panel of Fig. 2.20). Such an evolution could be indicative of a break in the balance of growth channels that maintains a constant M^* above $z \sim 1$ while Φ_1 increases. Two such channels include star formation and the rate of major mergers. Observations and simulations show that the contribution towards new stellar mass from both mechanisms falls at $z < 1.5$ in massive galaxies (Rodriguez-Gomez et al., 2015; Tomczak et al., 2016; Qu et al., 2017; O'Leary et al., 2020). The natural consequence of this would be to stabilise the high-mass end of the GSMF towards lower redshifts, as I find (e.g. the bottom right panel of Fig. 2.9 for $z < 0.75$).

2.6.2 The evolving quenched fraction of galaxies

The effects of the evolving in situ star formation rate and merger rates of galaxies results in the changing shape of the GSMF. In order to probe this further, I examine the population of massive galaxies over the redshift range $0.5 < z < 1.5$ in more detail. I separate my sample of massive galaxies into two broad specific star formation rate (SSFR) bins, based on the same SED template used to derive the stellar mass. I term these bins 'Star-forming' and 'Quenched', which are defined as bins of $-8.5 > \log_{10}(SSFR/\mathrm{yr}^{-1}) > -10.0$ and $-10.0 > \log_{10}(SSFR/\mathrm{yr}^{-1})$ respectively. The star forming bin is selected to cover the majority ($\sim$ 95 per cent) of the star forming main sequence at $z \sim 1.5$ as found in both observational and simulation based studies (e.g. Sparre et al., 2015; Tomczak et al., 2016). The quenched galaxies are thus defined as falling below this star forming main sequence. I do not enforce my definitions to be redshift dependent (i.e. a certain level above or below the evolving star-forming main sequence) to capture the global fall in star formation post-cosmic noon and produce results similar to those that use rest-frame colour selection (e.g. Davidzon et al., 2017a; McLeod et al., 2020). I

restrict the redshift range to $0.5 < z < 1.5$, as SSFR estimates for galaxies at redshifts $z < 0.5$ would become increasingly unreliable as the rest-frame UV exits the wavelength coverage of my observations.

I reproduce the GSMF for these two populations in Fig.2.21. My definitions for star-forming and passive galaxies exclude a small number of galaxies undergoing starbursts $-8.5 < \log_{10}(SSFR/\mathrm{yr}^{-1})$, and consequently the GSMFs in Fig.2.21 do not sum directly to the total GSMF presented in Fig.2.9. I find the change in total number density of galaxies with $M > M^*$ to be driven by a strong evolution in the passive population. As a consequence of this trend, the most massive component of the GSMF becomes increasingly dominated by more passive galaxies while the number densities of star-forming galaxies are more constant, agreeing with Davidzon et al. (2017a) and McLeod et al. (2020). Assuming a constant star formation rate, a borderline quenched galaxy with mass $M \geq M^*$, $\log_{10}(SSFR/\mathrm{yr}^{-1}) = -10$ at $z = 1.5$ would grow of order 0.15 dex in stellar mass during the ~ 4 Gyr that passes between $0.5 < z < 1.5$. The majority of the objects in my high mass sample will produce fewer stars than predicted from this overly simple estimation, as most start with SSFR values which are lower and there is a global trend for the SSFR to decrease with time. Consequently, the growth of galaxies through solely main sequence evolution is insufficient to produce the observed increase in high-mass galaxy number densities.

Since these quenched systems are not expected to form enough stars to move between bins of stellar mass ($\delta M \ll 0.25$ dex), the two primary ways of increasing the quenched number densities at high masses are consequently merger events and the addition of newly quenched systems taken from the star-forming population. As my star-forming number densities at high mass remain near constant, this implies that these galaxies must be replenished by slightly lower mass star-forming galaxies at approximately the same rate at which they undergo quenching. Major mergers ($\sim 1:4$ mass ratio) are capable of generating mass gains of order ~ 0.1 dex for passive galaxies on top of internal star formation. Mergers themselves can trigger starbursts in the aftermath of the event. However, at later times ($z < 1$), such

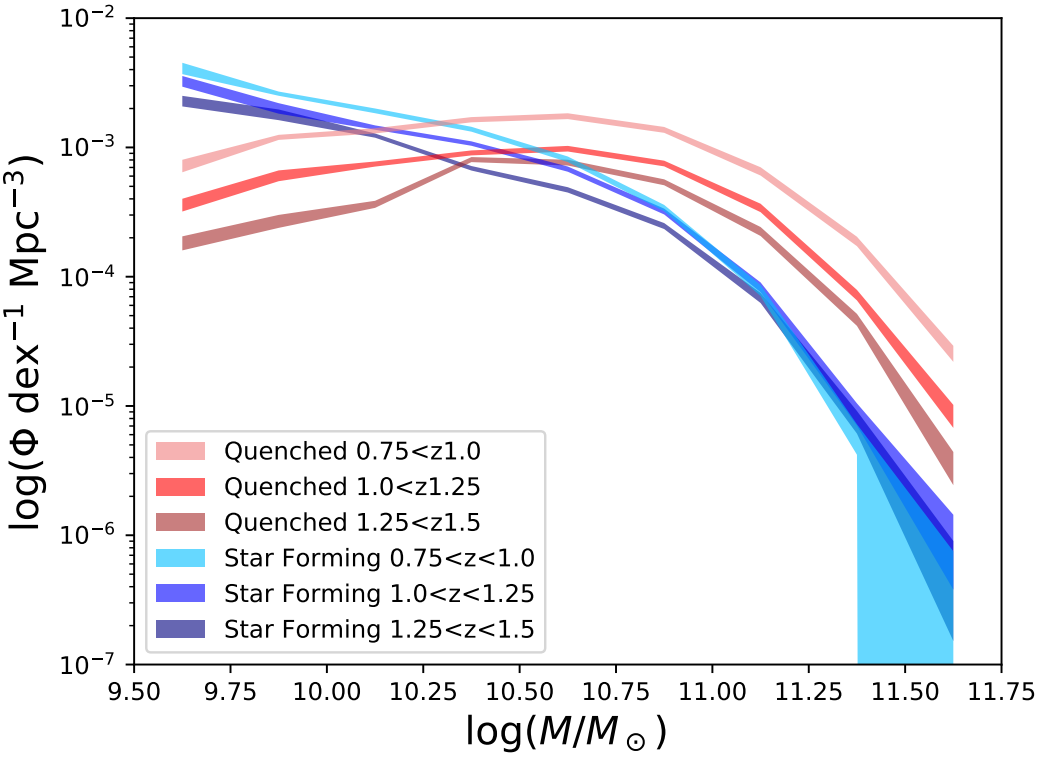


Figure 2.21: The GSMF when broken down into star forming and passive components by SSFR. The quenched population is shown with the red lines and the blue lines show the star-forming population. Darker shading indicates mass functions at higher redshifts. Shading widths indicates observational errors through both cosmic variance and V_{max}. The bin $0.5 < z < 0.75$ is not shown to reduce clutter and is found to be near identical to the bin $0.75 < z < 1.0$. The star-forming population is found to be near constant at higher masses while the passive population is found to evolve significantly between $1.5 < z < 0.75$.

mergers are increasingly 'dryer' as massive and gas poor galaxies tend towards residing in more dense environments and subsequently merge with each other (e.g. De Lucia & Blaizot, 2007; Lin et al., 2008; Davidzon et al., 2016; Tomczak et al., 2017). Merger events come at the cost of reducing the total number of galaxies and its effects will be embedded in the evolving shape of the GSMF. In addition to providing pathways for stellar mass growth of passive galaxies, merger events can also be used to solve issues surrounding the radial size of passive galaxies. At higher redshifts ($z \sim 1.5$), quenched galaxies are found to be compact relative to passive systems observed at $z = 0$ (e.g. Williams et al., 2010; Wuyts et al., 2011;

McLure et al., 2013). The passive nature of these galaxies requires that external processes must be responsible for the observed shift in galaxy radius. The study by McLure et al. (2013) finds that combinations of major and minor mergers can simultaneously reproduce both the growth in stellar mass and radius observed in these systems between $z \sim 1.5$ and today.

2.6.3 Comparison to simulations

In this section I compare my measured GSMF with the semi-analytic simulation (SAM) *SHARK* (Lagos et al., 2018) and the hydrodynamical simulations *Simba* (Davé et al., 2019) and EAGLE (Schaye et al., 2015). In Fig. 2.22 I show the GSMF results from these alongside my intrinsic (Eddington-bias corrected) double Schechter functions. The *SHARK*-SAM has the GSMF as one of physical measurements that it is tuned to reproduce at $z = 0$, 1 and 2. As a result it is unsurprising to find excellent agreement with my results at low and intermediate mass ranges ($\log_{10}(M^*/M_\odot) < 11.0$). However, there is an excess in the number density of the most massive objects ($\log_{10}(M/M_\odot) > 11.5$) in the simulation. At these high masses, I instead find that the hydrodynamical simulation EAGLE provides a greater match to observational results in this mass regime (discussed in more detail later in this section). With the close matching to all other components of the GSMF, the likely source for the high-mass discrepancy found in *SHARK* could be due to the choice of GSMF which was used to tune the simulation. The Lagos et al. (2018) models elect to calibrate against studies such as Muzzin et al. (2013) and Wright et al. (2018), all studies with M^* values on the higher end of the parameter space covered by observational studies ($10.8 < \log_{10}(M^*/M_\odot) > 11.0$). For my comparisons I utilise the GSMF derived from total masses provided from the primary SHARK run conducted in Lagos et al. (2018). Within that study, it is found that implementing a 30kpc aperture to measure stellar mass leads to minimal impact for galaxies with $\log_{10}(M/M_\odot) < 12.0$.

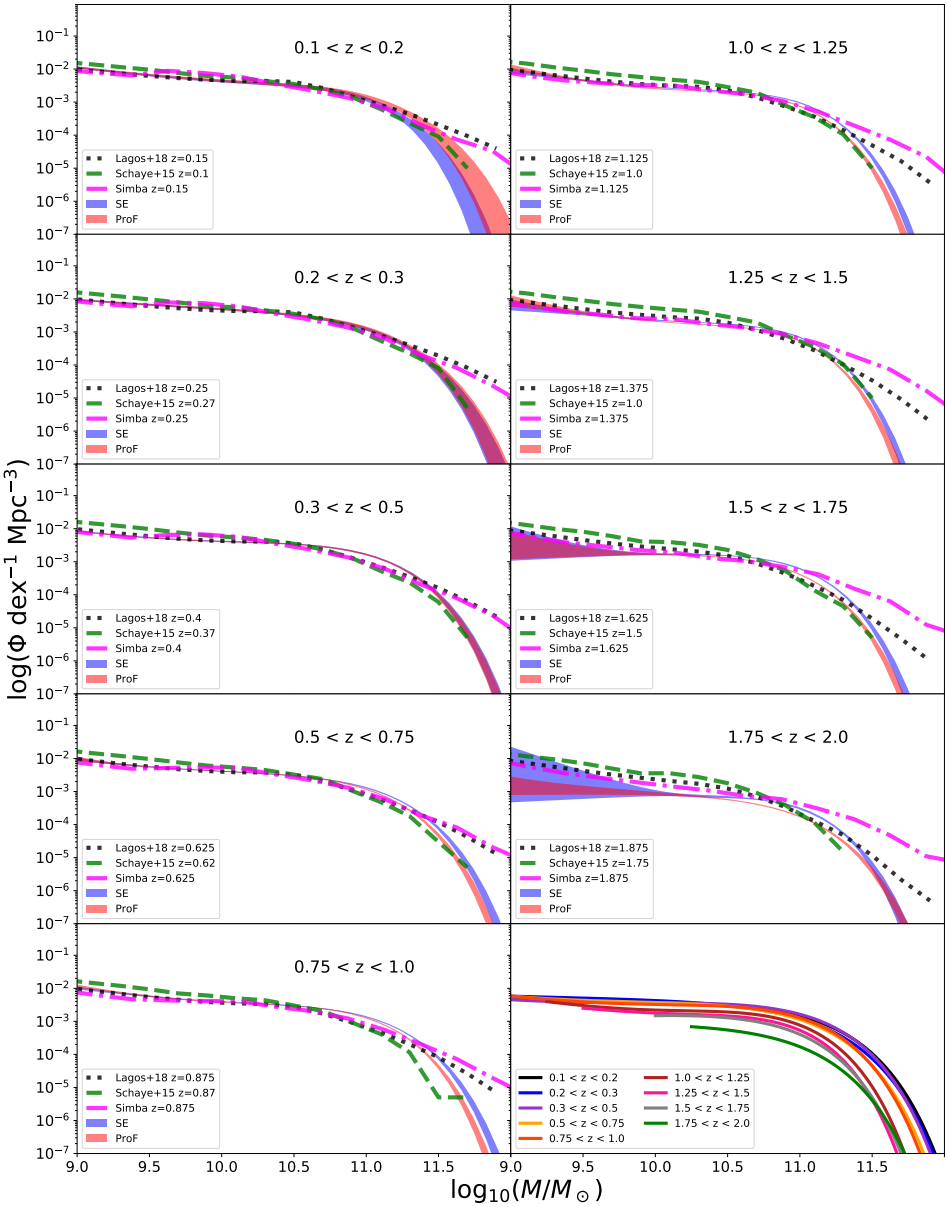


Figure 2.22: The intrinsic Galaxy Stellar Mass Function (GSMF) from Table 2.4 with Eddington bias removed. SExtractor derived results are in blue while ProFound is in red with the shading indicating the 1σ confidence interval. Presented alongside are examples of previously conducted simulations. For data from external sources, I present the closest redshift bin available for my comparisons. The studies used are the primary results from the *SHARK* run conducted in Lagos et al. (2018) as the black dotted lines, *EAGLE* (Schaye et al., 2015) as yellow dashed lines and *Simba* as the dot/dashed purple lines (Davé et al., 2019).

A similar situation is present when comparing against the results of the *Simba* hydrodynamical simulation (Davé et al., 2019). *Simba* implements a new, torque-limited accretion model (Anglés-Alcázar et al., 2017) for cold gas alongside more conventional Bondi accretion (Bondi, 1952) for hot gas. The energy built up in accretion is used to fuel feedback and quench galaxies. While this model for black hole growth and AGN activity was considered to be more physically motivated over previous simulations, it is found to still over produce galaxies of very high stellar mass. My observations reinforce the findings of Davé et al. (2019), where the largest discrepancy in high mass galaxies is at $z \sim 2$ (even when compared to the *SHARK* results) and that there is a slight underproduction of galaxies around M^* at low redshifts ($z < 1$). This results in a much shallower 'knee' than would be described with an exponential cut off. The authors explain possible causes of over-production of massive galaxies include over-merging of galaxies that blend, due to the use of a friend-of-friends (FoF) algorithm to count star particles, and the over-production of large halo masses.

The EAGLE simulation (Schaye et al., 2015) implements AGN feedback through inputting a fraction of the accreted gas as thermal energy into the local surroundings. It is found to produce mass functions with slightly higher low-mass normalisation but much sharper cut offs at the high mass end. This results in an over production of galaxies on the low mass slope by a factor of around 2, but a closer match at $M \geq M^*$. The inconsistencies at the low-mass end are thought to originate from the lack of quenching of lower mass galaxies at higher redshifts, or from the need to implement more burst-like star formation histories (Furlong et al., 2015). These GSMF measurements implement three-dimensional apertures of radius 30kpc when calculating stellar masses, the size of which is found to reduce the mass assigned to the largest, most massive systems in the simulation and prevent some of the over-merging issues described in Davé et al. (2019). This was found to reduce the number density of galaxies at $\log_{10}(M/M_\odot) \sim 11.5$ by around half a dex, imposing a steeper cut off in the number of high-mass systems. In the work conducted in Furlong et al. (2015), the *EAGLE* GSMF is fit with a double Schechter function

and is found to exhibit a strong evolution in M^* from $0.1 < z < 2.0$. The value of $\log_{10}(M^*/M_\odot) = 10.44 \pm 0.08$ at $z = 2$ is notably small, even lower than the results from McLeod et al. (2020) which lie at the lower extrema of the range of observational constraints. The increase from $\log_{10}(M^*/M_\odot) = 10.74 \pm 0.05$ at $z = 1$ to $\log_{10}(M^*/M_\odot) = 10.93 \pm 0.03$ at $z = 0.1$ matches my ProFound-based GSMF to within 1σ, while the value of M^* is lower at $z = 0.1$ from my SExtractor-derived GSMF. However, it is worth noting that Furlong et al. (2015) suggest that such an evolution in the characteristic mass could be the result of overly strong AGN feedback limiting the production of massive galaxies between $2 < z < 4$. It remains difficult to determine if faults lie within observations (mass errors, systematics etc.) or with the simulations (e.g. fine tuning AGN feedback) due to the sensitivity off all of these factors on the exponential break in number densities.

Within *EAGLE*, the origin of this stellar mass growth in the most massive galaxies arises from multiple sources. Qu et al. (2017) show that in the redshift range of $z < 1.5$, around 68 per cent of the massive galaxies ($\log_{10}(M/M_\odot) > 11$) in *EAGLE* undergo at least one major merger (at least 1:4 mass ratio, leading to > 0.1 dex growth) and have the average fraction of stellar mass originating from outside the primary galaxy increase by a factor of two (from a median of 10 per cent to 20 per cent), with most of this external contribution fuelled by these major merger events. At $z = 1$, over half of the simulated galaxies with $\log_{10}(M/M_\odot) > 11$ are defined as quenched ($\log_{10}(SSFR/\mathrm{yr}^{-1}) < -10$), falling significantly below *EAGLE* star-forming main sequence (Furlong et al., 2015). This leads to a similar fraction of quiescent galaxies to my sample. These consequently experience mass gains of less than 0.15 dex over the 4Gyr between $0.5 < z < 1.5$ under the simple assumption of a constant star formation rate. Creating this evolving value of M^* while maintaining a near constant value of Φ_1 thus likely requires a balance in merger events and internal star formation in order to sustain the growing number density of $M > M^*$ systems.

2.7 Conclusions

Utilising new photometric catalogues generated from optical and near-infrared data in the COSMOS and XMM-LSS fields, I have measured the GSMF over the redshift range of $0.1 < z < 2.0$, covering 60 per cent of the age of the Universe. The use of the two fields reduces the impact of cosmic variance by approximately a factor of two when compared to using the COSMOS field alone. This is due to both the wider area employed and the fact that the two fields are widely separated on the sky, subsequently allowing for tighter constraints on the GSMF at $M > M^*$. Simultaneously, the depth of the photometry available allows for constraints on on the GSMF of $\log_{10}(M/M_{\odot}) = 7.75$ for galaxies at $z \simeq 0.1$ and $\log_{10}(M/M_{\odot}) = 9.75$ for galaxies at $z \simeq 2.0$. I have investigated the use of ProFound and SExtractor source extraction software and also the impact of *Spitzer*/IRAC [3.6] and [4.5] photometry on the GSMF. My main conclusions are summarised as:

1. The inclusion of *Spitzer*/IRAC [3.6] and [4.5] photometry alleviates the degeneracies in select cases where the SEDs of red, massive galaxies in the redshift ranges $0.5 < z < 0.75$ and $1.0 < z < 1.25$ are confused with low redshift or stellar templates. Thus the inclusion of these data leads to fewer contaminants and a small decrease in the derived M^* (of up to 0.1 dex) in these redshift ranges.

2. Both SExtractor and Profound derived photometry produce consistent faint-end components of the GSMF. Differences between the mass functions at higher masses are greater when examining the extreme parts of my results (e.g. at the lowest and highest redshifts) but the resultant double Schechter fits are found to agree to within $\sim 1\sigma$.

3. The measured GSMF is found to disagree with the assumption that the characteristic mass M^* is constant with time between $0.1 < z < 2.0$ at the 3 (2) sigma level for the SExtractor (ProFound) derived results. Such an evolution in M^* between $0.5 < z < 1.5$ can be seen in some (but not

all) historical observations and also in the predictions from the *EAGLE* hydrodynamical simulation. However, significance is low and caveats in both the understanding of observational systematics and AGN feedback strengths in simulations (Furlong et al., 2015) mean a claim of an evolving M^* is presently very mild.

4. Eddington bias, and the methodology used to correct for it, is found to be highly influential on the shape of the high-mass end when attempting to retrieve the intrinsic GSMF from observations. There is presently still no consensus on the handling of observational and systematic errors that can impact stellar mass estimates (see also discussions within Grazian et al., 2015; Davidzon et al., 2017a). For my data, I find the correction required to be asymmetric and poorly described by commonly used analytic forms. When comparing to the observed GSMF, applying a simple Gaussian treatment to the Eddington bias is found to reduce M^* by around 0.05 dex. With the Lorentzian wings added into the description, the shift in M^* doubles to around 0.1 dex. Utilising a non-parametric form, based on the measured error distribution, leads to M^* shifts of 0.12 dex when compared to the observed GSMF.

5. When splitting my galaxy sample by specific star formation rate (SSFR), my results confirm the findings of previous studies that show increasing number densities of quenched galaxies are responsible for the rise in the GSMF for $M > M^*$ (e.g. Davidzon et al., 2017a; McLeod et al., 2020). Examining growth channels for stellar mass in the *EAGLE* simulation show this to be the result of a combination of internal star formation and merger events. This is because internal star formation alone would amount to less than 0.15 dex in stellar mass growth for the majority of the population of massive galaxies ($M > M^*$). Between $0.5 < z < 1.5$, the constant number densities of star-forming galaxies at $M > M^*$ indicate that these galaxies quench at approximately the same

rate that lower mass galaxies replace them through their own in situ star formation.

6. Comparisons to simulations reveal that the semi-analytic model *SHARK* and hydrodynamical simulation *Simba* both over-produce massive galaxies at low and intermediate redshifts. Likewise the *EAGLE* is found to over-predict the number of low-mass galaxies by a factor of around 2. This highlights that even in the era where such simulations can be fine tuned to better replicate observations, discrepancies still exist. It also further emphasises the conclusions drawn by these studies that additional work is required on all sides, from the observational data with which to calibrate/compare simulations against, to the development of the physical models used in simulations. However, I find the evolution of the high-mass component of the *EAGLE* simulations replicates my observations of an evolving M^*.

"Beauty is truth, truth beauty,"—that is all
Ye know on earth, and all ye need to know.

— John Keats' Ode on a Grecian Urn

3

Bridging the AGN and LBG UV Luminosity Funciton at $z \simeq 4$.

Contents

3.1 Introduction

The origins and subsequent evolution of galaxies at high-redshift remains a source of many unanswered questions and disputes in astrophysics. The rest-frame Ultraviolet (UV) Luminosity Function (LF) of galaxies (number density of galaxies as a function of intrinsic UV luminosity, defined here at $\lambda_{rest} = 1500$Å) is a fundamental observable that can be used to trace the time evolution of the overall galaxy population and the physical processes contained therein. The UV emission from these galaxies is a tracer of both young stellar populations from recent star formation and hot accretion disks around supermassive black holes (e.g. Wang & Heckman, 1996; Koratkar & Blaes, 1999; Adelberger & Steidel, 2000). By comparing the expected LFs derived from cosmological simulations to observations, it is possible to obtain insight into the physical processes occurring within galaxies, such as baryonic gas cooling, star-formation rates and feedback processes (e.g. Dekel & Silk, 1986; Nagamine et al., 2001; Benson et al., 2003; Powell et al., 2011; Silk & Mamon, 2012). A significant source of feedback impacting the bright-end of the LF has been attributed to Active Galactic Nuclei (AGN), which are postulated to hinder the growth of the most massive and luminous galaxies via kinetic feedback from jets as well as radiative feedback (e.g. Cole et al., 2002; Benson et al., 2003; Begelman, 2004; Bower et al., 2006). This results in a coupling of black-hole activity to the properties of the host galaxy (e.g. the $M_{\rm BH} - \sigma$ relation; Ferrarese & Merritt, 2000; Graham et al., 2011) and a predicted sharper high-luminosity turn-down of the LF (e.g. Ciotti & Ostriker, 1997; Silk & Rees, 1998; Kauffmann et al., 1999; Binney, 2004; Schawinski et al., 2007). This introduces a quantitative measure of when in cosmic time AGN are governing the shape of the LF or the galaxy mass function and how their influence may vary over time. Observationally, bright galaxies ($L > L^*$; where L^* is the characteristic luminosity of the knee in the LF) are rare however, and therefore constraining the bright end of the LF in the very high-redshift universe, and hence determining the impact of AGN feedback at this epoch, presents a challenge.

Both galaxies and AGN at high redshift can be selected via their strong Lyman-continuum break, a technique first implemented nearly three decades ago

by Guhathakurta et al. (1990) and Steidel & Hamilton (1992) (see also Steidel et al., 1996). The rest-frame UV emission is shifted into the optical regime of the observer at $z \geq 2.5$ and the break in the continuum can be observed in broadband photometry. The break allows for an approximate redshift determination either from a selection procedure using colour-colour cuts or through the implementation of photometric redshift fitting, resulting in the selection of so-called 'Lyman-break galaxies' (LBGs). To-date, many thousands of high-redshift LBGs have been found within deep *Hubble Space Telescope (HST)* data (e.g. Bouwens et al., 2015; Finkelstein et al., 2015). In addition, using shallower but much wider ground-based surveys, it has been possible to place the first constraints on the very bright-end of the galaxy LF, and the transition into the AGN regime (e.g. Bowler et al., 2014; Ono et al., 2018; Stevans et al., 2018).

Despite these advances, debate has persisted over the time evolution and functional form with which to describe the rest-frame UV LF at $z > 3$. Traditionally, the LBG LF has been fitted with a three-parameter Schechter function (Schechter, 1976), which is described by a normalisation (Φ^*), characteristic luminosity or absolute magnitude (L^* or M^*) and a faint-end slope (α). With the advent of wider-area data however, several works have found evidence for a departure from this form at the bright-end, with a double-power law (DPL) providing a better fit to the data at $z \simeq 4$–5 (Ono et al., 2018; Stevans et al., 2018) and $z \simeq 6$–7 (Bowler et al., 2014, 2015). The four parameter DPL fit allows for greater control of the slope at the bright-end (β), rather than the fixed exponential decline of a Schechter function. In parallel to discussions on the shape of the high-redshift LF is further debate on the redshift evolution of the parameters. Some previous works using predominantly *HST* imaging have favoured a fixed characteristic absolute magnitude of $M^* \simeq -21$ (e.g. Bouwens et al., 2015; Finkelstein et al., 2015), whereas other studies have suggested a rapid change in this parameter at $z > 5$ (e.g. McLure et al., 2009; McLure et al., 2013; Bunker et al., 2010; Wilkins et al., 2011; Oesch et al., 2012; Schenker et al., 2013; Schmidt et al., 2014; Bowler et al., 2015; Stevans et al., 2018; Viironen et al., 2018).

In parallel to studies of the LBG LF, there have been continued efforts to pin down the AGN/quasar LF out to high redshift. While previous LBG and AGN studies have been undertaken separately, the advent of wide-area LBG studies and deeper, smaller-area AGN studies, has resulted in a convergence of the samples. Around $M_{\rm UV} \simeq -23.5$ the number density of the sources of the two populations at $z \simeq 4$ are found to be comparable (Stevans et al., 2018; Ono et al., 2018). As the two populations can be selected in similar ways, contamination makes the derivation of the faint-end AGN slope and bright-end LBG slope difficult. Solving this problem is necessary to pin down the contributions of AGN and LBGs to the ionising background in the epoch of reionisation (EoR). In the process of measuring the AGN LF, debate has arisen between a number of studies over the steepness of the faint-end slope (e.g. Glikman et al., 2011; Masters et al., 2012; Giallongo et al., 2015; Ricci et al., 2017; Parsa et al., 2018; Akiyama et al., 2018; Boutsia et al., 2018; Stevans et al., 2018). With power-law values for the faint-end slope ranging from $\alpha \simeq -1.3$ to $\alpha \simeq -2.0$, this can have a significant impact on whether AGN are capable of sustaining reionisation on their own at this epoch and on where within the LF the dominant contributor changes between AGN and LBGs. To fully understand the shape of the faint-end of the AGN LF and the bright-end of the LBG LF, it is now apparent that both populations must be considered together.

In this work I determine the rest-frame UV LF at $z \simeq 4$ using the deep optical/near-infrared dataset available in the ground-based Cosmological Evolution Survey (COSMOS) and *XMM-Newton* Large-Scale Structure (XMM-LSS) fields. Combined, these fields provide $\sim 6\,\mathrm{deg}^2$ of data that enables the LF of $3.5 < z < 4.5$ galaxies and AGN to be measured precisely over a wide range in absolute UV magnitude ($-26 \lesssim M_{\rm UV} \lesssim -20$) using a coherent selection methodology. In comparison to previous studies at this epoch (e.g. Ono et al., 2018; Stevans et al., 2018), the inclusion of photometric data through to the K_s-band enables us to determine photometric redshifts with increased accuracy (4 to 5 per cent outlier rate), and provides a more complete selection than that provided by colour-cuts (up to 90 per cent).

This chapter is structured as follows. In Section 3.2 I provide a description of the data used in this study. The resultant photometric redshift estimations from template fitting, sample selection and completeness simulations are described in Section 3.3. In Section 3.4 I detail the final sample and method used for the measurement and fitting of the LFs and present the binned LF results. I then discuss the resulting functional fits in Section 3.5 in the context of past and present studies. I present my conclusions in Section 3.6. I assume a standard cosmology with $H_0 = 70\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, $\Omega_\mathrm{M} = 0.3$ and $\Omega_\Lambda = 0.7$. All magnitudes listed follow the AB magnitude system (Oke, 1974; Oke & Gunn, 1983).

3.2 Data

In this study I use a multi-wavelength data set consisting of 13-band photometry, with optical wavelength data from the Canada-France-Hawaii-Telescope Legacy Survey (CFHTLS) and the HyperSuprimeCam Strategic Survey Programme (HSC; Aihara et al., 2018b,a). Near-infrared data is provided by from the VISTA Deep Extragalactic Observations (VIDEO) survey (Jarvis et al., 2013) and UltraVISTA (McCracken et al., 2012) in the XMM-LSS and COSMOS fields respectively. The details of the available bands and depths for each field are provided in Table 3.1. The details of data/image post-processing are described in Bowler et al. (2020). Fields were selected on the basis of containing both wide (≈ 6 deg^2) and deep ($m_i > 25.4$ for 5σ detections) optical and near-infrared coverage. Successfully identifying $z \simeq 4$ galaxies requires deep imaging in the rest-frame UV band (shortward of the i-band in the observed frame) and in the surrounding bands in the optical and near-infrared to cover the Lyman and Balmer breaks of the galaxy spectrum, which are key features for determining photometric redshifts. I create object masks to exclude regions of missing or poor-quality data. Predominantly this involved masking the halos of foreground stars, image artefacts and the field edges. This leads to a removal of approximately 5 per cent of the total area available.

Table 3.1: Summary of the 5σ detection depths within the COSMOS and XMM-LSS fields. Depths are calculated in 2" diameter circular apertures, placed on empty regions of the image. Sources in the catalogues have a point-source aperture correction applied. The XMM-LSS field is split into three regions of $1.5\,\mathrm{deg}^2$ corresponding to the VISTA VIDEO tiles, and ordered from low to high RA. The first tile 'XMM1' contains the deeper HSC pointing and the third tile 'XMM3' contains the $1\,\mathrm{deg}^2$ CFHTLS D1 field.

Filter	COSMOS	XMM1	XMM2	XMM3	Origin
u^*	27.0	25.8	25.8	26.9	CFHT
g	27.1	–	–	27.0	CFHT
r	26.7	–	–	26.6	CFHT
i	26.4	–	–	26.4	CFHT
G	27.2	27.0	26.4	26.5	HSC
R	26.8	26.5	26.1	26.1	HSC
I	26.6	26.4	25.4	25.6	HSC
Z	25.9	26.3	24.6	24.8	HSC
y	25.5	25.6	24.1	24.1	HSC
Y	25.5	25.2	25.1	25.2	VISTA
J	25.3	24.7	24.7	24.7	VISTA
H	25.0	24.2	24.3	24.3	VISTA
K_s	24.8	23.8	23.9	23.9	VISTA

3.2.1 COSMOS

Covering $\sim 1.5\,\mathrm{deg}^2$ to a 5σ depth of $m_I = 26.6$, the COSMOS field contains the deepest data in my study. The optical imaging comes from HSC observations in the $GRIZy$ filters over the entire field. Additional optical data is provided by the CFHTLS-D2 field, which covers the central $1\,\mathrm{deg}^2$ of the field in the u^*, g, r, i bands. I use the UltraVISTA data release 3 (DR3) imaging in the near infrared (Y, J, H, K_s) over the full area used.

3.2.2 XMM-LSS

The XMM-LSS field is covered fully by optical imaging from HSC. One of the four pointings from HSC, which is centered on the Ultra Deep Survey from UK Infrared Telescope (UKIDSS; Lawrence et al., 2007), is deeper than the rest of the field. In addition to this, CFHTLS-D1 provides deep optical imaging in $1\,\mathrm{deg}^2$ of the field. As a result of this, the data can be grouped into three primary sub-fields of uniform optical data: 'HSC-DEEP' with $m_i = 25.4$ of area $1.48\,\mathrm{deg}^2$, 'HSC-

UDEEP' with $m_i = 26.3$ of area $1.92\,\mathrm{deg}^2$ and 'CFHTLS-D1' with $m_i = 26.43$ of area $0.97\,\mathrm{deg}^2$. All magnitudes are listed as 5σ detection limits. I extract u^* fluxes from the CFHTLS wide field survey in order to enable the whole XMM-LSS field to have full coverage in this filter. Near-infrared data is obtained from VIDEO and is uniform across the entire field. The total area of imaging that I utilise after masking is $5.88\,\mathrm{deg}^2$ over the two fields.

3.3 Methods

3.3.1 Photometric redshifts

In this section I outline the process used for estimating redshifts for the galaxies in my fields. With the wealth of multi-wavelength data available it is logical to make maximal use of the information available for each candidate galaxy. I thus elect to use a photometric redshift method to select my $z \simeq 4$ sample rather than traditional colour-colour cuts. Photometric catalogues were generated from my images using SExtractor (Bertin & Arnouts, 1996) in dual-image mode, with 2 " diameter circular apertures for photometry. The selection band was the i-band, from which forced photometry was performed on the remaining bands. Catalogues were cut at the 5σ detection limit of the i-band for the regions described in the previous section. All fluxes are corrected with a point-source aperture correction, where I model the point spread function (PSF) with PSFEx (Bertin, 2011) (See Bowler et al. (2020) for details). My observations are seeing limited ($\approx$ 0.7-0.8 "full-width-half-maximum, FWHM) meaning the image PSFs are the dominant contribution to measured sizes at the redshift range of interest, with mean sizes of 2kpc or $0.3''$ (e.g. Huang et al., 2013).

I estimate galaxy redshifts with the template fitting photometric redshift code LePhare (Arnouts et al., 1999; Ilbert et al., 2006). This publicly available code operates by minimising the χ^2 of galaxy spectral energy distribution (SED) templates fit to the multi-band photometry and uncertainties. I set the uncertainties of the photometry to a minimum of 5 per cent during the fitting process.

Template fitting

LePhare was run on all sources found by SExtractor over the $\sim$6 deg^2 of imaging. I used the COSMOS SED template set (Ilbert et al., 2009), where 32 templates are sourced from Polletta et al. (2007) with the GRASIL code (Silva et al., 1998) and from Bruzual & Charlot (2003). They cover a range of galaxy morphological classifications (E, S0, Sa, Sb, Sc, Sd, Sdm) and have the necessary rest-frame wavelength range to cover my optical and near-infrared dataset. Within the fitting process, each of these templates is allowed to be modified for the effects of dust attenuation using the Calzetti et al. (2000) attenuation law and an attenuation value in the range $E(B-V) = 0 - 1.5$. At each redshift, I use the Madau (1995) treatment for absorption by the inter-galactic medium (IGM). I experiment with the use of the model from Inoue et al. (2014) for the IGM in my photometric redshift methodology. The impact on the derived redshifts of my sample is a mean shift of 0.04 to higher redshift. As the systematic shift is much less than the photometric redshift errors I derive ($\delta z \simeq 0.15$ at $z \simeq 4$), the results of this work are unchanged with the use of this alternative IGM model. Alongside this, template spectra for AGN from Salvato et al. (2009) and stars from Hamuy et al. (1992, 1994); Bohlin et al. (1995); Pickles (1998); Chabrier et al. (2000) were fit and the χ^2 statistics were used to apply initial object classification and contamination control.

The AGN template set from Salvato et al. (2009) is created by developing hybrid SEDs by combining observed spectra of AGN and star forming galaxies with AGN domination in 10 per cent increments (e.g. one template is 80 per cent AGN and 20 per cent star forming galaxy). The base AGN templates are obtained from lower redshift ($z \ll 1$) targets using surveys such as SDSS, SWIRE and the *Infrared Space Observatory* PHT-S database (Polletta et al., 2007), while the galaxy templates are also obtained from Silva et al. (1998). The impact of the use of low redshift targets in the template development on the redshift determination for AGN sources at $z \sim 4$ (or $z \sim 5$ in Chapter4) is expected to be minimal. This is because these AGN are luminous in the rest frame ultraviolet, giving rise to apparent magnitudes of $m_I \sim 23$ and subsequently high SNR Lyman breaks with

which to reliably determine redshift. There has been evidence that the ultraviolet slope of the SED of AGN exhibits both luminosity and redshift dependence. This is a consequence of the time evolution of dust content within galaxies as well as black hole accretion rates (e.g. Carballo et al., 1999; Wu, 2011; Xie et al., 2015). This has the potential to reduce the quality of the template fits of high-redshift AGN across the rest-frame ultraviolet regime, reducing the reliability of solely using template fitting to classify objects as being AGN dominated. It may also lead to small differences in the measured value of $M_{\rm UV}$ at rest-frame 1500Å which has dependencies on i-band magnitude, ultraviolet slope and redshift.

Zero-points

One systematic that requires controlling is the photometric zero-points of the data. Zero-point errors can arise as a result of small systematics when modelling filter transmission functions, through biases within the choice of SED templates and from the calibration of the images. LePhare is capable of deriving a correction for this by fitting galaxies that have known spectroscopic redshifts. Small corrections are made to the zero-point of each filter to optimise the results on the sample of galaxies with known redshifts. For my data, a sample of 22,652 galaxies with spectroscopic redshifts from XMM-LSS and 21,990 galaxies in COSMOS from a mixture of the VVDS (LeFèvre et al., 2013), VANDELS (McLure et al., 2018; Pentericci et al., 2018), Z-COSMOS (Lilly et al., 2009), SDSS-DR12 (Alam et al., 2015), 3D-HST (Skelton et al., 2014; Momcheva et al., 2016), Primus (Coil et al., 2011; Cool et al., 2013), DEIMOS-10K (Hasinger et al., 2018) and the FMOS (Silverman et al., 2015) surveys were used. These redshifts are sourced from the same catalogue as used in Chapter 2 with two key differences. The VANDELS and DEIMOS-10K surveys are added to the catalogue in order better populate the high redshift regime and the flagging system to select high quality spectra is corrected for minor bugs. Only spectroscopic redshifts with flags indicating high-quality were used (confidence of ≥ 95 per cent). The inclusion of a large number of galaxies from different surveys in the zero-point calculations minimises the risk of the correction

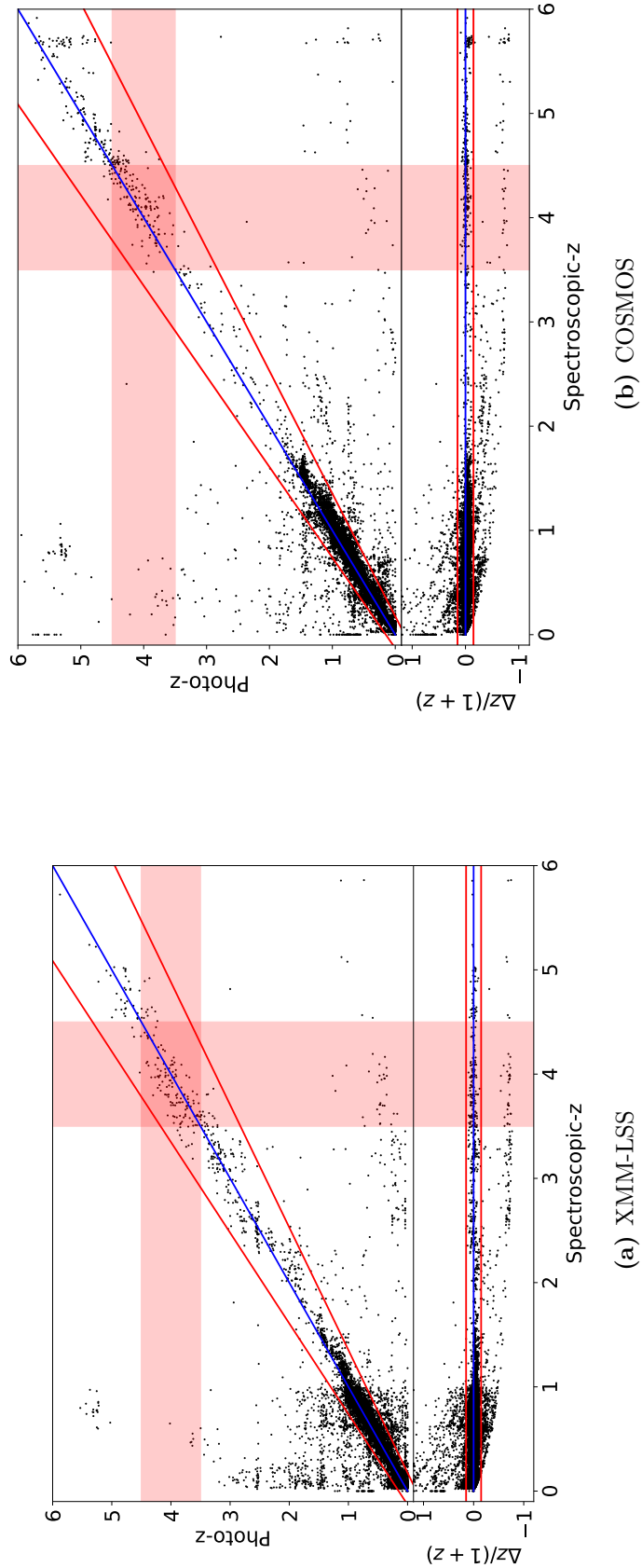
containing biases due to the individual surveys and their sample selections. Once the corrections were obtained, the zero-points were set and LePhare was re-run on the entire sample. The corrections themselves are minimal compared to the uncertainty in the photometry (< 0.08 mag).

Photometric redshift accuracy

Using the spectroscopic catalogues I are also able to assess the accuracy of my photometric redshifts which I show in Fig.3.1. Following Ilbert et al. (2009) and Jarvis et al. (2013) I evaluate the accuracy using the Normalised Median Absolute Deviation Hoaglin et al. (NMAD; 1983). In the case of photometric redshifts, this is $1.48 \times \mathrm{median}[|\Delta z|/(1+z)]$ and is used because it is resistant to extreme outliers. I define outlier photometric redshifts as being greater than 15 per cent different in $1+z$ to known spectroscopic redshifts. For my sample I find my photometric redshifts have an outlier rate of 5.4 per cent and a NMAD of 0.031 in the XMM-LSS field. For COSMOS I find a 3.9 per cent outlier rate and a NMAD of 0.027. Compared to results from Jarvis et al. (2013) with 3.3 per cent outlier rate and NMAD of 0.025 in XMM-LSS, my spectroscopic sample is significantly larger and extends across the whole field rather than just the deeper CFHTLS-D1 region, making it more representative of the whole field.

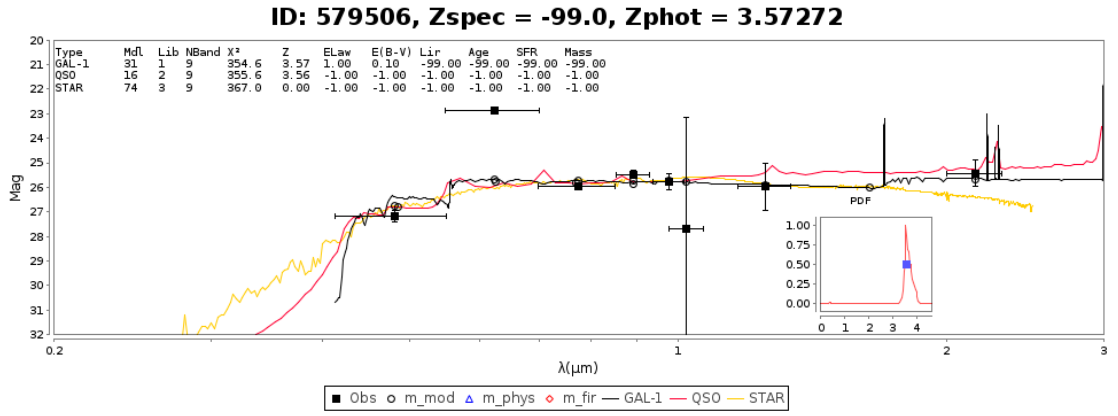
3.3.2 Sample selection

For my final $z \simeq 4$ sample I use galaxies that have the highest peak in their redshift probability distribution function within the redshift range $3.5 < z < 4.5$. I applied an initial cut of $\chi^2 < 1000$ in order to remove suspect contaminants which are poorly fit to my galaxy and AGN SED templates (removing the worst ~ 0.01 per cent of the population). 90 per cent of sources have χ^2 values of less than 10 and fewer than 1 per cent of galaxies have a χ^2 value greater than 100, inspection of random selections of these poor fit sources reveal some to be under the influence of artefacts, leading to spurious measurements in a single band. Additional issues which lead to inflated χ^2 include some cases where the optical measurements from

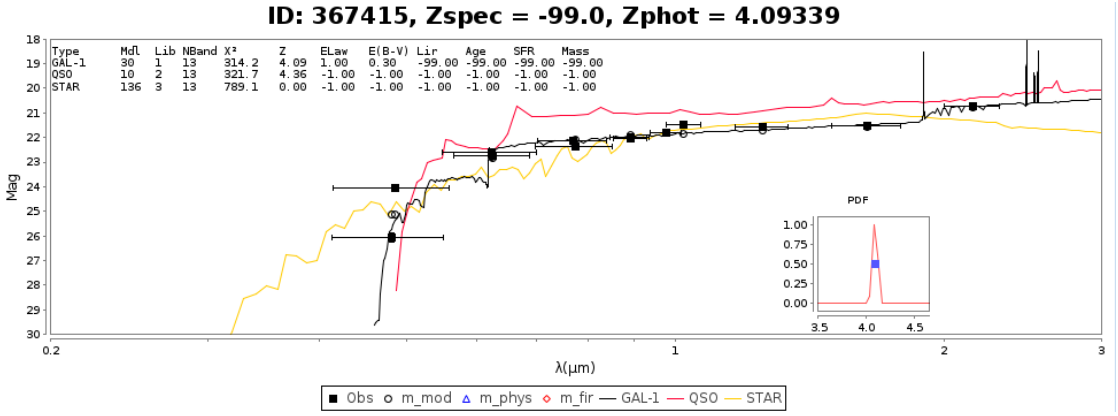


(a) XMM-LSS (b) COSMOS

Figure 3.1: A comparison of the photometric redshifts derived in this study to a compilation of spectroscopic redshifts. On the left I show the results for XMM-LSS and on the right, the results for COSMOS. The blue line shows the one-to-one correlation in the ideal case, and the upper and lower red lines show the 15 per cent margin in $1 + z$ that defines a significant outlier. The pink shading indicates the redshift bin $3.5 < z < 4.5$. The secondary-plot in each figure shows the same figure with the y-axis instead scaling with $\Delta z = (z_{\rm spec} - z_{\rm phot})$. The XMM-LSS field includes an additional 455 high-redshift objects from VANDELS while COSMOS includes 3839 objects from DEIMOS. These were used to test photometric redshift outlier rates in the selection regime and not used in zero-point calibrations.



(a)



(b)

Figure 3.2: Two examples of high χ^2 sources ($\chi^2 > 100$) with photometric redshifts within the range of interest. Object (a) is affected by an image artefact within the r-band while object (b) has a combination of disagreement between CFHT and HSC measurements in the optical regime and a slightly inflated Y-band measurement. Black lines indicate the best-fit galaxy template, red is the best fit AGN template and yellow is the best fit stellar template. Black data points are the measured photometry while the open circle data points are the expected measurement based on the best-fitting template. The probability distribution for the redshift is included as an insert plot within each sub-figure.

CFHT and HSC are offset from one another. Examples of two good $z \sim 4$ sources (well defined Lyman break and/or Balmer break) with high χ^2 values are displayed in Fig. 3.2 As a consequence of this, the χ^2 cuts in this chapter is more lenient than in other chapters at a potential cost of a small amount of potential contamination entering the sample. The application of these selection criteria provided an initial sample of 74,699 galaxies and AGN at $3.5 < z < 4.5$.

Several previous studies have used colour-colour cuts to select LBG samples.

While a conservative cut can provide a relatively pure sample of galaxies by taking advantage of the strong Lyman-break, they are often incomplete. In Fig. 3.3 I show the location of my $3.5 < z < 4.5$ sample in colour-colour space, along with the colour selection criteria used by van der Burg et al. (2010). I find that 60 per cent of my candidate $3.5 < z < 4.5$ galaxies would have been identified with such a colour-colour selection, in line with the 55–65 per cent completeness estimated by van der Burg et al. (2010). Also shown is an example evolution track of a Bruzual & Charlot (2003) model galaxy, which passes through my sample over the desired redshift range as expected.

To examine my ability to successfully identify $z \simeq 4$ objects I take a closer look at galaxies with a spectroscopic redshift in the redshift range of $3.5 < z < 4.5$ in the VANDELS and DEIMOS surveys. These were not used in the zero-point calibrations to avoid any favourable bias. There are a total of 147 objects with a spectroscopic redshift $3.5 < z < 4.5$ in COSMOS, 133 of these objects have photometric redshifts that successfully match the spectroscopic redshift within 15 per cent of $1 + z$. For the XMM-LSS sample, there are 149 objects with spectroscopic redshifts at $z \sim 4$ and 130 are successfully recovered within 15 per cent of $1 + z$. Half of the objects with an incorrect photometric redshift have bimodal solutions with the secondary solution being within the correct redshift region. Objects from the VANDELS survey are within my 'HSC-UDEEP' sub-field and have faint infrared magnitudes ($24 < m_H < 26$). This is thus an indication of the worst case within my data as the deep i-band in this sub-field selects these faint objects while the relatively shallow u^* and infrared coverage provides lower confidence of the detection of the Lyman break and galaxy continuum. I thus expect the completeness value of 87 per cent to be a lower bound of the photometric redshift completeness in my sample. Contamination from known low redshift objects ($z < 2$) having photometric redshifts measured within $3.5 < z < 4.5$ is found to be very low. I find ≤ 0.05 per cent of all other objects with spectroscopic redshifts are outliers that place them within the redshift range of interest ($3.5 < z < 4.5$). It is currently difficult to completely understand how many contaminant objects reach the final sample. With the simple assumption

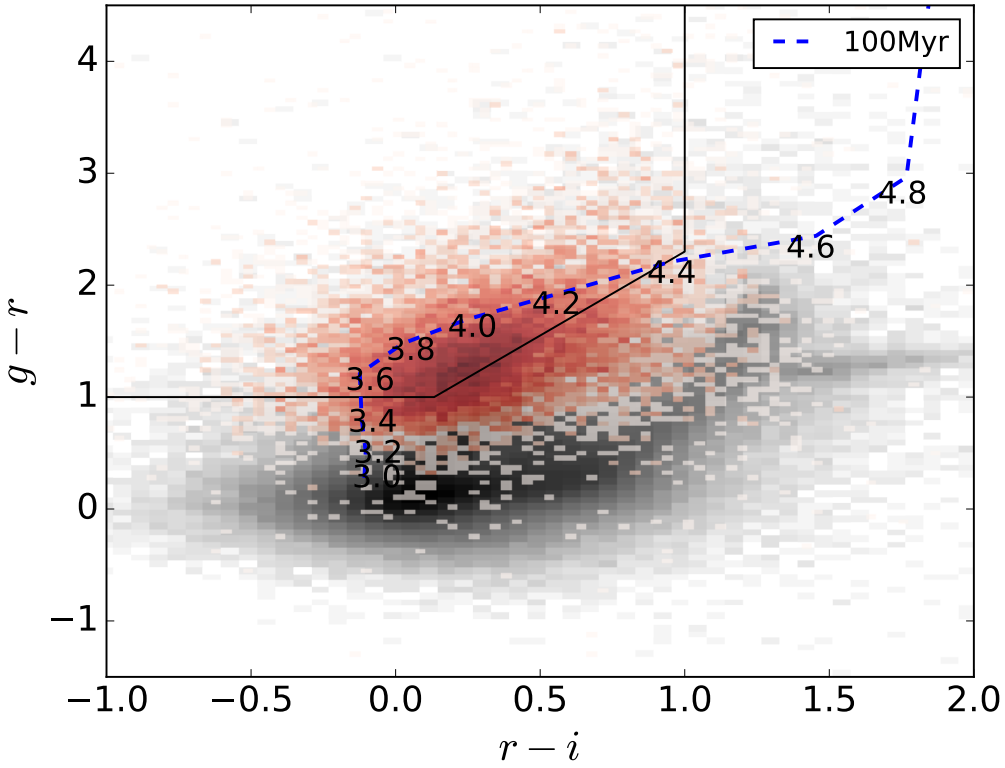


Figure 3.3: The g–r against r–i colour space, covering the Lyman break at $z \sim 4$. In red, I highlight my $3.5 < z < 4.5$ sample which was selected using the photometric redshift analysis described.
The background black distribution shows the colours of all other objects within the catalogue used in my study. The colour-colour cut used by van der Burg et al. (2010) is shown as the solid black line. When applied to my sample, this colour-cut is expected to select galaxies at $z \sim 4$ with around 60 per cent completeness. An example galaxy evolutionary track from Bruzual & Charlot (2003) is also displayed as the dashed blue line with the colour at each redshift labelled. The model has constant star formation, an age of 100 Myr and $E(B-V) = 0.15$.

that the measured rates of contamination from the spectroscopic sample can be expanded to the full photometric sample, this would lead to a contamination rate of low redshift objects in the final $z \simeq 4$ sample to be of order 1-2 per cent.

To assess the prospects of classifying AGN candidates and to further test the robustness of my photometric redshifts I match my COSMOS catalogue to the 16 objects identified spectroscopically as AGN at $z \sim 4$ in Boutsia et al. (2018). As four of these objects are located out of the UltraVISTA field, I expect twelve of these objects to be located in my catalogues. As expected, all twelve are found

in my catalogues. Ten are found to have the correct redshifts within errors and seven of those are found to have a better match to an AGN template than a galaxy template. I examine the SED fits for the two misclassified objects. I find that both have a suitable AGN fit at the correct redshift range, however a star template is found to have a lower χ^2 when fit to my photometry. One of these objects is heavily blended with a bright star (I account for this effect in my area and completeness calculations in section 3.3.3), while the other is located in a small region with no u^*-band data where the photometric redshifts are consequently expected to be of a slightly lower quality.

3.3.3 Selection completeness

In order to assess the completeness of the source extraction processes I simulated my selection procedure by injecting 3.6 million fake sources into my images and recovering these using SExtractor. My simulated galaxies have an assumed disk profile, with a fixed Sersic index of $n = 1$ (Sérsic, 1963; Conselice, 2014). For fainter magnitudes ($M_{\rm UV} > -22.5$) the UV luminosity of each fake galaxy is drawn from a posterior distribution that assumes the LF Schechter parameters and redshift evolution from Bouwens et al. (2015). This ensures I appropriately account for Eddington bias. The distribution is flat for brighter galaxies ($M_{\rm UV} < -22.5$) where the effect of this scattering is much weaker due to the higher confidence detections. The half-light radii of the galaxies are selected from the galaxy Size-Luminosity distribution of Huang et al. (2013). From this distribution the mean half-light radii of LBGs at $z \sim 4$ are generally smaller than 2kpc or $0.3''$, therefore the measured sizes in my imaging are dominated by the seeing. Simulated galaxies were sampled onto the pixel scale of the images, convolved with a model for the point-spread function (PSF) and scaled to the corresponding flux for the randomly assigned absolute magnitude. The PSF model is the same as that used in conducting the aperture correction to the $2''$ photometry and is derived by applying the software PSFEx (Bertin, 2011) to cut outs of stars and other point source-like objects in the original images.

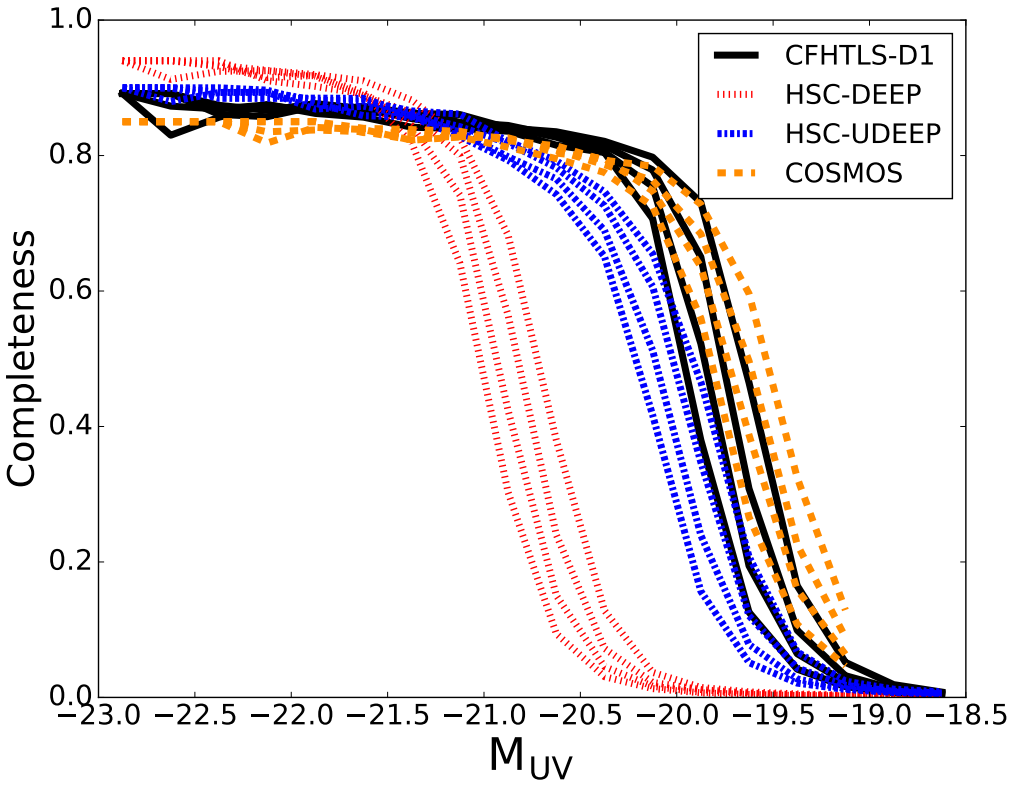


Figure 3.4: Completeness curves for my selection criteria as a function of absolute UV magnitude for the four subfields. Each subfield is represented by the colours black for CFHT-D1, red for HSC-DEEP, blue for HSC-UDEEP and yellow for COSMOS. For each sub-field I show four lines representing the variability of completeness across the full redshift bin $3.5 < z < 4.5$. Each line is a bin of 0.25 in redshift and from left to right is high redshift to low redshift, higher redshift bins being less complete due to the impact of distance on apparent luminosity. Deeper fields also appear more crowded, so have a higher probability of blending and hence converge to lower values of completeness at the bright-end.

To avoid unrealistic blending or image crowding, I added 2000 fake sources for each run in the CFHT images, and 4000 for the larger HSC images. To account for blending of sources, I assume that all fully-blended sources are unrecoverable. My simulated galaxies are prevented from being inserted into my images where the segmentation maps from SExtractor show the locations of pre-existing objects. This restriction applies to the central location of the simulated galaxy, so instances of partial blending (< 50 per cent overlap) are still present and accounted for. Due to the prevention of total blending in the simulations, my completeness curves converge to a value of 1.0 at the bright-end. These lines are then scaled down

by the fraction of the field which is occupied by pre-existing sources in order to account for the probability a source is lost to total blending. This is equivalent to reducing the total surveyed area by the proportion occupied by foreground objects. The corrected completeness functions then converge to the ratio of empty field area to total field area. The deeper the field the more dense it appears and so deeper fields converge to lower completeness values (see Fig. 3.4).

Simulated galaxies were placed into the image as faint as $M_{\rm UV} = -18.5$ in order to account for objects scattering above and below my selection limits in apparent magnitude. Each sub-field has its completeness function measured in bins of 0.25 in redshift in order to take into account changes in apparent luminosity with distance. Each redshift bin has a total of 200,000 simulated sources inserted into the images, totalling 3.6 million simulated objects across all sub-fields. I measured the completeness as the ratio of retrieved galaxies from SExtractor to the total number of galaxies injected into the simulation as a function of absolute magnitude. In Fig. 3.4 I show the completeness curves for the different regions of imaging I defined. The results of these simulations were used to set the faintest magnitude used in this work. I present the LF only in luminosity bins where the completeness is > 50 per cent in at least two sub-regions across the redshift range, to ensure that the contribution from each sub-field to the LF is not dominated by the completeness correction and to minimise cosmic variance (described in more detail in Section 3.4.1).

3.4 The Luminosity Function

The result of my photometric redshift selection procedure is a sample of 74,699 LBGs and AGN at $3.5 < z < 4.5$ from the XMM-LSS and COSMOS fields. Using this sample I proceed to measure the rest-frame UV LF at $z \simeq 4$.

3.4.1 The 1/Vmax method

I use the $1/V_{\rm max}$ (Schmidt, 1968; Rowan-Robinson, 1968) method to measure the LF. The maximum observable redshift of each object was found by taking the best

fit SED and redshift from LePhare and subsequently redshifting this further in steps of $\Delta z = 0.01$ to find the maximum redshift at which it would drop below the flux limit of my sample in each field. This is performed while accounting for the different i-band depths of each sub-field. The value of V_{max} then corresponds to the co-moving volume between the maximal observable redshift of the objects and a lower bound of $z = 3.5$. The final rest-frame UV LF, $\Phi(M_{\rm UV})$, for my complete sample of galaxies is then determined using:

$$\Phi(M)d\log(M) = \frac{1}{\Delta M}\sum_{i}^{N}\frac{1}{C_{i,f}V_{max,i}}, \tag{3.1}$$

where ΔM is the width of the magnitude bins and $C_{i,f}$ is the completeness correction for a galaxy i depending on its location within the sub-fields, f.

I estimate the uncertainty of the LF in each magnitude bin using:

$$\delta\Phi(M) = \frac{1}{\Delta M}\sqrt{\sum_{i}^{N}\left(\frac{1}{V_{max,i}}\right)^2}. \tag{3.2}$$

In the lower luminosity bins I adopt a bin size in absolute magnitude of $\Delta M = 0.25$. As I move towards brighter magnitudes, where I begin to probe the AGN LF beyond $M_{\rm UV} = -24$, I increase the bin size to first $\Delta M = 0.5$ and then a maximum of $\Delta M = 1.0$ if the number counts drop below 10. This is performed in order to maintain good number statistics in each bin. I determine the UV absolute magnitude $M_{\rm UV}$ using a top-hat filter of width 100Å , centred on 1500Å in the rest frame, using the best-fitting SED model for each object. After removing all objects in bins with less than 50 per cent completeness, I compute the final LF with a total of 46,904 galaxies and AGN.

Cosmic variance

With the LF fundamentally being a measurement of galaxy number densities, it is prone to the influence of large scale structure in the Universe when small survey volumes are used. The presence of voids, clusters and filamentary structures can skew LF results derived from single, small fields. Such an effect is commonly referred to as 'cosmic variance'. Making use of the Trenti & Stiavelli (2008) cosmic

variance calculator (which uses simulation volumes approximately the same order of magnitude in size as the survey volumes in this study $\sim 1 \times 10^7\,\mathrm{Mpc}^3$) with my galaxy samples, field areas and LF binning, I find that the error due to cosmic variance is consistent at the 3.5-4.5 per cent level across the whole LF. This is caused by the competing effects of low number density at brighter luminosities and the reduction in survey volume at lower luminosities, as fields fail to be 50 per cent complete. To be conservative I adopt a 5 per cent cosmic variance error across all bins, added in quadrature to my LF uncertainty from Equation 2.

Functional forms of the fits

Using broad-band photometry in my study, it is not possible to robustly separate $z \sim 4$ LBGs from AGN dominated systems, and hence my sample bridges both populations (see Section 3.3.2 where 7 out of 10 spectroscopically confirmed AGN are successfully classified as such). Due to the depth of my data, I do not fully constrain the very faint end of the LBG LF at $M_{\mathrm{UV}} \geq -20$. To compensate I elect to use the data points from Bouwens et al. (2015) for $M_{\mathrm{UV}} \geq -20$ to allow for stronger constraints on the LBG faint-end slope. Furthermore, although my sample contains objects as bright as $M_{\mathrm{UV}} \sim -26$, I do not have sufficient survey volume at this redshift to determine the position of the knee of the AGN LF at $M_{\mathrm{UV}} \simeq -26.5$ (Akiyama et al., 2018; Stevans et al., 2018). I therefore fit the AGN component of the LF with a single power law with the normalisation calculated at $M_{\mathrm{UV}} = -25.7$. I fit the data using different combinations of Schechter and DPL functions for the LBG component in addition to the single power law for the AGN component up to $M_{\mathrm{UV}} = -26.25$. When performing the fits, I make use of the Levenberg-Marquardt algorithm (Levenberg, 1944; Marquardt, 1963) from `scipy` to obtain a first order estimation of the fit parameters and their uncertainties. I then make use of a multi-dimensional χ^2 grid to derive final uncertainties and contours for the fit parameters.

Table 3.2: The rest-frame UV LF and its error margin at $3.5 < z < 4.5$. Column 1 shows the absolute UV magnitude at 1500Å ($M_{\rm UV}$). Column 2 shows the number density of objects and column 3 shows the errors in the number density which are calculated with equation 2. Both the value of the number density and its corresponding error are in a base 10 logarithmic scale.

$M_{\rm UV}$ [mag]	$\log_{10}(\Phi)$ [mag^{-1}Mpc^{-3}]	$\delta\log_{10}(\Phi)$ [mag^{-1}Mpc^{-3}]
−27.250	−7.792	0.997
−26.250	−7.330	0.336
−25.250	−6.949	0.183
−24.250	−6.511	0.103
−23.500	−6.256	0.109
−23.125	5.982	0.112
−22.875	5.646	0.087
−22.625	−5.216	0.050
−22.375	−4.825	0.036
−22.125	−4.524	0.030
−21.875	−4.160	0.026
−21.625	−3.861	0.024
−21.375	−3.588	0.023
−21.125	−3.391	0.023
−20.875	−3.219	0.023
−20.625	−3.095	0.023
−20.375	−2.975	0.022
−20.125	−2.848	0.022

3.4.2 The binned rest-frame UV LF

The completeness-corrected rest-frame UV LF from my sample is shown in Fig. 3.5, with the full list of binned data points given in 3.2. My results extend over a wide range in absolute UV magnitude, from the LBG dominated regime at $M_{\rm UV} = -20$, to the AGN dominated regime at $M_{\rm UV} \simeq -26$ and covers both the AGN-LBG transitional regime at $M_{\rm UV} \simeq -23.5$ and the LBG 'knee' at $M_{\rm UV} \simeq -21$. I present my results alongside examples of past results covering the full luminosity range. The studies of Akiyama et al. (2018) and Bouwens et al. (2015) cover the AGN and LBG luminosity functions respectively and the work by Stevans et al. (2018) covers a similar, but slightly brighter, UV luminosity range to this work. My measurements of the LF are consistent with these past studies at both ends of my luminosity range.

At $M_{\rm UV} \simeq -23.5$ I find a lower number density of objects when compared to those from Stevans et al. (2018), a feature I explore in more detail in Section 3.5.3.

The best-fit parameters for the Schechter and DPL functional forms for the LBG LF, alongside the power-law slope for the faint-end AGN, are shown in Table 3.3. My best performing functional form (minimal $\chi^2_{\rm red}$) to describe the LBG population is the Schechter function by a small margin, although a DPL is also acceptable. In Fig.3.6 and Fig.3.7 I present the 1σ and 2σ contours of the fit parameters for the Schechter and DPL parameters respectively for the LBG LF. These are both presented alongside the results from a small selection of past studies which have used the same functional forms to describe the $z \simeq 4$ LF. More detailed discussion on the comparisons between the studies is presented in Section 3.5.1.

In addition to the fits with the AGN LF included, I also show in Table 3.3 the results for fitting a Schechter and DPL function to just the data points fainter than $M_{\rm UV} = -23$, the magnitude where AGN contamination begins to become influential. This is performed to replicate the conditions of an LBG study that doesn't consider any significant AGN contamination in its fitting procedures and to observe what the consequences of this method would be. I find that the changes to the fits, when compared to my fits inclusive of AGN, are minimal and less than the 1σ level. However, there is a significant change to the χ^2 values for the fits where the DPL functional form stands out as being significantly better. This is likely driven by AGN beginning to contribute to the number counts at $M_{\rm UV} \sim -23$.

As the LBG LF approaches $M_{\rm UV} \simeq -23$ from the faint-end, the fraction of objects which are AGN begins to rise (e.g. Bian et al., 2013; Ono et al., 2018). When attempting to determine the functional form of the LBG LF, these AGN can act as contaminants, resulting in an excess at the bright-end of the LBG samples. If unaccounted for, this 'AGN contamination' results in the measurement of a shallower slope of the bright-end of the LBG LF when using the DPL functional form. In Fig.3.8 I show the fraction of objects that are AGN as a function of absolute UV magnitude derived from my best fit functions for both LBGs and AGN in Table 3.3, along with the results from Ono et al. (2018) that probed this regime with a

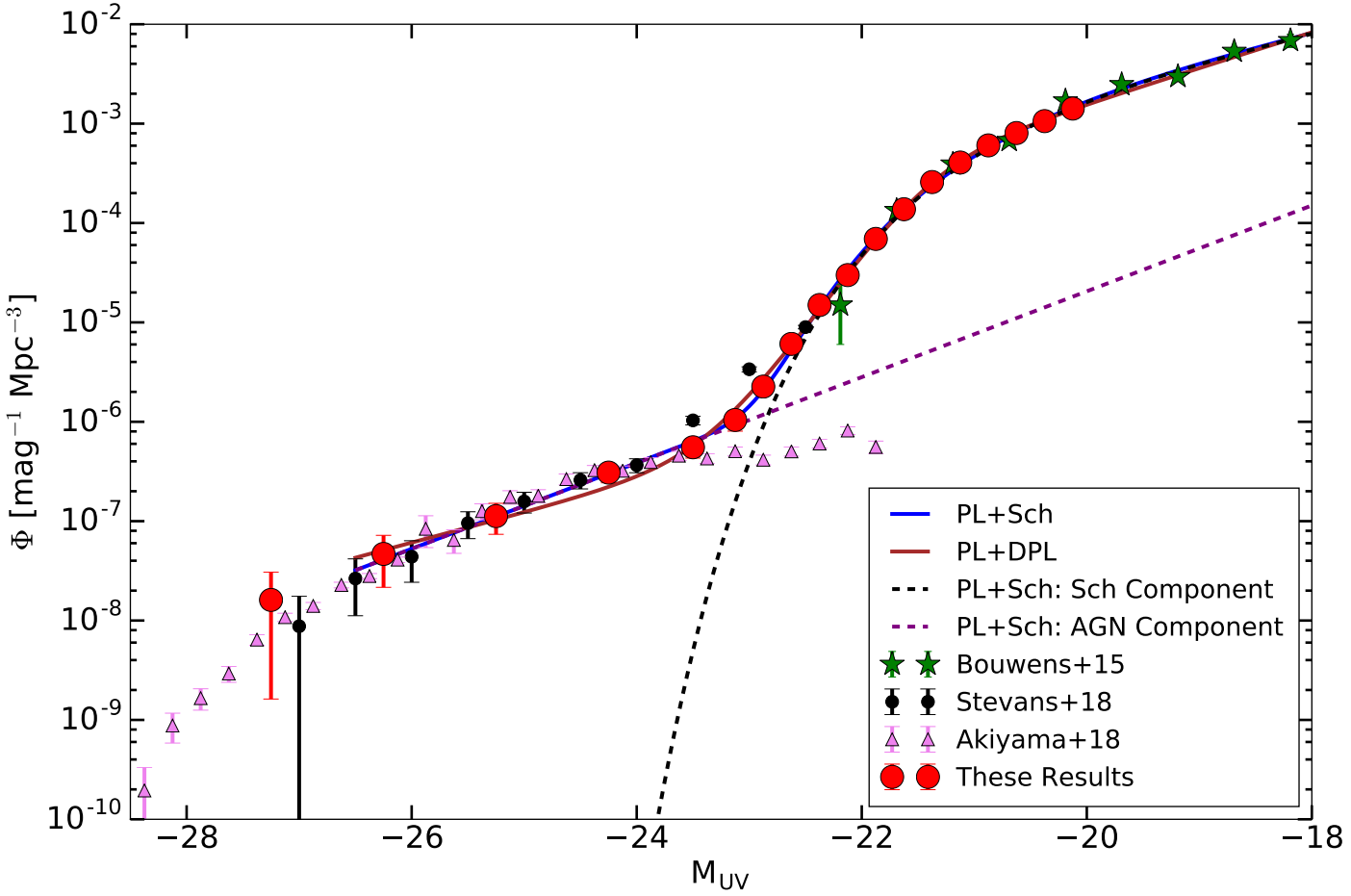


Figure 3.5: The $3.5 < z < 4.5$ UV LF derived in this study (red circles). The blue and red solid lines show the simultaneous fit of the AGN and LBG LF with a Schechter and DPL functional form for the LBGs respectively. The dashed lines show the individual AGN (purple) and LBG (black) contributions to the total LF for the PL+Sch case, combining to form the blue solid line. Shown alongside are results from previous studies by Bouwens et al. (2015) (green stars), Stevans et al. (2018) (black circles) and Akiyama et al. (2018) (violet triangles).

Table 3.3: The results of fitting to the $3.5 < z < 4.5$ rest-frame UV LF derived in this study. The first column lists the fitting parameterisation, with the next two columns showing the best-fit parameters for the AGN power law. Columns 4–8 show the LBG fitting results, either for a Schechter or DPL. The final columns shows the χ^2 and reduced χ^2 of the fit. The top two rows show the results of simultaneous fitting of the AGN and LBG components, where I fit a single power law for the AGN component and either a Schechter function or DPL for the LBGs. The bottom two rows show the results of fitting only the LBG component, where I include only data points fainter than $M_{\rm UV} \geq -23$. In all cases, at fainter than $M_{\rm UV} \geq -20$ I include the Bouwens et al. (2015) points in the fitting which reduces errors by a factor of ~ 2 by alleviating degeneracies on α. The AGN normalisation is calculated at $M_{\rm UV} = -25.7$ and the AGN LF does not make use of the data point at $M_{\rm UV} = -27.25$.

	AGN		LBGs					
	$\log_{10}(\Phi)$ [mag^{-1}Mpc^{-3}]	α	$\log_{10}(\Phi)$ [mag^{-1}Mpc^{-3}]	M^* [mag]	α	β	χ^2	$\chi^2_{\rm red}$
$\rm PL_{AGN}$+Sch	$-7.15^{+0.21}_{-0.35}$	$-2.09^{+0.32}_{-0.38}$	$-2.79^{+0.08}_{-0.08}$	$-20.89^{+0.12}_{-0.10}$	$-1.66^{+0.13}_{-0.08}$	–	25.8	1.36
$\rm PL_{AGN}$+DPL	$-7.07^{+0.29}_{-0.22}$	$-1.66^{+0.29}_{-0.58}$	$-3.30^{+0.08}_{-0.06}$	$-21.37^{+0.08}_{-0.11}$	$-1.92^{+0.07}_{-0.04}$	$-4.92^{+0.29}_{-0.25}$	25.3	1.41
Sch	–	–	$-2.85^{+0.06}_{-0.06}$	$-20.98^{+0.05}_{-0.06}$	$-1.71^{+0.06}_{-0.05}$	–	37.2	2.33
DPL	–	–	$-3.27^{+0.10}_{-0.08}$	$-21.34^{+0.10}_{-0.12}$	$-1.89^{+0.07}_{-0.06}$	$-4.77^{+0.22}_{-0.20}$	19.9	1.27

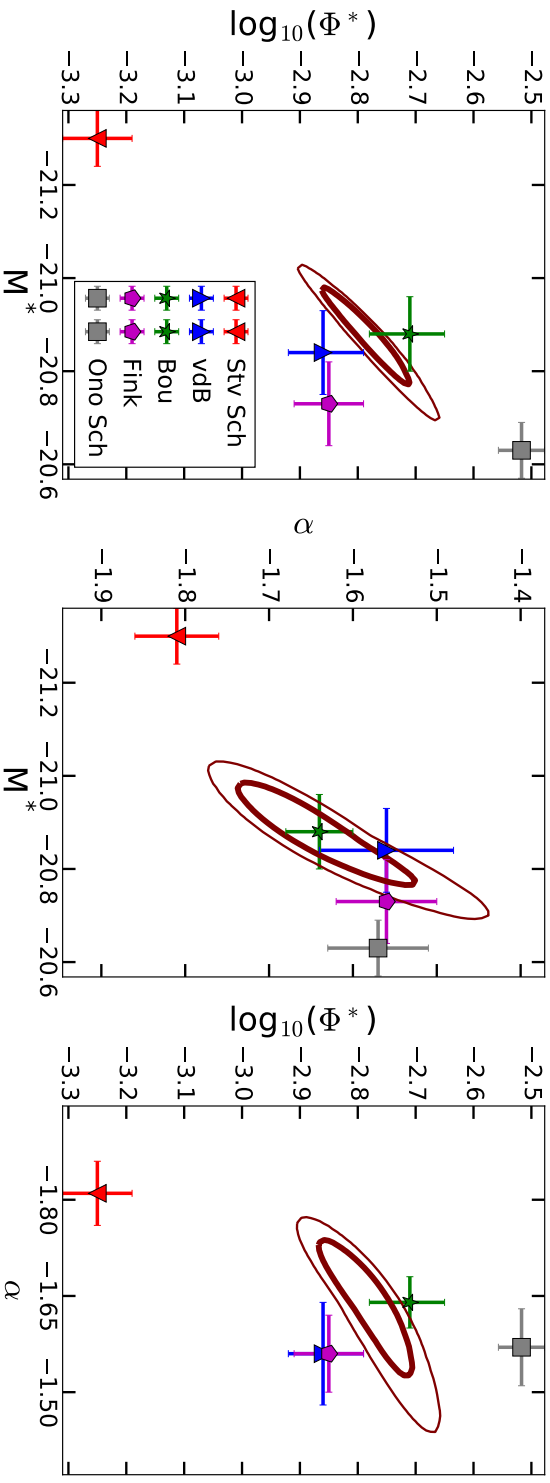


Figure 3.6: Contour plots showing the 1σ and 2σ confidence intervals for the LBG LF derived from my PL+Sch fit presented in Table 3.3. I compare my results to past studies that used the same functional form (van der Burg et al., 2010; Finkelstein et al., 2015; Bouwens et al., 2015; Ono et al., 2018; Stevans et al., 2018, displayed as vdB, Fink, Bou, Ono and Stv respectively)

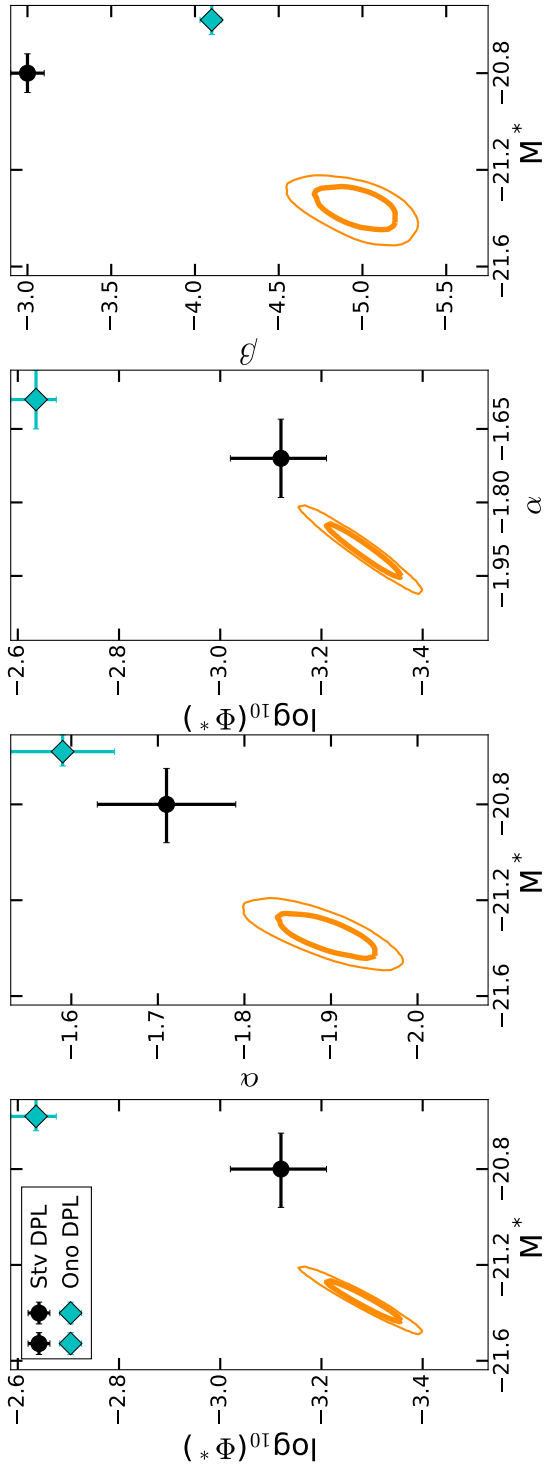


Figure 3.7: The same as Fig.3.6, but for the PL+DPL fit. I compare to the results of past studies using the same functional form for the LBG LF. (Ono et al., 2018; Stevans et al., 2018, displayed as Ono and Stv respectively)

spectroscopic sample. In both cases my results show that AGN make a significant contribution to the UV LF as faint as $M_{\rm UV} \simeq -23.0$. I note that the DPL fit has a smoother transition from LBG to AGN dominance than the Schechter function, which has an exponential cut off. At $M_{\rm UV} \simeq -23$ there is a very large disparity between the two functional forms, with an AGN fraction in the range $\sim$ 20–70 per cent depending on whether a DPL or Schechter function is used (see Fig.3.8).

3.5 Discussion

3.5.1 The LBG LF at z = 4

In Fig. 3.9 I show my results faintward of $M_{\rm UV} = -23$ and compare to the results of previous studies of $z \simeq 4$ LBGs. I find good agreement with the results of Bouwens et al. (2015) and Ono et al. (2018). Fainter than $M_{\rm UV} \geq -22$ I find that the studies of van der Burg et al. (2010) and Finkelstein et al. (2015) show a lower number density than this work. Stevans et al. (2018) suggest that the Finkelstein et al. (2015) sample may be lower than other studies due to the inclusion of *Spitzer* data in the mid-infrared, which assists in the removal of additional contaminants such as faint Milky Way brown dwarfs. I expect my photometric redshifts, which make use of full near-infrared coverage from VISTA and higher signal-to-noise ratio cuts, to also be robust to contamination from brown dwarfs. It is presently not clear why there is this offset between studies at the fainter-end of the LBG LF and more work is required to determine the cause.

Around $M_{\rm UV} \sim 22.5$, where the contribution of AGN begins to have an effect, I can compare my results to two previous studies. Ono et al. (2018) used HSC imaging to select a sample of LBGs and corrected for AGN contamination through their spectroscopic measurements presented in Fig. 3.8. Their final LBG LF is found to agree well with my DPL fit at the bright end ($M_{\rm UV} < -21$). I also find lower number counts in the $-24 < M_{\rm UV} < -22.5$ regime than Stevans et al. (2018), who utilised $18\,{\rm deg}^2$ of optical imaging from the Spitzer-HETDEX Exploratory Large-Area Survey (SHELA; Papovich et al., 2016). Within this magnitude range, my completeness is high, and the inclusion of two fields reduces the impact of cosmic

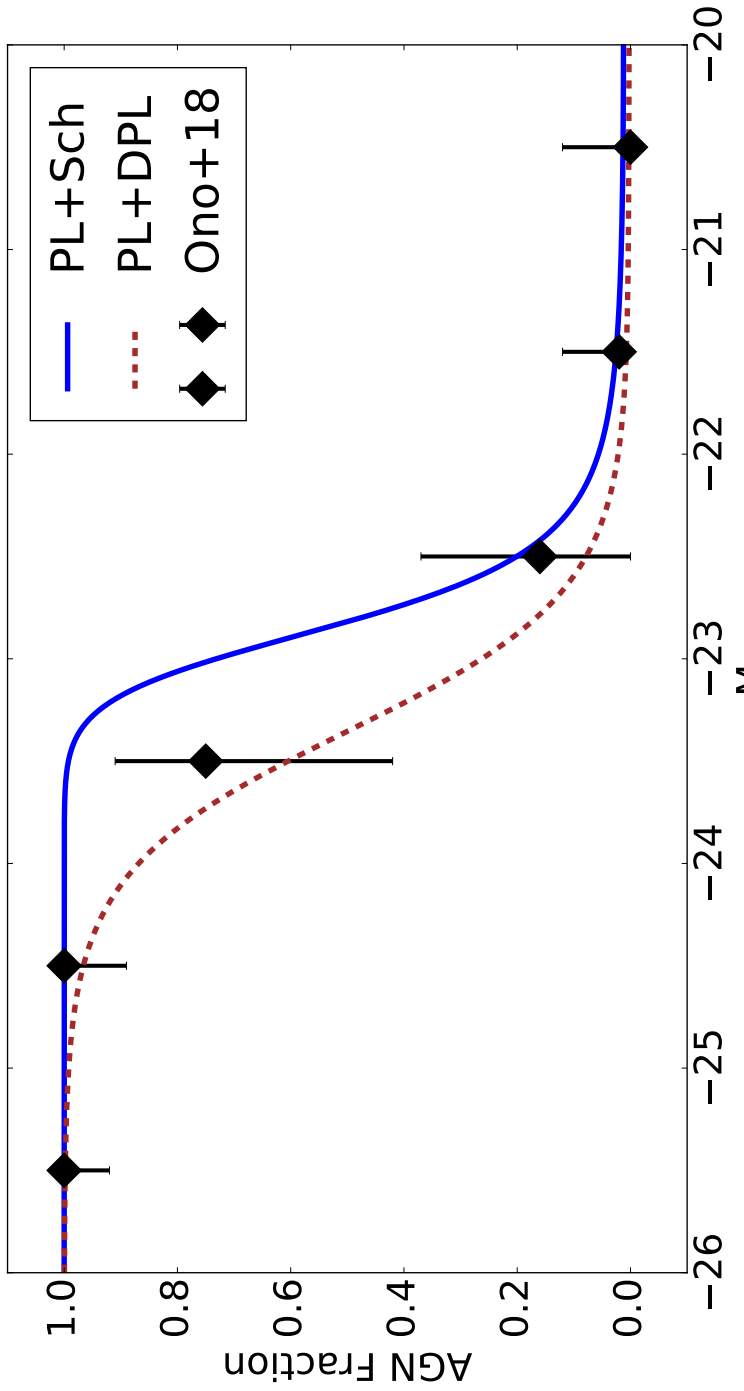


Figure 3.8: The fraction of objects estimated to be AGN as a function of absolute luminosity, found by taking the ratio of the best-fit AGN and LBG LF from Table 3.3. Also shown are the AGN fractions measured by Ono et al. (2018) who conducted a spectroscopic campaign of a subset of the objects contained within their sample.

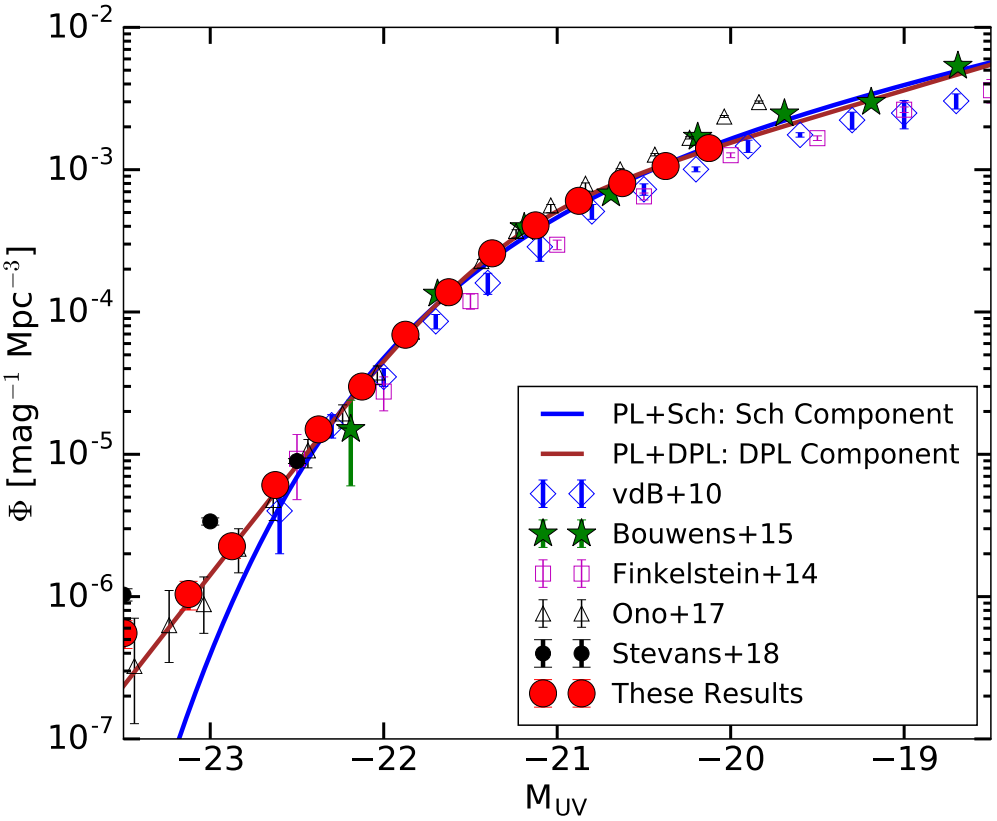


Figure 3.9: A zoom into the faint-end $M_{\rm UV} >= -23$ of the $z \sim 4$ LF. The blue/brown lines show the LBG component Schechter/DPL fits that are fit alongside AGN. Shown alongside are a myriad of past studies targeting the $z \sim 4$ population (van der Burg et al., 2010; Finkelstein et al., 2015; Bouwens et al., 2015; Ono et al., 2018; Stevans et al., 2018). At high luminosities, my results and the study by Stevans et al. (2018) are inclusive of AGN while Ono et al. (2018) attempted to remove them.

variance. The deep homogeneous multi-wavelength data gives us confidence in my results in this regime. I discuss the origin of this difference in Section 3.5.3

In Fig. 3.6 I compare the best-fit values and uncertainties for my Schechter fits against the aforementioned past studies. In this case my best-fit values are more closely matched to those of Bouwens et al. (2015) (within 1σ for all parameters) and centralised in the spread of the other results. It is because of this close matching to Bouwens et al. (2015) that I elect to use their data points to constrain the far faint-end $M_{\rm UV} > -20$. By relieving degeneracies on α, introducing these points reduces the errors on the best-fit LBG parameters by up to a factor of two while maintaining the same best-fit values to within $\sim 1\sigma$. The result from Ono et al.

(2018) provides a value of α which is in reasonable agreement with the studies of van der Burg et al. (2010),Finkelstein et al. (2015),Bouwens et al. (2015) and my own results, which are all in general agreement with each other. The degeneracies between M^* and Φ^* may be responsible for the slight disagreement in M^* and Φ^* measured by Ono et al. (2018). The resultant best-fit Schechter function from Stevans et al. (2018) produces values for all three parameters which are clear outliers to the collection of other past results; I believe this to be driven by their excess in the total number density around $-24 < M_{\rm UV} < -23$ (See Section 3.5.3 and Fig.3.5).

For the DPL fits in Fig.3.7 there are fewer studies to compare against, however, I find I disagree with both past studies that have attempted a DPL fit (Stevans et al., 2018; Ono et al., 2018). Although the linear trend observed in the two-dimensional contour plots show the large impact of degeneracies the DPL parameters can have. Compared to Stevans et al. (2018) the differences are primarily driven by my much steeper bright-end slope which drives M^* to brighter values. My DPL model is found to fit very well to the AGN corrected data points at the bright-end of the LBG LF from Ono et al. (2018) (see Fig. 3.9). The differences in the best-fit parameters are in this case primarily driven by the excess of objects that are found by Ono et al. (2018) at $M_{\rm UV} > -21$ that leads to a fainter derived M^*. At brighter UV luminosities ($M_{\rm UV} < -24$) my measured AGN number densities are found to closely match those of Akiyama et al. (2018) and Stevans et al. (2018).

3.5.2 Evolution of the rest-frame UV LF

I compare my results to the two linear redshift evolution models presented in Bouwens et al. (2015), derived from a mixture of results in the redshift range $4 < z < 8$. One model has a mildly-evolving value of M^* and one has a fixed value of M^*, justified by the evolving model producing such a shallow slope that M^* does not evolve with significance over the redshift range in which the model was created. I find agreement with the predictions to within $\sim 1\sigma$ for both the evolving and non-evolving M^* models respectively, suggesting little to no evolution of the shape of the bright-end of the LBG LF. This is unsurprising given my close

matching to their original data and their $z \sim 4$ bin being their most constrained and hence most influential on any evolutionary fit attempted.

I also compare my results to simple evolutionary models presented by Bowler et al. (2015). Within Bowler et al. (2015) a strong evolution in the Schechter function value of M^* is found from $z = 7$ to 5. My measured value of M^* is $-20.89^{+0.12}_{-0.10}$ which is 0.2 magnitudes fainter than their $z \sim 5$ measurement of $M^* = -21.07^{+0.09}_{-0.09}$ and much fainter than predicted if I extrapolate their proposed evolution to $z = 4$ ($M^* \sim -21.3$), suggesting a non-linear/slowing evolution. When compared to the DPL results in Bowler et al. (2015) I find a strong agreement with the shape of the bright end of the $z \sim 5$ fit, with their measured $M^* = -21.40^{+0.13}_{-0.12}$ and $\beta = -4.8^{+0.3}_{-0.4}$ all agreeing within $\sim 1\sigma$. Regardless of whether I assume a Schechter or DPL functional form (both are plausible with my data), these results suggest little to no evolution in the bright end from $z \simeq 4 - 5$.

3.5.3 Discrepancies in the AGN/LBG transitional regime

When considering the results of the DPL fits, the overall normalisation of the LBG fits have values similar to those measured by Stevans et al. (2018). However, a greater discrepancy is observed at the high-luminosity end where I observe a much steeper bright-end slope and brighter turn off with $\beta = -4.92^{+0.29}_{-0.25}$ and $M^* = -21.37^{+0.08}_{-0.11}$ with my DPL fit. Stevans et al. (2018) and Ono et al. (2018) find values of $\beta = -3.8$, $M^* = -20.8$ and $\beta = -4.1$, $M^* = -20.58$ respectively for the two parameters. While the discrepancies with Ono et al. (2018) can be somewhat explained by their use of colour-colour selection and the uncertainty in AGN contamination, the very clear discrepancy with Stevans et al. (2018) (see Fig.3.9) can be explained by examining the differences in the methodology between my two studies.

The selection criteria used in Stevans et al. (2018) is more complex than my own in an attempt to dig deeper into their shallower data while maintaining purity. They use a 3.5σ detection limit and cuts in both colour space and the redshift probability distribution. However, I believe that the observed discrepancy arises due to the fact that their $M_{\rm UV}$ values are calculated by converting the measured i-band apparent

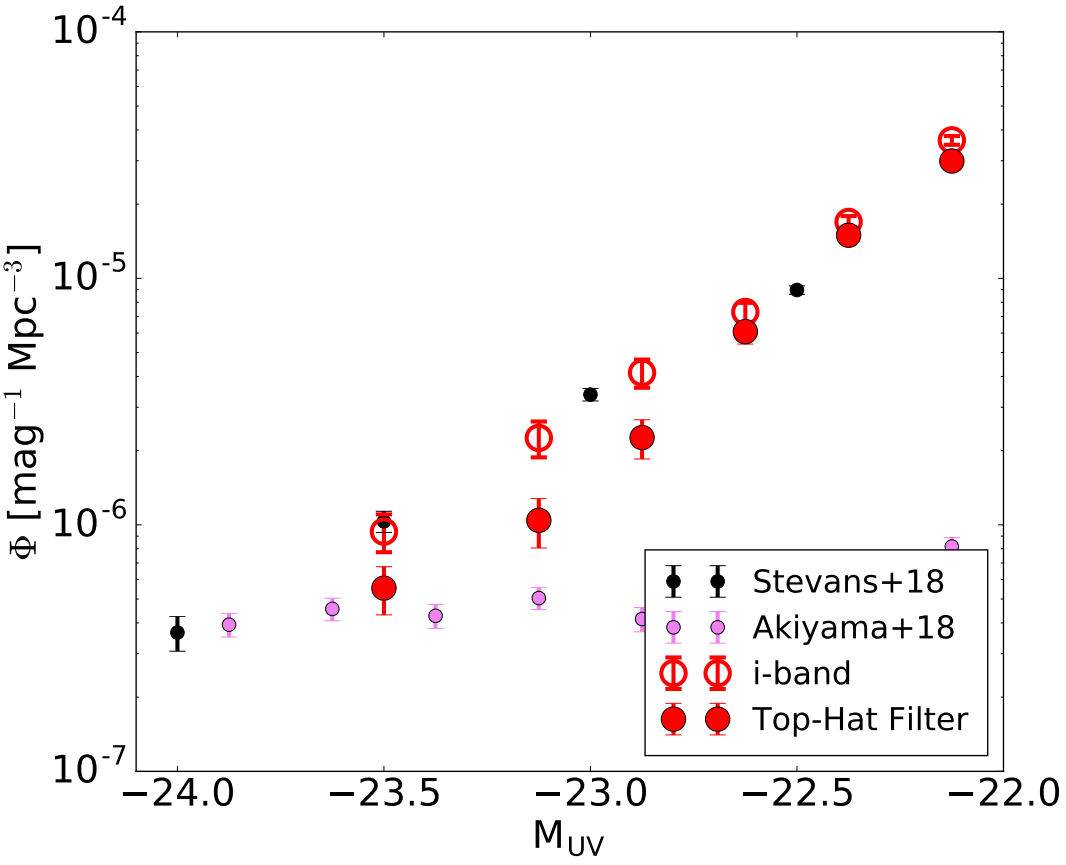


Figure 3.10: A zoom into the transition region at $-24 < M_{\rm UV} < -23$ showing how the UV LF changes depending on the method of measuring $M_{\rm UV}$. I note that when using the average i-band measurement that my results closely match those of Stevans et al. (2018). Also shown are the data from Akiyama et al. (2018) to show how the two methods converge onto the AGN LF in different places.

magnitude directly into an absolute magnitude using the estimated redshifts of each object. While the i-band contains the rest-frame 1500Å, its positioning within the filter is redshift dependant and the actual flux relative to the average i-band measurement is dependant on the UV slope of the galaxy SED which can vary from source to source. The ultraviolet slope of galaxies at high redshfit has been measured to follow a power law of approximately $\beta_\lambda \sim -2.0$, where $f_\lambda \propto \lambda^{\beta_\lambda}$ (Dunlop et al., 2012; Bouwens et al., 2012). AGN have been measured to have similar ultraviolet slopes at these redshifts, with median slopes of $\beta_\lambda \sim -1.6 \pm 0.3$ in the ultraviolet regime (Vanden Berk et al., 2001; Wu, 2011; Shull et al., 2012). This means that objects misclassified with AGN or galaxy templates will still have similar ultraviolet

slopes, leading to minimal impact on the measurement of $M_{\rm UV}$ at 1500Å.

The use of the i-band as a proxy for $M_{\rm UV}$ also leaves individual objects vulnerable to larger noise fluctuations in the i-band due to the lower confidence detections. These sources of potential inconsistency are accounted for in my methodology through the use of 5σ detection limits and the use of the best fit SED plus top-hat filter for calculating $M_{\rm UV}$. This makes my measurement of the absolute magnitude consistent between all galaxies and not as heavily reliant on the measurement in a single band, as the best-fit SED is constrained by all of the surrounding bands.

If I remeasure my luminosity function using the measured i-band flux as a proxy for $M_{\rm UV}$ I find a general trend of galaxies up-scattering to brighter bins. Galaxy numbers in every bin increase and the largest relative effect is found to be in the region where the LF is at its steepest (at $M_{\rm UV} \simeq -23$). When measuring $M_{\rm UV}$ in this way, the transition at $-24 < M_{\rm UV} < -23$ closely matches the results of Stevans et al. (2018) (see Fig.3.10). This highlights the importance of taking full consideration of the shape and redshift of the galaxy SED when measuring $M_{\rm UV}$.

3.5.4 The $z \sim 4$ AGN LF

A subject of ongoing debate with regards to the AGN LF at high redshifts is the steepness of the faint-end slope. Akiyama et al. (2018) measures a relatively flat faint-end slope of $\alpha = -1.3$, while in this study I measure $\alpha = -2.09^{+0.32}_{-0.38}/-1.66^{+0.29}_{-0.58}$ respectively for the Sch/DPL fits. This is in agreement with Stevans et al. (2018), however my AGN slope is slightly steeper when the DPL functional form is used for the LBGs. Between $-26 \leq M_{\rm UV} \leq -24$ the measured LF of the three studies agree very well, it is only when the transition to the LBG LF is reached that discrepancies arise. Both my work and the work of Stevans et al. (2018) make use of full template fitting and have data stretching into the near-infrared while the Akiyama et al. (2018) results are limited to just optical observations from HSC, resorting to stringent colour cuts and constraints on object morphology to minimise contamination. The morphological constraints require the sources to be dominated by nuclear emission and appear point-like. This is enforced by using

the second-order adaptive moment (derived in Hirata & Seljak, 2003), requiring the source differ in size and shape from the PSF of the selection images by no more than 10 per cent. By calculating $M_{\rm UV}$ with the i-band measurement as a proxy, the results from Stevans et al. (2018) show a greater number density of objects in the steep part of the LF in the transitional regime (see Section 3.5.3). Consequently, this leads to a shallower LBG bright-end slope, higher overall inferred galaxy fractions and a shallower AGN slope when using a DPL fit.

Another recent study by Boutsia et al. (2018) targeted the $z \sim 4$ AGN LF using an X-ray selected spectroscopic sample at $-24.5 < M_{\rm UV} < -23.5$. They find number densities of AGN to be high in this regime and similar to the values measured by Stevans et al. (2018), leading to their conclusion that the total UV LF should be dominated by AGN at $M_{\rm UV} < -23.5$. In Fig. 3.8 I show that both functional forms of my measured LBG LF give a high AGN fraction within this range of absolute luminosities and may be the dominant population in the LBG selection to absolute magnitudes as faint as $M_{\rm UV} = -23$, in agreement with Boutsia et al. (2018).

Purely photometric studies that focus on the AGN LF and rely on colour-colour selection and morphology (sources with a PSF-like profile) leave open the possibility of missing some AGN sources. Where the transition of the AGN to LBG luminosity function occurs, sources with weak AGN will have more significant contributions from the host galaxy. This can impact the measured profile of the source. Depending on the data used, selection criteria and extraction method, this could cause some objects to be missed through misclassification, leading to an underestimation of the AGN LF in such studies and making the completeness more challenging to model.

Within the results from the photometric study conducted by Akiyama et al. (2018), which uses morphology in AGN classifying, a shallow faint-end slope of $\alpha = -1.3$ is measured. However, when a cut is made to their data brighter than $M_{\rm UV} = -23$, -23.5 and -24 mags, where I now know the transition from AGN to LBG LFs occurs, this value steepens to $\alpha = -1.57$, -2.07 and -2.21 respectively and more closely matches the results of this study and that of Stevans et al. (2018). Other studies using additional selection criteria to extract pure AGN samples, such

as X-ray emission (e.g. Giallongo et al., 2015; Parsa et al., 2018), tend to have poorer number statistics and so uncertainties in the calculated faint-end slope of the AGN LF remains larger than the discrepancy between the various studies and provide no additional constraining power. In the work conducted by Marchesi et al. (2016), Chandra X-ray data is cross-matched to optical and NIR sources in the COSMOS field using a likelihood radio methodology (Sutherland & Saunders, 1992; Brusa et al., 2005). Cross-matching these optical/NIR sources with attributed X-ray emission to my own catalogue within a 1 arcsecond radius, I successfully recover the Boutsia et al. (2018) sample along with an extra source at $z_{phot} = 4.36$. A redshift which is outside of the selection range of $3.6 < z < 4.2$ implemented by the Boutsia et al. (2018) study. At $M_{\rm UV} \lessapprox -23.5$, 66 per cent of my sources have an X-ray counterpart with that fraction dropping off rapidly at fainter UV magnitudes. Together, this highlights that measuring the faint-end of the AGN LF remains challenging at luminosities fainter than the transition into LBG dominance ($M_{\rm UV} \gtrsim -23$) and that the key to solving issues around the faint-end AGN slope relies on secure object classification with well understood completeness (examples including spectroscopic surveys, use of multiwavelength signatures and Baldwin, Phillips & Terlevich diagrams; Baldwin et al., 1981).

To examine the origin of the AGN LF in greater detail, the work presented in this chapter was expanded upon in Bowler et al. (2021), where a 'toy' model was developed to explore where the shape of the AGN LF comes from. The model takes the UV LF of galaxies and uses observed relations and scatter between UV luminosity to galaxy stellar mass (Stark et al., 2013; Duncan et al., 2014; Salmon et al., 2015; Song et al., 2016b; Tacchella et al., 2018) and subsequently galaxy stellar mass to central black hole mass to produce a distribution of black holes (Targett et al., 2012; Venemans et al., 2016; Volonteri & Reines, 2016; Venemans et al., 2017). These are then assigned a distribution of accretion rates and a fraction (60 per cent) then selected to be obscured (Willott et al., 2010; Kelly & Shen, 2013; Ueda et al., 2014; Vito et al., 2018). Not all black holes are in a state of active accretion however. Instead, AGN are commonly found to have bursts of activity

lasting up to of order 100's of millions of years (Sternberg & Soker, 2009; Wise et al., 2017; Morganti, 2017; Delvecchio et al., 2020). The number density of UV luminous AGN can thus be a proxy measurement for the duty cycle of AGN activity. In the Bowler et al. (2021) toy model, the fraction of UV-luminous AGN is a variable left free to fit the normalisation of the AGN LF. The results provide a value of for the fraction of UV luminous black holes to be $f_{\rm active} \sim 0.003$ and produce a shape for the AGN LF that matches remarkably well to the results presented in this chapter (see Fig.9 within Bowler et al., 2021).

3.6 Conclusions

I exploit deep optical/NIR data from the COSMOS and XMM-LSS fields to measure the rest-frame UV LF at $z \simeq 4$. The combination of depth and area allows us to measure the LF from $-27 < M_{\rm UV} < -20$ using 46,904 objects selected through a photometric redshift method. My conclusions on the resultant LFs are:

1. When fit alongside AGN, I find I am unable to confidently discern between the two LBG functional forms of a Schechter function or DPL. When I fit an LBG LF to only those data points fainter than $M_{\rm UV} = -23$ I find that the DPL stands out as being the better descriptor. However, this is mostly driven by the inclusion of mild AGN contamination in the regime of $-23 < M_{\rm UV} < -22$ inflating the LF and highlighting the need to properly handle AGN when measuring the LBG LF.

2. My best-fit values for both Schechter and DPL functional forms of the LBG LF are found to be consistent with the $z \sim 5$ measurements from Bowler et al. (2015) and the mild linear evolution models of the LBG LF from Bouwens et al. (2015). These findings suggest that the shape of the bright-end of the UV LBG LF does not evolve significantly in the redshift range $3 < z < 5$.

3. I suggest that discrepancies found between studies at the magnitude range where the LBG LF transitions into the AGN LF ($-24 < M_{\rm UV} < -23$) can

be explained through differing definitions and methods of measuring $M_{\rm UV}$. My proposed method of measuring $M_{\rm UV}$ through the use of the best fit SED to photometry and a thin top-hat filter positioned at the rest-frame 1500Å allows for a more robust and consistent determination of $M_{\rm UV}$ for each object that is not reliant on a single measurement nor impacted by the combination of redshift and UV slope of the galaxy spectrum on broad-band filters.

4. I measure the transition between AGN/LBG domination in the UV LF and find that regardless of the functional form used to fit for the LBGs, the 50 per cent AGN fraction occurs within the range of $-23.5 < M_{\rm UV} < -23$, in agreement with recent results from Boutsia et al. (2018).

5. I find agreement with recent studies suggesting a steep faint-end slope for the AGN UV LF at $z \sim 4$ with $\alpha_{AGN} = -2.09^{+0.35}_{-0.38}$ ($-1.66^{+0.29}_{-0.58}$) when fit simultaneously with a Schechter (DPL) for the LBGs. These results support the conclusion from Stevans et al. (2018) who suggest that AGN, while not the dominant source of ionising photons, could sustain re-ionisation of Hydrogen on their own at this epoch. My results highlight the importance of simultaneously fitting the two populations of LBG and AGN together. Future insight into the nature of the sources at the transition (e.g. spectroscopic follow up with *VLT, JWST*) will shed light onto the astrophysics at play in shaping the bright end of the galaxy population.

The HUBBLE has given us nothing less than an ontological awakening, a forceful reckoning with what is. The telescope finally compels the mind to contemplate space and time on a scale just shy of the infinite.

— Ross Anderson

4

The Evolving UV Luminosity Function Post-Reionization

Contents

4.1 Introduction

Over the past 30 years, survey programmes exploiting various combinations of area and depth have allowed for the rest-frame ultraviolet (UV, 1500Å) luminosity function (LF) of galaxies and active galactic nuclei (AGN) to be confidently measured across a wide range of luminosity and cosmic time. However, at high redshifts ($z > 4$) there are many outstanding questions regarding the true density of galaxies brighter than the characteristic luminosity ($M_{\rm UV} < M^*$) and faint/obscured AGN ($-24 < M_{\rm UV}$). The luminosity regime $-24 < M_{\rm UV} < -22.5$ is especially challenging as this regime is where the galaxy and AGN comoving space densities become comparable (Ono et al., 2018; Stevans et al., 2018; Adams et al., 2020; Bowler et al., 2021). Here, the space density is low enough that *Hubble* Space Telescope (HST) programmes do not provide enough cosmic volume to produce statistically significant samples of objects (e.g. Oesch et al., 2010; McLure et al., 2013; Finkelstein et al., 2015; Bouwens et al., 2015; Parsa et al., 2016; Ishigaki et al., 2018; Bouwens et al., 2021). Meanwhile, ground-based observations cover larger survey volumes, but lower resolution makes morphological differentiation between galaxies and AGN (e.g. Masters et al., 2012; Akiyama et al., 2018; Matsuoka et al., 2019; Niida et al., 2020) more challenging. This redshift/luminosity regime is also often lacking in spectroscopic completeness (e.g. Ikeda et al., 2012; McGreer et al., 2013, 2018; Ono et al., 2018).

These complications result in contrasting results for the faint-end slope of the AGN Luminosity Function, with slopes ranging from $-2.1 < \alpha_{AGN} < -1.3$, as well as an ongoing debate over whether the bright end of the galaxy population is better described with a double power law (DPL) or Schechter function (Schechter, 1976) beyond some characteristic luminosity (Bowler et al., 2015; Ono et al., 2018; Stevans et al., 2018; Adams et al., 2020; Bowler et al., 2020). The primary difference between the two functions is the gradient of the slope brightwards of the characteristic luminosity or 'knee', leading to differing estimations for the comoving space density of the most luminous systems. However, constraining the numbers of these galaxies with observations proves difficult, as the steepness of the bright-end slope is such

that galaxies are quickly outnumbered by AGN-dominated sources. A route to solving these problems is to simultaneously fit both the AGN and galaxy UV LF (e.g. Stevans et al., 2018; Adams et al., 2020). This ensures that all sources are accounted for, and that AGN that have significant contributions from their host galaxy are not discounted through various selection criteria, allowing us to consider all sources of UV emission in the Universe (Bowler et al., 2021).

Constraining the comoving space density of highly luminous galaxies and UV-faint AGN are key to understanding several physical issues in extragalactic astronomy at high redshifts. These include: the contribution of AGN towards the budget of ionising photons in the latter stages of reionisation (e.g. Madau & Haardt, 2015; Yoshiura et al., 2017; Bosch-Ramon, 2018; Hassan et al., 2018; Parsa et al., 2018; Dayal et al., 2020), AGN activity timescales/black hole growth in the early Universe (e.g. Bañados et al., 2018; Wang et al., 2019; Yang et al., 2020; Wang et al., 2021) and the growth rates of the most luminous and massive galaxies at early times (e.g. Ouchi et al., 2009; Bowler et al., 2017; Stefanon et al., 2019; Bowler et al., 2020; Forrest et al., 2020; Neeleman et al., 2020). At $z \sim 5$, the Universe is approximately 500Myr post-reionisation, yet studies have shown there to be a rapid increase in the number of galaxies exhibiting strong AGN emission across this time period (McGreer et al., 2013; Jiang et al., 2016; McGreer et al., 2018; Kulkarni et al., 2019; Niida et al., 2020). If the space density of faint AGN rise at a greater rate than UV luminous galaxies, the transition between AGN dominated emission and galaxy dominated emission would be expected to shift faintwards with time.

Many recent measurements of the UV LF at $z \sim 5$ have made use of colour-colour cuts in order to select galaxy samples (van der Burg et al., 2010; Bouwens et al., 2015; Finkelstein et al., 2015; Ono et al., 2018; Bouwens et al., 2021). While colour-colour cuts can be effective in selecting high redshift samples using the Lyman continuum break at $\lambda_{rest} < 1216$Å (Guhathakurta et al., 1990; Steidel & Hamilton, 1992; Steidel et al., 1996), there are significant compromises that must be made regarding completeness and contamination rates. This is because regions of colour-colour space are shared with lower redshift galaxies and Milky Way brown dwarf stars

(Stanway et al., 2008; Bowler et al., 2014; Wilkins et al., 2014). As a consequence of this, sample completeness is often around the 60 per cent margin in these studies and have the volume required to constrain the bright-end of the galaxy population (e.g. van der Burg et al., 2010; Ono et al., 2018). Extensive near-infrared observations that complement optical observations, provide the opportunity to improve upon selection criteria by modelling the wider SED of each source rather than a select few colours. Previous studies around this redshift range have shown that modelling photometric redshifts using both optical and NIR data can lead to an increase in completeness to $\sim$ 80–90 per cent while minimising the potential for contamination (McLure et al., 2009; Bowler et al., 2015; Stevans et al., 2018; Adams et al., 2020).

In this study, I measure the $z \sim 5$ UV LF with the aim of mapping the bright-end galaxy population and the transition into AGN dominated UV emission. This is performed with the aim of constraining the comoving space density of UV-faint AGN and UV-luminous galaxies in order to better understand the rise in AGN after reionisation and determine if the bright-end of the galaxy population can be better described with a Schechter function or DPL functional form. These are important in the context of understanding black hole growth rates, the ionising photon budget, AGN feedback and dust obscuration in highly star-forming galaxies. I conduct this measurement with the use of modern, deep optical and near-infrared photometry available in the COSMOS and XMM-LSS fields. The surveys covering these fields provide the ideal depth-area combination, enabling for rare sources as bright as $M_{\rm UV} \sim -25$ to be detected while being simultaneously deep enough to detect sources down to $M_{\rm UV} \sim -20.5$. I combine these measurements with results of additional studies that use other datasets to expand the luminosity range to $-28.5 < M_{\rm UV} < -16$ and conduct the first simultaneous modelling of the entire luminosity function at $z \sim 5$.

In Section 4.2, I describe the optical and NIR photometric data used in this study and how these data are used to estimate photometric redshifts for sample selection. In Section 4.3 I assess the completeness of the sample, measure and fit the the UV LF. In Section 4.4, I detail the results of my fitting procedure and comment

on the performance of the different models. In Section 4.5 I expand discussion to contextualise my results with past and present studies from the literature. I present my conclusions in Section 4.6. We assume a standard cosmology with $H_0 = 70\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, $\Omega_\mathrm{M} = 0.3$ and $\Omega_\Lambda = 0.7$. All magnitudes listed follow the AB magnitude system (Oke, 1974; Oke & Gunn, 1983).

4.2 Data and Sample Selection

4.2.1 Photometry

In this study, I use a multiwavelength dataset spanning 14 photometric bands covering 0.4–2μm across two extragalactic fields (XMM-LSS, COSMOS). Measurements in these bands are derived from the Canada-France-Hawaii-Telescope Legacy Survey (CFHTLS; Cuillandre et al., 2012) and the HyperSuprimeCam Strategic Survey Programme (HSC; Aihara et al., 2018b,a, 2019) in the optical regime. Near-infrared data is provided by the final data release of the VISTA Deep Extragalactic Observations (VIDEO) survey (Jarvis et al., 2013) and UltraVISTA DR4 (McCracken et al., 2012) in the XMM-LSS and COSMOS fields respectively. A breakdown of the available filters and the depths within each subfield is provided in Table 4.1.

Using HSC-SSP DR2, the two extragalactic fields can be broken down into three subregions of equal depths in the z-band, which contains the rest-frame UV (1500Å) at $z \sim 5$. I designate these regions COSMOS (COS), XMM-Deep (XMMD) and XMM-UltraDeep (XMMU) in longhand (shorthand). The XMMU sub-region is centred on the Ultra Deep Survey from UK Infrared Telescope (UKIDSS; Lawrence et al., 2007). CFHT provides coverage in the u^*-band for over 90 per cent of the survey area used in this study and I also include the other optical bands within the CFHT D1 and D2 fields which each cover $1\,\mathrm{deg}^2$ within XMM-LSS and COSMOS. Photometry was extracted using SExtractor (Bertin & Arnouts, 1996) in dual image mode, with the HSC's z-band used for selection and flux extracted in 2 " diameter circular apertures in the other bands. These fluxes are then corrected with a point-source aperture correction, where I model the point spread function (PSF) with PSFEx (Bertin, 2011) (see Bowler et al. (2020) for details on catalogue

Table 4.1: Summary of the 5σ detection depths within the COSMOS and XMM-LSS fields. Depths are calculated in 2" diameter circular apertures, placed on empty regions of the image. The XMM-LSS field is split into two regions. XMMU contains the deeper HSC pointing, and 1 $\deg^2$of COSMOS and XMMD contain coverage from CFHT-griz bands.

Filter	COSMOS	XMMU	XMMD	Origin
u^*	27.1	25.7	$25.7 - 27.1$	CFHT
g^*	27.3	–	27.4	CFHT
r^*	26.9	–	26.9	CFHT
i^*	26.6	–	26.4	CFHT
z^*	25.5	–	25.4	CFHT
g	27.4	26.9	26.7	HSC
r	27.1	26.4	25.9	HSC
i	26.9	26.3	25.6	HSC
z	26.5	25.7	25.4	HSC
y	25.7	24.9	24.2	HSC
Y	25.4	25.1	25.1	VISTA
J	25.3	24.7	24.7	VISTA
H	25.1	24.1	24.2	VISTA
K_s	25.0	23.8	23.9	VISTA

generation). Survey areas for each subfield are calculated based on the areas in which the CFHT, HSC & VISTA observations overlap and large artefacts (e.g. stellar ghosting) have been masked out. The COSMOS region has a survey area of 1.55 $\deg^2$ and 5σ z-band depth of $m_z = 26.5$ in a 2 " aperture, while the XMMD region is 2.70 $\deg^2$ and has a 5σ z-band depth of $m_z = 25.4$ and XMMU is 1.70 $\deg^2$ with a 5σ z-band depth of $m_z = 25.7$.

4.2.2 Photometric redshifts

With such a wealth of multiwavelength data in these extragalactic fields, I select my sample following a spectral energy distribution (SED) template fitting procedure. Object classification and redshift estimates are made using the template fitting photometric redshift code LePhare (Arnouts et al., 1999; Ilbert et al., 2006). This code minimises the χ^2 of various spectral energy distribution (SED) templates for galaxies, AGN and Milky Way stars using the multi-band photometry and uncertainties. I set the uncertainties of the photometry to a minimum of 5 per cent during the fitting process. The template sets used in this study are the galaxies

covered by the COSMOS SED template set (Ilbert et al., 2009), AGN/QSO-like objects from Salvato et al. (2009) and stellar templates from a combination of results from Hamuy et al. (1992, 1994); Bohlin et al. (1995); Pickles (1998); Chabrier et al. (2000). Additional stellar templates are also added from the SpecX[1] brown dwarf sample. These brown dwarf templates were added to the sample because M and L-class dwarfs can have very similar colours to those of $z \sim 5$ galaxies. In addition, because high-redshift AGN and Milky Way stars both exhibit point spread function-like morphologies, no morphological cuts to the sample were made in order to prevent loss of potential AGN sources. (see Section.4.2.3 for further discussion on brown dwarfs). The treatment derived in Madau (1995) is used for absorption by the inter-galactic medium (IGM).

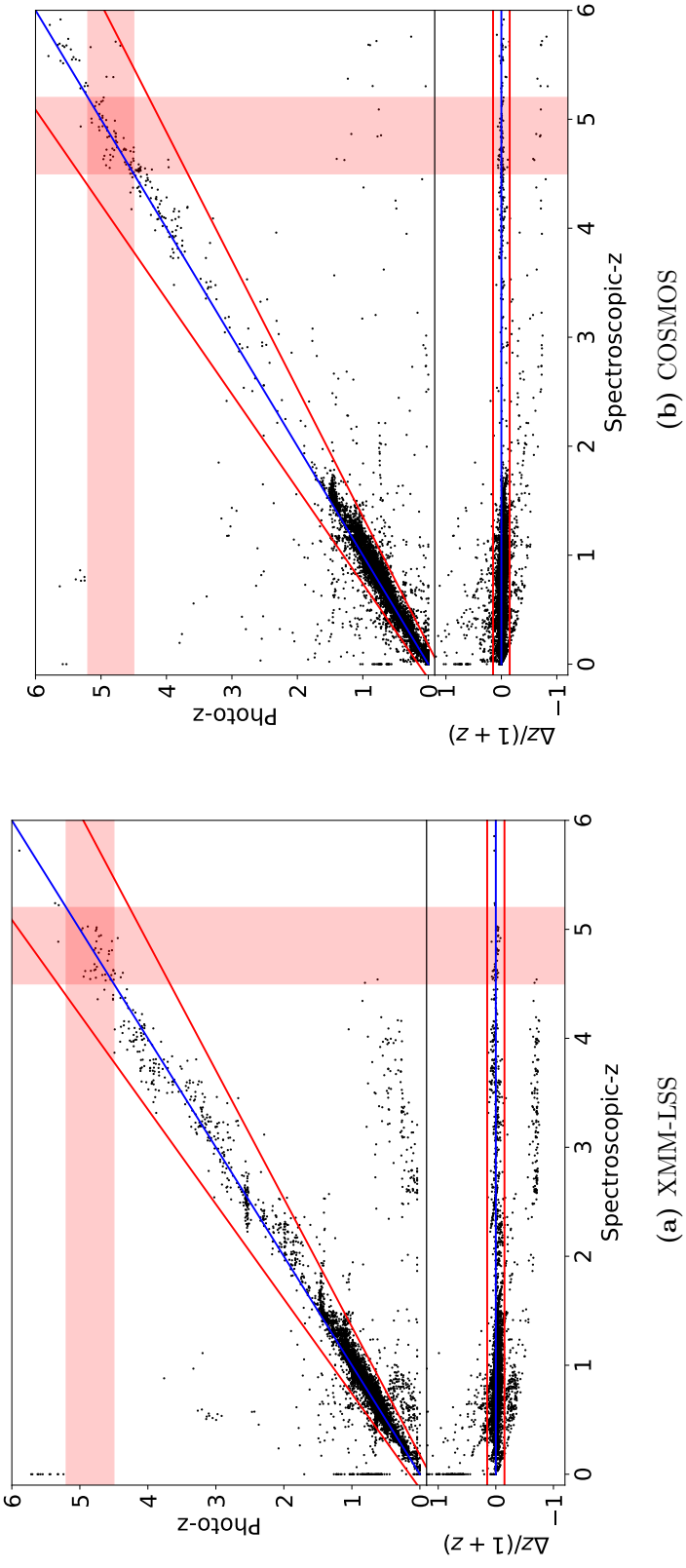
The SED fitting process was conducted in two stages. In the first stage, I cross-match my catalogues to a spectroscopic sample which compiles results from the VVDS (LeFèvre et al., 2013), VANDELS (McLure et al., 2018; Pentericci et al., 2018; Garilli et al., 2021), Z-COSMOS (Lilly et al., 2009), DEIMOS-10K (Hasinger et al., 2018), VIPERS (Scodeggio et al., 2018) and FMOS (Silverman et al., 2015) surveys. Only objects with spectroscopic redshifts that have flags indicating a greater than 95 per cent confidence are used, providing a total spectroscopic sample of 14811 sources in XMM-LSS and 18811 sources in COSMOS. Within these two fields, 43/42 objects have a spectroscopic redshift between $4.5 < z < 5.5$ in XMM/COSMOS and 41/38 have a 5σ detection in my selection band (HSC-z). Compared to Chapters 2 and 3, these spectroscopic catalogues have been remade from the ground up with a number of improvements. The catalogue made by the HSC team implemented a mechanism where if one source had multiple spectroscopic redshifts from different surveys, the value assigned to it was the average of the surveys. This was found to lead to a number of potentially erroneous redshifts. Examining the individual (rather than averaged) redshift measurements often found one which agreed with the photometric redshifts derived in this work. Potential erroneous redshifts were found to be primarily driven by PRIMUS based results

[1]http://pono.ucsd.edu/ adam/browndwarfs/spexprism/

and so this survey is removed from consideration in the work conducted in this chapter. In addition, the VIPERS survey is updated to DR2 in this version of the spectroscopic catalogue. These catalogues are run through LePhare in its AUTO_ADAPT mode, which makes iterative adjustments to the zero-points of the photometric filters in order to optimise the results against the spectroscopic sample. A diverse spectroscopic sample is required in order to prevent these zero-point modifications from being biased towards a limited set of galaxy colours. The results of this process provides small offsets of 0–0.1mags across all bands. These offsets are then applied to the photometry and LePhare is run a second time on the full sample to obtain object classification and redshifts.

Comparisons to the spectroscopic sample can provide indications of the reliability of the photometric redshift estimations. This can be broken down into two numerical values. 1) The outlier rate: the fraction of photometric redshifts which disagree with the spectroscopic redshift by more than 15 per cent in $(1+z)$, and 2) The Normalised Median Absolute Deviation Hoaglin et al. (NMAD; 1983): a measurement of the spread of the photometric redshifts around the ground truth in a manner that is resistant to the few extreme outliers that are present, this is defined as $1.48 \times \text{median}[|\Delta z|/(1+z)]$. Across the COSMOS field I find an outlier rate of 3.1 per cent and a NMAD of 0.029, in XMM-LSS the outlier rate is 4.5 per cent and the NMAD is 0.031. Fig. 4.1 shows the spectroscopic redshifts against my photometric redshift estimates.

Focusing on the high redshift regime, I successfully recover (within 15 per cent of $1+z$) 41 out of 41 objects with a spectroscopic redshift between $4.5 < z < 5.5$ and a 5σ detection in my selection band within the XMM-LSS field. Within COSMOS I recover 36 out of 38 objects with the same criteria. Examining the SEDs of these two objects that were unsuccessfully selected shows that both have significant CFHT-u^* band detections as a consequence of blending, and hence do not have an observed Lyman break. Instead, the SED fit misidentifies the drop in flux in the I-band as being a Balmer break instead of the start of the Lyman Break and the objects are assigned a redshift around $z \sim 0.8$. Such objects are accounted for



(a) XMM-LSS (b) COSMOS

Figure 4.1: Comparisons between the photometric redshifts derived in this study and a large compilation of spectroscopic redshifts. On the left I show the results for XMM-LSS and on the right, the results for COSMOS. The blue line shows the one-to-one correlation in the ideal case, and the red lines define the 15 per cent margin in $1 + z$ that defines a significant outlier. The pink shading indicates the redshift bin $4.5 < z < 5.2$. The secondary-plot in each figure shows the same figure with the y-axis instead scaling with $\Delta z = (z_{\rm spec} - z_{\rm phot})$.

as part of my completeness simulations which are detailed in Section 4.3.1. There are also found to be no low redshift objects from the spec-z catalogue that have photoz's between $4.5 < z < 5.2$, with a small number creeping in at the upper end of the bin $5.2 < z < 5.5$. I discuss this further in the next subsection.

4.2.3 Sample selection

To produce a sample of $z \simeq 5$ galaxies, I first select 15614 objects which have a best-fitting SED (minimum χ^2) as being a galaxy or QSO within the redshift range of $4.5 < z < 5.5$, have a 5σ detection in the HSC-z band ($\sim 5\sigma$ rest frame UV detection) and a $< 3\sigma$ CFHT-u^* measurement, if the object lies within the 90 per cent of the area covered by this filter. The strong Lyman break exhibited by galaxies at these redshifts should result in a non-detection in the CFHT-u^* band and so such a cut removes the potential for lower redshift contaminants. However, it was immediately obvious that these cuts alone were insufficient to produce a pure sample of $z \simeq 5$ galaxies. Examining the redshift distribution of the sample revealed a large spike in number counts for luminous objects ($M_{\rm UV} < -22$) with $z > 5.2$. Such a spike in number counts can be attributed to Milky Way brown dwarfs, particularly M-class dwarfs whose optical/NIR colours become degenerate with high-redshift galaxies at $z > 5.2$ (see Fig. 4.2 for an example of the colours of $z \sim 5$ galaxies and brown dwarf stars around the redshifted Lyman Break). The inclusion of the brown dwarf templates from the SpecX dataset was found to reduce the number of ultra-luminous sources at $z > 5.2$ by greater than a factor of two, but a significant spike in luminous objects remained.

A possible solution to this problem is the inclusion of *Spitzer* IRAC data, which can break the degeneracy between galaxies and brown dwarf stars. This is because the brown dwarf stars broadly follow Planck's law and decrease in luminosity towards the mid-infrared, while galaxies transition into a dust-driven increase in luminosity. However, deep Spitzer data has a larger point spread function and the subsequent source blending issues can further complicate selection. A second solution would be to introduce an upper limit of $z < 5.2$ to the UV LF. Such a cut

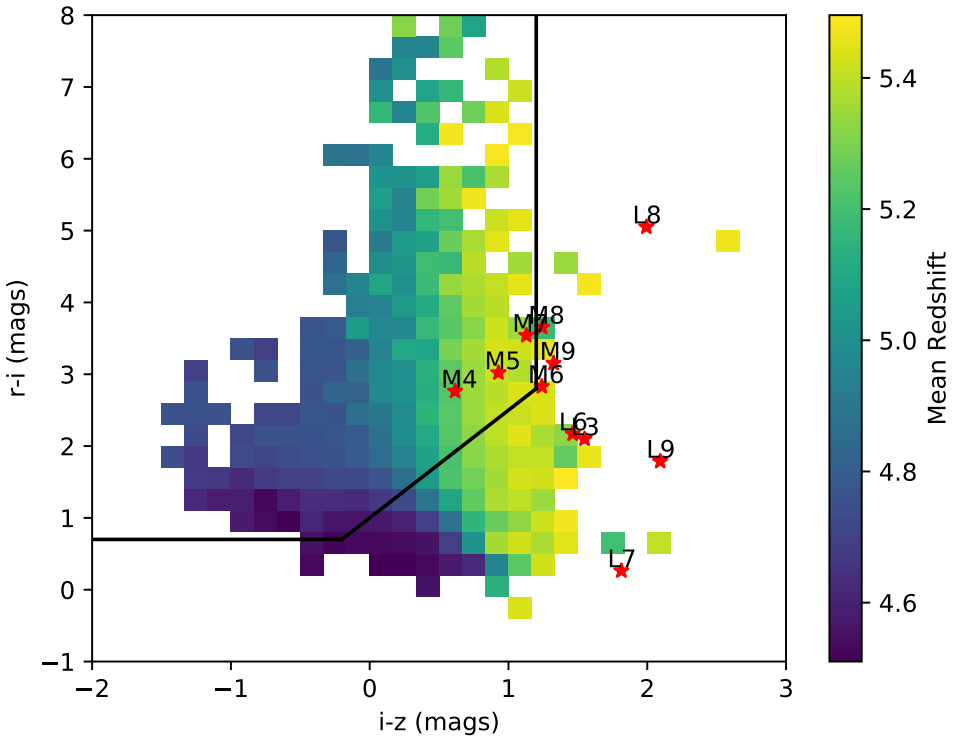


Figure 4.2: A colour-colour diagram showing the initial $4.5 < z < 5.5$ photometric redshift sample around the photometric bands that span the Lyman Break. I group the sample into a 30x30 grid to show how the mean redshift changes across the colour space. Overplotted in solid black lines is the selection criteria employed by the Ono et al. (2018) to select galaxies at $z \sim 5$ which selects sources in the upper left quadrant. I also show the colours of typical M and L-class brown dwarf stars as the red points and highlight that they overlap with the colour space probed by galaxies with redshifts towards the upper limits ($z > 5.2$).

only causes a small shift in the mean redshift of the UV LF from $\overline{z} = 4.9$ to $\overline{z} = 4.8$ and greatly minimise the overlap in colour space between galaxies and brown dwarfs. I thus proceed with restricting the sample to $4.5 < z < 5.2$ and apply an additional cut of $\chi_{\rm best} < 100$ which removes the worst fit ~ 1 per cent of objects from the sample. Because of the potential threat of brown dwarf contamination, the χ^2 cut is more stringent than in the previous chapters. This results in a total of 12,253 galaxy and AGN candidates with which I measure the UV LF.

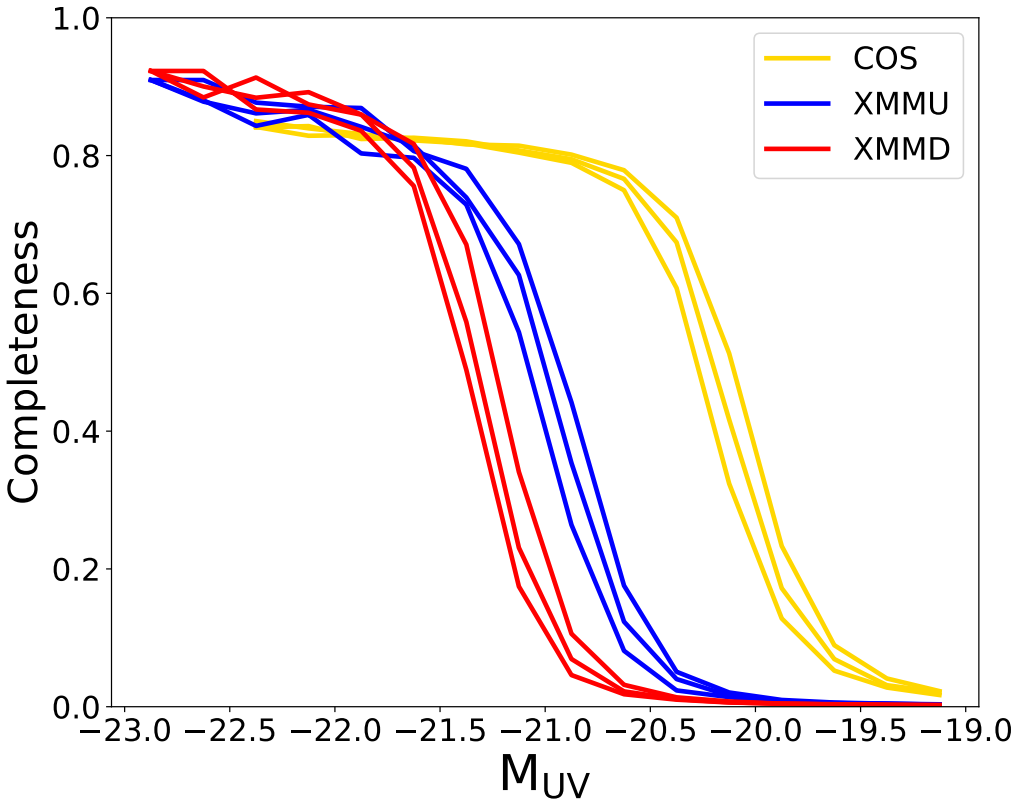


Figure 4.3: The completeness curves as derived from my completeness simulation in each of the three sub-fields, red for XMMD, blue for XMMU and yellow for COSMOS. For each sub-field I show three lines representing the variability of completeness across the full redshift bin $4.5 < z < 5.25$. Each line is a bin of 0.25 in redshift and from left to right is high redshift to low redshift. Higher redshift bins are less complete due to the impact of distance on apparent luminosity. Deeper fields are also more dense due to the presence of fainter sources, so have a higher probability of blending and hence converge to lower values of completeness at the bright-end.

4.3 Methods

4.3.1 Selection completeness

The incompleteness of the sample selection can be described primarily through two effects. The first of these is the probability that an object is lost due to partial or total blending with a secondary source in the image plane. The second is the dependence on $M_{\rm UV}$, redshift and location within the imaging data on whether a source meets the magnitude cut corresponding to the average 5σ depth in the HSC-z band.

The first effect can be corrected for by examining the segmentation map generated

by SExtractor for each image. The ratio between the number of unoccupied sky pixels can be compared to the total number of unmasked pixels to provide the percentage chance that a randomly positioned source would have significant overlap with another. In this situation, I assume that a highly blended source (one where the centre of the source lies on a pixel already occupied by another) is irretrievable. To correct for the second effect, I conduct a simple simulation where 3.6 million fake sources are injected into the images (in iterations of a few thousand to prevent artificially making the image over-dense). These sources have an assumed profile that is described by a Sèrsic index of $n = 1$ (Sérsic, 1963; Conselice, 2014) and have an intrinsic ultraviolet luminosity that is drawn from the redshift evolution of the UV LF derived in Bouwens et al. (2015). This ensures that I account for Eddington bias, where the larger number of faint sources have a chance of being scattered above the 5σ detection limit. Simulated sources have a half-light radius that is sampled from the results by Huang et al. (2013) at $z \sim 5$ and are convolved with the PSF model. My observations however, are seeing-dominated and so this make negligible difference to my results.

We restrict the simulation from placing galaxies with a central coordinate that is occupied by another source. This prevents double counting the first effect and allows instances of partial ($< 50\%$) blending to occur. The recovery rate as a function of redshift and UV luminosity is then scaled by the ratio derived for the loss of galaxies due to heavy/total blending in the image. The derived completeness curves are presented in Fig. 4.3.

The photometric redshift calculations and additional sample refinements produce a sample of 12,253 galaxies and AGN within the range $4.5 < z < 5.2$ across $5.95\,\mathrm{deg}^2$ of the combined XMM-LSS and COMSOS fields. I now proceed to measure the UV LF of these sources.

4.3.2 The 1/Vmax method

I use the $1/V_{\mathrm{max}}$ (Schmidt, 1968; Rowan-Robinson, 1968) to measure the UV LF of my sample. I calculate the maximum observable redshift z_{det} by iteratively

shifting the best-fit SED of each source in small steps of $\delta z = 0.01$ and convolve this with the selection filter (HSC-z) until the galaxy falls below the 5σ detection threshold. A maximum volume in which the object could have been detected (V_{max}) is thus the co-moving volume contained within the range $z_{\rm min} < z < z_{\rm max}$, where $z_{\rm min} = 4.5$ and $z_{\rm max} = \min(5.2, z_{\rm det})$. From this, the rest frame UV LF ($\Phi(M)$) is be calculated using:

$$\Phi(M) d\log(M) = \frac{1}{\Delta M} \sum_{i}^{N} \frac{1}{C_{i,f} V_{max,i}}, \tag{4.1}$$

where ΔM is the width of the magnitude bins and $C_{i,f}$ is the completeness correction for a galaxy i depending on its location within the sub-fields, f.

The measured uncertainty of the LF is given by:

$$\delta\Phi(M) = \frac{1}{\Delta M} \sqrt{\sum_{i}^{N} \left(\frac{1}{V_{max,i}} \right)^2}. \tag{4.2}$$

To balance number statistics with resolution in luminosity, I utilise three bin widths of $\Delta M = 0.25, 0.5, 1.0$. Bin widths of 1.0 are used within the AGN UV LF at $M_{\rm UV} < -23.5$; widths of 0.5 are used in the intermediate regime of $-23.5 < M_{\rm UV} < -23$ and widths of 0.25 are used for $M_{\rm UV} > -23$ in the galaxy luminosity function. Following Chapter 3, the absolute UV magnitude ($M_{\rm UV}$) is calculated by convolving a 100Å top-hat filter, centred at 1500Å within the rest frame, on the best fitting SED of each object. In the final measurement, I do not consider luminosity bins where the completeness across each sub-field over the full redshift range is below 50 per cent. While less than 50 per cent complete, bins fainter than my imposed limits contain large number of candidates due to the shape of the UV LF, this results in the loss of a few thousand galaxies and a final sample of 8309 sources used in measuring the UV LF.

Cosmic variance

As the UV LF is measured using a finite volume of the Universe and two sight-lines, it can be susceptible to biases resulting from the large-scale structure of the Universe. Finite sight-lines can cause non-representative conclusions to be

drawn and such an effect is commonly referred to as 'cosmic variance'. I follow the treatment described by Trenti & Stiavelli (2008) to determine the effects of this on my galaxy number counts. While number counts of galaxies increase towards fainter luminosities, the usable survey area decreases as components of the XMM field have different magnitude cuts. This leads to a fairly constant cosmic variance of 6.0-7.5 per cent across the range of absolute luminosities. To be conservative, I adopt the upper value found of 7.5 per cent across the UV LF, which is added in quadrature to my LF uncertainty from Equation 2. The $1/V_{\rm max}$ values calculated from the above processes are presented in Table 4.3.

4.3.3 Fitting the Luminosity Function

In this work, I simultaneously fit for the UV emission of galaxies and AGN instead attempting to separate them. The work conducted by Bowler et al. (2021) and summarised in Section 3.5.4 shows this method effectively enables for the transition between galaxy dominated UV emission and AGN dominated UV emission to be mapped without the need to rely on spectroscopy or morphology to conduct selections. To the AGN population, I fit the commonly used double power law (DPL). To the galaxy population, I trial both the Schechter and DPL functional forms in order to asses the quality of the fits and what impact their use has on the luminosity dependent fraction of sources with AGN dominated emission. It is worth noting that both the DPL and Schechter functional forms are not physically motivated and are used simply because they have been found to describe the basic shape of luminosity functions. The above methodology assumes that objects fall onto either the AGN LF or the galaxy LF. In reality, sources classed as AGN are galaxies with inflated UV emission and subsequently, these small numbers of host galaxies will be missing from the galaxy LF.

To conduct the model fits, I combine my LF results with those of other studies that probe luminosity regimes beyond what is possible with the dataset used here. I use Niida et al. (2020) and Bouwens et al. (2021) for the brightest AGN and faintest galaxies respectively. These are selected on the basis of having similar

mean redshifts for their samples ($\overline{z} = 4.9$) compared to my own and their ability to constrain the UV LF beyond the luminosity range probed by my own results. I use data points from Niida et al. (2020) for $M_{\rm UV} < -24.5$ and data points from Bouwens et al. (2021) for $-20.2 < M_{\rm UV}$, providing total coverage of $-28.5 < M_{\rm UV} < -16$. I fit this total LF with the sum of a DPL for the AGN component and either a Schechter or DPL for the galaxy population.

The two models are fit to the data using a Markov-Chain Monte Carlo (MCMC) implemented using `emcee` (Foreman-Mackey et al., 2013). The MCMC has 500 walkers which each burn in for 100,000 steps before mapping the posterior distribution for a further 20,000 steps. The walkers are initially distributed uniformly over a wide parameter space and priors for each parameter are set to be wide and uniform.

4.4 Results

In Fig. 4.4 I show the the $1/V_{\rm max}$ points and the MCMC fits over the range in which my data is able to constrain the UV LF. In Fig. 4.7 I extend the luminosity range to cover the full AGN and galaxy UV LF. I provide the tabulated list of my data points, measured from the combined COSMOS and XMM-LSS fields, in Table 4.3. The best-fit parameters derived from my model fitting process are provided in Table 4.2. The sample spans $-25 < M_{\rm UV} < -20.5$ and covers the transition between AGN and galaxy dominance as well as the 'knee' of the galaxy LF. My measured number densities for the galaxy population are in close agreement with results from recent literature where there is overlap, particularly those from Bouwens et al. (2021) and Ono et al. (2018). In the AGN regime, the lower cosmic volumes in my fields limit my ability to probe the AGN LF brighter than $M_{\rm UV} < -25$. At lower luminosities, where the data overlap, I find AGN space densities slightly lower than those found by Niida et al. (2020), although the significance is low.

4.4.1 The Galaxy LF

Comparing the quality of the two functional forms trialled to fit the galaxy UV LF, the DPL is found to perform better than the Schechter function. However, breaking

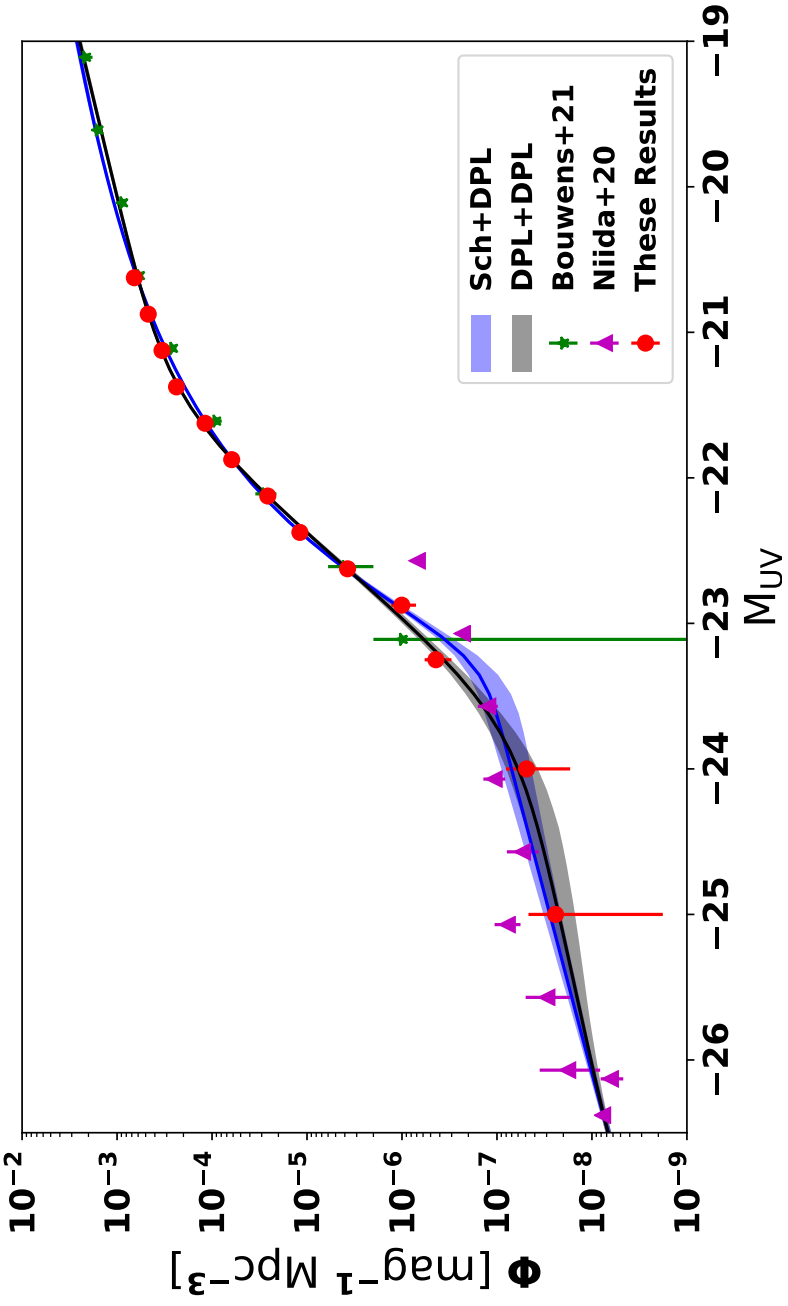


Figure 4.4: The $4.5 < z < 5.2$ UV LF measured in this study (red circles) and focusing on the transition regime between UV-bright galaxies and UV-faint AGN. The blue and black lines show the simultaneous fit of the AGN and LBG LF with a Schechter and DPL functional form for the LBGs respectively. The shaded regions indicate the 1σ uncertainty in the fit as found by sampling the posterior distribution. Other symbols are used to indicate data points from other studies used in the fitting of the UV LF, these include Bouwens et al. (2021) (green stars) and Niida et al. (2020) (violet triangles).

Table 4.2: The results of the MCMC fitting applied to the total UV LF. The first column lists the fitting parameterisation used. Columns 2–5 are the best fit DPL parameters for the AGN LF. Columns 6–9 show the best fit LBG parameters, either for a Schechter function or DPL. The final columns show the χ^2 and reduced χ^2 of the fit. The top two rows show primary results of the fitting process, where all variables are left free with wide priors. The bottom four rows show the impact of fixing the bright-end slope of the AGN LF to two different values that lie within the commonly found range in other studies.

	AGN				LBG				
	$\log_{10}(\Phi_{\rm AGN})$	$M^*_{\rm AGN}$	$\alpha_{\rm AGN}$	$\beta_{\rm AGN}$	$\log_{10}(\Phi)$	M^*	α	β	χ^2	$\chi^2_{\rm red}$
Sch+DPL	$-8.55^{+0.16}_{-0.13}$	$-27.49^{+0.18}_{-0.14}$	$-1.98^{+0.20}_{-0.14}$	$-5.99^{+1.00}_{-1.13}$	$-2.91^{+0.04}_{-0.04}$	$-20.92^{+0.05}_{-0.05}$	$-1.58^{+0.06}_{-0.05}$	–	42.52	1.46
DPL+DPL	$-8.43^{+0.19}_{-0.16}$	$-27.38^{+0.22}_{-0.17}$	$-1.79^{+0.29}_{-0.20}$	$-5.59^{+0.88}_{-1.15}$	$-3.55^{+0.05}_{-0.05}$	$-21.59^{+0.05}_{-0.05}$	$-1.91^{+0.04}_{-0.04}$	$-5.52^{+0.19}_{-0.17}$	24.31	0.87
Sch+DPL	$-8.18^{+0.14}_{-0.16}$	$-27.01^{+0.18}_{-0.20}$	$-1.71^{+0.39}_{-0.22}$	-4.0	$-2.92^{+0.04}_{-0.04}$	$-20.93^{+0.05}_{-0.05}$	$-1.59^{+0.05}_{-0.05}$	–	48.39	1.67
DPL+DPL	$-8.08^{+0.11}_{-0.14}$	$-26.89^{+0.14}_{-0.17}$	$-1.38^{+0.48}_{-0.34}$	-4.0	$-3.55^{+0.05}_{-0.05}$	$-21.58^{+0.05}_{-0.05}$	$-1.91^{+0.04}_{-0.04}$	$-5.47^{+0.17}_{-0.18}$	28.58	1.02
Sch+DPL	$-7.40^{+0.19}_{-0.18}$	$-25.62^{-0.28}_{+0.28}$	$-1.29^{-0.37}_{+0.30}$	-3.0	$-2.92^{+0.04}_{-0.04}$	$-20.92^{+0.05}_{-0.05}$	$-1.59^{+0.05}_{-0.05}$	–	62.72	2.16
DPL+DPL	$-7.44^{+0.18}_{-0.17}$	$-25.68^{+0.28}_{-0.25}$	$-1.06^{+0.33}_{-0.32}$	-3.0	$-3.55^{+0.05}_{-0.05}$	$-21.59^{+0.05}_{-0.05}$	$-1.91^{+0.04}_{-0.04}$	$-5.49^{+0.16}_{-0.18}$	41.71	1.49

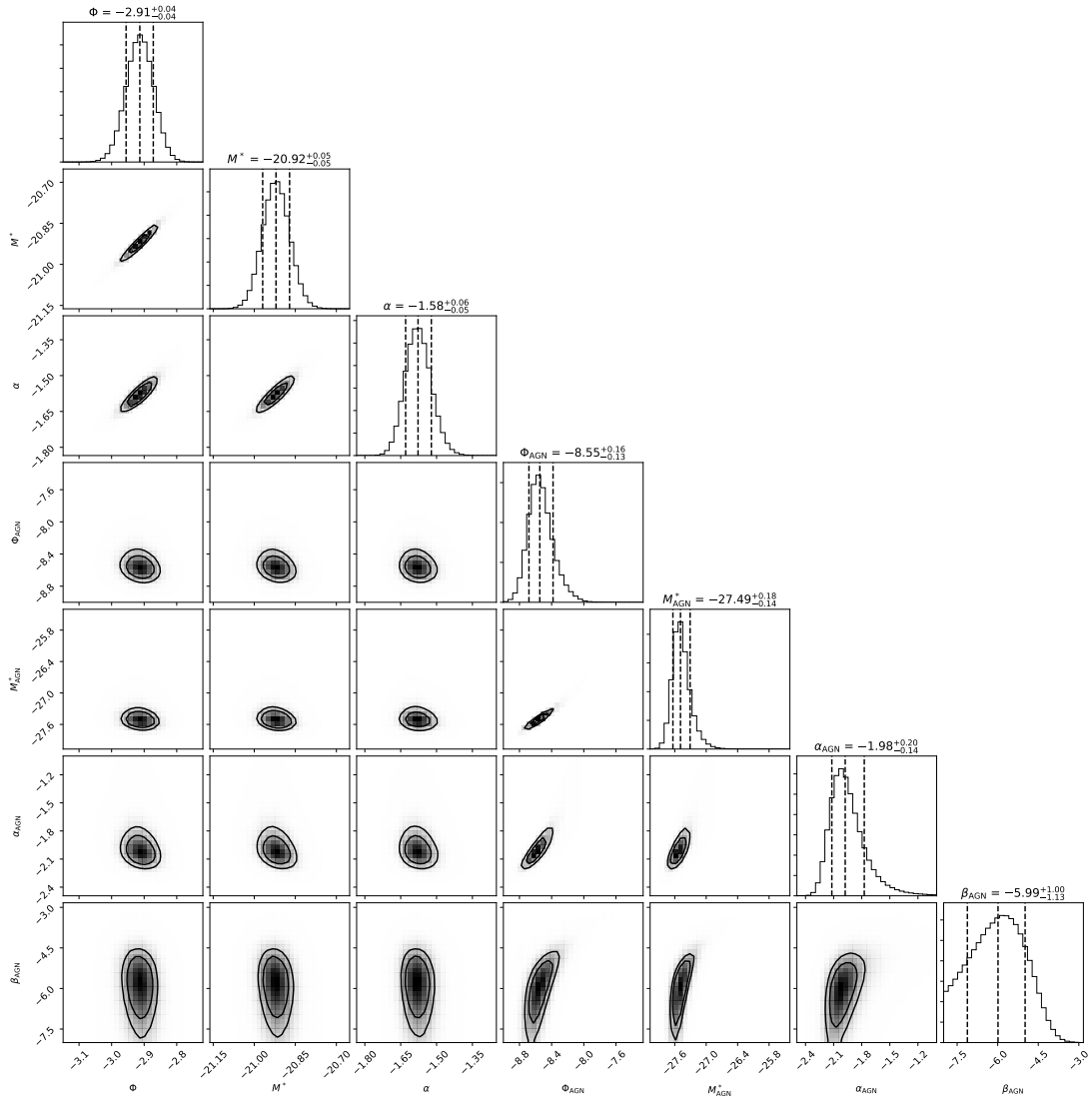


Figure 4.5: Contour plots showing the results of the MCMC process on the combined LF measurements from this work and those of Niida et al. (2020) and Bouwens et al. (2021). Displayed are the 1σ and 2σ confidence intervals for 1-D and 2-D posterior distributions for the Sch+DPL functional form. The parameter set labelled with the subscript 'AGN' refers to the DPL parameters for the AGN luminosity function. The remaining parameters describe the Schechter function for the galaxy luminosity function.

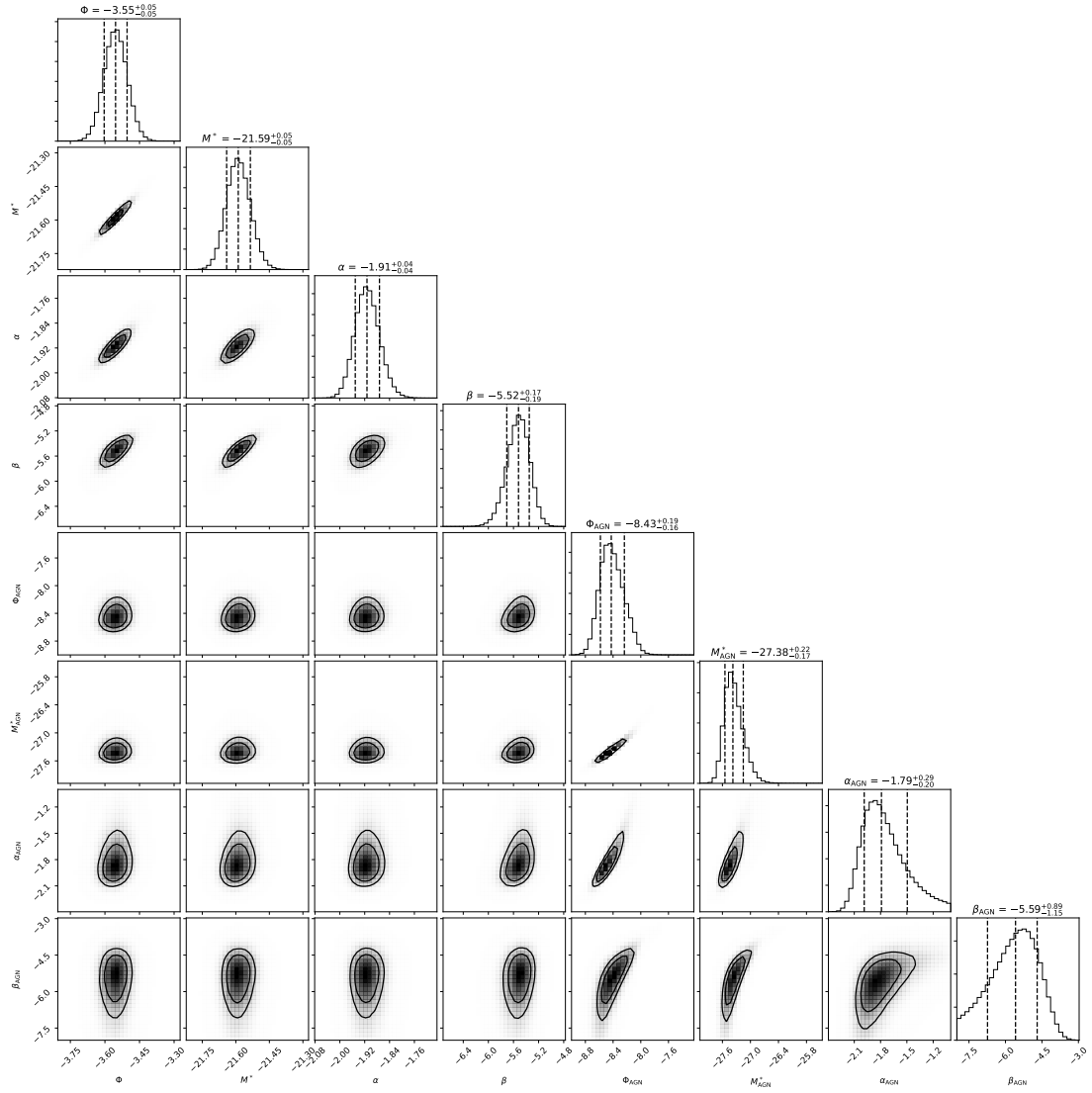


Figure 4.6: Contour plots showing the results of the MCMC process on the combined LF measurements from this work and those of Niida et al. (2020) and Bouwens et al. (2021). Displayed are the 1σ and 2σ confidence intervals for 1-D and 2-D posterior distributions for the DPL+DPL functional form. The parameter set labelled with the subscript 'AGN' refers to the DPL parameters for the AGN luminosity function. The remaining parameters describe the DPL function for the galaxy luminosity function.

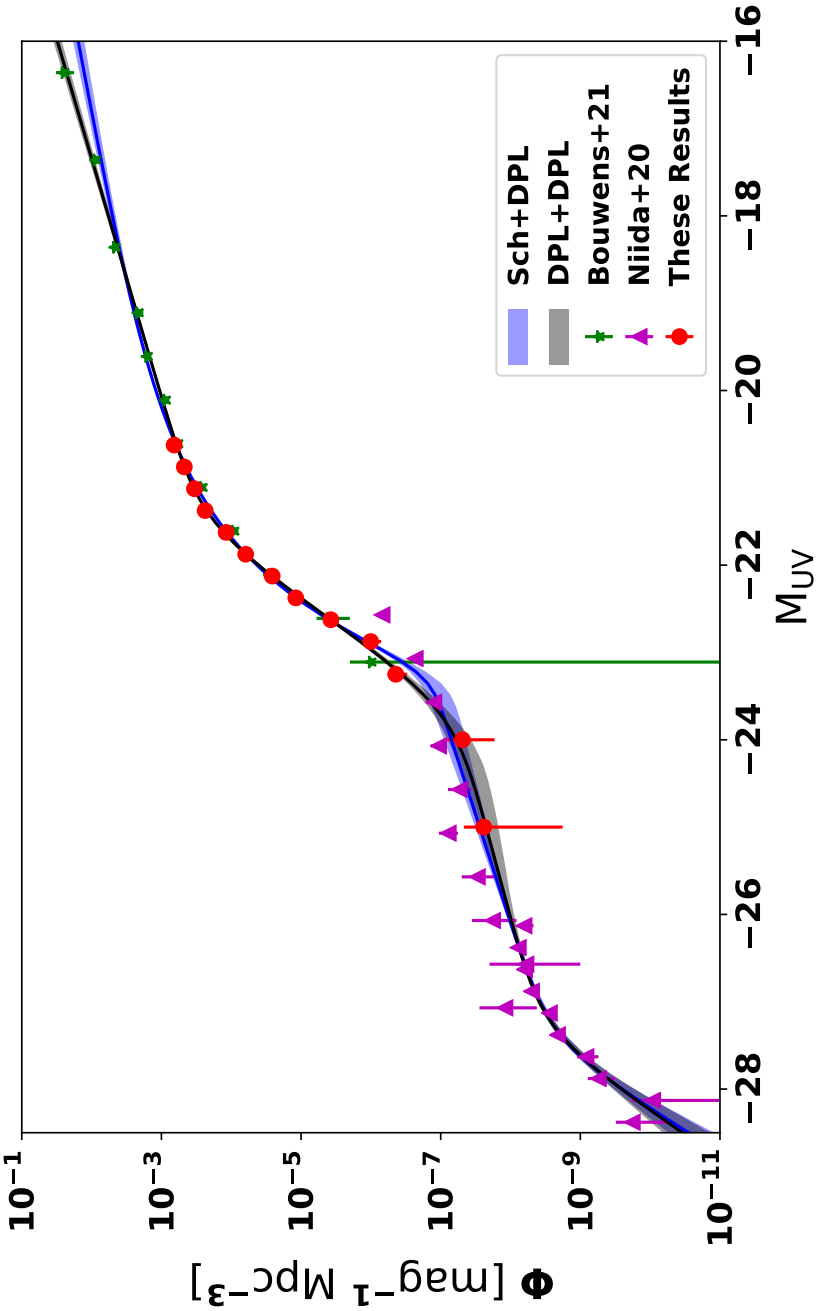


Figure 4.7: The full UV LF at $4.5 < z < 5.2$ measured in this study. The blue and black lines show the simultaneous fit of the AGN and LBG LF with a Schechter and DPL functional form for the LBGs respectively. The shaded regions indicate the 1σ uncertainty in the fit line as found by randomly sampling the posterior distribution 10,000 times. Symbols are the same as used in Fig.4.4.

Table 4.3: The measured rest-frame UV LF and its error margin at $4.5 < z < 5.2$. Column 1 shows the absolute UV magnitude at 1500Å. Column 2 shows the number density of objects and column 3 shows the errors in the number density which are calculated with equation 2.

$M_{\rm UV}$ [mag]	$\Phi(10^{-4})$ [mag^{-1}Mpc^{-3}]	$\delta\Phi(10^{-4})$ [mag^{-1}Mpc^{-3}]
−25.000	0.00024	0.00022
−24.000	0.00049	0.00032
−23.150	0.0044	0.0014
−22.875	0.010	0.003
−22.625	0.037	0.006
−22.375	0.189	0.013
−22.125	0.259	0.024
−21.875	0.621	0.051
−21.625	1.184	0.094
−21.375	2.364	0.186
−21.125	3.364	0.270
−20.875	4.697	0.369
−20.625	6.575	0.510

down the contributions from each data point towards the total χ^2 value reveals that the largest contributions arise from data points much fainter than the transition regime. Significant χ^2 contributions arise from the Bouwens et al. (2021) data points faintward of $M_{\rm UV} > -19$. In this luminosity regime, my Sch+DPL fit is found to underestimate the number density of these faint galaxies. The shallow faint-end slope is found to be driven by the data points between $-20.75 < M_{\rm UV} < -19$, where there is a transition between the use of my own data and that of Bouwens et al. (2021). A small discontinuity as a result of the slightly different mean redshifts, cosmic variance or similar effects resultant from combining two different studies could thus drive a shallower slope and be the reason the Sch+DPL functional form fails to replicate observations at the very faint end.

Contrarily, the DPL+DPL functional form is found to be capable of successfully replicating observations across the full luminosity range. The greatest contributor to the total χ^2 of the DPL+DPL fit is the Niida et al. (2020) data point located at $M_{\rm UV} = -25.05$. From $M_{\rm UV} \sim -23.25$, the UV LF data points are found to be in excess of the result obtained from the Sch+DPL fit but well matched

by the DPL+DPL fit. This contrast is the result of the DPL+DPL functional form providing greater numbers of highly luminous galaxies $M_{\rm UV} < -22.5$. By $M_{\rm UV} \sim -23.5$ the number density of galaxies found by both functional forms differ by an order of magnitude. The result of this is a transition which is brighter and more gradual between galaxy and AGN dominated UV emission in the DPL+DPL case.

4.4.2 The AGN LF

Our fits to the AGN LF favour a bright value for the characteristic luminosity ($M^*_{\rm AGN} \sim -27.5$) and a steep bright-end slope, in agreement with the results from McGreer et al. (e.g. 2013); Yang et al. (e.g. 2016); McGreer et al. (e.g. 2018) but in contrast to results which favour a fainter characteristic luminosity of AGN$^* \sim -25$ (e.g. Niida et al., 2020). However, the constraints on the bright-end AGN population ($M_{\rm UV} < -28$) are poor and the bright-end slope is subsequently measured to have very large errors. Past measurements of the AGN LF at $z \sim 5$ have measured the bright-end slope $\beta_{\rm AGN}$ to be within the range of $-4 < \beta_{\rm AGN} < -2.5$ and many studies have elected to fix $\beta_{\rm AGN}$ as part of their fitting processes (McGreer et al., 2013, 2018; Niida et al., 2020). To examine the impact of fixing the value of $\beta_{\rm AGN}$, I repeat my fits with $\beta_{\rm AGN}$ fixed to two common values, -4 and -3. The results using these fixed β values are presented alongside the main result in Table 4.2.

These fits show that fixing the bright-end slope $\beta_{\rm AGN}$ to shallower values causes a significant shift in the remaining DPL parameters. In particular, the characteristic luminosity shifts fainter and the faint-end slope becomes flatter. This degeneracy between the DPL parameters explains the difference between my results and those found by Niida et al. (2020), showing that a fainter value of $M^*_{\rm AGN}$ can be obtained by fixing $\beta_{\rm AGN}$ to shallower values. While the choice to fix the bright-end slope can lead to large shifts in the DPL parameters, the consequence of this on the AGN fractions displayed in Fig. 4.8 is found to be negligibly small. Significant deviations in number densities only occur when extrapolating to the most extreme of luminosities ($M_{\rm UV} \ll -28$ and $M_{\rm UV} \gg -22$). This will be important consideration when estimating AGN contributions to the budget of ionising photons.

Figure 4.8: The fraction of objects estimated to be AGN as a function of absolute luminosity, found by taking the ratio of the best-fit AGN and LBG LF from Table 4.2. Solid lines are the results when using a Schechter function to describe the galaxy LF and the dashed lines are when a DPL functional form is used. The blue lines are the results from this study and the gray lines are the results at ~ 4 from Chapter 3 (Adams et al., 2020). Also shown is the AGN fraction at $z \sim 5$ measured by spectroscopic follow up of a fraction of the sources identified in Ono et al. (2018).

4.5 Discussion

4.5.1 The transition between AGN and LBG dominance

Within the luminosity range $-24, M_{\rm UV} < -23$ there is a clear transition in the UV LF from the bright-end galaxy population to the faint-end AGN population. While the two functional forms to describe this transition remain difficult to distinguish, in Fig. 4.8 I show that they provide notably different predictions for how the AGN fraction of objects changes with absolute UV luminosity. The transition of galaxy to AGN dominance in UV emission is much sharper in the case where a Schechter function is used to describe the bright-end of the galaxy population than in the case where a DPL is used. The 50 per cent crossover is found to occur at $M_{\rm UV} = -23.65$ in the DPL+DPL case and $M_{\rm UV} = -23.1$ in the Sch+DPL case. Compared to results from $z \sim 4$ in Chapter 3, the transition from AGN to galaxy dominance is shifted brightwards by around 0.25 mags at $z \sim 5$. This is primarily a result of the significant drop in the normalisation of the AGN LF between these two redshift bins, where the AGN number density is measured to be around 0.8 dex lower at $z \sim 5$ compared to $z \sim 4$, while evolution in the galaxy population is much lower (0.1-0.2 dex).

Distinguishing between the two functional forms is necessary when considering the number density of ultra-luminous galaxies ($M_{\rm UV} < -23$). By $M_{\rm UV} \sim -23.5$, the two functional forms provide galaxy number densities that contrast by ~ 1 dex. Because galaxy dominated UV emission is extended a further 0.25mags brighter at $z \sim 5$ compared to $z \sim 4$, it should be easier for observations to determine if the galaxy LF follows a Schechter or DPL functional form as more of the bright-end slope is available without significant AGN contamination. My observations around and brighter than the 'knee' of the galaxy LF suggest that the DPL+DPL functional form is a better fit to the total UV LF. In particular, the DPL+DPL functional form is capable of successfully reproducing my observations in the key luminosity regime of $-24 < M_{\rm UV} < -22.5$.

The evolving comoving space density of galaxies with $M_{\rm UV} < -22.5$ provides insight into the build up of the most massive/luminous objects in the early universe and the impact of high-luminosity sources on cosmic reionisation (e.g. Ouchi et al., 2009; Bowler et al., 2017; Stefanon et al., 2019; Bowler et al., 2020; Forrest et al., 2020; Neeleman et al., 2020). Such objects are also rare enough that hydrodynamical simulations do not yet provide sufficient volumes to recover them (Feng et al., 2016; Vijayan et al., 2021), meaning wide area surveys are presently the primary way to gain insight into these objects.

As a consequence of the strong evolution in the AGN space density from $z \sim 4$ to $z \sim 5$ (see also McGreer et al., 2013; Jiang et al., 2016), number statistics at $M_{\rm UV} < -23.5$ are relatively poor compared to those at $z \sim 4$ from Chapter 3. Past AGN focused studies that probe faint AGN ($M_{\rm UV} > -24$) with wide survey areas (e.g. McGreer et al., 2013; Niida et al., 2020) all employ criteria that favourably select AGN which are unobscured, which are typically characterised as being point-source-like with broad emission lines. Work by Bowler et al. (2021) shows that for $M_{\rm UV} > -24$, light from the host galaxies becomes increasingly important for sources containing an AGN. This means that the use of the data points from these AGN studies that employ very wide areas for luminosities $M_{\rm UV} > -24$ would not solve my number statistics issue since their own selection methods would exclude AGN with large host-galaxy contributions or could potentially include bright but compact galaxies. In order to further improve number statistics within the luminosity regime of $-24 < M_{\rm UV} < -23$, a replication of the methodology undertaken here on additional extragalactic fields will be required.

4.5.2 Evolution in the galaxy population

To explore the time evolution of the galaxy LF, I compare the results derived in this study to those obtained at $z \sim 4$ in Chapter 3. Focusing on the case where the galaxy UV LF follows a Schechter function, the measured values of M^* are found to be consistent with the M^* values measured at $z \sim 4$. The lack of evolution in the characteristic luminosity over this time frame agrees with the findings of

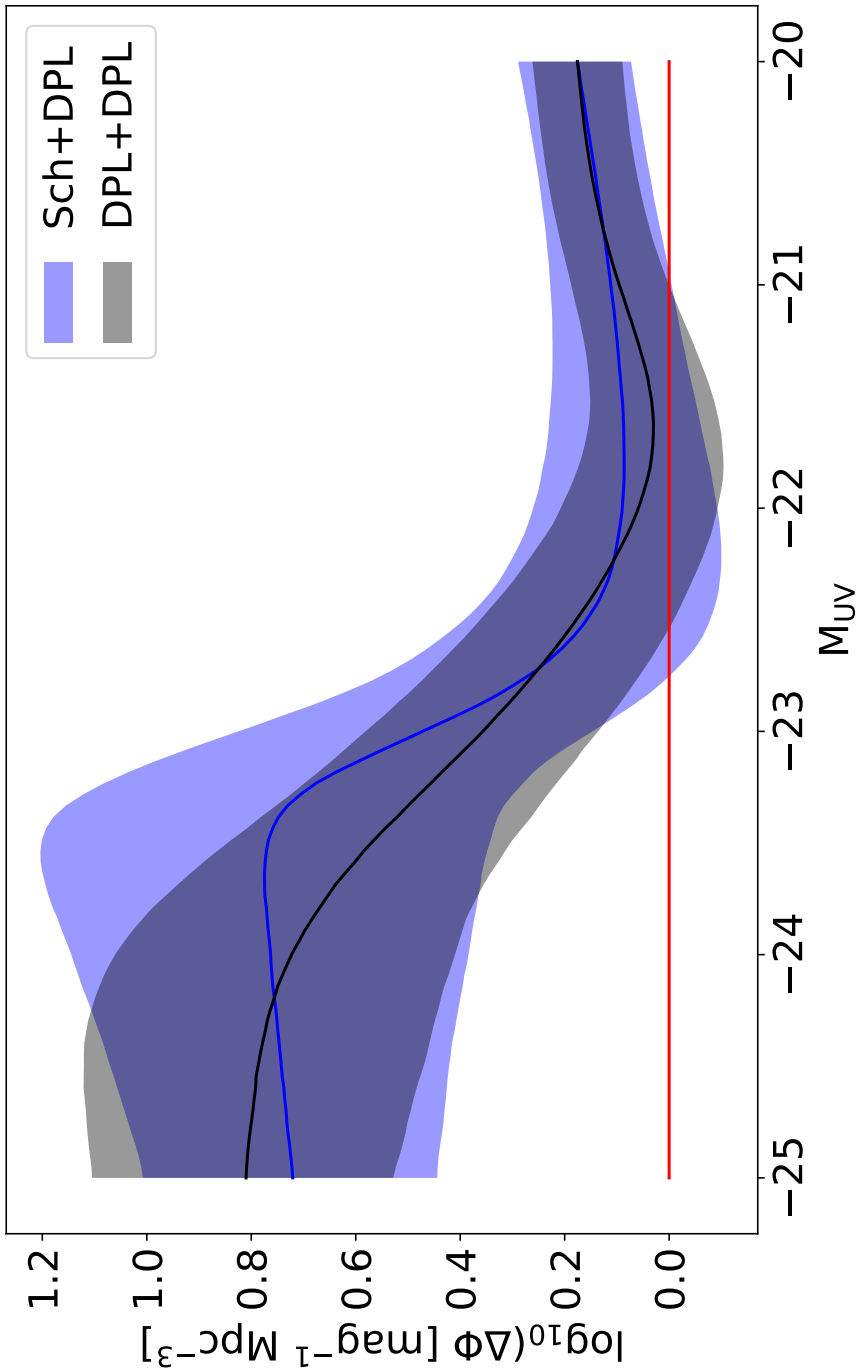


Figure 4.9: The logarithmic change in the number density of all sources from $z \sim 5$ to $z \sim 4$ using the measured total luminosity function as measured in this study and Chapter 3 for $z \sim 4$. The two functional forms are shown using the same colour scheme as Fig. 4.4, with Sch+DPL in blue and DPL+DPL in grey. Contours are generated by calculating the ratio between 10,000 randomly sampled PDFs of the $z \sim 4$ LF and $z \sim 5$ LF. The red line highlights the zero line, which indicates no evolution between $z \sim 5$ and $z \sim 4$. Here it is clear that the AGN population exhibits the strongest evolution between these two time periods, with mean increase of 0.8dex while the evolution of the galaxy population is more mild (< 0.2 dex).

other studies using a Schechter function in this redshift regime (e.g. Bouwens et al., 2015; Finkelstein et al., 2015; Bouwens et al., 2021). Two primary limitations for producing highly luminous ultraviolet galaxies are AGN feedback and dust obscuration. AGN feedback quenches massive galaxies, leading to a reduction in their star formation rates and ultraviolet emission (e.g. Cole et al., 2002; Benson et al., 2003; Begelman, 2004; Bower et al., 2006), while dust absorbs ultraviolet light from highly star-forming galaxies and re-emits in the far infrared (e.g. Gonzalez-Perez et al., 2013; Ma et al., 2019). A constant characteristic mass indicates that the impact of these two processes on the galaxy population is unchanging over the redshift range $3.5 < z < 5.2$. Supporting evidence for this conclusion includes the measured stability in the impact of dust obscuration on the ultraviolet output of highly star-forming galaxies at high redshifts (Bouwens et al., 2009; Reddy et al., 2010). In addition, studies of stellar mass functions of passive and star-forming galaxies, using both observations and simulations, also find that additional quenching arising from AGN feedback are not required at early times ($z > 3$) (e.g. Kaviraj et al., 2017; Davidzon et al., 2017b) and supports my findings.

When examining the stellar mass function of galaxies during this time, it is found that massive galaxies ($\log_{10}(M_*/M_\odot) > 9.5$) increase in space density by 0.1–0.2dex between $z \sim 5$ and $z \sim 4$. (e.g. Duncan et al., 2014; Song et al., 2016b). Under the observed relation between stellar mass and UV emission (M_* vs $M_{\rm UV}$; Duncan et al., 2014; Song et al., 2016b), it is these galaxies that are expected to contribute towards the space density of UV luminous galaxies ($M_{\rm UV} < -21$). Since the increase in the space density of massive galaxies is of the same magnitude as the increase in the space density of UV luminous galaxies in my sample, there has not yet been a significant mass-dependent shift in star formation rates. Thus, I find no evidence that AGN feedback has begun to significantly quench these galaxies, even though the AGN space density has increased significantly during this time period.

The faint-end slope (α) of the Schechter function is also found to be consistent between $z \sim 5$ and $z \sim 4$. This lack of evolution in α contrasts against the findings in Bouwens et al. (2015); Finkelstein et al. (2015); Bouwens et al. (2021),

who find α to increase with cosmic time. An increase in α (a flattening of the faint-end slope) is proposed to be primarily due to the evolution in low-mass dark matter halos in the halo mass function (e.g. Bouwens et al., 2008; Trenti et al., 2010; Tacchella et al., 2013; Bouwens et al., 2015). Examining my fits in the faint galaxy regime ($M_{\rm UV} > -19$) reveals that the Schechter function fit in this work underestimates the numbers of galaxies compared to Bouwens et al. (2021). This shallow slope is driven by the data points that cover the transition from my own work and the use of the data from Bouwens et al. (2021) ($-20.75 < M_{\rm UV} < -19$). Consequently, the lack of α evolution between $z \sim 4$ and $z \sim 5$ could be the result of a discontinuity between my data points and those of Bouwens et al. (2021), caused by slightly different selection criteria, mean redshifts, underestimated cosmic variance or a combination of all the above.

However, the results from this study indicate that the DPL functional form provides the better fit to the galaxy LF. In this case, M^* is found to evolve strongly with time, unlike results with the Schechter function. My fits match the results of Ono et al. (2018) and Bowler et al. (2020), both of whom fit the DPL functional form to the UV LF at $z > 4$. Instead, the evolution of the LF is described by a brightening of the characteristic luminosity and a steepening of the bright-end slope with decreasing redshift. An explanation for the steepening bright-end slope and build-up of galaxies at M^* is a gradual rise in dust content and, later on, AGN-driven quenching within high-mass galaxies over the time period of $z \sim 9$ to $z \sim 4$ (Peng et al., 2010; Michałowski, 2015; McLure et al., 2018). In my fit, both the characteristic luminosity and bright-end slope (β) match the predictions made by Bowler et al. (2020), where they derive a linear evolutionary model. Extending the model to lower redshifts, the steepening of the bright-end slope will bring the DPL into closer agreement with the predictions of a Schechter function. The findings of Chapter 3 at $z \sim 4$ show that both the DPL and Schechter function are comparable in the goodness of fit and maintain agreement with the Bowler et al. (2020) evolutionary model within the uncertainties to $\sim 2\sigma$. At lower redshifts, studies such as Parsa et al. (2016) and Moutard et al. (2020) have shown that the

Schechter function provides a good fit to the bright end of the galaxy UV LF at $z < 3$ and that the space density of ultra-luminous galaxies ($M_{\rm UV} < -23$) begin to decrease, leading to a fall in the value of M^*.

Based on these results, there will be an observed transition from a DPL-favoured model, with a rising M^* for the UV LF at high redshifts, into one that more closely matches a Schechter function with a decreasing M^* at lower redshifts. At early times, this evolution is likely driven by rising number counts of galaxies that become increasingly dusty as time advances. AGN feedback then becomes increasingly dominant at $z < 4$, leading to mass-dependent quenching and reducing the numbers of ultra-luminous galaxies $M_{\rm UV} < -23$.

4.5.3 Evolution in the AGN population

An evolution in the shape of the AGN LF is capable of providing information on the changing black hole population. This is because the AGN LF is dependent on both the mass function of active black holes and the distribution of accretion rates (Hopkins et al., 2005). A change in the shape of the AGN LF can thus be indicative of a change in either of these two factors. My measurements of the faint-end slope of the AGN LF indicate that there is no significant evolution in $\alpha_{\rm AGN}$ from $z \sim 5$ to $z \sim 4$ when compared to results from Chapter 3. The simplest interpretation of this result is that the rise in the space density of luminous AGN can be attributed to an increase in the space density of active black holes. This is because a global increase in accretion rates would reduce the space density of the faintest AGN and ultimately flatten the faint-end slope. While the mass function of active black holes has proven difficult to measure at high redshifts (see e.g. Shankar et al., 2009; Cao, 2010; Kelly & Shen, 2013; Habouzit et al., 2021), there is a consensus that the normalisation in the mass function of active black holes does increase from high redshifts down to $z \sim 2$ (e.g. Croom et al., 2009; Labita et al., 2009; Ikeda et al., 2012; Niida et al., 2016). Therefore, my results suggest that the increase in the space density of AGN at this time is fuelled by an increasing number of active black holes, rather than a change in the behaviour of the active population.

To compare the relative growth in the space density of AGN with time, I can employ an evolution parameter (k: McGreer et al., 2013, 2018; Niida et al., 2020), which is defined as the difference in dex between two luminosity functions at two different redshifts z_1 and z_2 ($k = \log_{10}(\Phi_{z_1}/\Phi_{z_2})/(z_1 - z_2)$). To calculate k in this time period, I sample the posterior distributions for the $z \sim 5$ and $z \sim 4$ luminosity functions 10,000 times and calculate the ratio of the resultant luminosity functions at these two time periods. I find that the space density of luminous AGN decreases from $z \sim 4$ to $z \sim 5$ by $k \sim -0.72 \pm 0.29$ at $M_{\rm UV} \sim -25$ and $k = -0.75 \pm 0.38$ at $M_{\rm UV} \sim -24$ when using the Sch+DPL model. In the DPL+DPL model, I find that $k = -0.82 \pm 0.28$ at $M_{\rm UV} \sim -25$ and $k = -0.68 \pm 0.24$ at $M_{\rm UV} \sim -24$. Applying this method of calculating k to the AGN LF derived in Niida et al. (2020) for $z \sim 5$ and Akiyama et al. (2018) for $z \sim 4$, the evolution parameter is found to be approximately $k \sim -0.5 \pm 0.12$ over the same luminosity range. My results are thus consistent with the findings of the evolving faint end found in Niida et al. (2020), while folding in the extra uncertainty due to contamination from bright galaxies. The measured value of k also matches the findings of studies towards higher redshifts $z \sim 6$ which report $k = -0.8 > k > 0.95$ at the faint end (Niida et al., 2020) and $k = -0.72 \pm 0.11$ in the bright end (Jiang et al., 2016), thus indicating that a strong evolution in the space density across the entire AGN LF persists over the redshift range $4 < z < 6$.

4.5.4 The number density of galaxies around the 'knee' at $z \sim 5$

To explore the number density of galaxies further, I break down my UV LF into the three constituent sub-fields in order to explore the impacts of cosmic variance on my measurements. Between my own fields, I find that relative to the COSMOS field the XMMU sub-field is overdense at $z \sim 5$ by less than 0.1dex while the XMMD sub-field is underdense by around 0.15dex. Recalculating the cosmic variance for each subfield individually provides values of 11-15 per cent across the luminosity function. Meaning the differences observed between these fields are

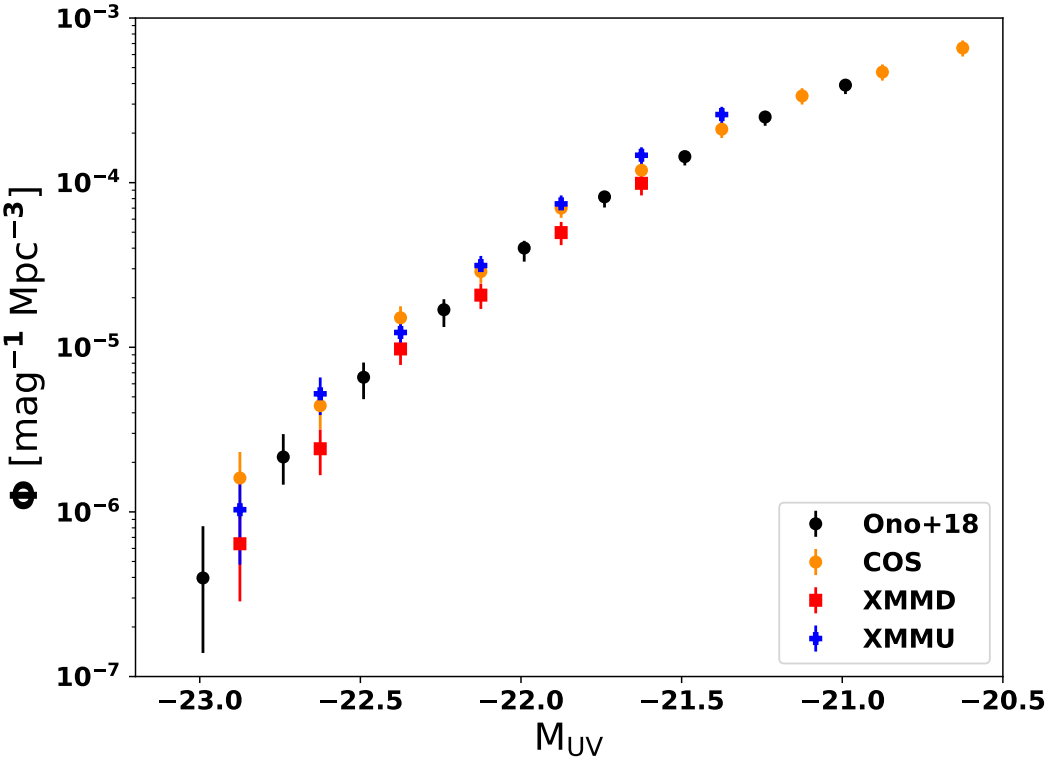


Figure 4.10: A zoom into the 'knee' of the $z \sim 5$ LF. I beak down my measured LF into the consituent sub-fields to show the scale of the impact of cosmic variance. Shown alongside are the results from Ono et al. (2018), the study most comparable to my own in depth/area combinations.

consistent to within 2σ of expected variability due to large scale structure. Two $1\,\mathrm{deg}^2$ fields used in van der Burg et al. (2010) are contained within my dataset (one in XMMD and one in COSMOS). I note that in both my work and in van der Burg et al. (2010), the COSMOS field is found to be comparatively higher in density at $z \sim 5$ than that of XMMD.

The primary study which has previously probed the bright-end of the galaxy population ($M_{\mathrm{UV}} < M^*$) has been Ono et al. (2018) which makes use of the shallower components of the HSC-SSP survey. Agreement with this study is found to be very strong across the regime of bright galaxies with $M < -20.5$. While the Ono et al. (2018) data points in Fig. 4.10 are corrected for the AGN fractions shown in Fig. 4.8, agreement between my studies remains as a consequence of relatively low AGN fractions for $M_{\mathrm{UV}} > -23$. The Ono et al. (2018) study utilises

colour-colour selections in the optical bands provided by HSC in order to identify galaxies at $z \sim 5$. As a consequence of this, completeness is comparatively low (around 60 per cent) and detailed studies involving a multi-band version of the completeness simulation conducted in this study and the degradation of their deepest images to examine scatter around selection limits were required. The use of VISTA data and photometric redshifts in my study helps to both reduce sources of contamination and increase completeness.

4.6 Conclusions

We utilise the final data release of VIDEO and DR4 of the UltraVISTA survey programme, alongside the optical photometry available in the XMM-LSS and COSMOS fields, to measure the rest-frame UV LF at $z \sim 5$ from $-25 < M_{\rm UV} < -20.5$ with 8309 galaxies and AGN selected using a comprehensive photometric redshift treatment. I combine these results with work from recent literature to increase the luminosity range available in order to model the total UV LF for sources in the range $-28 < M_{\rm UV} < -16$. My conclusions based on this are as follows:

1. The transition from AGN dominated UV emission to galaxy dominated emission is found to occur ~ 0.25mags brighter than the transition found at $z \sim 4$. This shift is primarily the result of the significant drop in the comoving space density of AGN from $z \sim 4$ to $z \sim 5$ (~ 0.8 dex) verses the fall in the space density of bright galaxies (< 0.2 dex).

2. Distinguishing which functional form better describes the galaxy population (Schechter or DPL) remains difficult in this redshift regime. I find that the DPL functional form provides a better fit and is capable of replicating observations of the galaxy LF across the full range of data available. The fit parameters agree strongly with those derived from Bowler et al. (2015) and Ono et al. (2018) and match the strong evolutionary trends found at higher redshifts ($z > 6$) by Bowler et al. (2020). The steepening of the bright-end slope with time explains why studies at lower redshifts ($z < 3$; Parsa et al.,

2016; Moutard et al., 2020) find that a Schechter function well describes observations. It also shows that elevated space densities of luminous galaxies at earlier times $z > 3$, compared to an exponential cutoff, are due to a lack of dust obscuration and quenching from AGN feedback.

3. Cosmic variance between the sub-fields used is found to agree within 2σ of predictions made using the treatment developed in Trenti & Stiavelli (2008). The use of $6\,\mathrm{deg}^2$of sky enables for uncertainty as a consequence of cosmic variance to be reduced to a level of 7.5 per cent and it remains the dominant source of uncertainty in the galaxy LF.

4. Fits to the faint-end slope of the AGN LF are found to be steep ($\alpha < -1.75$) and match those measured in Stevans et al. (2018) and Chapter 3 at $z \sim 4$. The reason for this is likely the non-exclusive nature of the sample selection procedure, enabling the sample to contain AGN which have significant UV contributions from their host galaxies as found in Bowler et al. (2021). A lack in the evolution of the faint-end slope can be explained by a simple rise in the normalisation of the active black hole mass function at this time.

5. Degeneracies between the DPL parameters used to describe the AGN LF are found to be significant within the commonly adopted values of the bright-end slope ($-4 < \beta < -2.5$). Contrasting results between past studies are likely a result of electing to fix β based on an independent handling of the bright-end before extending the model to the rest of the LF. I find that fixing the bright-end slope of the AGN LF leads to a negligible change in the luminosity dependent AGN fraction over the transition regime of $-24 < M_{\mathrm{UV}} < -23$. However, when it comes to extrapolating the AGN LF fainter still (e.g. to calculate AGN contributions to reionising photons), improved number statistics of ultra-bright AGN ($M_{\mathrm{UV}} < -27$) are also required in order more accurately determine the respective values of β and M^* and relieve the degeneracy on the faint-end slope.

I think that inside every adult is the heart of a child. We just gradually convince ourselves that we have to act more like adults.

— Shigeru Miyamoto

5 Conclusions and the Future

Contents

5.1 Overall Conclusions

In this thesis, I have presented the results from a series of studies targeting the evolving population of luminous and massive galaxies over 90 per cent of the Universe's lifetime. This has included measuring the stellar mass function of galaxies from $0.1 < z < 2.0$ and the ultraviolet luminosity function of both galaxies and AGN from $3.5 < z < 5.2$ in order to provide strong observational limits with which to further the development of theories of galaxy formation and evolution, in particular on AGN feedback, merger rates of massive galaxies and dust obscuration in the early Universe.

In Chapter 2, I provided a robust examination of the stellar mass function using $\sim 320,000$ galaxies at $0.1 < z < 2.0$. In this work I found that the use of *Spitzer* 3.6μm and 4.5μm photometry did not have a significant impact on the redshift distribution of sources calculated with photometric redshift template fitting methods. However, it was found to be influential in removing some of the degeneracy between Milky Way stars and massive galaxies at redshifts $0.5 < z < 0.75$ and $1.0 < z < 1.25$. I also used photometry derived from a new software package `ProFound` and show that the choice of source extraction software led to no significant change in the results ($< 1\sigma$ on each of the parameters of the double Schechter function). The use of 6deg^2 of survey area over two fields drastically reduces the impact of cosmic variance compared to previous studies and provides much greater number statistics for the high-mass end of the stellar mass function. This combination of depth and area enables the measurement of an evolving characteristic mass M^* that grows by over 0.2dex between $z \sim 1.5$ and $z \sim 0.5$, matching predictions from the hydrodynamical simulation EAGLE. I show that the massive population of galaxies is still growing over this redshift range. This is likely due to the merging of galaxies, since the high-mass end is found to be dominated by quenched galaxies with star formation rates of more than a factor of 10 lower than the star forming main sequence.

In Chapter 3, I use 46,904 sources at $z \sim 4$, with ~ 90 per cent completeness, to measure the transition between AGN- and galaxy-dominated ultraviolet emission in the ultraviolet luminosity function. The key finding was that both a Schechter function and DPL parameterisation were capable of describing the observed luminosity function, however they are found to quickly deviate at the high-luminosity end of the UV LF. For galaxies with luminosity $M_{\rm UV} < -23.5$, the number density between the two functional forms contrast by greater than an order of magnitude. This leads to a drastic difference in the luminosity-dependent AGN fraction, meaning that a study capable of accurately classifying sources with luminosity in the range $-24 < M_{\rm UV} < -23$ should solve whether the ultraviolet luminosity function of galaxies is better described by a DPL or Schechter function. This would be an important advancement because the true number densities of

ultra-luminous galaxies are necessary to constrain the impact of feedback processes on star formation, as well as determining the amount of dust in galaxies. Alongside these findings, my results show that the faint-end slope of the AGN luminosity function is steep $\alpha_{\rm AGN} < -1.7$, which indicates that faint or obscured AGN activity is more prevalent than originally thought at this time.

In Chapter 4, the work from the previous chapter is expanded to a higher redshift sample at $4.5 < z < 5.2$. In this work, I find that the DPL is better fit to the galaxy UV LF than a Schechter function. This distinction is possible because of the lower space density of AGN at this time compared to $z \sim 4$, enabling for more of the bright end of the LF to be captured before significant AGN contamination becomes an issue. The DPL functional form shows that there are more galaxies of very high UV luminosity ($M_{\rm UV} < -22.5$) at $z \sim 5$ than otherwise predicted in the past with Schechter functions. The observed evolution the the ultraviolet luminosity function of galaxies can be explained by a population of galaxies that become increasingly dusty with time. The onset of AGN feedback at lower redshifts ($z < 3$) explains why observations of the luminosity function at later times more closely follows the exponential fall off of a Schechter function. I also find that the AGN population exhibits a much stronger evolution in number density than the galaxy population between $z \sim 5$ and $z \sim 4$. This can be explained by a rise in the space density of active black holes and it results in a 0.25mag shift in the UV luminosity-dependent AGN fraction, with the transition between AGN and galaxy dominated emission occurring at brighter magnitudes at $z \sim 5$ compared to $z \sim 4$.

Combined, this work on deep optical and near-infrared surveys shows that there is considerable evolution in the number density of the most luminous and massive galactic systems throughout most of the lifespan of the Universe. In order to develop this work further, greater number statistics are required, alongside a deeper understanding of the systematic and observational errors behind stellar mass estimations. These will enable for better constraints of the bright/massive population of galaxies and AGN, and will also counter the effects of the Eddington

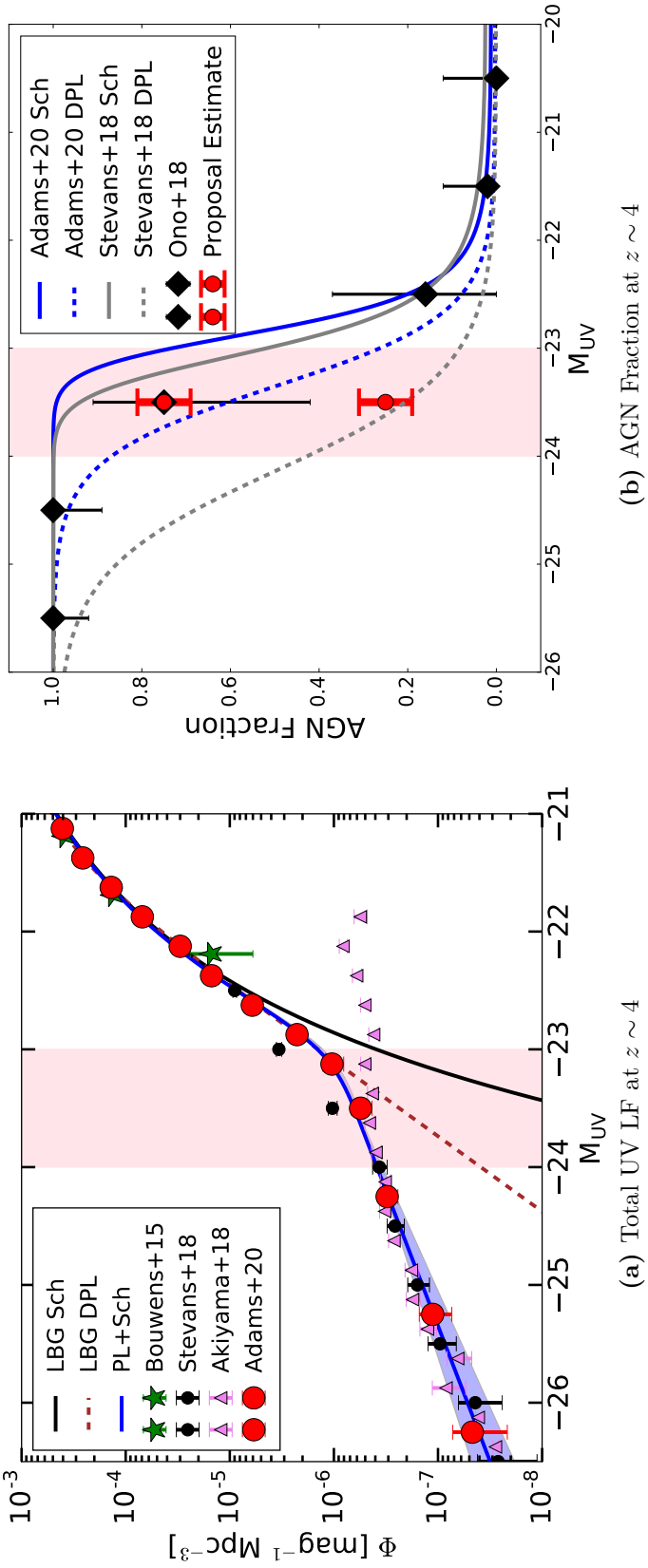
bias in the stellar mass function, which remains an important roadblock for future studies to consider.

5.2 Future Work

5.2.1 A spectroscopic study of the AGN-galaxy transition regime $-24 < M_{\mathbf{UV}} < -23$ at high redshifts

The luminosity dependent AGN fractions derived in Chapters 3 and 4 (see Fig.3.8 and Fig.4.8) show that there is a clear distinction between the predicted AGN fractions when using a Schechter function or a double power law in the luminosity range $-24 < M_{\rm UV} < -23$. As a consequence of this, a high-completeness spectroscopic survey could be capable of successfully answering whether the galaxy ultraviolet luminosity function at high redshifts is better described by a Schechter function or a DPL. With such a survey, I would be able to measure common AGN diagnostic lines in the rest-frame ultraviolet (e.g. CIV, HeII), enabling for the accurate classification of sources in this key luminosity range. The number of sources present in the COSMOS and XMM-LSS fields (45 between $-24 < M_{\rm UV} < -23$) will provide high enough number statistics to enable the AGN fraction to be constrained to within 5 per cent based on the binomial confidence interval of separating the two populations.

In Fig. 5.1 I present the ultraviolet luminosity function at $z \sim 4$ as measured in Chapter 3 and highlight the region of interest. Alongside, I show the luminosity-dependent AGN fraction and an example of the limits that could be set by such a spectroscopic study. The work will require spectroscopic measurements to be taken for 36 sources of the total sample of 45, the remainder of the galaxies already have spectra that were obtained by Boutsia et al. (2018) in the COSMOS field. The sources have apparent magnitudes $22 < m_i < 23.5$, which mean secure line diagnostics could be obtained with exposure lengths of 20-50 minutes per source on an 8 meter class telescope (e.g. XSHOOTER/VLT or MOONS). The use of a spectrograph with a wavelength range that extends into the near-infrared would also enable the MgII line to be observed and a black hole mass estimate could then be conducted for those sources identified as AGN. The MgII emission line



(a) Total UV LF at $z \sim 4$

(b) AGN Fraction at $z \sim 4$

Figure 5.1: Details behind a spectroscopic survey of sources in the transition regime of the total UV LF. In panel a), I show the total UV LF at $z \sim 4$ as measured in Chapter 3 and in panel b) I show the estimated AGN fraction based on models fit to the total UV LF in Chapter 3 and those from Stevans et al. (2018). The red data points in panel b) show two example scenarios where my sample is used to measure the AGN fraction, one shows the result for a 25 per cent AGN fraction and the other a 75 per cent AGN fraction. These highlight the low anticipated error margins, which will enable for the functional form best describing the galaxy population to be answered.

is emitted as part of the broad line region (BLR) of AGN, where high rotational velocities of orbiting gas around the black hole lead to significant red/blueshifting of the emitted light. Work by McLure & Jarvis (2002) has shown that the breadth of the line correlated with black hole masses and that black hole mass estimations can be made with some simple assumptions regarding the orbital geometry of the BLR. Comparisons with other methodologies to obtaining black hole masses, such as reverberation mapping, have shown this method to be a good estimator.

Such a spectroscopic study would provide a definitive answer regarding the number densities of galaxies brighter than $M_{\rm UV} < -23$ at high redshifts. These galaxies will be hosts to high levels of unobscured star formation. The combination of spectroscopy and template fitting will be used to explore the space density, stellar mass and star formation rate of these dust-poor systems and examine their potential impact on the reionisation process. In addition, the results will constrain how common weak or obscured AGN activity was when the Universe was less than 10 per cent of its current age. Obtaining the numbers of these sources at high redshifts are necessary in order to understand how black holes have grown, why there is this rapid increase in activity from $4 < z < 6$ and what impacts AGN feedback may have had on star formation at earlier times.

5.2.2 Synergy between the Optical/NIR surveys and the upcoming MIGHTEE radio survey

The catalogue of galaxy photometry and photometric redshifts produced as a part of this work has the potential to be of additional scientific use outside of the studies conducted within this thesis. In particular, the MeerKAT International GigaHertz Tiered Extragalactic Exploration survey (MIGHTEE) will be targeting the very same fields with new, deep radio imaging. The survey has already begun and will continue to take data for the next four years. There will consequently be substantial overlap of observations between these new radio data and the optical/NIR data that already exists in these fields. This will allow for near-UV to near-infrared diagnostics of galaxies to be compared to their radio counterparts.

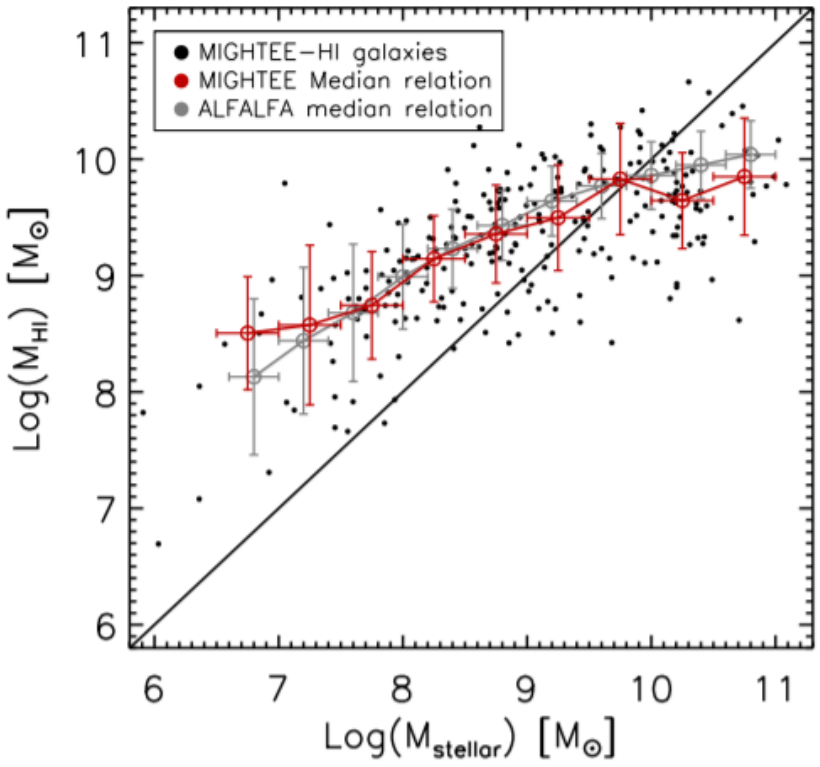


Figure 5.2: An example of the results produced by combining the work conducted in this thesis with MIGHTEE data. Here, the mass of a galaxy in the form of stars and in the form of neutral hydrogen (HI) are compared against each other. My work contributed the stellar mass measurements used for the x-axis of this diagram. Source:Maddox et al. (2021)

One example of the work that could be undertaken is the comparison of mass in HI (neutral hydrogen in the cold gas reservoir) and the stellar mass of galaxies. Shown in Fig. 5.2 is the result of an early attempt to carry out this type of work using stellar masses from my catalogues, as well as early science results from the MIGHTEE survey in the COSMOS field (Maddox et al., 2021). HI detections from MIGHTEE are available only for very low redshift galaxies, thus stellar mass calculations are required to be conducted with the use of total photometry, as opposed to aperture photometry used throughout most of this thesis, due to their large angular sizes. As with the ultraviolet luminosity function, a radio luminosity function can also be derived using the combination of my photometric redshift estimates and radio flux measurements. Radio flux can be generated due to synchrotron emission from electrons accelerated from supernova explosions or from

AGN. The long wavelengths of radio waves mean they do not interact with dust and thus, they provide a way to estimate unobscured star formation and AGN activity. This work was conducted in Malefahlo et al. (2020) using 1.4GHz data from the Very Large Array (VLA). Combining my work with MIGHTEE will enable for a deeper understanding the connection between star formation (radio continuum) and the cold gas reservoir (HI) to be made, as well as the impact of dust obscuration on measurements made in the rest-frame ultraviolet and optical regimes, through comparisons between UV and radio-based estimations of star formation rates.

5.2.3 The next generation of multiwavelength, wide-deep surveys

Beyond the survey programmes used in this thesis, in the next decade the Vera Rubin Observatory, *Euclid* Space Telescope and Nancy Grace Roman Space Telescope will seek to vastly increase the survey areas available with deep optical and near-infrared imaging. This will lead to multiple tens of square degrees with depths comparable to that of the COSMOS and XMM-LSS fields. This data will significantly reduce the impact of cosmic variance (the largest source of error for faint-end/low-mass galaxies) and provide greater number statistics for the bright-end/high-mass galaxies (with Poisson statistics being the dominant source of error for these objects) in the luminosity/mass functions at high redshifts. The precise functional form of the ultraviolet luminosity function has important consequences for the predicted yields of high-redshift galaxies from these survey programmes, since the DPL and Schechter functional forms deviate by greater than 1dex for $M_{\rm UV} < -23.5$. Combined, these surveys will expand the work presented in this thesis to higher redshifts $z > 5$ and larger luminosity ranges. I will use this data to provide greater insight into the evolution of rarest, most luminous sources present in the early Universe. Science goals include identifying galaxies hosting the largest supermassive black holes at $z > 7.5$, providing the necessary observational constraints to further understand the growth and formation of these mysterious objects, and finding UV-luminous galaxies embedded in ionised bubbles during the Era of Reionisation $z > 7$.

5.2.4 Towards the first galaxies and stars with JWST

The launch of the James Webb Space Telescope (JWST) is anticipated to occur before the end of 2021. This new facility is equipped with state-of-the-art near-infrared imaging and spectroscopic facilities, with a key goal to search for galaxies with extreme redshifts $z > 7$. Survey programmes planned with the JWST will enable for measurements fainter than the knee of the ultraviolet luminosity function to be conducted for galaxies in this early time period. I will use the narrower but deeper observations provided by this facility to complement observations that will be undertaken by the Vera Rubin Observatory and *Euclid* Space Telescope. Combining observations from these facilities will provide a full picture of ultraviolet emission and galaxy growth in the Era of Reionisation. Within the next few years, I aim to utilise JWST data to measure the evolution in the UV LF at redshifts $7 < z < 12$. The galaxies identified at these redshifts will be used to map the growth of stellar mass and impacts of dust obscuration in this early time. Follow up programmes using spectroscopy will be developed to measure black hole masses for AGN candidates and conduct searches for evidence of Population III stars (e.g. using strong HeII emission, di Serego Alighieri et al. (2008)).

5.3 Final Remarks

The work presented in this thesis is one step among many to understanding the formation and evolution of galaxies and AGN. Within the next decade, a series of large scale projects on the ground and in space will vastly improve data quality and volume, enabling astronomers to once again break the record for the most distant galaxy found, the oldest supermassive black hole discovered or the earliest reionisation bubbles. However, with these new powerful instruments it is the unexpected which excites us all the most. It is these that will provide daring new challenges to our understanding and maintain that passion humanity has for exploration.

Figure 5.3: The Hubble Ultra Deep Field (HUDF), one of the deepest images of the sky taken to date. Credit: NASA, ESA, R. Ellis (Caltech), and the HUDF 2012 Team

Scathing eyes ask that we be symmetrical, one sided and easily processed. Yet every misshapen spark's unseen beauty is greater than its would be judgement.

— Monty Oum

Bibliography

Abramowicz M. A., Fragile P. C., 2013, Living Reviews in Relativity, 16, 1

Abramowicz M. A., Czerny B., Lasota J. P., Szuszkiewicz E., 1988, ApJ, 332, 646

Adams N. J., Bowler R. A. A., Jarvis M. J., Häußler B., McLure R. J., Bunker A., Dunlop J. S., Verma A., 2020, MNRAS, 494, 1771

Adelberger K. L., Steidel C. C., 2000, ApJ, 544, 218

Aihara H., et al., 2018a, PASJ, 70, S4

Aihara H., et al., 2018b, PASJ, 70, S8

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

Akiyama M., et al., 2018, PASJ, 70, S34

Alam S., et al., 2015, ApJS, 219, 12

Alarcon A., et al., 2020, arXiv e-prints, p. arXiv:2007.11132

Albareti F. D., et al., 2017, ApJS, 233, 25

Alexander D. M., Swinbank A. M., Smail I., McDermid R., Nesvadba N. P. H., 2010, MNRAS, 402, 2211

Andrews S. K., Driver S. P., Davies L. J. M., Kafle P. R., Robotham A. S. G., Wright A. H., 2017, MNRAS, 464, 1569

Anglés-Alcázar D., Davé R., Faucher-Giguère C.-A., Özel F., Hopkins P. F., 2017, MNRAS, 464, 2840

Arav N., Moe M., Costantini E., Korista K. T., Benn C., Ellison S., 2008, ApJ, 681, 954

Arnouts S., Cristiani S., Moscardini L., Matarrese S., Lucchin F., Fontana A., Giallongo E., 1999, MNRAS, 310, 540

Ashby M. L. N., et al., 2013, ApJ, 769, 80

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

Atek H., et al., 2015, ApJ, 800, 18

Atek H., Richard J., Kneib J.-P., Schaerer D., 2018, MNRAS, 479, 5184

Bañados E., et al., 2018, Nature, 553, 473

Bahé Y. M., McCarthy I. G., Balogh M. L., Font A. S., 2013, MNRAS, 430, 3017

Bahé Y. M., et al., 2019, MNRAS, 485, 2287

Balbus S. A., Hawley J. F., 1998, Reviews of Modern Physics, 70, 1

Baldry I. K., et al., 2012, MNRAS, 421, 621

Baldwin J. A., Phillips M. M., Terlevich R., 1981, PASP, 93, 5

Barro G., et al., 2019, ApJS, 243, 22

Beckmann R. S., et al., 2017, MNRAS, 472, 949

Begelman M. C., 2004, in Ho L. C., ed., Coevolution of Black Holes and Galaxies. p. 374 (arXiv:astro-ph/0303040)

Begelman M. C., Volonteri M., Rees M. J., 2006, MNRAS, 370, 289

Bellstedt S., et al., 2020, arXiv e-prints, p. arXiv:2005.11215

Benítez-Llambay A., Navarro J. F., Abadi M. G., Gottlöber S., Yepes G., Hoffman Y., Steinmetz M., 2013, ApJ, 763, L41

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

Bernardi M., Nichol R. C., Sheth R. K., Miller C. J., Brinkmann J., 2006, AJ, 131, 1288

Bertin E., 2011, in Evans I. N., Accomazzi A., Mink D. J., Rots A. H., eds, ASP Conference Series Vol. 442, Astronomical Data Analysis Software and Systems XX. p. 435

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

Bian F., et al., 2013, ApJ, 774, 28

Bigiel F., Leroy A., Walter F., Brinks E., de Blok W. J. G., Madore B., Thornley M. D., 2008, AJ, 136, 2846

Binney J., 2004, MNRAS, 347, 1093

Binney J., Tabor G., 1995, MNRAS, 276, 663

Birnboim Y., Dekel A., 2003, MNRAS, 345, 349

Blandford R., Meier D., Readhead A., 2019, ARA&A, 57, 467

Blumenthal G. R., Faber S. M., Primack J. R., Rees M. J., 1984, Nature, 311, 517

Bohlin R., Colina L., Finley D., 1995, ApJ, 110, 1316

Bondi H., 1952, MNRAS, 112, 195

Bongiorno A., et al., 2016, A&A, 588, A78

Bonning E. W., Cheng L., Shields G. A., Salviander S., Gebhardt K., 2007, ApJ, 659, 211

Bosch-Ramon V., 2018, A&A, 617, L3

Bournaud F., Jog C. J., Combes F., 2005, A&A, 437, 69

Boutsia K., Grazian A., Giallongo E., Fiore F., Civano F., 2018, ApJ, 869, 20

Bouwens R. J., Illingworth G. D., Franx M., Ford H., 2008, ApJ, 686, 230

Bouwens R. J., et al., 2009, ApJ, 705, 936

Bouwens R. J., et al., 2012, ApJ, 754, 83

Bouwens R. J., et al., 2014, ApJ, 793, 115

Bouwens R. J., et al., 2015, ApJ, 803, 1

Bouwens R. J., Oesch P. A., Illingworth G. D., Ellis R. S., Stefanon M., 2017, ApJ, 843, 129

Bouwens R. J., et al., 2021, arXiv e-prints, p. arXiv:2102.07775

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

Bowler R. A., et al., 2014, MNRAS, 440, 2810

Bowler R. A., et al., 2015, MNRAS, 452, 1817

Bowler R. A. A., Dunlop J. S., McLure R. J., McLeod D. J., 2017, MNRAS, 466, 3612

Bowler R. A. A., Jarvis M. J., Dunlop J. S., McLure R. J., McLeod D. J., Adams N. J., Milvang-Jensen B., McCracken H. J., 2020, MNRAS, 493, 2059

Bowler R. A. A., Adams N. J., Jarvis M. J., Häußler B., 2021, MNRAS, 502, 662

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

Bridge C. R., et al., 2007, ApJ, 659, 931

Brinchmann J., Charlot S., White S. D. M., Tremonti C., Kauffmann G., Heckman T., Brinkmann J., 2004, MNRAS, 351, 1151

Bromm V., 2013, Reports on Progress in Physics, 76, 112901

Brooks A. M., Governato F., Quinn T., Brook C. B., Wadsley J., 2009, ApJ, 694, 396

Brown M. J. I., Duncan K. J., Landt H., Kirk M., Ricci C., Kamraj N., Salvato M., Ananna T., 2019, MNRAS, 489, 3351

Brusa M., et al., 2005, A&A, 432, 69

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

Bruzual A. G., 1983, ApJ, 273, 105

Bunker A. J., et al., 2010, MNRAS, 409, 855

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

Cao X., 2010, ApJ, 725, 388

Capak P., et al., 2012, SPLASH: Spitzer Large Area Survey with Hyper-Suprime-Cam, Spitzer Proposal

Caproni A., Mosquera Cuesta H. J., Abraham Z., 2004, ApJ, 616, L99

Carballo R., González-Serrano J. I., Benn C. R., Sánchez S. F., Vigotti M., 1999, MNRAS, 306, 137

Carr B. J., Bond J. R., Arnett W. D., 1984, ApJ, 277, 445

Cattaneo A., Dekel A., Devriendt J., Guiderdoni B., Blaizot J., 2006, MNRAS, 370, 1651

Cattaneo A., et al., 2009, Nature, 460, 213

Chabrier G., 2003, PASP, 115, 763

Chabrier G., Baraffe I., Allard F., Hauschildt P., 2000, ApJ, 542, 464

Ciotti L., Ostriker J. P., 1997, ApJ, 487, L105

Cirasuolo M., et al., 2012, in Proc. SPIE. p. 84460S (`arXiv:1208.5780`), doi:10.1117/12.925871

Clayton G. C., Martin P. G., 1985, ApJ, 288, 558

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

Cole S., et al., 2001, MNRAS, 326, 255

Cole S., Lacey C. G., Baugh C. M., Frenk C. S., 2002, MNRAS, 319, 168

Conselice C. J., 2014, ARA&A, 52, 291

Conselice C. J., Chapman S. C., Windhorst R. A., 2003, ApJ, 596, L5

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

Croom S. M., et al., 2009, MNRAS, 399, 1755

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

Cuillandre J.-C. J., et al., 2012, in Peck A. B., Seaman R. L., Comeron F., eds, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 8448, Observatory Operations: Strategies, Processes, and Systems IV. p. 84480M, doi:10.1117/12.925584

Daddi E., et al., 2007, ApJ, 670, 156

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

Davidzon I., et al., 2016, A&A, 586, A23

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

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

Davies L. J. M., et al., 2015, MNRAS, 452, 616

Davies L. J. M., et al., 2018, MNRAS, 480, 768

Dayal P., et al., 2020, MNRAS, 495, 3065

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

Dekel A., Birnboim Y., 2006, MNRAS, 368, 2

Dekel A., Silk J., 1986, ApJ, 303, 39

Dekel A., et al., 2009, Nature, 457, 451

Delvecchio I., et al., 2020, ApJ, 892, 17

Di Matteo T., Springel V., Hernquist L., 2005, Nature, 433, 604

Dijkstra M., 2014, PASA, 31, e040

Driver S. P., et al., 2006, MNRAS, 368, 414

Driver S. P., et al., 2011, MNRAS, 413, 971

Driver S. P., et al., 2019, The Messenger, 175, 46

Duc P. A., Mirabel I. F., Maza J., 1997, A&AS, 124, 533

Duncan K., et al., 2014, MNRAS, 444, 2960

Dunlop J. S., McLure R. J., Robertson B. E., Ellis R. S., Stark D. P., Cirasuolo M., de Ravel L., 2012, MNRAS, 420, 901

Dunn J. P., et al., 2010, ApJ, 709, 611

Dyson J. E., Williams D. A., 1997, The physics of the interstellar medium, doi:10.1201/9780585368115.

Ebeling H., Stephenson L. N., Edge A. C., 2014, ApJ, 781, L40

Eddington A. S., 1913, Monthly Notices of the Royal Astronomical Society, 73, 359

Efstathiou G., 2000, MNRAS, 317, 697

Elbaz D., et al., 2007, A&A, 468, 33

Elsner F., Feulner G., Hopp U., 2008, A&A, 477, 503

Fabian A. C., 2012, ARA&A, 50, 455

Farouki R., Shapiro S. L., 1981, ApJ, 243, 32

Feldmann R., Mayer L., 2015, MNRAS, 446, 1939

Feldmann R., Carollo C. M., Mayer L., 2011, ApJ, 736, 88

Feng Y., Di-Matteo T., Croft R. A., Bird S., Battaglia N., Wilkins S., 2016, MNRAS, 455, 2778

Ferrarese L., Merritt D., 2000, ApJ, 539, L9

Ferreras I., Lisker T., Pasquali A., Khochfar S., Kaviraj S., 2009, MNRAS, 396, 1573

Finkelstein S. L., 2016, PASA, 33, e037

Finkelstein S. L., et al., 2015, ApJ, 810, 71

Fontana A., et al., 2004, A&A, 424, 23

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

Forrest B., et al., 2020, ApJ, 890, L1

Fritz J., Franceschini A., Hatziminaoglou E., 2006, MNRAS, 366, 767

Furlong M., et al., 2015, MNRAS, 450, 4486

Gabor J. M., Bournaud F., 2014, MNRAS, 441, 1615

Garilli B., et al., 2021, arXiv e-prints, p. arXiv:2101.07645

Gaskell C. M., 2009, New A Rev., 53, 140

Giallongo E., et al., 2015, A&A, 578, A83

Glikman E., Djorgovski S. G., Stern D., Dey A., Jannuzi B. T., Lee K. S., 2011, ApJ, 728, L26

Gnedin N. Y., 1998, MNRAS, 294, 407

Gonzalez-Perez V., Lacey C. G., Baugh C. M., Frenk C. S., Wilkins S. M., 2013, MNRAS, 429, 1609

Gordon K. D., Clayton G. C., Misselt K. A., Landolt A. U., Wolff M. J., 2003, ApJ, 594, 279

Graham A. W., Onken C. A., Athanassoula E., Combes F., 2011, MNRAS, 412, 2211

Grazian A., et al., 2015, A&A, 575, A96

Guhathakurta P., Tyson J. A., Majewski S. R., 1990, ApJ, 357, L9

Gunn J. E., Gott J. Richard I., 1972, ApJ, 176, 1

Habouzit M., et al., 2021, MNRAS, 503, 1940

Haehnelt M. G., Rees M. J., 1993, MNRAS, 263, 168

Hale C. L., Robotham A. S. G., Davies L. J. M., Jarvis M. J., Driver S. P., Heywood I., 2019, MNRAS, 487, 3971

Hall P. B., et al., 2002, ApJS, 141, 267

Hammer D. M., Hornschemeier A. E., Salim S., Smith R., Jenkins L., Mobasher B., Miller N., Ferguson H., 2012, ApJ, 745, 177

Hamuy M., Walker A., Suntzeff N., Gigoux P., Heathcote S., Phillips M., 1992, PASP, 104, 533

Hamuy M., Sunteff N., Heathcote S., Walker A., Gigoux P., Phillips M., 1994, PASP, 106, 566

Hasinger G., et al., 2018, ApJ, 858, 77

Hassan S., Davé R., Mitra S., Finlator K., Ciardi B., Santos M. G., 2018, MNRAS, 473, 227

Henriques B. M. B., White S. D. M., Thomas P. A., Angulo R. E., Guo Q., Lemson G., Springel V., 2013, MNRAS, 431, 3373

Hirata C., Seljak U., 2003, MNRAS, 343, 459

Ho S. H., Martin C. L., Turner M. L., 2019, ApJ, 875, 54

Hoaglin D. C., Mosteller F., Tukey J. W., 1983, Understanding Robust and Exploratory Data Analysis (New York: Wiley)

Hopkins P. F., Elvis M., 2010, MNRAS, 401, 7

Hopkins P. F., Hernquist L., Cox T. J., Di Matteo T., Robertson B., Springel V., 2005, ApJ, 630, 716

Hopkins P. F., et al., 2009, MNRAS, 397, 802

Huang K. H., Ferguson H. C., Ravindranath S., Su J., 2013, ApJ, 765, 68

Ikeda H., et al., 2012, ApJ, 756, 160

Ilbert O., et al., 2006, A&A, 457, 841

Ilbert O., et al., 2009, ApJ, 690, 1236

Ilbert O., et al., 2010, ApJ, 709, 644

Ilbert O., et al., 2013, A&A, 556, A55

Inoue A. K., Shimizu I., Iwata I., Tanaka M., 2014, MNRAS, 442, 1805

Ishigaki M., Kawamata R., Ouchi M., Oguri M., Shimasaku K., Ono Y., 2018, ApJ, 854, 73

Ivezić Ž., et al., 2019, ApJ, 873, 111

Jáchym P., et al., 2019, ApJ, 883, 145

Janka H.-T., 2012, Annual Review of Nuclear and Particle Science, 62, 407

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

Jiang L., et al., 2016, ApJ, 833, 222

Kapferer W., Sluka C., Schindler S., Ferrari C., Ziegler B., 2009, A&A, 499, 87

Kashlinsky A., Rees M. J., 1983, MNRAS, 205, 955

Kauffmann G., Colberg J. M., Diaferio A., White S. D. M., 1999, MNRAS, 303, 188

Kauffmann G., White S. D. M., Heckman T. M., Ménard B., Brinchmann J., Charlot S., Tremonti C., Brinkmann J., 2004, MNRAS, 353, 713

Kaviraj S., et al., 2017, MNRAS, 467, 4739

Kawata D., Mulchaey J. S., 2008, ApJ, 672, L103

Kelly B. C., Shen Y., 2013, ApJ, 764, 45

Kereš D., Katz N., Weinberg D. H., Davé R., 2005, MNRAS, 363, 2

Kereš D., Katz N., Fardal M., Davé R., Weinberg D. H., 2009, MNRAS, 395, 160

Khochfar S., Silk J., 2009, MNRAS, 397, 506

King A., 2003, ApJ, 596, L27

King A., Pounds K., 2015, ARA&A, 53, 115

Kleiner D., Pimbblet K. A., Jones D. H., Koribalski B. S., Serra P., 2017, MNRAS, 466, 4692

Konigl A., Kartje J. F., 1994, ApJ, 434, 446

Koratkar A., Blaes O., 1999, PASP, 111, 1

Kormendy J., Richstone D., 1995, ARA&A, 33, 581

Koudmani S., Sijacki D., Bourne M. A., Smith M. C., 2019, MNRAS, 484, 2047

Krasnopolsky R., Li Z.-Y., Blandford R., 1999, ApJ, 526, 631

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

Kroupa P., 2001, MNRAS, 322, 231

Kulkarni G., Worseck G., Hennawi J. F., 2019, MNRAS, 488, 1035

Labita M., Decarli R., Treves A., Falomo R., 2009, MNRAS, 396, 1537

Lagos C. d. P., Tobar R. J., Robotham A. S. G., Obreschkow D., Mitchell P. D., Power C., Elahi P. J., 2018, MNRAS, 481, 3573

Laigle C., et al., 2016, ApJS, 224, 24

Larson R. B., 1974, MNRAS, 169, 229

Larson R. B., 1998, MNRAS, 301, 569

Larson R. B., Tinsley B. M., Caldwell C. N., 1980, ApJ, 237, 692

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

LeFèvre O., et al., 2013, A&A, 559, A14

Leauthaud A., et al., 2016, MNRAS, 457, 4021

Leja J., Speagle J. S., Johnson B. D., Conroy C., van Dokkum P., Franx M., 2020, ApJ, 893, 111

Leroy A. K., Walter F., Brinks E., Bigiel F., de Blok W. J. G., Madore B., Thornley M. D., 2008, AJ, 136, 2782

Levenberg K., 1944, Quart. Appl. Math., pp 164–168

Lilly S. J., et al., 2009, ApJS, 184, 218

Lin L., et al., 2008, ApJ, 681, 232

Lin L., et al., 2010, ApJ, 718, 1158

López-Corredoira M., Vazdekis A., Gutiérrez C. M., Castro-Rodríguez N., 2017, A&A, 600, A91

Lotz J. M., Jonsson P., Cox T. J., Primack J. R., 2010, MNRAS, 404, 575

Lu J., 1991, Progress in Astronomy, 9, 271

Ma X., et al., 2019, MNRAS, 487, 1844

Madau P., 1995, ApJ, 441, 18

Madau P., Dickinson M., 2014, ARA&A, 52, 415

Madau P., Haardt F., 2015, ApJ, 813, L8

Madau P., Rees M. J., 2001, ApJ, 551, L27

Maddox N., et al., 2021, A&A, 646, A35

Magdis G. E., Rigopoulou D., Huang J. S., Fazio G. G., 2010, MNRAS, 401, 1521

Magorrian J., et al., 1998, AJ, 115, 2285

Malefahlo E. D., Jarvis M. J., Santos M. G., White S. V., Adams N. J., Bowler R. A. A., 2020, arXiv e-prints, p. arXiv:2012.09797

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

Maraston C., 2005, MNRAS, 362, 799

Maraston C., Nieves Colmenárez L., Bender R., Thomas D., 2009, A&A, 493, 425

Marchese E., Della Ceca R., Caccianiga A., Severgnini P., Corral A., Fanali R., 2012, A&A, 539, A48

Marchesi S., et al., 2016, ApJ, 817, 34

Marchesini D., van Dokkum P. G., Förster Schreiber N. M., Franx M., Labbé I., Wuyts S., 2009, ApJ, 701, 1765

Marks M., Kroupa P., Dabringhausen J., Pawlowski M. S., 2012, MNRAS, 422, 2246

Marquardt D. W., 1963, SIAM Journal on Applied Mathematics, 11, 431

Martin D. C., et al., 2019, Nature Astronomy, 3, 822

Masters D., et al., 2012, ApJ, 755, 169

Matsuoka Y., et al., 2019, ApJ, 883, 183

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

McCarthy I. G., Frenk C. S., Font A. S., Lacey C. G., Bower R. G., Mitchell N. L., Balogh M. L., Theuns T., 2008, MNRAS, 383, 593

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

McDowell J. C., 1986, MNRAS, 223, 763

McGreer I. D., et al., 2013, ApJ, 768, 105

McGreer I. D., Fan X., Jiang L., Cai Z., 2018, AJ, 155, 131

McKellar A., 1941, Publications of the Dominion Astrophysical Observatory Victoria, 7, 251

McLeod D. J., McLure R. J., Dunlop J. S., Cullen F., Carnall A. C., Duncan K., 2020, arXiv e-prints, p. arXiv:2009.03176

McLure R. J., Jarvis M. J., 2002, MNRAS, 337, 109

McLure R. J., Cirasuolo M., Dunlop J. S., Foucaud S., Almaini O., 2009, MNRAS, 395, 2196

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

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

Mehta V., et al., 2018, ApJS, 235, 36

Meurer G. R., Heckman T. M., Calzetti D., 1999, ApJ, 521, 64

Michałowski M. J., 2015, A&A, 577, A80

Mihos J. C., Harding P., Feldmeier J., Morrison H., 2005, ApJ, 631, L41

Moe M., Arav N., Bautista M. A., Korista K. T., 2009, ApJ, 706, 525

Moll R., et al., 2007, A&A, 463, 513

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

Morales M. F., Wyithe J. S. B., 2010, ARA&A, 48, 127

Morganti R., 2017, Nature Astronomy, 1, 596

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

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

Moutard T., Sawicki M., Arnouts S., Golob A., Coupon J., Ilbert O., Yang X., Gwyn S., 2020, MNRAS, 494, 1894

Muzzin A., et al., 2013, ApJ, 777, 18

Naab T., Johansson P. H., Ostriker J. P., 2009, ApJ, 699, L178

Nagamine K., Fukugita M., Cen R., Ostriker J. P., 2001, MNRAS, 327, L10

Narayan R., Quataert E., 2005, Science, 307, 77

Natarajan P., Pringle J. E., 1998, ApJ, 506, L97

Neeleman M., Prochaska J. X., Kanekar N., Rafelski M., 2020, Nature, 581, 269

Negroponte J., White S. D. M., 1983, MNRAS, 205, 1009

Niida M., Nagao T., Ikeda H., Matsuoka K., Kobayashi M. A. R., Toba Y., Taniguchi Y., 2016, ApJ, 832, 208

Niida M., et al., 2020, ApJ, 904, 89

Noeske K. G., et al., 2007, ApJ, 660, L43

O'Leary J. A., Moster B. P., Naab T., Somerville R. S., 2020, arXiv e-prints, p. arXiv:2001.02687

Oesch P. A., et al., 2010, ApJ, 725, L150

Oesch P. A., et al., 2012, ApJ, 759

Oke J. B., 1974, ApJS, 27, 21

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

Ono Y., et al., 2018, PASJ, 70, S10

Ouchi M., et al., 2009, ApJ, 706, 1136

Ouyed R., Pudritz R. E., 1997, ApJ, 482, 712

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

Parsa S., Dunlop J. S., McLure R. J., Mortlock A., 2016, MNRAS, 456, 3194

Parsa S., Dunlop J. S., McLure R. J., 2018, MNRAS, 474, 2904

Peebles P. J. E., 1982, ApJ, 263, L1

Peng Y.-j., et al., 2010, ApJ, 721, 193

Pentericci L., et al., 2018, A&A, 616, A174

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

Pérez-González P. G., et al., 2008, ApJ, 675, 234

Pickles A. J., 1998, PASP, 110, 863

Pillepich A., et al., 2018, MNRAS, 475, 648

Planck M., Masius M., 1914, The Theory of Heat Radiation. Blakiston, `http://books.google.co.uk/books?id=2PR_AAAAMAAJ`

Planck Collaboration et al., 2016, A&A, 596, A108

Planck Collaboration et al., 2020, A&A, 641, A1

Poggianti B. M., Barbaro G., 1997, A&A, 325, 1025

Polletta M., et al., 2007, ApJ, 663, 81

Powell L. C., Slyz A., Devriendt J., 2011, MNRAS, 414, 3671

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

Prevot M. L., Lequeux J., Maurice E., Prevot L., Rocca-Volmerange B., 1984, A&A, 132, 389

Pritchard J. R., Loeb A., 2012, Reports on Progress in Physics, 75, 086901

Proga D., 2003, ApJ, 585, 406

Puglisi A., et al., 2021, Nature Astronomy, 5, 319

Qu Y., et al., 2017, MNRAS, 464, 1659

Racca G. D., et al., 2016, in Proc. SPIE. p. 99040O (`arXiv:1610.05508`), doi:10.1117/12.2230762

Rahman N., et al., 2012, ApJ, 745, 183

Ramos Almeida C., Ricci C., 2017, Nature Astronomy, 1, 679

Rawlings S., Saunders R., 1991, Nature, 349, 138

Reddy N. A., Erb D. K., Pettini M., Steidel C. C., Shapley A. E., 2010, ApJ, 712, 1070

Reed D., Gardner J., Quinn T., Stadel J., Fardal M., Lake G., Governato F., 2003, MNRAS, 346, 565

Rees M. J., Ostriker J. P., 1977, MNRAS, 179, 541

Reines A. E., Volonteri M., 2015, ApJ, 813, 82

Ricci F., Marchesi S., Shankar F., Franca F. L., Civano F., 2017, MNRAS, 465, 1915

Richards G. T., et al., 2006, ApJS, 166, 470

Richings A. J., Faucher-Giguère C.-A., Stern J., 2021, MNRAS, 503, 1568

Robotham A. S. G., Davies L. J. M., Driver S. P., Koushan S., Taranu D. S., Casura S., Liske J., 2018, MNRAS, 476, 3137

Rodriguez-Gomez V., et al., 2015, MNRAS, 449, 49

Rowan-Robinson M., 1968, MNRAS, 138, 445

Rubin V. C., Ford W. Kent J., 1970, ApJ, 159, 379

Salmon B., et al., 2015, ApJ, 799, 183

Salpeter E. E., 1955, ApJ, 121, 161

Salvato M., et al., 2009, ApJ, 690, 1250

Scannapieco C., Tissera P. B., White S. D. M., Springel V., 2006, MNRAS, 371, 1125

Schawinski K., Thomas D., Sarzi M., Maraston C., Kaviraj S., Joo S. J., Yi S. K., Silk J., 2007, MNRAS, 382, 1415

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

Schechter P., 1976, ApJ, 203, 297

Schenker M. A., et al., 2013, ApJ, 768

Schmidt M., 1968, ApJ, 151, 393

Schmidt K. B., et al., 2014, ApJ, 786

Schreiber C., et al., 2015, A&A, 575, A74

Schwarzschild M., Spitzer L., 1953, The Observatory, 73, 77

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

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

Sérsic J. L., 1963, Boletin de la Asociacion Argentina de Astronomia La Plata Argentina, 6, 41

Shankar F., Weinberg D. H., Miralda-Escudé J., 2009, ApJ, 690, 20

Shi J., et al., 2020, ApJ, 893, 139

Shull J. M., Stevans M., Danforth C. W., 2012, ApJ, 752, 162

Silk J., Mamon G. A., 2012, Research in Astronomy and Astrophysics, 12, 917

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

Silva L., Granato G. L., Bressan A., Danese L., 1998, ApJ, 509, 103

Silverman J. D., et al., 2015, ApJS, 220, 12

Simon J. D., 2019, ARA&A, 57, 375

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

Smith S., 1936, ApJ, 83, 23

Sokasian A., Yoshida N., Abel T., Hernquist L., Springel V., 2004, MNRAS, 350, 47

Somerville R. S., Davé R., 2015, ARA&A, 53, 51

Song M., et al., 2016a, ApJ, 825, 5

Song M., et al., 2016b, ApJ, 825, 5

Soucail G., Mellier Y., Fort B., Mathez G., Cailloux M., 1988, A&A, 191, L19

Sparre M., et al., 2015, MNRAS, 447, 3548

Springel V., Di Matteo T., Hernquist L., 2005, MNRAS, 361, 776

Stanway E. R., Bremer M. N., Lehnert M. D., 2008, MNRAS, 385, 493

Stark D. P., Ellis R. S., Bunker A., Bundy K., Targett T., Benson A., Lacy M., 2009, ApJ, 697, 1493

Stark D. P., Schenker M. A., Ellis R., Robertson B., McLure R., Dunlop J., 2013, ApJ, 763, 129

Stefanon M., et al., 2019, ApJ, 883, 99

Steidel C. C., Hamilton D., 1992, AJ, 104, 941

Steidel C. C., Giavalisco M., Pettini M., Dickinson M., Adelberger K. L., 1996, ApJ, 462, L17

Sternberg A., Soker N., 2009, MNRAS, 398, 422

Stevans M. L., et al., 2018, ApJ, 863, 63

Strateva I., et al., 2001, AJ, 122, 1861

Sutherland W., Saunders W., 1992, MNRAS, 259, 413

Tacchella S., Trenti M., Carollo C. M., 2013, ApJ, 768, L37

Tacchella S., Bose S., Conroy C., Eisenstein D. J., Johnson B. D., 2018, ApJ, 868, 92

Targett T. A., Dunlop J. S., McLure R. J., 2012, MNRAS, 420, 3621

Thorne J. E., et al., 2020, arXiv e-prints, p. arXiv:2011.13605

Tomczak A. R., et al., 2016, ApJ, 817, 118

Tomczak A. R., et al., 2017, MNRAS, 472, 3512

Toomre A., 1977, in Tinsley B. M., Larson Richard B. Gehret D. C., eds, Evolution of Galaxies and Stellar Populations. p. 401

Tremonti C. A., Moustakas J., Diamond-Stanic A. M., 2007, ApJ, 663, L77

Trenti M., Stiavelli M., 2008, ApJ, 676, 767

Trenti M., Stiavelli M., Bouwens R. J., Oesch P., Shull J. M., Illingworth G. D., Bradley L. D., Carollo C. M., 2010, ApJ, 714, L202

Tumlinson J., et al., 2013, ApJ, 777, 59

Ueda Y., Akiyama M., Hasinger G., Miyaji T., Watson M. G., 2014, ApJ, 786, 104

Umemura M., Loeb A., Turner E. L., 1993, ApJ, 419, 459

Vanden Berk D. E., et al., 2001, AJ, 122, 549

Venemans B. P., Walter F., Zschaechner L., Decarli R., De Rosa G., Findlay J. R., McMahon R. G., Sutherland W. J., 2016, ApJ, 816, 37

Venemans B. P., et al., 2017, ApJ, 837, 146

Verma A., Lehnert M. D., Förster Schreiber N. M., Bremer M. N., Douglas L., 2007, MNRAS, 377, 1024

Viironen K., et al., 2018, A&A, 614, A129

Vijayan A. P., Lovell C. C., Wilkins S. M., Thomas P. A., Barnes D. J., Irodotou D., Kuusisto J., Roper W. J., 2021, MNRAS, 501, 3289

Vito F., et al., 2018, MNRAS, 473, 2378

Vogelsberger M., Marinacci F., Torrey P., Puchwein E., 2020, Nature Reviews Physics, 2, 42

Volonteri M., Reines A. E., 2016, ApJ, 820, L6

Walsh D., Carswell R. F., Weymann R. J., 1979, Nature, 279, 381

Wang B., Heckman T., 1996, ApJ, 457, 645

Wang F., et al., 2019, ApJ, 884, 30

Wang F., et al., 2021, ApJ, 907, L1

Warren S. J., Hewett P. C., Irwin M. J., McMahon R. G., Bridgeland M. T., 1987a, Nature, 325, 131

Warren S. J., Hewett P. C., Osmer P. S., Irwin M. J., 1987b, Nature, 330, 453

Weigel A. K., Schawinski K., Bruderer C., 2016, MNRAS, 459, 2150

Wetzel A. R., White M., 2010, MNRAS, 403, 1072

Weymann R. J., Morris S. L., Foltz C. B., Hewett P. C., 1991, ApJ, 373, 23

White S. D. M., Frenk C. S., 1991, ApJ, 379, 52

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

Wilkins S. M., Bunker A. J., Lorenzoni S., Caruana J., 2011, MNRAS, 411, 23

Wilkins S. M., Stanway E. R., Bremer M. N., 2014, MNRAS, 439, 1038

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

Williams R. J., Quadri R. F., Franx M., van Dokkum P., Toft S., Kriek M., Labbé I., 2010, ApJ, 713, 738

Willott C. J., et al., 2010, AJ, 140, 546

Wise J. H., 2019, arXiv e-prints, p. arXiv:1907.06653

Wise M. W., et al., 2017, in AAS/High Energy Astrophysics Division #16. p. 105.01

Witt A. N., Bohlin R. C., Stecher T. P., 1984, ApJ, 279, 698

Wright A. H., et al., 2017, MNRAS, 470, 283

Wright A. H., Driver S. P., Robotham A. S. G., 2018, MNRAS, 480, 3491

Wu J., 2011, PhD thesis

Wuyts S., et al., 2011, ApJ, 742, 96

Wyder T. K., et al., 2005, ApJ, 619, L15

Xie X., Shen S., Shao Z., Yin J., 2015, ApJ, 802, L16

Yang J., et al., 2016, ApJ, 829, 33

Yang J., et al., 2020, ApJ, 897, L14

Yoshiura S., Hasegawa K., Ichiki K., Tashiro H., Shimabukuro H., Takahashi K., 2017, MNRAS, 471, 3713

Zhuang M.-Y., Ho L. C., Shangguan J., 2018, ApJ, 862, 118

Zubovas K., King A., 2012, ApJ, 745, L34

Zwicky F., 1933, Helvetica Physica Acta, 6, 110

Zwicky F., 1937, ApJ, 86, 217

de Barros S., Schaerer D., Stark D. P., 2014, A&A, 563, A81

di Serego Alighieri S., Kurk J., Ciardi B., Cimatti A., Daddi E., Ferrara A., 2008, in Hunt L. K., Madden S. C., Schneider R., eds, Vol. 255, Low-Metallicity Star Formation: From the First Stars to Dwarf Galaxies. pp 75–78 (arXiv:0807.4634), doi:10.1017/S1743921308024605

van der Burg R. F. J., Hildebrandt H., Erben T., 2010, A&A, 523, A74