Marcelo Ciani Vicentin

The environment of QSOs triplets at $1 \lesssim z \lesssim 1.5$

São Paulo
2020

MARCELO CIANI VICENTIN

The environment of QSOs triplets at $1 \lesssim z \lesssim 1.5$

Versão Corrigida. O original encontra-se disponível na Unidade.

Dissertação apresentada ao Instituto de Astronomia, Geofísica e Ciências Atmosféricas da Universidade de São Paulo como requisito parcial para a obtenção do título de Mestre em Ciências.

Área de concentração: Astronomia extragalática.

Orientador: Prof. Dr. Laerte Sodré Júnior.

São Paulo
2020

É preciso se desfazer do feito para continuar a fazer-se.
Scarlett Marton

Agradecimentos

Durante esta trajetória, diversas pessoas e instituições foram fundamentais para que o projeto proposto fosse desenvolvido com todo cuidado e dedicação necessária a um trabalho científico. Em tempos obscuros, a ciência pavimenta e ilumina caminhos para uma sociedade melhor, onde não há espaço para mitos e fantasias arquitetadas por aqueles que defendem apenas o benefício próprio. As universidades públicas brasileiras são sinônimo de excelência ao redor do mundo, e mesmo a duros golpes ao longo dos anos, conseguem se destacar nos topos das mais diversas áreas em termos quantitativos e qualitativos, deixando mais do que claro, que os cientistas brasileiros devem ser respeitados e devem ter mais reconhecimento pelo seu trabalho por parte de muitos governantes, que insistem em negligenciá-los.

Dito isso, primeiramente gostaria de agradecer ao Instituto de Astronomia, Geofísica e Ciências Atmosféricas da Universidade de São Paulo (IAG/USP), juntamente com todo o seu corpo docente, administrativo e funcional, me proporcionou um excelente ambiente de trabalho, providenciando todas as necessidades com agilidade e qualidade.

À *Fundação de Amparo à Pesquisa do Estado de São Paulo* (FAPESP), que acreditou neste projeto e concedeu um auxílio financeiro por dois anos e, devido à pandemia, generosamente ampliou a concessão por mais dois meses, o que foi crucial para minha dedicação exclusiva durante todo o período.

A experiência de passar por uma universidade deste calibre é algo verdadeiramente engrandecedor. Além de ter tido a oportunidade de aprender muito sobre a minha área específica e a astronomia como um todo, o contato pessoal com grandes cientistas e aspirantes, me rendeu novas perspectivas que, sem sombra de dúvidas, me tornaram uma pessoa melhor. Os meus sinceros agradecimentos ao meu orientador e amigo, Professor Laerte Sodré Jr., pelos conselhos e trocas de ideia. Me ajudou a aliar objetividade com rigor, e também trabalho com leveza. Ao meu grande parceiro e amigo, Pablo Araya-Araya, por todo o trabalho que compartilhamos juntos e, claro, aos papos regados à cerveja. Também agradeço o meu relator, Professor Eduardo Cypriano, que contribuiu significativamente com sugestões que foram adotadas neste projeto. Da mesma forma, o Professor Roderik Overzier, referência na área específica a qual este trabalho faz parte. E o amigo Erik V. R. de Lima, que sempre foi cordial e solícito em esclarecer as dúvidas relacionadas aos *redshifts* fotométricos, sua área de pesquisa.

Sem relação direta ao trabalho, mas não menos importante, também dedico meus agradecimentos a todos amigos do departamento de astronomia do IAG/USP, todos os professores que dedicaram o seu tempo para nos ensinar sobre esta bela ciência, e, finalmente, a todos meus amigos e familiares que estiveram/estão presentes na minha vida. Só sou, porque sois.

Resumo

O presente trabalho se trata de uma análise do ambiente onde se localizam tripletos de quasares em $1 \lesssim z \lesssim 1.5$. Para isso, foram utilizadas seis imagens multi-bandas no óptico (r, i, z, ou g, r, i), obtidas pela Megacam do telescópio CFHT, tendo como foco principal investigar se esses sistemas estão associados com protoaglomerados de galáxias. As imagens foram processadas e os objetos detectados foram mensurados fotometricamente e organizados em catálogos através do *software SExtractor*. O efeito de extinção causado pela poeira da Galáxia foi corrigido e as magnitudes instrumentais foram calibradas no sistema do *SDSS*. Em seguida, foram estabelecidos critérios fotométricos e morfológicos para separar objetos pontuais e extensos, uma vez que apenas as galáxias são relevantes para a nossa análise. Além das três bandas do óptico, foi utilizado o catálogo *unWISE* para incluir medidas em outras duas bandas do infra-vermelho médio (W1 = 3.6μm e W2 = 4.5μm), que foram fundamentais para estimar os *redshifts* fotométricos através de métodos de *machine learning* (MLM). Esta estimativa foi feita utilizando o catálogo de galáxias *COSMOS2015* – o qual possui *redshifts* fotométricos altamente acurados – como conjunto de treinamento do MLM, que foi aplicado com o programa `ANNz2`. Nesta configuração, foi possível atingir bons resultados com dispersões da ordem de 4%. A análise da distribuição das galáxias nos redshifts dos tripletos foi feita a partir de mapas de significância de sobredensidade, produzidos com ténicas de estimadores de densidade (*Kernel Density Estimator*, KDE) Gaussianos em duas dimensões, e a significância da sobredensidade foi calculada. Todos os critérios adotados na nossa análise foram reproduzidos identicamente, ou da forma mais justa possível, em dados de simulações. Isto nos permitiu analisar a eficiência da detectabilidade de protoaglomerados de acordo com as limitações observacionais. Em apenas 1 dos nossos campos, o tripleto se encontra no centro de uma sobredensidade significativa e, portanto, indica uma provável associação com um protoaglomerado de galáxias em z $\sim$ 1.12.

Palavras-chave: quasares: geral, galáxias: estruturas, galáxias: estatística, galáxias: fotometria, galáxias: distâncias e redshifts, catálogos, surveys, métodos: análise de dados

Abstract

We present an analysis of the environment of six QSO triplets at $1 \lesssim z \lesssim 1.5$ by analyzing multiband (r, i, z, or g, r, i) images obtained with Megacam at the CFHT telescope, aiming to investigate whether they are associated or not with galaxy protoclusters. This was done by using photometric redshifts trained using the highly accurate photometric redshifts of the *COSMOS2015* catalogue. To improve the quality of our photometric redshift estimation, we included in our analysis mid-infrared photometry (3.6 and 4.5μm) from the *unWISE* survey available for our fields and the COSMOS survey. This approach allowed us to obtain good photometric redshifts, with dispersions of $\sim$0.04 for our six fields. Our analysis setup was reproduced on lightcones constructed from the Millennium Simulation data and the latest version of the L-GALAXIES semi-analytic model to verify the protocluster detectability in such conditions. The density field in a redshift slab containing each triplet was then analyzed with a Gaussian kernel density estimator. We find that all but one of our triplets tend to avoid overdense regions. In one case the triplet is associated with a possible protocluster at $z = 1.12$.

Keywords: quasars: general, galaxies: structures, galaxies: statistics, galaxy: photometry, galaxies: distances and redshifts, catalogues, surveys, methods: data analysis.

Contents

List of Figures

List of Tables

1 Introduction

1.1 | From General Relativity to the Expansion of The Universe: A Historical Perspective

Given the greatness of the theme, this Section aims to introduce the characters and briefly describe their deeds that laid the foundation for our current comprehension of the universe. The purpose of this topic is to provide a historical perspective and to contextualize the emergence and importance of its future developments. This work as a whole is a modest example of this legacy.

The main references for this Section are Jones (1997), Mörtsell (2016), and, also, several websites for general consultations.

1.1.1 | General Relativity

The 20th century was a temporal stage for great scientific discoveries and especially the birth of a new understanding of the origin of the universe. The seed for this revolution of thought was sowed by Albert Einstein, who, from 1908 to 1915, developed perhaps his most brilliant theory: The General Relativity. After almost two and a half centuries of the complete sovereignty of Newton's universal gravitational law (1687), a new philosophical and mathematical concept has emerged. General relativity conceives a four-dimensional universe (3 spatial and 1 temporal – the "space-time") where the gravitational field is described geometrically by field equations (Equation 1.4). Matter deforms space-time due to its self-gravity property, and, therefore, the geometry of the universe is directly affected by its distribution:

$$Geometry \propto Matter \tag{1.1}$$

$$G_{\mu\nu} \propto \kappa T_{\mu\nu} \tag{1.2}$$

where $G_{\mu\nu}$ and $T_{\mu\nu}$ are tensors related to the geometry and the energy-momentum, respectively.

After a few years, in 1919, Eddington headed the first observational mission to confirm one

of Einstein's theory predictions. The idea was based on taking advantage of the solar eclipse that occurred that year to measure the deflection of light from a star whose position, in the observer's line-of-sight, was close to the Sun in the sky plane. That is, the distortion of space-time caused by the Sun would bend the trajectory of the light emitted by the star according to the relativistic mathematics, and, thus, the observations would indicate an apparent shift in its position. The result was in good agreement with the predicted by general relativity. Currently, several predictions have already been confirmed making this theory a consolidated basis for several fields of physics.

1.1.2 | Cosmological Implications

Black Holes – Schwarzschild Radius

One of the first successful solutions to the field equations came about in the middle of the first world war with German astronomer Karl Schwarzschild. He devised a point source of mass, i.e., a situation with spherical symmetry, and described it with Einstein's equations. The solution for this problem has, at a given finite distance from the center of mass, a singularity. Nowadays, this distance is known as *Schwarzschild Radius*, and any structure with sufficient compact mass can distort the space-time into a singularity. Later, in 1958, David Finkelstein interpreted the Schwarzschild radius as a spherical region limit in space-time that, beyond it, nothing can escape, including light. This conception led to what we know today as an *event horizon*, which is the theoretical frontier of a *black hole* that can be simply calculated as

$$R_s = \frac{2GM}{c^2} \tag{1.3}$$

where R_s is the Schwarzschild radius centered in a body with mass M, G is the gravitational constant, and c the speed of light.

The *black hole* term was coined in 1967 by the American astronomer John Wheeler, and, only in the '70s, astronomers spotted these objects in the universe, reinforcing Einstein's general relativity theory with another prediction being validated.

Einstein's Model

In 1917, the creator of the theory presented his first attempt to give a cosmological solution to his equations. For this, Einstein based himself on three hypotheses:

1. The universe is homogeneous and isotropic in large-scales – today's known as *cosmological principle*

2. The universe has a closed spatial geometry – i.e., the three spatial dimensions have positive and uniform curvature, the universe is finite, however, it has no borders or limits, to be consistent with the first hypothesis.

3. The universe is static – i.e., its large-scale properties do not vary over time.

Based on these hypotheses, Einstein realized that the solutions found were meaningless since, without an entity to balance the gravity from matter, the universe would collapse. To get around this problem, he introduced a new term to his original equations – the *cosmological constant*

$$G_{\mu\nu} - \Lambda g_{\mu\nu} = \kappa T_{\mu\nu} \tag{1.4}$$

which, in Newtonian terms, would be a force with a module equivalent to gravitational, but opposite directions. This term ensured the third hypothesis.

Almost a decade later, when an expanding universe was conceived and observed, Einstein got frustrated with the modification in the initial theory – since he could have predicted this property with the original equations – this is what he declared to be the "biggest blunder" of his life. Later, when it was observed through distant supernovae the accelerated expansion of the universe, and the idea of "dark energy" was conceived as the repulsive force responsible for this expansion, what Einstein called an error ended up by becoming one of the most important variables for the today's adopted *standard cosmological model*, which will be presented later in this dissertation.

De Sitter's Model

In the same year (1917), Willem De Sitter came up with new theoretical solutions for the field equations, which neglects the ordinary matter. Therefore, the universe is dominated by the cosmological constant and, to keep it invariant in space-time, it bends. Thus, the radius of geometric curvature is given by $R^2 = \Lambda/3$.

The proper distance between two events in the same spatial frame can be described as

$$ds^2 = \left(1 - \frac{r^2}{R^2}\right) c^2 dt^2 \tag{1.5}$$

where r is the distance between the observer and the event and c the speed of light. This relation shows that the higher the r, the lower is the ds^2. If, for example, the event is the transit from one tick to another in a clock, the distance between the ticks (ds) will appear to be lower for an observer at a distance r, and, hence, for the same observer, the time slower during the transition.

In the case of a light emission from a given star, it is reasonable to think how the observer would capture the light in terms of frequencies or wavelengths

$$\frac{\delta\lambda}{\lambda_0} = \frac{r^2}{2R^2} \tag{1.6}$$

The redshifted wavelength that an observer at a distance r in a static universe that an observer would verify, became known as *de Sitter's effect*, and, which is a quadratic relation between distance and velocity – objects separated from each other at higher distances, extend it in a quadratic proportion faster.

Friedmann-Lemaître-Robertson-Walker – A More Realistic Approach

Until then, the cosmological theory was making significant advances with the relativistic treatment of time-space, but it still lacked a connection with the real universe, both in theory and in observations. At this point, in the 1920s, four scientists were fundamental to take important steps towards the standard cosmological model – a treatment with geometrical concepts of a universe with matter and redshifts.

In 1922, Alexander Friedmann derived solutions for field equations assuming a universe where the cosmological principle is valid, with matter, and a time-dependent scale factor. Also, the universe is filled with a perfect fluid with density (ρ) and pressure (p). Combining all of these assumptions, we have the Friedmann's equations

$$\frac{\dot{a}^2 + kc^2}{a^2} = \frac{8\pi G\rho + \Lambda c^2}{3} \tag{1.7}$$

$$\frac{\ddot{a}}{a} = \frac{-4\pi G}{3}\left(\rho + \frac{3p}{c^2}\right) + \frac{\Lambda c^2}{3} \tag{1.8}$$

where a is the time-scale factor, G is the Newtonian gravitational constant, and k is the curvature of the universe, which can assume +1, 0, -1 values, for a close, flat, and open universe, respectively, due to the cosmological principle. Five years later, the Belgian monk scientist George Lemaître, came not only with a more general solution for the equations but also with a new astronomical interpretation in his work "Homogeneous universe of constant mass and increasing radius, explaining radial velocities of extragalactical nebulae". Lemaître was responsible for conceiving for the first time the idea of an expanding universe. In his words

> Velocities of repelling of extragalactical nebulae represent the cosmical effect of the expansion of the Universe.

1929 was a year of tremendous importance both in theory and in observations. Howard P. Robertson and Arthur Geoffrey Walker independently derived solutions to the field equations based on Friedmann and Lemaître assumptions, and, introducing a new geometric formalism,

they gave birth to what is known today as the Friedmann-Lemaître-Robertson-Walker (FLRW) metric (Eq. 1.9), which is the most used model to describe a homogeneous and isotropic universe.

$$ds^2 = -dt^2 - a(t)^2 \left[\frac{dr^2}{1 - kr^2} + r^2(d\theta^2 + sin^2\theta \; d\phi^2) \right] \tag{1.9}$$

The Observational Proof

In the same year, based on the discoveries of Henrietta Leavitt about the period-luminosity scale (1912), Edwin Hubble published his analyses of the Cepheid's observations and realized that the recession velocity was directly proportional to the distance. Hubble's findings became an observational proof of an expanding universe where galaxies move away from each other according to a law that bears his name (Eq. 1.10). These results went down in history as one of the greatest discoveries of the 20th century. Although lots of credit is given to Hubble, as pointed out in the previous sections, his conclusion was not a single man's work.

$$v = H_0 \; D \tag{1.10}$$

where v is the recession velocity, D is the distance from the object to the observer, and H_0 is the Hubble's constant. The later is a fundamental parameter in the cosmological models since it is related to the universe's expansion rate.

Besides the brilliant scientists discussed briefly throughout this section, the complete history of these discoveries is enormous and goes further. Therefore, it would be impossible to describe it in detail and quote all the names and deeds of those involved in this dissertation. So, to be as fair as possible, I will list a few more cases

- 1912: Vesto Slipher noticed shifted lines in M31's spectrum.

- 1924: Knut Lundmark showed observational evidence of the expansion of the universe estimating accurate extragalactic distance.

- 1928: Howard P. Robertson predicted theoretically the linear expansion law – observed by Hubble one year later.

1.2 | The Formation and Evolution of Matter Structures

This section includes some important cosmological concepts that are a direct consequence of the Big Bang theory. It begins with a description of the Cosmic Microwave Background (CMB) and the Lambda-CDM model. Then, I present a brief discussion of the primordial energy density field, how the perturbations in this field evolve, and, consequently, arriving at the hierarchical model of structure formation.

The main references that I used were the *NASA* educational website[1], the *Extragalactic Astronomy and Cosmology: An Introduction* book (Schneider, 2006), the class notes of Professor Gastão B. Lima Neto (*Astronomia Extragaláctica e Cosmologia* – available at http://www.astro.iag.usp.br/~gastao/extragal/ExtraGal2020.pdf), and the *ADVANCED TOPICS IN COSMOLOGY: A PEDAGOGICAL INTRODUCTION* from Professor T. Padmanabhan (available at https://ned.ipac.caltech.edu/level5/March06/Padmanabhan/Nabhan3.html).

1.2.1 | Cosmic Microwave Background

Theoretically predicted in 1948 by George Gamow, Ralph Alpher, and Robert Herman, the CMB represents one of the best foundations for the Big Bang theory. In the early times, the universe was composed of a hot plasma of photons, electrons, and baryons, where the particles of energy were constantly interacting with the matter ones through the Compton Effect. Due to the expansion of the universe, this plasma was cooling throughout time, and, around 3000K, the combination of electrons and protons became possible, which gave rise to the first H and He atoms. With the particles of matter clumping together, the photons were able to move more freely, i.e., a larger *free mean path*. It is estimated that at this time the universe was approximately 380,000 years old (z = 1088).

Since then, the universe has continued to expand and, consequently, these photons are losing their energy and cooling. The CMB is this radiation. In other words, what we observe today on Earth, is the result of the cooling process of the recombination epoch photons, which the temperature is now approximately 2.7 K. At this level, the wavelength is quite long, and, therefore, observed in the microwave and radio spectral range.

The first measure was taken in 1965 unexpectedly by Arno Penzias and Robert Woodrow Wilson, from Bell Telephone Laboratories. In order to carry out experiments in radio astronomy and satellite communication, they used a radiometer device, that detected an excessive thermal noise of 3.5 K. It turns out that this noise was the prediction of Gamow, Alpher, and Herman,

[1] https://lambda.gsfc.nasa.gov/education/graphic_history/

which yielded the 1978 Nobel Prize in Physics to Penzias and Wilson.

The CMB shows anisotropies caused by the inflationary epoch in the first moments after the Big Bang. This aspect is a key element in cosmological models and the formation and evolution of matter structures scenarios, as will be seen in the following sections.

1.2.2 | The ΛCDM Cosmological Model

In the last three decades, observational technology development made it possible to perform cosmology through a large volume of more precise and accurate data. The three main observation fronts in this regard were

- Microwave sky mapping with satellites (COBE: Smoot et al. 1992; Bennett et al. 1996; WMAP: Bennett et al. 2003, 2013; Planck: Planck Collaboration et al. 2014, 2015), and ground-based telescopes (e.g., DASI: Halverson et al. 2002; SPT: Keisler et al. 2011) with the instrumental optimized for cosmological purposes.

- The use of Type Ia Supernovae to measure the expansion rate of the universe (Riess et al. 1998; Perlmutter et al. 1999; SDSS-II SN survey: Sako et al. 2008; SNLS: Sullivan et al. 2011).

- The development of photometric and spectroscopic large-area surveys of galaxies and quasars (e.g., SDSS: Schneider and SDSS Collaboration 2001; Cannon et al. 2003; 2dFGRS: Colless et al. 2001; 6dFGS: Jones et al. 2004; WiggleZ: Blake et al. 2010; COSMOS2015: Laigle et al. 2016).

With the evolution of the observational astrophysics, the theoretical models were refined and, currently, the ΛCDM (CDM – Cold Dark Matter) is the most accepted to explain the consequences of a universe originated by the Big Bang. For this reason, it became known as the *standard model*.

This model is a mathematical parametrization of Big Bang cosmology, assuming GR as the theory which explains gravity on cosmological scales, and the FLRW metric and hypothesis, for the solution of the field equations. It can be completely specified with six independent parameters. This set of six parameters can be somewhat freely specified, and from them, others can be derived. Examples are the physical baryon density parameter ($\Omega_b h^2$), physical dark matter density parameter ($\Omega_c h^2$), age of the universe (t_0), scalar spectral index (n_s), scalar fluctuation amplitude (A_s), and reionization optical depth (τ) – all of them can be constrained with the observations described above. ΛCDM assumes that the universe is composed of photons, neutrinos, baryonic matter, dark matter, and dark energy. The latter is responsible for explaining the expansion of the universe, observed by Hubble, and is modeled with the cosmological constant (Λ).

As much as the ΛCDM model has obtained and continues to be successful in several tests throughout its history, there are still challenges to achieve. The biggest problems are directly related to the nature of energy and dark matter, which are, together, responsible for $\sim 95\%$ of the universe, according to this model. In addition to not knowing what they are, some measurements still show inconsistencies. Examples of this are the fine-tuning problem and the coincidence problem. The former-one rises with the measurement of Λ, which is too low than what is expected by the particle physics prediction. The later is related to the fact that dark matter and dark energy densities are almost equal for no obvious physical reason.

Therefore, advances in technology, data processing, and analysis techniques have been the main tools to go further in this investigation. In the last decades, the ΛCDM kept its position as the best and most used cosmological model, while its parameters continue to be refined (see Planck Collaboration et al. 2015), and its inconsistencies ascertained.

1.2.3 | The Hierarchical Structure Formation and Evolution Model

The hierarchical structure formation scenario is a natural outcome of the ΛCDM cosmological model. Small density fluctuations collapse to give rise to the first stars and galaxies. These galaxies continue to grow through gas infall and the accretion or merger of other smaller or similar mass structures (e.g., White and Rees, 1978), at the same time that the first groups and clusters of galaxies assemble (e.g., Kravtsov and Borgani, 2012). All of these visible structures made up of baryonic matter follow the development of underlying dark matter halos clumps. Therefore, when a large structure is observed, like a cluster of galaxies, it is known that there is an immense amount of dark matter bounding all the galaxies together – a huge dark matter halo. The explanation of this behavior is the gravitational evolution of a density perturbation, where overdense and underdense regions will evolve to even higher contrast density across time (1.11).

$$\delta(x,t) = \frac{\rho(x,t) - \bar{\rho}(t)}{\bar{\rho}(t)} \tag{1.11}$$

This process is known as *gravitational instability*. Any inhomogeneity in the energy density distribution in the very early universe, directly means differences in the density field as well, and, thus, gravitational instability will take place to always amplify these contrasts. In practice, this means the formation of increasingly larger structures of dark and ordinary matter.

The primordial energy field

Currently, the most accepted model to explain the origin of the inhomogeneities is based on the idea that the early universe has gone through a very rapid inflation phase that leads to the primordial energy density field fluctuations. A common approach to better understand this inflation scenario is based on model construction in which the fluctuations are described by a Gaussian random field, and characterized by an initial power spectrum as a function of the wavenumber k in the form $P_0(k) = A\ k^{n_s}$, with $n_s \sim 1$. The values of the amplitude A and the scalar spectral index n_s can be constrained by the temperature fluctuations measurements at the *cosmic microwave background* (CMB) (Figure 1.1). The inflation perturbations were the responsible for the creation of all particles known in the universe from the stored energy of the inflation field – the reheating process. Since the perturbations in all other quantities were originated from the same progenitor, they are adiabatic, and the number densities are the same for all quantities

$$\delta\left(\frac{n_i}{n_\gamma}\right) = 0 \implies \frac{\delta n_i}{n_i} = \frac{\delta n_\gamma}{n_\gamma} \tag{1.12}$$

the numeric density of a given specie (n_i) is related with the matter density as $\rho_i = m_i n_i$. Also, the continuity equation correlates photon number density (n_γ) with temperature (T) as $n_\gamma \propto T^3$, and the photon energy density as $\rho_\gamma \propto T^4$ (e.g., Kurki-Suonio, 2010).

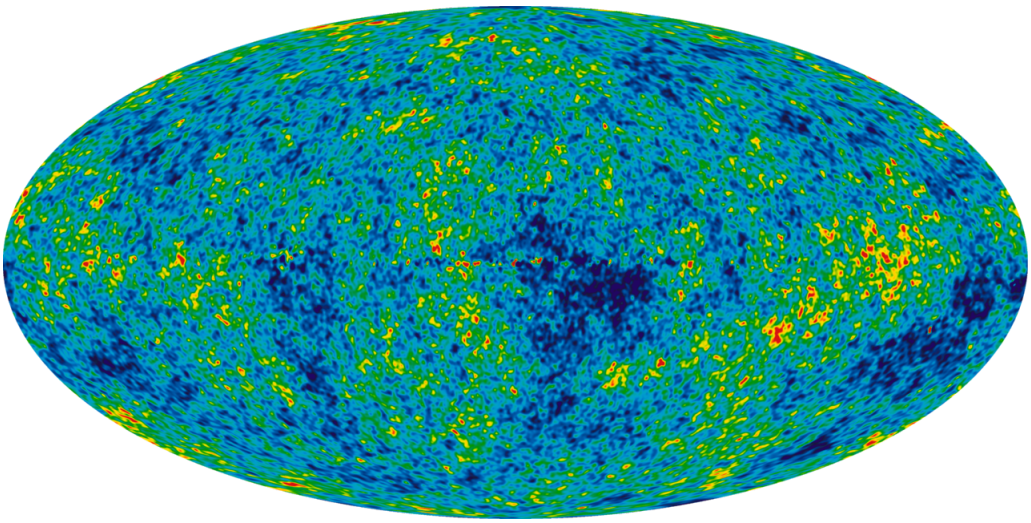

Figure 1.1: The Cosmic Microwave Background temperature fluctuations from the 7-year Wilkinson Microwave Anisotropy Probe. The colors represent the temperature fluctuations, with differences ~ 0.0002 K between the blue (colder) and the red (warmer). The average temperature is 2.725 K. Public domain image. Available at https://wmap.gsfc.nasa.gov/media/101080/

Depending on the density field perturbation size, it is known which theory must be invoked to explain their evolution. For scales where the density contrast $\delta(x,t) \sim 1$, the linear perturbation theory is enough. However, for higher values, the non-linear theory is necessary.

The crucial moment that defines when linear theory is no longer valid is the equipartition. This is when matter becomes dominant and the radiation pressure is no longer strong enough to

contain gravity. Therefore, structures begin to form due to the infall of baryons on to dark matter halos. The power spectrum will also change since the universe horizon size at the equipartition moment (L_0) is a critical length for how a structure with a given size (L) will evolve

$$\begin{aligned} L > L_0 &\implies P(k) \propto k^1 \\ L < L_0 &\implies P(k) \propto k^{-3} \end{aligned} \tag{1.13}$$

The transfer function

Another fundamental information to explain the evolution of the density field perturbations is the transfer function $T(k)$. It will take into account the physical properties associated with the ΛCDM universe and, thus, dictate how the power spectrum will change. These properties are

- The dark matter is *"cold"*. Otherwise, the small initial density fluctuations would not evolve to form bigger structures because even if only a fraction were *Hot Dark Matter* (HDM), the *free-streaming* could not be prevented[2].
- Before equipartition, the radiation was the dominant component of the universe. This aspect leads to the differences in the universe's expansion law. So, if the growth of the universe itself changes, the growth of matter structures will change as well.
- There is a comoving cosmic horizon ($r_{H,c}(t)$). This determines the volume where the physical interactions can occur. If $L \sim k/2\pi \gtrsim r_{H,c}(t)$, the GR, and the linear perturbation theory, are necessary to explain the growth of structures.

In the limiting cases, the transfer function can assume the asymptotic form

$$\begin{aligned} L \gg L_0 &\implies T(k) \propto 1 \\ L \ll L_0 &\implies T(k) \propto \frac{1}{(kL_0)^2} \end{aligned} \tag{1.14}$$

The power spectrum ($P(k) \propto k^{n_s} T(k)^2$) associated with the asymptotic form of the transfer function, explains the behavior of the perturbation evolution in the equipartition epoch, described, approximately, by the 1.13. Qualitatively, perturbations that enter the horizon ($L < L_0$) at $z_{enter} > z_{eq}$, will have their growth suppressed by the dominant radiation and, therefore, will only grow again when $z < z_{eq}$. However, this growth occurs at a lower rate than the perturbations that entered the horizon after the matter dominance ($L > L_0$ and $z < z_{eq}$). From this, a general property of structure formation in the CDM universe is derived – the *hierarchical structure formation*, or "bottom-up"-scenario, in which the small perturbations collapse first since the time for this process is also smaller.

[2]This phenomenon happens when the potential wells are not enough to hold the dark matter particles bounded.

The spherical collapse

The mass variation in a density field characterized by a power law can be derived as

$$\frac{\delta M}{M} \propto M^{-\frac{n_s+3}{6}} \tag{1.15}$$

Replacing the 1.13 relations, it is possible to replace the n_s variable depending on the scale L of the structure. Thus, the results obtained show that the smaller structures will reach the non-linear regime first, and, therefore, collapse before the major perturbations.

The non-linear regime starts when the density contrast becomes close to unit. From this moment on, a simple way to deal with the evolution of perturbations is through the spherical collapse model. In this idealization, a perturbation is spherical, with constant total mass M, and, up to a radius of R, the density profile also remains constant. After R, the density drops abruptly to zero – the *Top-Hat profile* (Figure 1.2). To describe the evolution of this sphere, we can use the energy law as

$$E = \frac{1}{2}\left(\frac{dr}{dt}\right)^2 - \frac{Gm}{r} \tag{1.16}$$

where E is the total energy per unit of mass.

In this model, initially, the condition $E < 0$ must be satisfied for a gravitationally bound sphere, which starts to collapse. Then, it is possible to derive a parametric solution for the diferential equation (1.16) as

$$\begin{aligned} r &= A(1 - cos\theta) \\ t &= B(\theta - sin\theta) \end{aligned} \tag{1.17}$$

where $\theta \in [0, 2\pi]$. With this solutions, first the sphere will expand ($\frac{dr}{dt} > 0$). However, with a lower rate across time ($\frac{d^2r}{dt^2} < 0$) since the universe expansion is faster than the sphere expansion, and also because the gravitational component increases with the aggregated matter. The sphere expands until it reaches a maximum radius (r_m) in a given time (t_m). As Figure 1.2 shows, at this point, it starts to shrink ($\frac{dr}{dt} < 0$), and a minimum radius is reached, at t_{col}. Then the value will float around half of r_m, where the Virial equilibrium happens. This oscillation is known as *violent relaxation* since it happens much faster than the two-body dynamical relaxation. The contrast density of the initial perturbation at this moment in an universe like ours ($\Omega_m \sim 0.27$ and $\Omega_d \sim 0.73$) is $\delta \sim 356$.

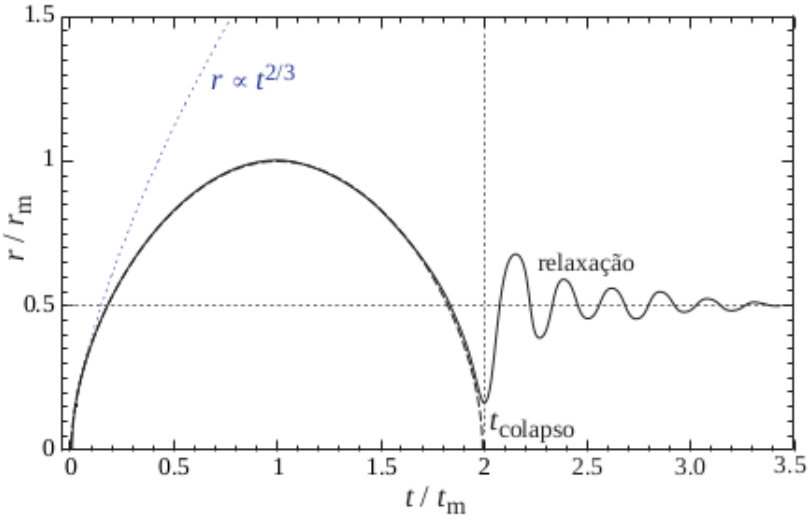

Figure 1.2: The evolution of the perturbation radius assuming the spherical collapse model with the Top-Hat profile. The dotted line is the linear approximation evolution, and the continuous line, the non-linear evolution regime obtained by numerical simulations. This Figure was extracted from Prof. Gastão B. Lima Neto class notes.

1.3 | Galaxy Clusters

The history of galaxy clusters begins at the end of the 18th century. However, the concept of "galaxy" would only flourish in the middle of the 20th century and, therefore, it was an unknown structure. Two important characters were William Herschel and Charles Messier, who observed concentrations of "nebulae" in the constellations of Coma Berenice and Virgo, respectively. Later, in the '20s, with Hubble's discoveries, it became known that "nebulae" are objects which reside outside the Milky Way, and similarly, with The galaxy, they also should be galaxies. It turns out that the concentration of "nebulae" noticed by Herschel and Messier, then, came to be understood as a concentration of several galaxies – a cluster of galaxies. Nowadays, the clusters observed at the end of the 18th century are known as Coma and Virgo clusters, respectively. Among the closer ones, the first is one of the largest and more massive clusters, and, the second, the richest. An important galaxy catalogue receives the name of Messier in honour of the french Astronomer Charles Messier.

A great difficulty in studying this type of structure lies in the fact that it does not have a clear observable physical boundary. The galaxies that are members of a cluster have different physical-chemical characteristics, which adds more complexity to the observations, considering that the electromagnetic emission varies according to the phenomena and types of objects that compose the cluster. Besides, they are distant structures, and it is possible to observe a large number of these objects only in surveys that cover a large volume of the universe. Given these reasons, systematic studies of galaxy clusters have only become possible in last decades, due to the development of observational techniques, the instrumental technology of telescopes and satellites, and computational to treat and analyze large volumes of data.

Currently, the wide or deep-angle multi-band photometric surveys, like the *Sloan Digital Sky Survey* (SDSS), enabled the identification of thousands of groups and galaxy clusters. Due to the extremely hot nature of the intra-cluster environment, instruments optimized for X-ray observations (Chandra and XMM-Newton) are of fundamental importance bringing detailed knowledge of the structure and dynamics of the clusters. The Sunyaev–Zeldovich effect is another important phenomenon that is also associated with the thermal characteristic of the intra-cluster medium since the energy difference associated with a photon from the cosmic microwave background (due to the inverse Compton process) provides information about the thermodynamic quantities of the structure. Finally, the gravity of such massive structures distorts the time-space as predicted by the general relativity, and, consequently, causing the gravitational lensing effect, which can also be used to identify and for measurements as the total mass and the size of the cluster.

An important parameter for these structures is the richness, which by the most traditional Abell definition (Abell, 1958), measures the number of galaxies brighter than $m_3 + 2$[3] within a clustercentric radius of R $\sim$ 1.5 h^{-1} Mpc. A cluster with at least 30 members that fall under these constraints is considered to be a rich cluster. Following this definition, typical rich clusters present a line-of-sight velocity dispersion of $\sim$ 750 km s^{-1}, which corresponds to a typical total mass range of $\sim$ 2 x 10^{14} – 10^{15} $M_\odot$, divided into $\sim$ 85, 10, 5%, of dark matter, gas, and stellar masses, respectively. Furthermore, the galaxy members of a cluster live in an extremely hot and rarefied medium - the intracluster medium (IGM). The gas in this environment has temperatures ranging from 2 – 14 keV, and an electron density of $\sim 10^{-3}$ electrons cm^{-3}. For this reason, IGM is responsible for strong X-ray emissions $\sim L_x = 10^{44}$ erg s^{-1}, due to bremsstrahlung thermal energy radiation. Table 1.1, adapted from Bahcall (1996), summarizes the typical properties of rich and poor galaxy clusters.

Despite being colossal structures, galaxy clusters are relatively rare, with only $\sim 10^{-5}$ clusters Mpc^{-3}. As the redshift increases, many clusters were still in their formation process, i.e., structures at high redshifts were not virialized, and, therefore, were more diffuse in space. Thus, detect this structures is a challenge, as will be shown in the next sections.

[3] m_3 is the magnitude of the third brightest galaxy in the cluster

Table 1.1: Galaxy clusters properties (from Bahcall 1996).

Property	Rich Clusters	Groups and Poor Clusters
Richness[a] [galaxies]	30 – 300	3 – 30
Radius[b][h^{-1} Mpc]	1 – 2	0.1 – 1
Radial Velocity Dispersion[c] [km s^{-1}]	400 – 1400	100 – 500
Mass[d] ($r \leqslant 1.5h^{-1}Mpc$) [h^{-1} $M_\odot$]	10^{14} – 2 $\times 10^{15}$	$10^{12.5}$ – 2 $\times 10^{14}$
Luminosity[e] (L_B) [h^{-2} $L_\odot$]	6 $\times 10^{11}$ – 6 $\times 10^{12}$	$10^{10.5}$ – 10^{12}
Mass-to-light ratio[f] (M/L_B) [$M_\odot/L_\odot$]	$\sim$ 300	$\sim$ 200
X-ray Temperature[g] [keV]	2 – 14	$\lesssim 2$
X-ray Luminosity[h] [h^{-2} erg s^{-1}]	$10^{42.5}$ – 10^{45}	$\lesssim 10^{43}$
Cluster Number Density[i] h^3 Mpc^{-3}	10^{-5} – 10^{-6}	10^{-3} – 10^{-5}

[a] *The richness of the cluster following the Abell's criteria.*
[b] *The radius where the cluster shows the highest concentration of galaxies. Above it, the remaining galaxies members represent $\sim$ 1% of the total.*
[c] *The range of radial galaxies velocity dispersions.*
[d] *The dynamical mass range within $1.5h^{-1}$ Mpc radius.*
[e] *The luminosity range in the B band within $1.5h^{-1}$ Mpc radius.*
[f] *The mean mass-to-light ratio $^d/_e$.*
[g] *The Intracluster Medium (IGM) temperature measured in X-ray radiation.*
[h] *The X-ray luminosity range.*
[i] *The number of galaxies within a megaparsec cubic volume.*

1.3.1 | Clusters in Different Redshifts

Studying the formation and evolution of galaxy clusters, through observations at different redshifts, is one of the main approaches to constrain our knowledge on the large-scale structures. At $z = 0$, cluster galaxies show the direct impact of living in such dense environments over billions of years. Examples include the high incidence of quiescent galaxies and the relationship between the morphological types and local galaxy density, indicating that the environment affects the evolution of a galaxy and its stellar population. Such characteristics are evidenced, for example, by the "red sequence" (Visvanathan and Sandage, 1977; Bower et al., 1992) in the color-magnitude diagrams and the 'morphology-density' relation (Dressler, 1980). Many studies seek to identify the molecular gas quenching level in different redshifts to infer the possible evolutionary stage of the galaxies. Those which are part of groups/clusters tend to lose their gas more quickly throughout their history and, consequently, no longer form stars. This can be caused, for example, by the interaction between different members of the group/cluster ('galaxy harassment', e.g., Moore et al., 1996). Another possibility is the falling of a galaxy into a cluster, where the ICM causes a rapid gas depletion through 'ram pressure stripping' (e.g., Gunn and Gott, 1972). There is also the phenomenon known as 'strangulation' (e.g., van den Bosch et al., 2008), where the stripping of the hot circumgalactic medium gas will interrupt the flow of cold gas into the galaxy, which ceases the fuel available for the formation of new stars. All of these processes are widely accepted as the main causes of star formation quenching and, consequently, reddening of galaxies that live in the densest environments in the universe (galaxy quenching, e.g., Koutsouridou and Cattaneo, 2019; Liu et al., 2019; Trussler et al., 2020; Joshi et al., 2020; Gu

et al., 2020; Lu et al., 2020; Tiley et al., 2020). Among others, these recent studies take advantage of this effect by using measurements as the quenching efficiency and fraction to compare field galaxies with clusters members.

At low redshifts ($z \lesssim 1$), most high mass clusters are virialized structures that can be identified through a concentration of red galaxies[4]. Additionally, their hot intra-cluster medium helps their identification through either their high X-ray emissions (Rosati et al., 2002; Mullis et al., 2005; Stanford et al., 2006; Mehrtens et al., 2012) or by the Sunyaev-Zeldovich effect (Zel'dovich and Syunyaev, 1972; Bleem et al., 2015; Hilton et al., 2018; Huang et al., 2020). At higher redshifts, $z \gtrsim 1$, most of the observed clusters are still in their formation process, and are often called high-z clusters or protoclusters (see Overzier, 2016, for a review).

1.4 | Protoclusters

The definition of galaxy clusters is still ambiguous. Basically two different interpretations are used by different authors:

1. At z = 0, virialized structures with $M \gtrsim 10^{14}$ $M_\odot$ are considered "clusters". Whereas, those with $10^{13} \lesssim M \lesssim 10^{14}$ $M_\odot$, are called "groups" (e.g., Bower and Balogh, 2004).
2. Above a certain mass limit, all structures are considered as "groups", or "clusters" (e.g., Yang et al., 2007).

The first definition is more commonly used, as it has the advantage of facilitating the evolutionary study of these objects due to the clear limits that can be applied to compare with computer simulations, and theoretical studies. Therefore, adopting this definition of galaxy cluster, Overzier (2016) wrote the definition of "protocluster" as

> A protocluster is a structure that will, at some stage, collapse into a galaxy cluster (i.e., a virialized object of $\gtrsim 10^{14}$ $M_\odot$ at $z \geqslant 0$).

And he also points out that

> This definition has the convenient implication that the combined space number density of clusters and protoclusters at any redshift is equal to the abundance of clusters today.

Thus, protoclusters are the progenitors of today's galaxy clusters (Figure 1.3).

When studying protoclusters, a very interesting question is "based on its characteristics and redshift, this structure will evolve into a rich and massive, average, or a poor cluster?". To answer this question, the numerical simulations are a very useful tool since it is possible to obtain

[4]redMaPPer: http://risa.stanford.edu/redmapper/

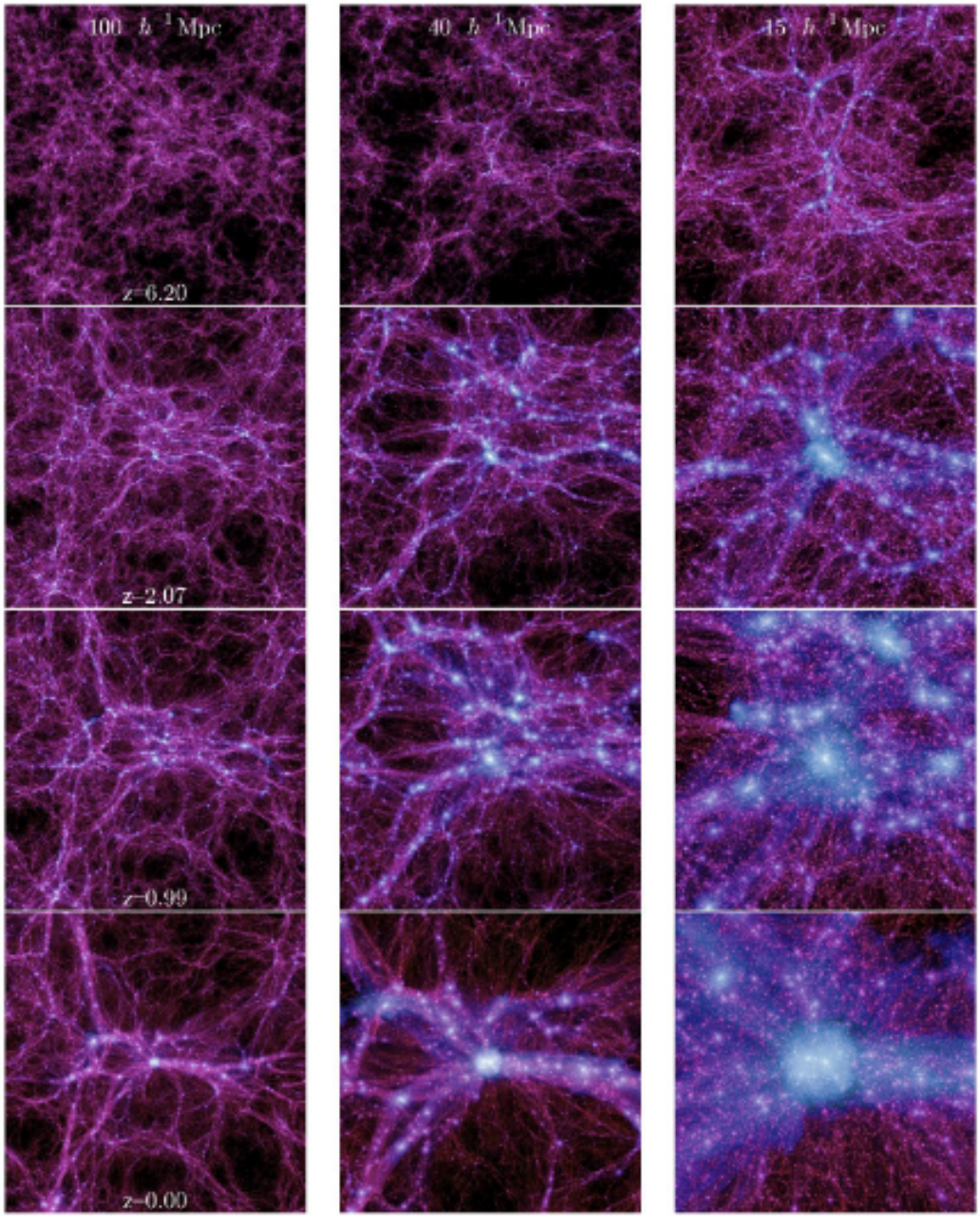

Figure 1.3: The formation of a massive protocluster from the Millennium II dark matter simulation. The panels show the evolution through different redshifts (6.20, 2.07, 0.99, 0, from the top to the bottom), and the visualization in different scales (100, 40, 15 h^{-1} Mpc, from the left to the right). This figure was extracted from Boylan-Kolchin et al. (2009)

protoclusters evolutionary tracks. Based on that, Chiang et al. (2013) introduced a criterion to classify the structure formed at z = 0. He used the Millennium Run simulation to study in detail the properties of $\sim$ 3000 clusters sample (Figure 1.4). They were divided according to the structure's descendant mass (at z = 0) using three famous observed clusters as reference: Fornax, Virgo, and Coma (see Table 1.2). The left panel of Figure 1.4 shows that, at z $\sim$ 0.5-0.6, most of the progenitors of the three different classes of clusters had already reached the 10^{14} $M_{\odot}$ limit. On the other hand, at z $\geqslant$ 2, all the structures were below this mass. Besides, it is also interesting to observe, at the right panel of the same figure, that the effective radius stabilizes

(indicating the structure's virialization) at z $\sim$ 0.6, 0.5, and 0.4, for Fornax, Virgo, and Coma-like, respectively. These results show the complexity of defining a dynamical structure at high redshifts, like a protocluster.

As will be presented in section 1.5, in this work we searched for high concentrations of galaxies in 6 different fields at z = 1 – 1.5, depending on the field. Within this range, almost all the Coma-like protoclusters, the average of the Virgo-like, and no Fornax-like, reached 10^{14} $M_{\odot}$. Moreover, no structure is virialized at these redshifts according with the models used in this study.

Table 1.2: The protocluster type division based on its descendant mass.

Cluster type	Mass range
Fornax-like	1 – 3 x 10^{14} $M_{\odot}$
Virgo-like	3 – 10 x 10^{14} $M_{\odot}$]
Coma-like	> 10^{15} $M_{\odot}$]

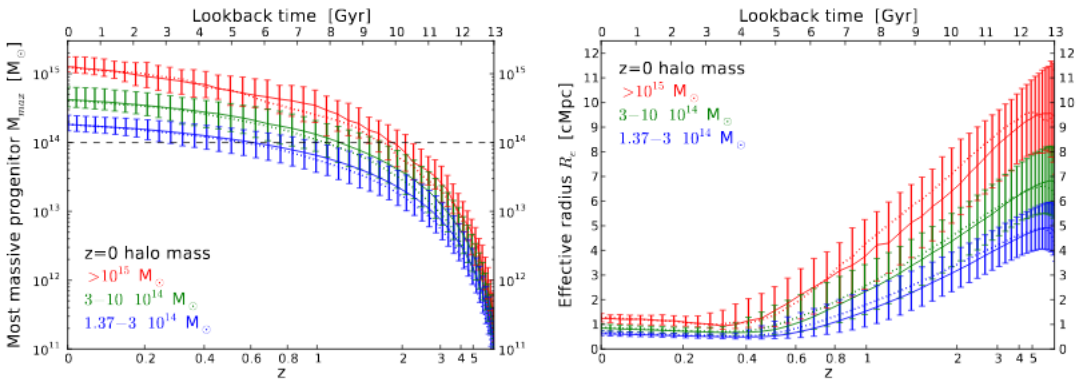

Figure 1.4: Evolutionary track for mass (left panel) and effective radius (right panel) in function of the redshift and lookback time in the lower and upper horizontal axis, respectively. This Figure was extracted from Chiang et al. (2013).

1.4.1 | Discovering Protoclusters

These structures harbor important clues about the galaxy formation process and the star formation history in such environments (White and Frenk, 1991; Bekki, 1998; Kodama et al., 1998; van Dokkum, 2005; Mei et al., 2006). However, their identification is not trivial. Since they do not present many of the observational properties of low-z, virialized clusters (see Section 1.3.1), the search for galaxy overdensities at high redshifts is one of the most effective ways to find them (Overzier, 2016).

In this investigation, it is important to distinguish whether a given overdensity is real or a projected structure through the analysis of the redshift distribution, using either spectroscopic or photometric redshifts (e.g., Adami et al., 2010, 2011; George et al., 2011; Wen and Han, 2011; Jian

et al., 2014). At $z > 1$, spectroscopic redshift samples are quite limited or incomplete. However, the increasing number of photometric surveys in different broad, medium, and narrow-bands have made it possible to obtain high-quality photometric redshifts using template-fitting or machine learning techniques, useful for the search of high redshift galaxy clusters. For example, Chiang et al. (2014) reported 36 new candidates in the COSMOS field; by analyzing data from the Wide-field Infrared Survey Explorer (WISE) mission on the Pan-STARRS and SuperCOSMOS surveys, Gonzalez et al. (2019) found an amazing 1787 new high redshift clusters; at $z \sim 4$, Toshikawa et al. (2018) reported another 210 candidates in the wide layer area of the HSC-SSP; Martinache et al. (2018) also present new candidates for protoclusters by studying *Herschel* and *Planck* sources at $1.3 < z < 3$.

Photometric Redshifts

The discovery of the relationship between astronomical distances and radiation flux at different wavelengths has become one of the greatest tools for the astronomical community. From then on, the transformation between an observable obtained by a telescope (flux), into one of the most important quantities within many different areas of this science became possible. In the last seven decades, the extragalactic astrophysics and cosmology fields used extensively the differences between flux measurements in different wavelengths (colors) of galaxies to estimate cosmological distances - the *photometric redshifts*.

The great advantage of the photometric redshifts is that with only photometric images, it becomes possible to obtain distances for all the detected objects while obtaining spectroscopic data is, in all senses, much more expensive. The counterpart is that the photometric redshifts have a significant larger associated error that depends on the method used for the estimates, and, the quantity and quality of the photometric measurements. That is why it is always important to evaluate the performance of the estimates using metrics such as the *bias*, outlier fraction, and *normalized median absolute deviation* (NMAD).

According to Salvato et al. (2019), in the last two decades, refereed works that include the term "photometric redshift" have grown by a factor of ten, which can be partly explained by the rise of multi-wavelength surveys. One of the fields most benefited from this advent is the formation and evolution of galaxies studies. A large number of galaxies with flux measurements in several photometric bands, are the basis for high-quality photometric redshifts and, also, large samples of all these quantities together, allow statistical analysis approaches. Some examples are studies related to the evolution of galactic properties in cosmic time (e.g., Fontana et al., 2000), as well as the identification of high-density regions combined with the possible formation of galaxy groups or clusters (e.g., MacKenzie et al., 2020) and the evolution of the members (e.g., Shi et al., 2020) or the cluster as a whole structure, and, also, the relation between these baryonic objects with the underlying dark matter halo (e.g., Coupon et al., 2015).

Biased Tracers of Protoclusters

There are some galaxies or systems with special physical characteristics that may be used to find overdensities, such as Lyman Break Galaxies (LBGs) and Lyman Alpha Emitters (LAEs) (e.g., Overzier et al. (2006, 2008); Chiang et al. (2015); Bădescu et al. (2017); Higuchi et al. (2019b)), Hydrogen Alpha Emitters (HAEs) (Hatch et al., 2011b; Hayashi et al., 2012), submillimeter galaxies (Daddi et al., 2009; Capak et al., 2011; Rigby et al., 2014; Dannerbauer et al., 2014), radio galaxies (Venemans et al., 2002, 2007; Hatch et al., 2011a; Hayashi et al., 2012; Wylezalek et al., 2013; Cooke et al., 2014), isolated QSOs (Sánchez and González-Serrano, 1999, 2002; Stott et al., 2020) and systems of QSOs (Boris et al., 2007; Onoue et al., 2018). These are all *biased tracers* and are useful to find massive systems in formation (Overzier, 2016), since the bias between the distribution of baryonic matter and dark matter haloes in structure formation scenarios (Kaiser, 1984) implies that high-z objects form in large high-z density fluctuations.

Lyman Alpha Emitters The Lyα emission analysis is important to understand the reionization state of the intergalactic medium (IGM) at high redshifts (2–7). LAEs photons with energy higher than the Lyman limit (912 Å)[5] are almost completely absorbed by the HI atoms of the IGM, causing its ionization (Figure 1.5). Also, photons with lower but enough energy excite the HI, and, hence, when the electrons return to the fundamental state from n = 2, the Lyα emission happens. Therefore, the HI density is essential to calculate how much of the Lyα emission line and its blue side damping wing will be affected in the photon's trajectory from the emitter to the observer. It is still unclear what kind of object is a LAE, but it has been suggested that distant low-mass galaxies with an intense star formation rate are good candidates (Higuchi et al., 2019a). LAEs may indicate high-density regions because it is expected that large ionized bubbles form in a volume with a high concentration of star-forming galaxies (Ishigaki et al., 2016; Overzier, 2016; Chiang et al., 2017).

Lyman Break Galaxies The population of galaxies that belong to this group reside in high redshifts, they are relatively massive, and with an evolved stellar population (Iye, 2011). This name is related to a peculiar spectral characteristic of an abrupt fall in the blue part of the spectra right after the redshifted Lyα line (Figure 1.5), which means that these galaxies can only be observed at wavelengths redder than this line. This is caused by the intrinsic absorption of the stellar atmospheres and, also, of the neutral hydrogen in the galaxy or in the IGM. Therefore, as explained for the LAEs, such aspects are correlated with high-density regions. It is also interesting to notice that this spectral feature of flux intensity attenuation at the bluer side of the Lyα line gave rise to a very efficient technique for determining redshifts using only photometric images known as "dropout galaxies" (e.g., Madau et al., 1996; Toshikawa et al., 2016). According to this technique, objects must be observed in different photometric filters which cover a wavelength range that includes the redshifted Lyα emission line. The filter at the bluer side of the line in the observer frame will have a flux intensity much lesser than at

[5]The Lyman limit wavelength correspond to the energy of the ionization potential of the hydrogen.

the red side. Hence, it would be barely seen in photometric bands bluer than Lyα. With this information, it is possible to estimate the photometric redshift. Figure 1.6 shows an example of this effect.

Multi-band images of a field containing high redshift galaxies can be used to identify those objects that have very red colors as a result of the redshifted Lyman limit falling between any two filters.

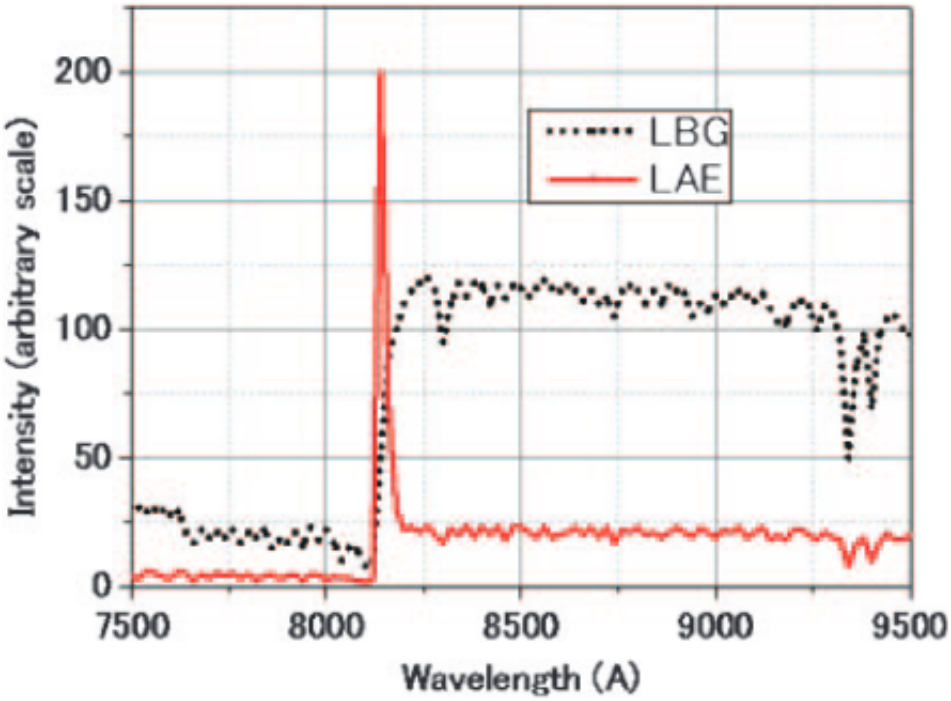

Figure 1.5: An example of a typical LAE and LBG spectra. LAE objects present a very clear Lyα emission line and, then, the flux falls abruptely in the bluer side. This same intense attenuation happens for the LBGs, but without the Lyα emission. This figure was extracted from Iye (2011).

Hydrogen Alpha Emitters Some surveys that were carried out on already confirmed protoclusters at z $>$ 2 show that there is an excess in the numerical density of Hα emitter galaxies which are part of these structures (e.g., Hatch et al., 2011a; Tanaka et al., 2011). Also, some studies with mid-IR photometry, which measures the radiation re-emitted by dust, show that in these environments there is a star formation activity similar to the general field at the same redshift, and the higher the redshift, the more the formation rate increase (Hayashi et al., 2012).

Radio Galaxies In several studies over the past six decades, it has been shown that powerful radio-loud AGNs tend to reside in the most luminous regions of the color-magnitude diagrams red sequence galaxies (e.g., Matthews et al., 1965; Griffith and Stern, 2010). Based on this aspect, several studies have turned their attention to identify protoclusters where such objects are located, which resulted in a high success rate and, therefore, radio sources have become an important tracer of galaxy protoclusters (e.g., Pentericci et al., 2000; Mayo et al., 2012). From a quantitative perspective, Galametz et al. (2012) found 23% of their radio galaxy sample, from

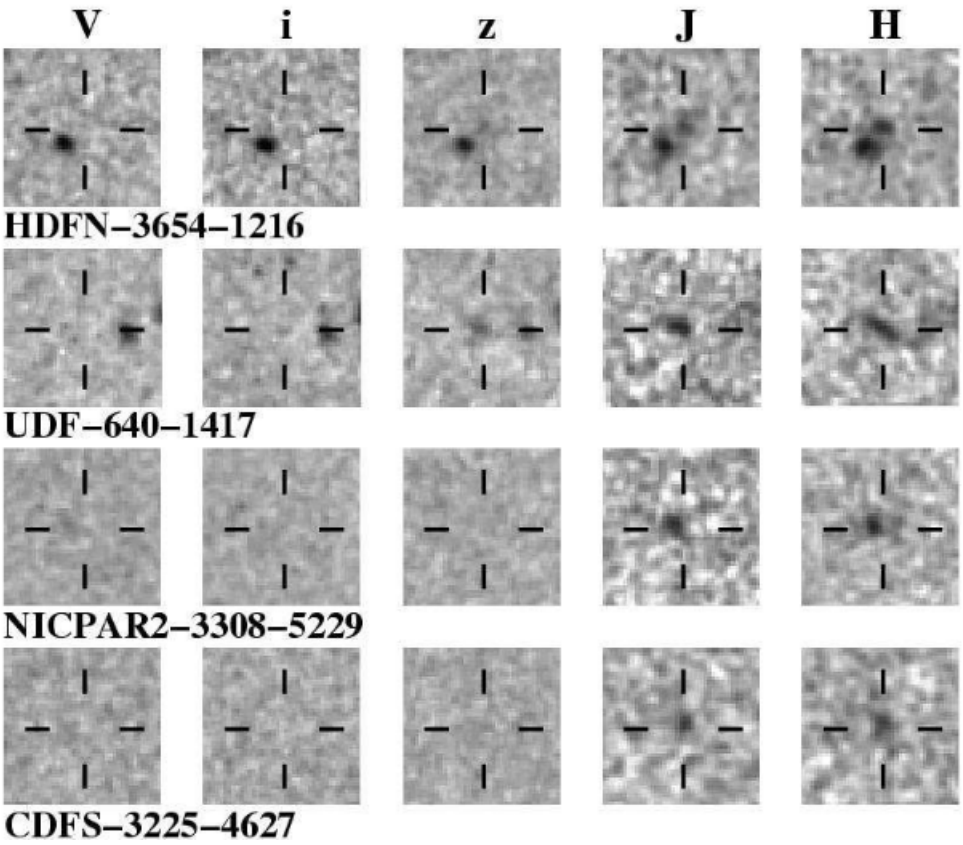

Figure 1.6: This is Figure shows a sub-sample of Bouwens et al. (2008).

SWIRE survey (Lonsdale et al., 2003), being part of overdense regions. This is very significant when compared with the 0.3% high densities of random control fields of the same survey with similar sample conditions. Noirot et al. (2016) reported two spectroscopic cluster confirmations from the CARLA survey sample at z $\sim$ 2, and Cooke et al. (2016) also confirmed one mature cluster at z = 1.58 from a candidate sample of the same survey. Using *low luminosity radio galaxies* (LLRG) at z = 1 – 2, Castignani (2017) found an amazing 70% out of the 32 LLRG sample from the COSMOS survey inhabiting rich groups of galaxies, or protoclusters.

Isolated and Systems of Quasars

Quasars are luminous objects that inhabit the center of distant galaxies. Their high luminosity is due to the interactions of a supermassive black hole with the matter particles of an accretion disk orbiting it. The greater the mass of the central black hole, the greater the system's spin and the kinetic energy of the particles. Shocks in the disk are the responsible for transforming part of this energy into thermal, which highly increases the bolometric luminosity of this object. At this stage, it is possible to occur the formation of jets in the centre and perpendicular to the system *black hole + accretion disk* plane.

These characteristics qualify quasars as a type of active galaxy nucleus (AGN), and, due

to their high luminosity, they can be observed at very high redshifts. The formation of jets is associated with quasars with the most massive black holes ($\gtrsim 10^8 M_\odot$) and luminosities ($\gtrsim 10^{45}$ergs/s). In such cases, there is radio emission (due to the acceleration of charged particles in the jet, emitting synchrotron radiation), and they are classified as *radio-loud quasars*. Given their high mass and luminosity, it has been suggested that radio-loud quasars inhabit high-density regions and, like radio galaxies, they are often investigated as tracers of high redshift clusters, or protoclusters.

Recently, Stott et al. (2020) studied 12 fields containing UV luminous QSOs at $1 < z < 2$. The data were obtained with the *Hubble Space Telescope* WFC3 G141 grism spectroscopy, and 2/3 of the sample showed significant galaxy overdensities associated to this type of quasar. Cheng et al. (2020) measured overdensities of 850μm-submillimetric sources in a sample of 46 protoclusters candidates selected from the Planck high-z catalogue (PHz, Planck Collaboration et al., 2016) and the Planck Catalogue of Compact Sources (PCCS) that were followed up with Herschel-SPIRE (Planck Collaboration et al., 2015) and SCUBA-2 (Geach et al., 2017), finding that 25 are associated with significant overdensities.

It is still a matter of debate whether quasar systems are associated with overdense regions (here we will use QSO or quasar indistinctly to designate objects with active nuclei). Previous studies with QSOs pairs have been inconclusive, with some finding association of these systems to protoclusters and others, not (e.g., Boris et al., 2007; Farina et al., 2011; Green et al., 2011; Sandrinelli et al., 2014; Eftekharzadeh et al., 2017; Onoue et al., 2018).

1.5 | Overview of This Work

In this work, we analyzed six fields containing QSOs triplets, where at least one of the members is a radio-loud object – the radio signature at high redshift may be considered as an indication of a massive galaxy in a dense environment (e.g., Sánchez and González-Serrano, 1999, 2002; Kuiper et al., 2012; Hatch et al., 2014). The images were obtained with CFHT/Megacam in three different optical filters. We also include mid-IR information from the *unWISE* catalogue to improve the accuracy of photometric redshifts estimated through machine learning algorithms.

This work is organized as follows: in Chapter 2 we present our observations and the reduction procedure, as well as the supplementary data (from the *unWISE*, COSMOS, SDSS surveys, and simulated data) used in this work. In Chapter 3 we discuss our methods to estimate the photometric redshifts, for the analysis of the triplets environment through an evaluation of the galaxy overdensity significance field where they reside, and for the protocluster detectability with the mock. Also, the quantitative results are addressed in this same chapter. A discussion about our analysis is presented in Chapter 4. Finally, in Chapter 5 we summarize our findings and present our future perspectives. In Figure 1.7, we present a flow chart of the data treatment and analysis to get our results. Throughout this work we adopt a ΛCDM concordance cosmology,

with recent cosmological parameters from Planck Collaboration et al. (2014): $h = 0.673$, $\Omega_{\rm m} = 0.315$ and $\Omega_{\Lambda} = 0.685$.

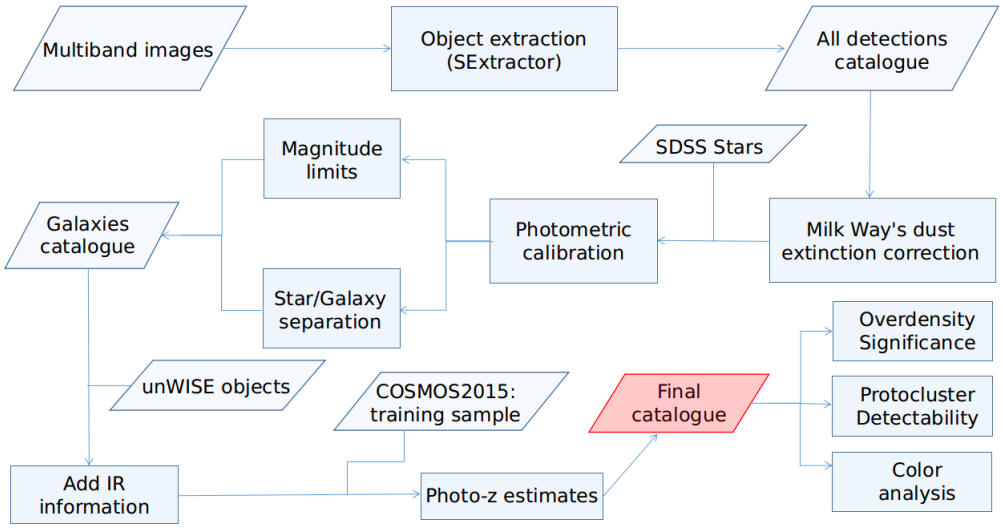

Figure 1.7: A flow chart from the CFHT/Megacam images data to our results.

2 Observations and Data Reduction

In this Chapter we describe how we have selected the sample of triplets for imaging at CFHT and the tools and procedures adopted for the image analysis, aiming to identify and extract the objects along with several photometric and morphological parameters. The detected objects were then corrected by galactic dust extinction and photometric calibrated in the SDSS system. We then discuss the star/galaxy separation, necessary for the analysis of the environment of the triplets, and finishes with a description of the mid-IR data extracted from the unWISE catalogue, which will be used later in this dissertation to improve estimates of photometric redshifts.

2.1 | Sample selection

In this work, we investigate six fields containing triplets of quasars. These systems were identified in the 13^{th} edition of the Véron-Cetty and Véron (2010) quasar catalogue where we searched for quasar triplets in the redshift interval $1 \lesssim z \lesssim 1.5$, with at least one object being radio-loud (RL), with separation between pairs in spectroscopic redshift less than 0.05, and in angular coordinates less than 6 arcmin. Using a friends-of-friends algorithm, we found 21 systems, from which 6 were chosen for observational follow-up.

We obtained, for each triplet, multi-band images with MegaCam at the CFHT telescope[1] during the period 2014A (PI: Roderik Overzier). MegaCam is appropriate for this kind of work because it covers a large field-of-view, 1 x 1 square degree, with a resolution of 0.187 arcsecond per pixel. In Figure 2.1, we present a RGB composition of the central part of the images. Each field is centered on the triplet's centroid and have widths corresponding to 20×20 cMpc.

At Table 2.1, we present the seeing of the observations, the number and the exposure time for each imaging, and the correspondent total amount of time. Table 2.2 summarizes some additional relevant information: the field identification, the SDSS ID (J2000.0) of the triplet members, their redshifts, the difference between the mean and individual QSO redshifts (Δv), their reddening corrected magnitude in the reference band, the mean redshift of the triplet, the filters adopted for each field, the magnitude interval in the reference magnitude used in our analysis (for the selection criteria, see Section 2.6), and the slab width in km/s (see Section 3.2 for the definition). The velocity intervals Δv and $\Delta v_{\rm slab}$ are given as $c(\Delta z)/(1+\bar{z})$ (Danese et al., 1980). We also included the minimum stellar mass $(M^*_{\rm min})$[2] that we are probing for each

[1] https://www.cfht.hawaii.edu/Instruments/Imaging/Megacam/

[2] We consider as $M^*_{\rm min}$ the stellar mass of the percentile 1% of the stellar mass distribution to avoid outliers.

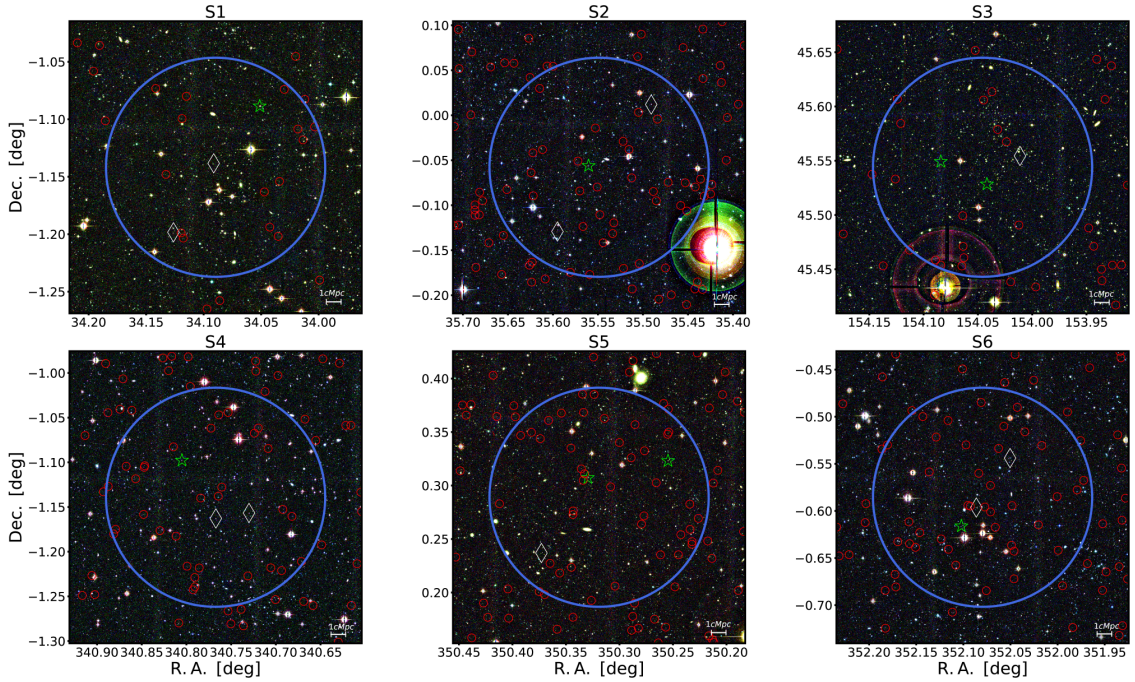

Figure 2.1: RGB composed images of our six fields with widths corresponding to 20×20 cMpc . White diamonds and green stars stand for radio-quiet and radio-loud quasars, respectively; the blue circle has a projected radius of 7.5 cMpc; red circles are galaxies with probability larger than 50% to be in the redshift slab of the triplet (see section 3.2).

field within our magnitude limits and the redshift slabs of the quasars, obtained by Araya-Araya et al. (2020) with mock catalogs generated from the Millennium Simulation (Springel et al., 2005) data and the latest version of the L-GALAXIES semi-analytic model (Henriques et al., 2015) (Figure 2.2) – the mocks subject will be addressed again in the Section 2.3.3 and Section 3.3. Finally, the bolometric luminosity and the virial black hole mass information, obtained by spectral analysis, were included from the Shen et al. (2011) catalogue[3] since massive and luminous quasars ($L_{\rm bol} \gtrsim 10^{15}$ erg/s and $M_{\rm BH} \gtrsim 10^{8}\ M_{\odot}$) are more expected to be members of greater structures (as we mentioned in Section 1 citing other works). All quasars in our sample are above these values. Radio-quiet (RQ) objects in the Véron-Cetty and Véron (2010) catalogue are indicated in the table with '*'. In Appendix A, we present the available QSOs spectra. For one object in S5 and two in S6 we do not have this information.

[3] http://quasar.astro.illinois.edu/BH_mass/dr7.htm

Table 2.1: Sample observational information.

Field[a]	Photometric band[b]	Seeing[c] (FWHM) (arcsec)	n x Exposure time[d] (s)	Total exposure time[e] (s)
S1	r	0.66	5 x 300	1500
	i	0.63	6 x 310	1860
	z	0.62	10 x 300	3000
S2	g	0.59	6 x 300	1800
	r	0.56	6 x 300	1800
	i	0.57	7 x 300	2100
S3	r	0.64	5 x 300	1500
	i	0.64	6 x 310	1861
	z	0.64	3 x 300	900
S4	g	0.66	6 x 300	1800
	r	0.51	6 x 300	1800
	i	0.69	7 x 300	2100
S5	r	0.49	5 x 300	1500
	i	0.46	12 x 310	3720
	z	0.53	10 x 300	3000
S6	g	0.57	6 x 300	1800
	r	0.51	6 x 300	1800
	i	0.55	7 x 300	2100

[a] *The name of the field that is adopted in this work.*
[b] *The three photometric bands that were used for each field observation.*
[c] *The FWHM seeing measured in arcseconds for each filter.*
[d] *The number of observations and the exposure time of each one.*
[e] *Total amount of exposure time that the a field in a certain band was observed.*

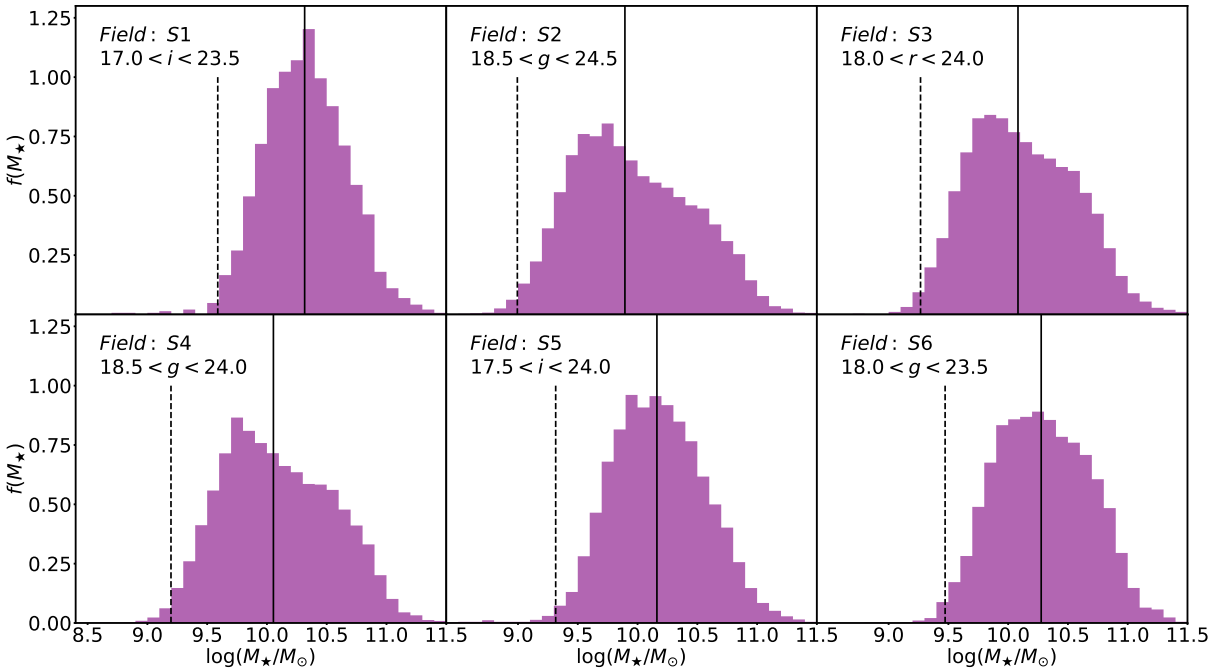

Figure 2.2: The stellar mass distribution within our magnitude limits for each field obtained with mock data. The solid line denotes the median, and, the dashed line, the mean of the percentile 1% of the distribution.

Table 2.2: Sample physical information.

Field	QSO[a]	$z^b_{\rm QSO}$	$\bar{z}^c$	$\Delta v^d[km/s]$	Reference Band Magnitude[e]	Adopted Limits[f]	$log(M^*_{\rm min})^g$ $[M_\odot]$	$\Delta v^h_{\rm slab}$ [km/s]	$log(L_{\rm bol})^i$ (erg/s)	$log(M_{\rm BH})^j$ $[M_\odot]$
	SDSSJ021612-010519	1.480		-3112	$i = 17.325$				47.04	9.22
S1	SDSSJ021622-010818*	1.518	1.506	+1437	$i = 19.956$	$17 < i < 23.5$	9.589	±15000	46.01	9.14
	SDSSJ021630-011155*	1.521		+1796	$i = 20.205$				45.96	8.98
	SDSSJ022158+000043*	1.042		-1607	$g = 18.790$				46.29	9.54
S2	SDSSJ022214 -000322	1.066	1.053	+1900	$g = 19.250$	$18.5 < g < 24.5$	8.995	± 12582	46.04	9.33
	SDSSJ022223-000745*	1.051		-292	$g = 21.095$				45.65	8.87
	SDSSJ101603+453316*	1.369		-884	$r = 19.923$				45.95	8.92
S3	SDSSJ101610+453142	1.380	1.376	+505	$r = 19.039$	$18 < r < 24$	9.267	± 15326	46.36	9.24
	SDSSJ101620+453257	1.379		+379	$r = 17.626$				46.93	9.53
	SDSSJ224255-010924*	1.033		-1615	$g = 19.347$				45.83	8.82
S4	SDSSJ224304-010946*	1.044	1.044	0	$g = 19.269$	$18.5 < g < 24$	9.197	± 10779	45.96	8.99
	SDSSJ224313-010552	1.054		+1615	$g = 19.342$				46.12	9.21
	SDSSJ232101+001923	1.362		+1275	$i = 18.881$				46.34	8.70
S5	SDSSJ232119+001826	1.327	1.352	-3189	$i = 20.358$	$17.5 < i < 24$	9.315	± 16715	-	-
	SDSSJ232129+001413*	1.368		+2041	$i = 20.320$				45.58	8.80
	SDSSJ232812-003238*	1.083		-4679	$g = 20.997$				-	-
S6	SDSSJ232821-003547*	1.136	1.116	+2835	$g = 20.280$	$18 < g < 23.5$	9.473	± 10219	-	-
	SDSSJ232824.5-003658	1.128		+1701	$g = 20.116$				45.79	8.35

Note: Radio-quiet QSOs are denoted by "*".

[a] *The SDSS identifier of the quasars.*
[b] *The redshift of the quasars.*
[c] *The mean redshift of the triplet.*
[d] *The radial velocity interval between the quasar and the triplet's mean.*
[e] *The reference band magnitude of the quasars.*
[f] *The reference band limits adopted. The lower to avoid saturated objects, while the upper to ensure the completeness of the sample.*
[g] *The minimum protocluster mass detected with the mocks within the magnitude range adopted.*
[h] *The velocity interval within the triplet's redshift slab.*
[i] *The quasar's bolometric luminosity.*
[j] *The quasar's black hole mass.*

2.2 | Object extraction

2.2.1 | SExtractor

SExtractor is a software created to process astronomical images in order to identify objects, measure their parameters, and organize them into catalogues. It was developed especially for galaxy-surveys data reduction, but it also works in moderately at crowded star fields. Some of the main features of this program among others are:

- Image processing speed of 50 Mpixel/s (equivalent to 10k sources/s) in a 3 GHz processor.
- Work with very large images (up to 2Gx2G pixels).
- Robust deblending tools for extended objects.
- Pixel-a-pixel photometry in dual-image mode. [4]

One of the main SExtractor tools that we use in this work is the dual-image mode. It provides an efficient way to process multi-band images. To use this tool, the images obtained in different filters must have been taken exactly in the same position, apertures, and must have the same Cartesian dimensions. The image that has the largest number of objects is chosen as the reference one for the others. That is, the sources are identified only in this image, and the photometric measurements are made also in the others. Therefore, the catalogues in different photometric bands will have exactly the same objects.

2.2.2 | SExtractor operation

The CHFT science frames were processed with the package THELI (Erben et al., 2005; Schirmer, 2013). The images were bias/overscan-subtracted, trimmed, flat-fielded, and registered to a common pixel and sky coordinate positions using SCAMP (Bertin, 2006). A combined astrometric solution for all three filters (g, r, i, or r, i, z) was derived using the SDSS DR9 catalog. The resulting astrometric calibrated individual frames were then sky subtracted, re-sampled to a common position, and stacked with SWARP (Bertin, 2010, SWarp: Resampling and Co-adding FITS Images Together, Astrophysics Source Code Library).

Then, we ran `SExtractor` (Bertin and Arnouts, 1996) along with `PSFex` (Bertin, 2011), which performs PSF fitting photometry, on the combined frames of each band, to detect and extract

[4] https://sextractor.readthedocs.io/en/latest/Introduction.html

objects and then create the photometric catalogues containing all the measurements that are necessary for this work: celestial coordinates (RA and DEC, J2000), photometric and morphological information such as magnitudes and their errors, the image width `FWHM`, `CLASS_STAR`, the maximum surface brightness $\mu_{\rm max}$ and the `SPREAD_MODEL`. The `CLASS_STAR` parameter is a point/extended source classifier that estimates the a posteriori probability that a detection made by `SExtractor` is a point source. `CLASS_STAR` relies on a multilayer feed-forward neural network trained using supervised learning to generate the estimates. Objects with `CLASS_STAR` close to 1 are likely point sources, whereas those close to 0 are probably galaxies.

`SPREAD_MODEL` is a powerful morphological parameter obtained with `PSFex`. `PSFex` does not work directly on images. Instead, it operates on `SExtractor` catalogues, which have a small sub-image ("vignette") recorded for each detection. This makes things much easier since one does not have to handle the detection and deblending processes. The catalogue files read by `PSFEx` must be in the `SExtractor` `FITS_LDAC` binary format, which allows the software to have access to the original image header content. In order to use this tool, we performed the following steps for each field/band:

1. Run `SExtractor` in *single mode* with only the necessary parameters that will be used by `PSFex`. The result is a catalogue (`FITS_LDAC` binary format file) containing all the identified objects with the parameters information.

2. Run `PSFex` to generate the `PSF` for each object. In this step, the output of (i) is used as input, and the result will be the `".psf"` files.

3. Run `SExtractor` again including the output of (ii), which is responsible for the `SPREAD_MODEL` parameter and its corresponding error. Also, once we know which filter has detected more objects (from (i)), we set `SExtractor` in *dual mode* to use this filter as reference for the others.

In the end of this process, there is one catalogue for each field containing all the detected objects with the required measurements. Appendix B shows one example of the scripts for the items described above.

2.3 | Complementary databases

Besides the CFHT images, we have used some complementary databases to do the photometric calibration and to perform estimates of photometric redshifts (photo-z_s). They are described below.

2.3.1 | SDSS

The photometric calibration was made using the *ugriz SDSS* system (Fukugita et al., 1996) since our images are in similar filters. This survey has a photometric accuracy of 2-3% and astrometric accuracy better than 0.1" (Pier et al., 2003).

We have used a query search in the *Catalog Archive Server Jobs System (CasJobs)*[5] for *SDSS-DR14* to find the unsaturated stars in the same area of our images. The catalogues generated by *CasJobs* contain the coordinates of each object and the photometric measurements in the same filters of our images. We present in Section 2.5 details of the photometric calibration.

2.3.2 | COSMOS2015

In Section 3.1 we present our machine-learning approach to estimate photometric redshifts for our 6 fields. Since there is a lack of a representative and homogeneous sample of galaxies with spectroscopic redshift at the triplet's redshift to train our algorithm, we have used data from the *COSMOS2015* (Laigle et al., 2016) catalogue, supplemented with unWISE photometry. *COSMOS2015* is a public catalogue with more than a million objects over a 2deg^2 of the COSMOS field. The main motivation to use photometric redshifts from this catalogue to train our own photo-z estimator is because half of its objects have very accurate photo-z_s, $\sigma_{\Delta z}/(1+z_s) \simeq 0.007$, through a comparison with the *zCOSMOS*-bright spectroscopic redshifts (spec-z_s). This high accuracy is due to a large number of photometric bands (>30) obtained by several surveys in this area, from the UV to the mid-IR (e.g., UltraVISTA-DR2, Subaru/Hyper-Suprime-Cam, IRAC/*Spitzer* (SPLASH)).

The *COSMOS2015* catalogue catalogue provides magnitudes in filters (B, V, r, i+, and z++, from Suprime-Cam/Subaru) similar to those of our images, with a depth of $i \sim 26.5$. In the case of the g filter, we applied the transformation proposed by Jester et al. (2005), using apparent magnitudes in the B and V bands:

$$g = V + 0.64 \times (B - V) - 0.13 \tag{2.1}$$

We also (re)calibrated the magnitudes of *COSMOS2015* in bands equal or similar to those of our fields in the *SDSS* photometric system, as described in Section 2.5.

[5] https://skyserver.sdss.org/casjobs/

2.3.3 | Simulated Data

The main goal of this work is to explore the QSO triplets environment. To quantify the possibility to detect protoclusters at their redshifts, given our observational constraints, we construct 20×4 *PCcones*, using the technique presented in Araya-Araya et al. (2020). These particular mocks consist of structures from the Millennium Simulation (Springel et al., 2005) placed in the center of the mocks at a certain redshift. The `L-GALAXIES` (Henriques et al., 2015) semi-analytical model is used to populate the dark matter halos of the Millennium simulation with galaxies and to calculate their apparent magnitudes through the *star formation histories (SFH)* information. The original area of each lightcone is π deg^2, but we constrained them to 1 deg^2, which is the field-of-view of our observations.

We placed 20 different structures at the center of the lightcone in 4 different redshifts: $z =$ 1.04, 1.12, 1.36 and 1.51. Following the definition of Chiang et al. (2013), we have 8 Fornax-type ($M_{z=0} = 1.37 - 3.00 \times 10^{14}\ M_\odot$), 6 Virgo-type ($M_{z=0} = 3 - 10 \times 10^{14}\ M_\odot$) and 6 Coma-type ($M_{z=0} \geqslant 10^{15}\ M_\odot$) protoclusters at these redshifts.

Our observational data is not homogeneous in the sense that we have different photometric bands for each field. Additionally, the magnitude limits are not the same for the whole sample. To address these differences, we have mimic real observations from the photometric catalogs of our 80 lightcones, by applying to them the adopted magnitude limits presented in Table 2.2. In summary, we have 20 lightcones for each field.

2.4 | Correction of the galactic extinction

The magnitudes generated by `SExtractor` were first corrected by the extinction caused by the Milky Way's dust, which causes a slight increase in the magnitude of each object, depending on its celestial coordinates (Figure 2.3, (extracted from Mörtsell, 2013)). This correction was determined using a routine in python, which reads the value of E(B-V) corresponding to the coordinates of each object on the Schlegel et al. (1998) maps, and calculates the extinction with the Cardelli et al. (1989) law for each photometric band[6]. These extinctions were then subtracted from the corresponding magnitude for each object.

[6]The packages used were `sfdmap` and `extinction`

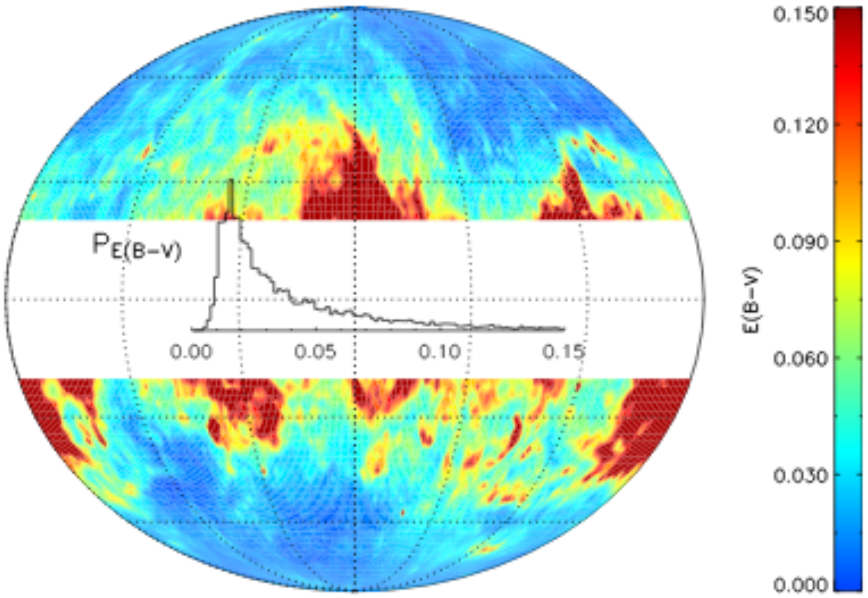

Figure 2.3: The colored map represents the E(B-V) dust extinction for galactic latitudes $|b| > 20°$. The coordinate $|b|$ of this work fields are $\sim$ 57, 55, 54, 50, 55, and 57°, from S1 to S6, respectively. The plot in the middle of the figure is the probability density function of E(B-V) done by Mörtsell (2013) with Schlegel et al. (1998) data. The PDF shows $\sim$ 0.048, $\sim$ 0.016, and $\sim$ 0.028, for the mean, peak, and median, respectively.

2.4.1 | E(B-V) extinction maps

One of the most common approaches to obtain the extinction (E (B-V)) caused by the milky way's dust is through the Schlegel et al. (1998) maps. For this, they composed a 100 μm all-sky map using previous observations of IRAS (*Infrared Astronomy Satellite*)/ISSA (*IRAS Sky Survey Atlas*), which were calibrated with the DIRBE experiment (*Diffuse InfraRed Background Experiment*) carried out by the COBE (*Cosmic Background Explorer*) satellite.

The IRAS satellite was launched in 1983 and was responsible for the first all-sky coverage map with a $\sim$ 5 arcmin beam in four IR-broadband (12, 25, 60, and 100 μm) to capture the diffuse background radiation. This technology was developed in order to optimize the detection of small angular extent sources (bei, 1988), however, with its data, Wheelock et al. (1994) built images of large areas for what was called as the ISSA project. The DIRBE experiment uses 10 photometric broadbands in a range of 1–240μm, with a $\sim$ 0.7° beam, which is a much lower resolution than IRAS. However, it has two major advantages over the latter, which is the very best control over absolute calibration and two channels that go further in the submillimeter (140 and 240μm)

To preserve the best qualities of both surveys, Schlegel et al. (1998) performed a series of corrections and calibrations, such as remove the zodiacal foreground emission and the striping artifacts from the DIRBE and IRAS/ISSA maps, respectively; subtract the confirmed point sources; and, finally, both maps were combined. The dust color temperature is based on the ratio of 100 to 240 μm, which gives the necessary information to understand the 100μm emission

in function of the column density of radiating dust. The construction of these maps have difficulties to deal with extinction in low galactic latitudes ($|b| < 20°$), where the dust concentration is much denser and dominated by several radiation sources with different characteristics depending on the emitting phenomenon. In such environments, these maps may neglect the actual extinction caused by the dust columns. However, for all other directions, where the emissions are, approximately, from a single environment and the same radiation field, the measurements on the extinction maps are valid.

The reddening calculation due to the galactic dust can be expressed as

$$E(B-V) = pD^T \tag{2.2}$$

where p is a calibration coefficient, and D^T is the column density of galactic dust that radiates in the 100–240μm interval, measured in the map with all corrections and calibrations described above.

2.4.2 | Extinction law

It is very complex to define an extinction law for Galaxy absorption since it varies with each line of sight, and this variation also depends on the wavelength of the radiation being emitted in the region that is being observed. One approach to deal with this problem is through the ratios between absolute extinctions at different wavelengths, $A(\lambda)/A(V)$ (the reference in the V band is an arbitrary and conventional choice).

Cardelli et al. (1989) used photometric information from the same stars in the UV (Fitzpatrick and Massa, 1986, 1988, observations), optical, and near infrared (NIR), from different lines of sight with different types of environment (diffuse dust, molecular clouds, and HII regions), to compare and better understand the relationships between the extinction at different wavelengths. This study used as a parameter the total-to-selective extinction ratio, $R_v = A(V)/E(B-V)$. Figure 2.4 left (taken from Cardelli et al., 1989) shows the A_{lambda}/AV ratio as a function of the inverse of the R_v^{-1}. It is possible to observe a clear linear relationship for different wavelengths. Environments with diffuse dust from the ISM have lower values ($R_v \sim 3.1$) than denser molecular regions ($4 < R_v < 6$). The R_v parameter is estimated quantitatively, so particular cases can be left out of more general predictions like the values exposed above.

To exemplify the application of this analytical formula, Figure 2.4 right, shows the function obtained through the $A(\lambda)/A(V) - R_v$ ratios, and compares them with observations. The discrepancies between the analytical and observed curves are attributed to the dispersion between the measures shown in Figure 2.4 left. The deviations are greater as the wavelengths increase, that is, in the interval analyzed in this work, in the ultraviolet region.

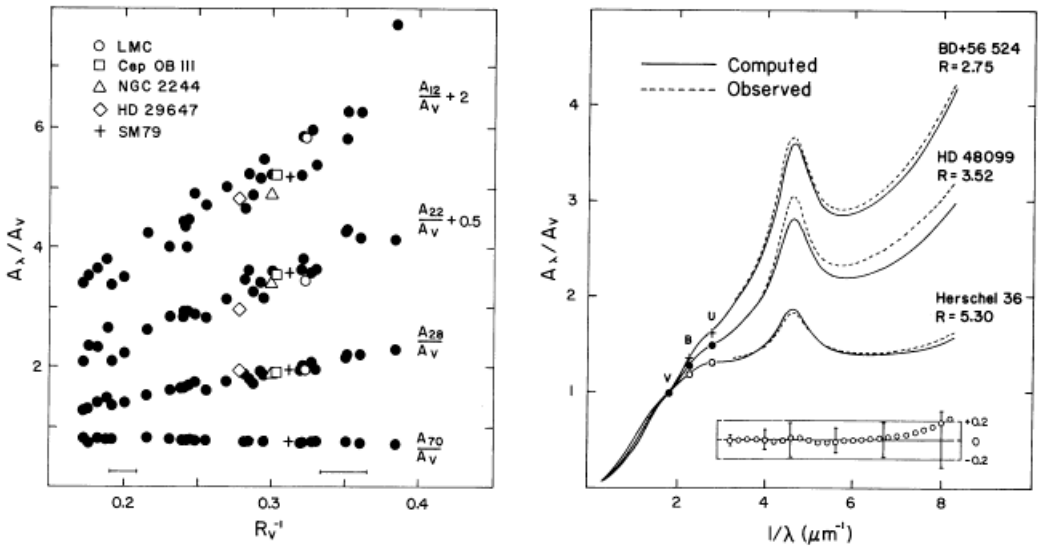

Figure 2.4: Left: The observations of $A(\lambda)/A(V)$ in function of R_V^{-1}; black dots are observational data; the subscribed number at the A means the wavelength (e.g.: A_{12} for 1200Å). Right: Three cases of the extinction law obtained by Cardelli et al. (1989). For more details about the figures, please check Cardelli et al. (1989)'s Figure 1 and 2.

2.5 | Photometric calibration

The photometric calibration was done using isolated and non-saturated stars in the *SDSS* Data Release 14. We have ran a `SQL` script at *CasJobs* using appropriate flags as specified in the Tutorial Flags at the *SDSS* website[7]. After, we applied additional constraints to this data, as illustrated in Table 2.3, to ensure that the selected stars have indeed good photometry.

Table 2.3: Constraints applied for selecting unsaturated stars for the photometric calibration.

Parameter	*Selection*
$\mu_{\rm max}$	$\mu_{\rm sat} + 0.5 < \mu_{\rm max} < \mu_{\rm sat} + 3.5$
`FWHM`	< 4.2
`CLASS_STAR`	> 0.95
`SPREAD_MODEL`	< 0.0002
`MAG_ERR`	< 0.03

These criteria were based on visual inspection of some diagrams that relate the morphological parameters to the magnitudes of the objects (see Figure 2.5 for an illustration of the procedure applied to S1, and, Appendix C, for the other fields). To eliminate the saturated objects we chose a lower magnitude limit using the $\mu_{\rm max}$ parameter (*bottom-right* in this figure) since the saturated objects at the brighter magnitude-end have essentially a constant value for this parameter ($\mu_{\rm sat}$). This value defines the lower and upper values for $\mu_{\rm max}$ according to Table 2.2. We also set limits for other morphological parameters to make sure that would be no inconsistencies in the

[7] https://www.sdss.org/dr14/tutorials/flags/

selection.

The next step was a cross-match of celestial coordinates between our fields and *SDSS*, assuming a tolerance in the angular separation of the sources $\leqslant 1$ arcsec, and a 3σ clipping to remove the outliers. Our final sample for calibration can be checked in the diagrams of Figure 2.5 and Appendix C (pink objects).

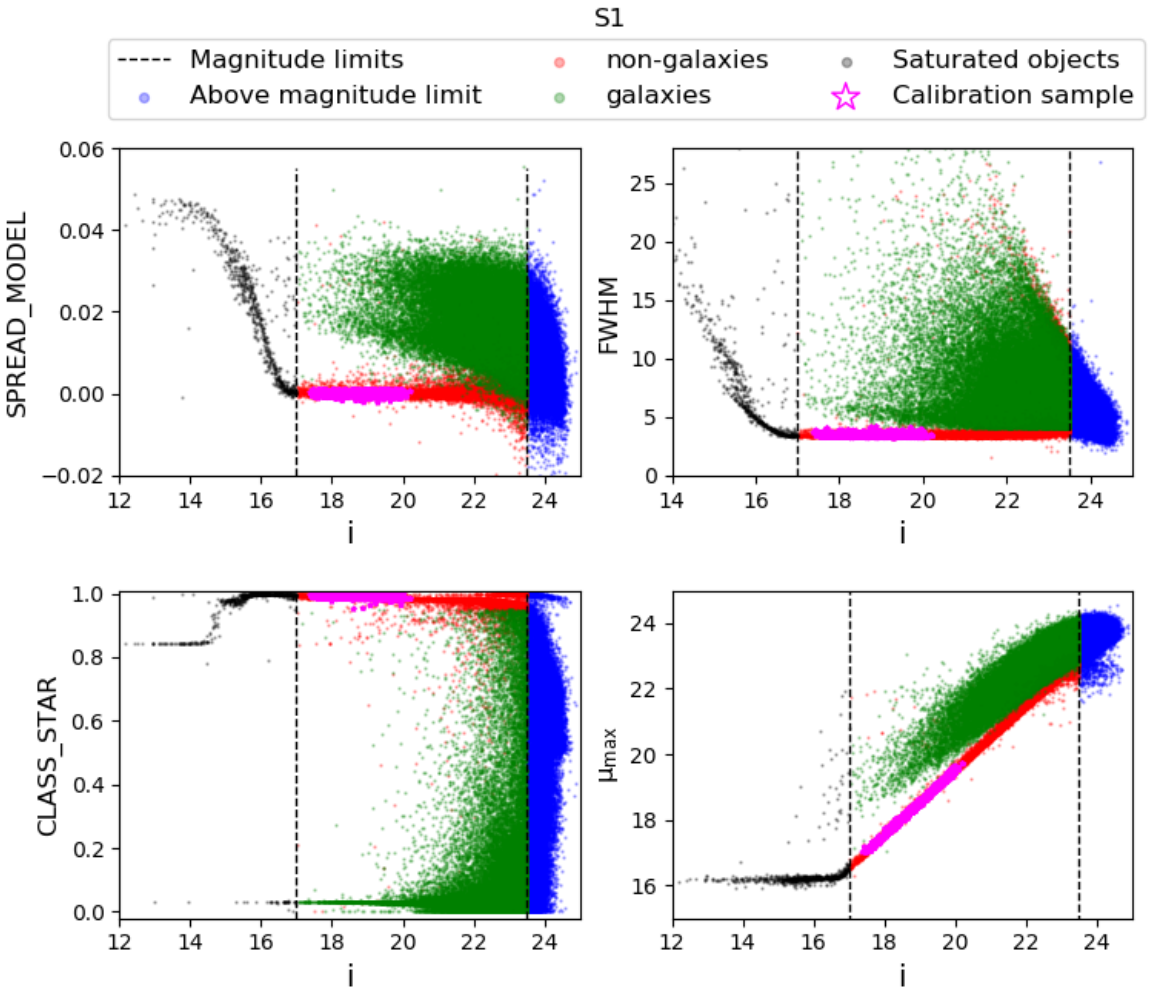

Figure 2.5: Object selection for the S1 field. Green objects are considered galaxies; red are non-galaxies; pink are the stars selected for calibration; black are objects considered as saturated; and blue objects are those above the upper magnitude limit (see text).

After these steps, we transformed our instrumental magnitudes to the *SDSS* photometric system using a polynomial fit of degree one or two, with the choice based on a χ^2 decision test. Figure 2.6 shows this calibration for one of our fields. These transformations were applied to all bands in all fields, including the corresponding magnitudes of the *COSMOS2015* catalogue (Table 2.4). Typical errors in this calibration are, approximately, in the range 0.02 to 0.04 for r, i, and z bands; and, 0.06 to 0.08 for the g band.

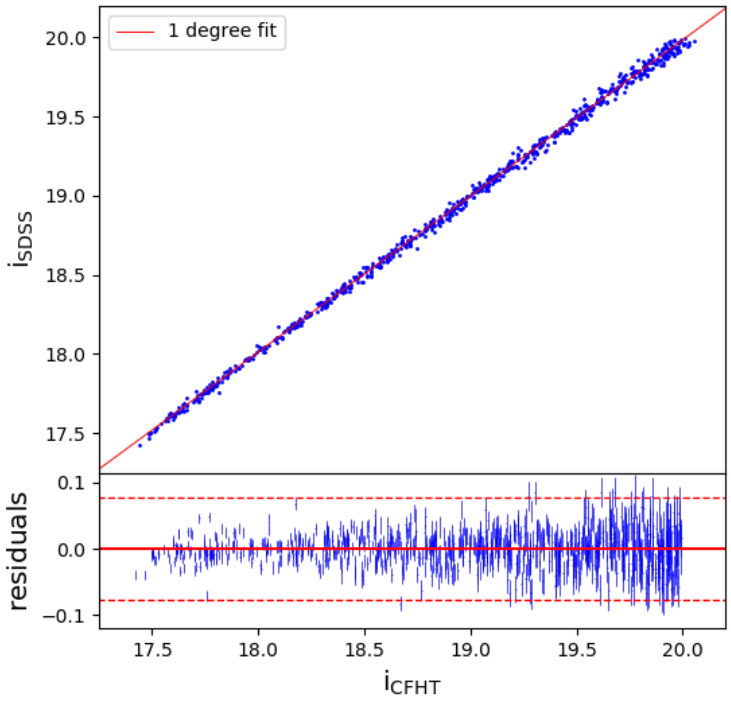

Figure 2.6: Photometric calibration for the i-band in the S1 field. Above: the linear calibration applied in this case. Below: the residuals of the fit as a function of the calibrated magnitude. The dashed horizontal red lines represent $\pm 3\sigma$.

Table 2.4: Fit results to transform the SExtractor instrumental magnitudes ($m_{\rm inst}$) to the SDSS photometric system ($m_{\rm sdss}$).

Field	Filter[a]	Fit[b]	σ[c]
S1	r	$m_{\rm sdss} = 0.99 m_{\rm inst} + 0.12$	0.022
	i	$0.98 m_{\rm inst} + 0.24$	0.026
	z	$0.98 m_{\rm inst} + 0.31$	0.035
S2	g	$0.99\ m_{\rm inst} + 0.03$	0.075
	r	$0.99 m_{\rm inst} + 0.19$	0.021
	i	$0.98 m_{\rm inst} + 0.28$	0.024
S3	r	$0.99 m_{\rm inst} + 0.11$	0.023
	i	$0.99 m_{\rm inst} + 0.24$	0.026
	z	$0.99 m_{\rm inst} + 0.25$	0.034
S4	g	$-0.04 m_{\rm inst}^2 + 2.4 m_{\rm inst} - 13.2$	0.078
	r	$0.99 m_{\rm inst} + 0.23$	0.020
	i	$0.99 m_{\rm inst} + 0.24$	0.022
S5	r	$0.99 m_{\rm inst} + 0.56$	0.027
	i	$0.98 m_{\rm inst} + 0.65$	0.028
	z	$0.98 m_{\rm inst} + 0.44$	0.041
S6	g	$-0.009 m_{\rm inst}^2 + 1.3 m_{\rm inst} - 3.4$	0.06
	r	$0.99 m_{\rm inst} + 0.17$	0.022
	i	$0.98 m_{\rm inst} + 0.29$	0.027
COSMOS	g	$0.99 m_{\rm inst} + 0.16$	0.079
	r	$0.99 m_{\rm inst} + 0.07$	0.039
	i	$0.98 m_{\rm inst} + 0.48$	0.061
	z	$0.99 m_{\rm inst} + 0.13$	0.052

[a] *The photometric band that the calibration was applied.*
[b] *The minimum least squares fit to obtain the magnitudes in the SDSS photometric system.*
[c] *The mean dispersion between the SDSS magnitude and the calibrated one.*

2.6 | Star/galaxy separation

Given the objectives of this work, we need to identify the galaxies in our fields, and, for this, any point source present in our catalogues will be considered as a “non-galaxy". In this section, we describe our approach to galaxy selection.

First of all, we defined the magnitude interval of interest, considering limits due to image saturation at the bright side ($mag_{\mathrm{ref,min}}$, such that $\mu_{\mathrm{max}} = \mu_{\mathrm{sat}} + 0.5$ in the reference band) and the upper limit at the faint side as ∼0.1 below the peak of the magnitude distribution in the reference band; see Figure 2.7. These limits are also presented in Table 2.2 for the reference band of each field.

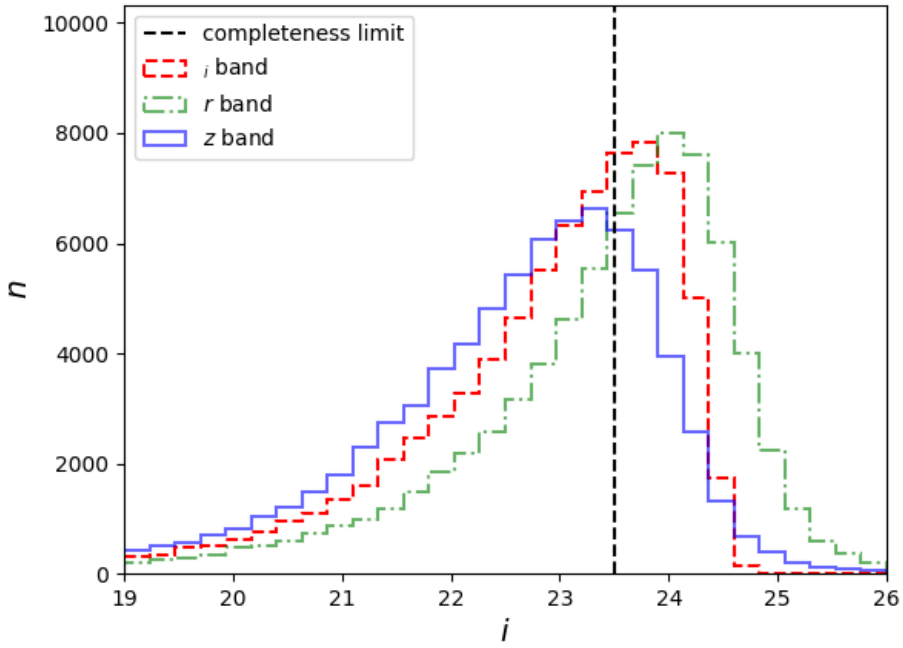

Figure 2.7: Magnitude distributions for the S1 field. Objects fainter than the upper magnitude limit (the black dashed line) are discarded.

Within this magnitude interval in the reference band, we consider galaxies as extended objects, identified using three morphological parameters, `SPREAD_MODEL`, `CLASS_STAR`, and `FWHM` (pixel), as a function of the magnitude in the detection band.

As mentioned before, the `SPREAD_MODEL` parameter provides a powerful measurement of the width of an object, so we used it along with its error. By taking in to account this error, the efficiency of the galaxy selection (true galaxies classified as such, over total true galaxies) increases considerably at the faint-end[8].

We reinforce this selection by adopting limits for the `CLASS_STAR` and `FWHM` parameters based

[8] https://cdcvs.fnal.gov/redmine/projects/des-sci-verification/wiki/A_Modest_Proposal_for_Preliminary_StarGalaxy_Separation

on the *star sequence* identified in plots like those shown in Figure 2.5, where the sequence of unsaturated stars can be well distinguished through visual inspection at low values of these parameters. Since `CLASS_STAR` estimates the probability that an object is a point source, we expect that extended objects avoid high values for this parameter. Besides, the magnitude-$\mu_{\max}$ diagram also presents nice discrimination between extended and point source objects. Based on the distribution of the objects in Figure 2.5, we have set the following limits on these parameters for an object to be considered as a galaxy:

$$(\texttt{SPREAD\_MODEL} + 3 \times \texttt{SPREADERR\_MODEL}) \geqslant 0.003$$
$$\textbf{and } \texttt{CLASS\_STAR} \leqslant 0.95$$
$$\textbf{and } \texttt{FWHM} \geqslant 4$$

The different colors in Figure 2.5 show the different classes of objects after applying these constraints. We also show in this figure the resulting classification in the magnitude-$\mu_{\max}$ diagram.

2.7 | Including W1 and W2 from *unWISE*

The *unWISE* catalogue (Schlafly et al., 2019) contains more than 2 billion sources all over the sky in the 3.6 and 4.5μm (W1 e W2) bands from the analysis of the *unWISE* coadds of the *WISE* images[9]. The *unWISE* coaddition is described in Lang (2014) and Meisner et al. (2017b,a, 2018).

Concerning its predecessor *ALLWISE*, this catalogue presents deeper images, since it involves the coaddition of all publicly available 3–5 microns *WISE* imaging. This procedure is equivalent to increasing the total exposure time by a factor of 5 and allows the detection of magnitudes ~ 0.7 fainter (at a 5σ level). This doubles the number of detections between redshifts 0 and 1, and triples between 1 and 2, totalizing more than half a billion galaxies. Another advantage of this new catalogue is the improvement in the modeling of the crowded regions with the `crowdsource`[10] analysis pipeline (Schlafly et al., 2018), which optimizes the positions, fluxes, and the background sky to minimize the differences between observation and model. Our images and *COSMOS2015* area are inside the *unWISE* coadds, which makes it possible for us to do a cross-match between catalogues to include the mid-IR information in our analysis.

The accuracy of photometric redshifts depends strongly on the number of available photometric bands. Our observations were done in only three bands per field. For this reason, we included in our sample (as well as in the *COSMOS2015* catalogue) the *W1* and *W2* bands from the *unWISE*.

[9] http://catalog.unwise.me/
[10] https://github.com/schlafly/crowdsource

The field of view of each of our fields correspond to two or four images (coadds) of the *WISE* survey, and we have used *topcat* (Taylor, 2005) to perform the matching of the objects detected in *unWISE W1* and *W2* bands with our catalogues. In this process, we removed the coadds overlapping edges to avoid object repetition, concatenated the data of each image, and, finally, made a cross-match between the resulting *unWISE* catalogue and ours with a projected distance of $\leqslant 2.75$ arcsec. It is also important to point out that, in all fields, including *COSMOS*, about 60% of the cross-matched galaxies have measurement only in W1 or W2. These galaxies with missing photometry reduce the quality of the estimated photometric redshifts (Section 3.1) by 1–2%, since less information is being used as input data in the machine learning process. However, given the expressive amount of galaxies with missing mid-IR photometry, we decided to keep them for further analysis. We also checked the galaxies without photometry in the optical filters. They represent ∼1% of the total sample of objects measured with SExtractor, and, after the cross-match with the *unWISE* galaxies, only ∼0.02% remains. Since this ratio is very low, and most of these galaxies already do not have measurements in one of the mid-IR filters, we discarded them.

The flux measurements of *unWISE* are in VEGA nanomaggies units (Finkbeiner et al., 2004). Therefore, we use the expressions provided by Schlafly et al. (2019) to convert them into AB magnitudes:

$$m_{\mathrm{W1,AB}} = m_{\mathrm{W1,Vega}} + 2.699$$
$$m_{\mathrm{W2,AB}} = m_{\mathrm{W2,Vega}} + 3.339$$

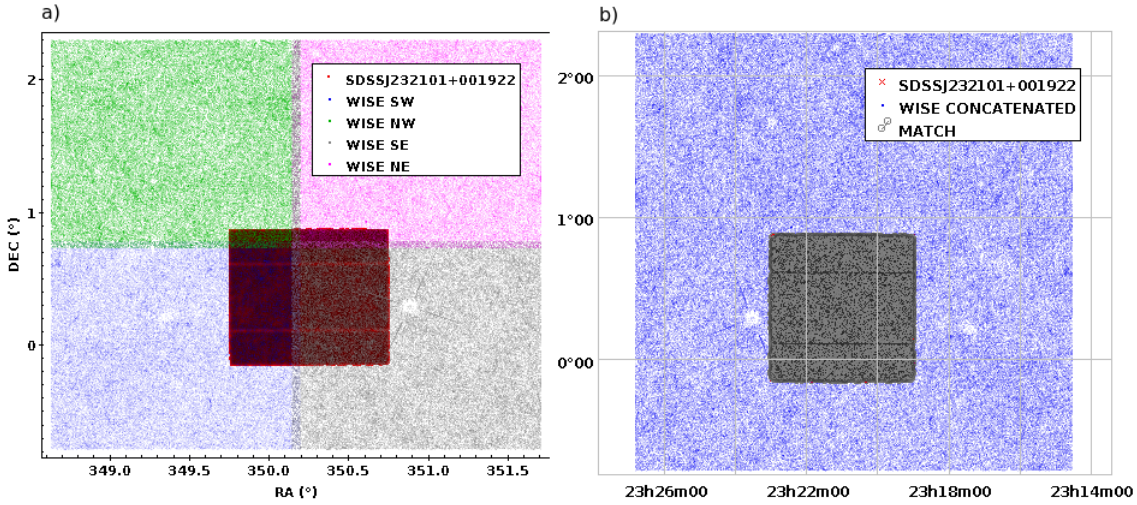

Figure 2.8: ***Left:*** *Green, pink, blue and gray points are the unWISE catalog objects extracted from WISE coadds. Dark red are the objects extracted from S4 field.* ***Right:*** *Blue points are the sum of the four coadds that composes the total area of S4, and gray points are the cross-matched objects of S4 field and unWISE.*

3 Analysis

At this stage, we have galaxy catalogues with photometric measurements in five bands (3 in the optical, from the CFHT images; and 2 in the mid-IR, from unWISE). In this Chapter, we discuss, initially, the methods that we adopted for the photometric redshifts estimation for each field, the procedure to evaluate the galaxy density field in a redshift interval containing each one of our triplets, and the comparison with the mock data.

3.1 | Photometric redshifts

There are basically two ways to calculate photo-z_s: the *template fitting* approach (Benítez, 2000; Arnouts et al., 2002; Ilbert et al., 2006; Tanaka, 2015) and machine learning methods (MLMs: e.g., Collister and Lahav, 2004; Almosallam et al., 2016; Sadeh et al., 2016). The first one relies on empirical (Coleman et al., 1980) or synthetic spectra (Bruzual and Charlot, 2003; Maraston, 2005; Charlor and Bruzual, 2007) of different types of galaxies that are processed along with the photometric information of the observations, taking into account the telescope response and the filters characteristics. MLMs work with *training data sets* comprised of objects with known redshifts to derive a relationship between the photometric measurements and the photo-z_s. For this reason, they do not require physically motivated models, which helps to incorporate new observables into the inference and mitigates systematic errors (Sadeh et al., 2016).

We use `ANNz2`[1] (Sadeh et al., 2016) to estimate photo-z_s. This is a new implementation of the `ANNz1` package (Collister and Lahav, 2004). `ANNz2` uses the `ROOT C++` software framework (Brun and Rademakers, 1997), that contains the *Toolkit for Multivariate Data Analysis (TMVA)* package (Hoecker et al., 2007), allowing the choice of different algorithms to train MLMs. For this reason, unlike its predecessor, `ANNz2` incorporates different MLMs in its code (e.g., *artificial neural network* (ANN), *boosted decision trees* (BDT), among others), providing larger versatility in the photo-z_s inference, as well as probabilistic errors estimation. Like all MLMs, a sample with reliable redshifts is required and it is split into *training*, *validation* and *test sub-samples*. During training, the *validation sub-sample* is used step-by-step to check the convergence of the solutions and evaluate the mapping between photometry and redshifts. The *test sub-sample* – which is not part of the training process – allows an independent evaluation of the performance of the algorithm.

In addition, it is possible to operate `ANNz2` in *single* or *randomized regression*. The first one

[1] https://github.com/IftachSadeh/ANNZ

is the simplest configuration and generates a similar nominal product as `ANNz1`. However, Sadeh et al. (2016) shows that `ANNz2` has slightly superior performance. They also point out that the method used to calculate uncertainties has been significantly improved. The original version uses the propagation of input uncertainties to obtain those for the estimated redshifts, through the chain rule; `ANNz2` uses a data-driven method where it takes into account that objects with similar photometric properties should also have similar uncertainties in the photo-z_s. For this, `ANNz2` uses the *k-nearest neighbours* (KNN) method, which was demonstrated to be an efficient training-based error estimator (see Oyaizu et al. (2008)). On the other hand, *randomized regression* uses several combinations of MLMs to compute photo-z_s probabilistic distribution functions (PDF) for each galaxy. These combinations can differ in several ways, for example using randomized seeds of initialization, changing the combination of parameters of a given MLM, as the number of neurons, or hidden layers of an ANN, or the number of decision trees of the BDT method. Then, these combinations are ranked following some metrics, like *bias*, *outlier fraction* and *scatter*.

Another challenge for the current work is to find a homogeneous sample with a large enough number of galaxies up to spec-$z_s \sim 1.5$. High-z surveys with a large number of objects are in general biased in specific redshift slices, depending on the survey objectives, resulting in inhomogeneities that can bias their use as a training set in MLMs. In this work, we have used the *COSMOS2015* catalogue, since it has photo-z_s with high precision and homogeneous distribution up to $z \sim 1.5$, as described in section 2.3.2. The training was done with `BDT` in the *single regression* mode, since the randomized and `ANN` algorithms have a larger computational cost and the results are very similar when compared through the $Bias =< z_{in} - z_{out} >$, and the σ_{NMAD} metrics, proposed by Molino et al. (2017), which evaluates the accuracy of the estimated redshift (z_{out}) with respect to the *input* (z_{in}) value:

$$\sigma_{NMAD} = 1.48 \times MEDIAN\left(\frac{(|\Delta z - MEDIAN(\Delta z)|)}{1 + z_{in}}\right), \tag{3.1}$$

where, $\Delta z = z_{out} - z_{in}$.

To obtain an unbiased sample when running `ANNz2` it is also important that the training, validation, and test samples (divided as 70%, 15%, and 15%, respectively) be consistent between them and with our sample. Since the *COSMOS2015* catalogue is deeper than our optical observations, we selected only galaxies inside the same interval of magnitudes applied at our images (see Table 2.2). Thus, even though COSMOS is deeper than our images, there is no significant statistical difference between the magnitude errors of the two samples (Figure 3.1).

We present the magnitude distributions in the 5 available photometric bands for the *S1* field and the *COSMOS* catalogue in Figure 3.2. We also tested including cuts in other bands to better converge the faint-end between the two counts. However, we decided to keep cuts only in the reference band to avoid bias, and also because the difference in the results is not large and does not affect the final result qualitatively. For all magnitude distributions, see Appendix D.

We tested different input configurations (only colors between adjacent bands; adding the reference band; and adding more bands with different combinations) comparing which one gives the best results based on the metrics described above. The best results were obtained with all magnitude measurements and colors: *r*, *i*, *z*, *W1*, *W2*, *r-i*, *i-z*, *z-W1*, *W1-W2*, for the S1, S4, and S6 fields; and *g*, *r*, *i*, *W1*, *W2*, *g-r*, *r-i*, *i-W1*, *W1-W2*, for S2, S3, and S5. To check the

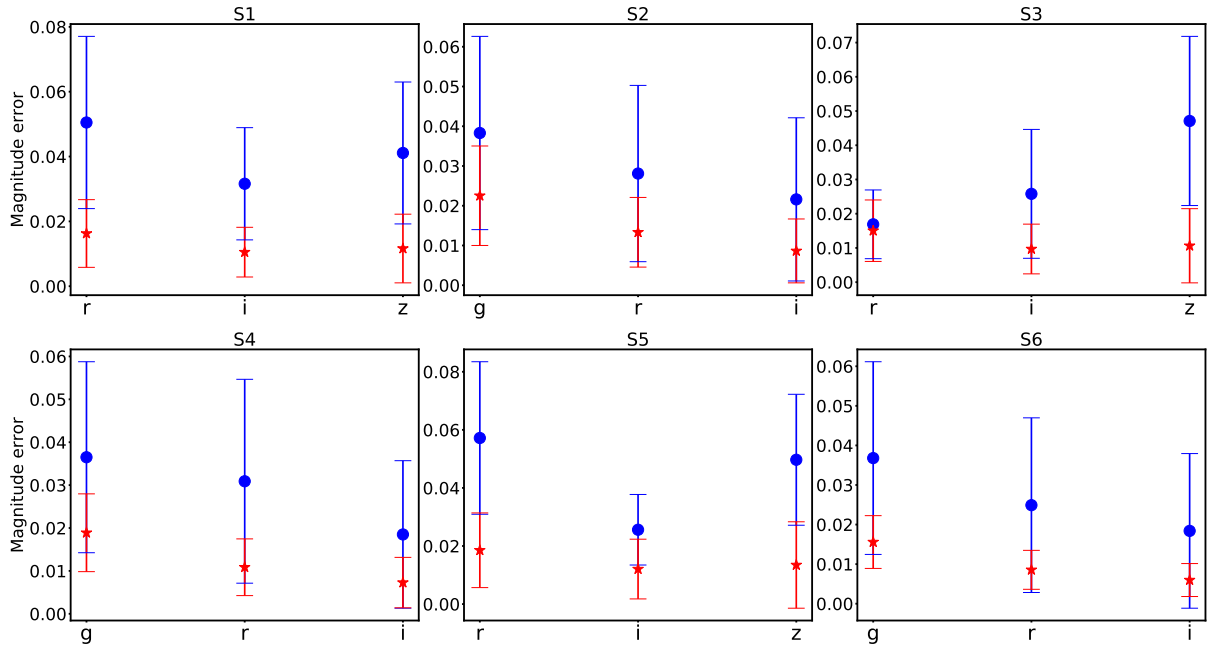

Figure 3.1: The scatter points are the median magnitude errors with a 1σ error bar. Blue stands for our fields data set, while red for the COSMOS.

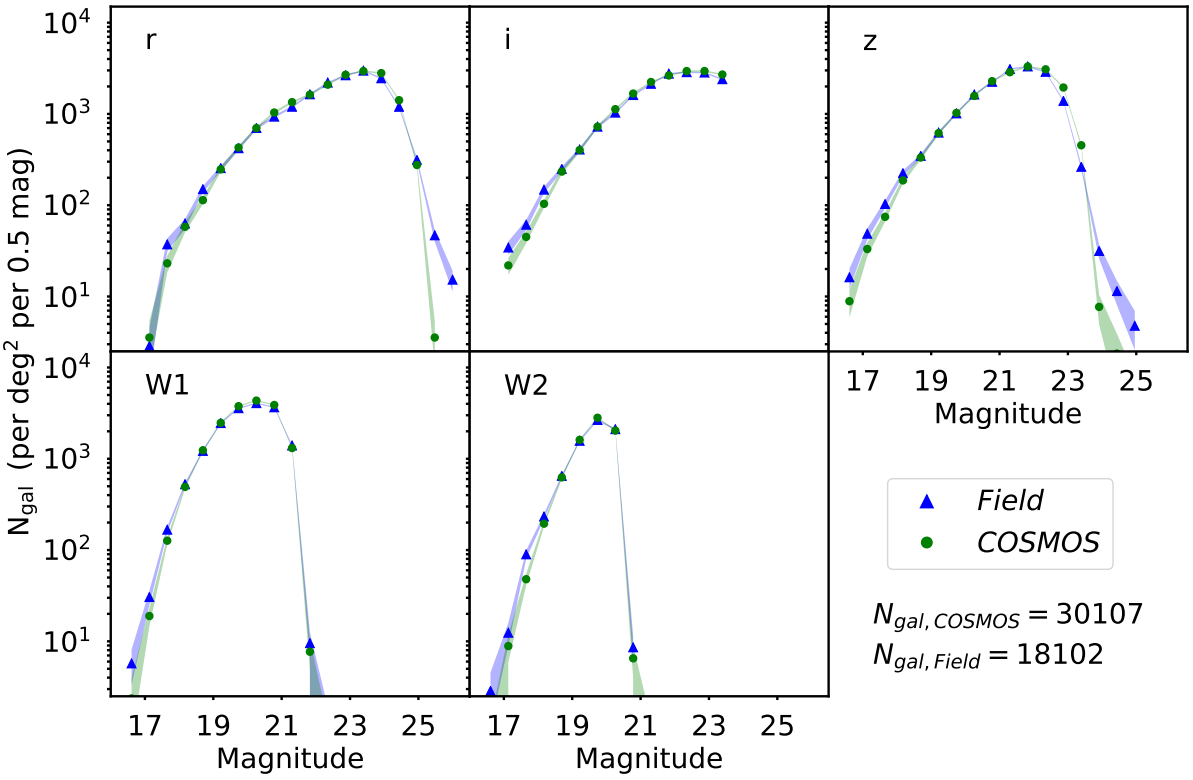

Figure 3.2: Galaxy counts for the S1 field (blue) and COSMOS (green).

performance of the algorithm, we plot $z_{\rm in}$ versus $z_{\rm out}$ for the test sub-samples, as well as the *Bias* and the $\sigma_{\rm NMAD}$ values in Figure 3.3. Figure 3.4 presents the redshift distributions of each field as well as for the COSMOS training and test samples.

For all cases in Figure 3.4, the photo-z$_s$ distributions between the COSMOS training and test samples follow similar shapes, which is another indication (in addition to the photo-z metrics) that the network training process was successful. For the distributions of photometric redshifts estimates for our fields, it is noticed that the S4 and S6 have different shapes, when compared with the COSMOS samples. To better understand this, we can check their magnitude distributions (Appendix D). The S6 field shows differences in the optical r and i bands, and this may explain the divergence in the estimates. These features indicate field peculiarities that may be associated with the concentration of galaxies and the evolutionary stage of the structures in these redshifts. This excess of galaxies can be considered as one indicator of a protocluster at the triplet's redshift.

It is also important to say that we also tested the estimates without and with the galaxies that have no measurements in W1 or W2 (as described in 2.7), obtaining $\sigma_{\rm NMAD}$ ranges of 2.4–3.8% and 3.4–5.5%, respectively. Galaxies that have no measurements in one of these two bands represent 60% of the sample, so we decided to keep them even if this represents a cost in the performance of `ANNz2`. These values are obviously much higher than the accuracy achieved by the *COSMOS2015* photo-z estimation, since the amount of information (inputs) available is much less. Still, our $\sigma_{\rm NMAD}$ values are comparable with those obtained in other works (e.g., van der Burg et al., 2020; Jian et al., 2020).

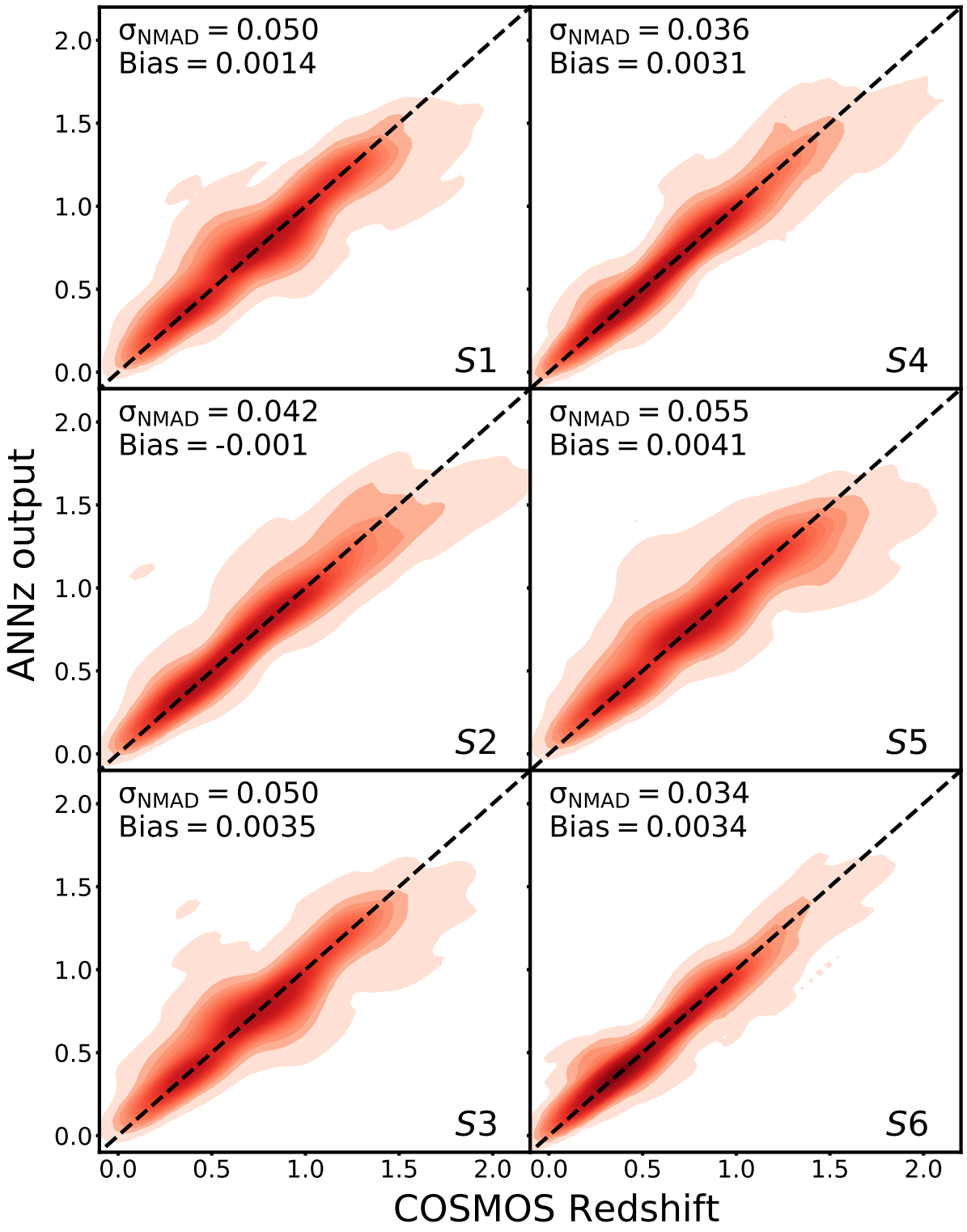

Figure 3.3: Photometric redshift estimation using `ANNz2` *for the validation and test data sets.*

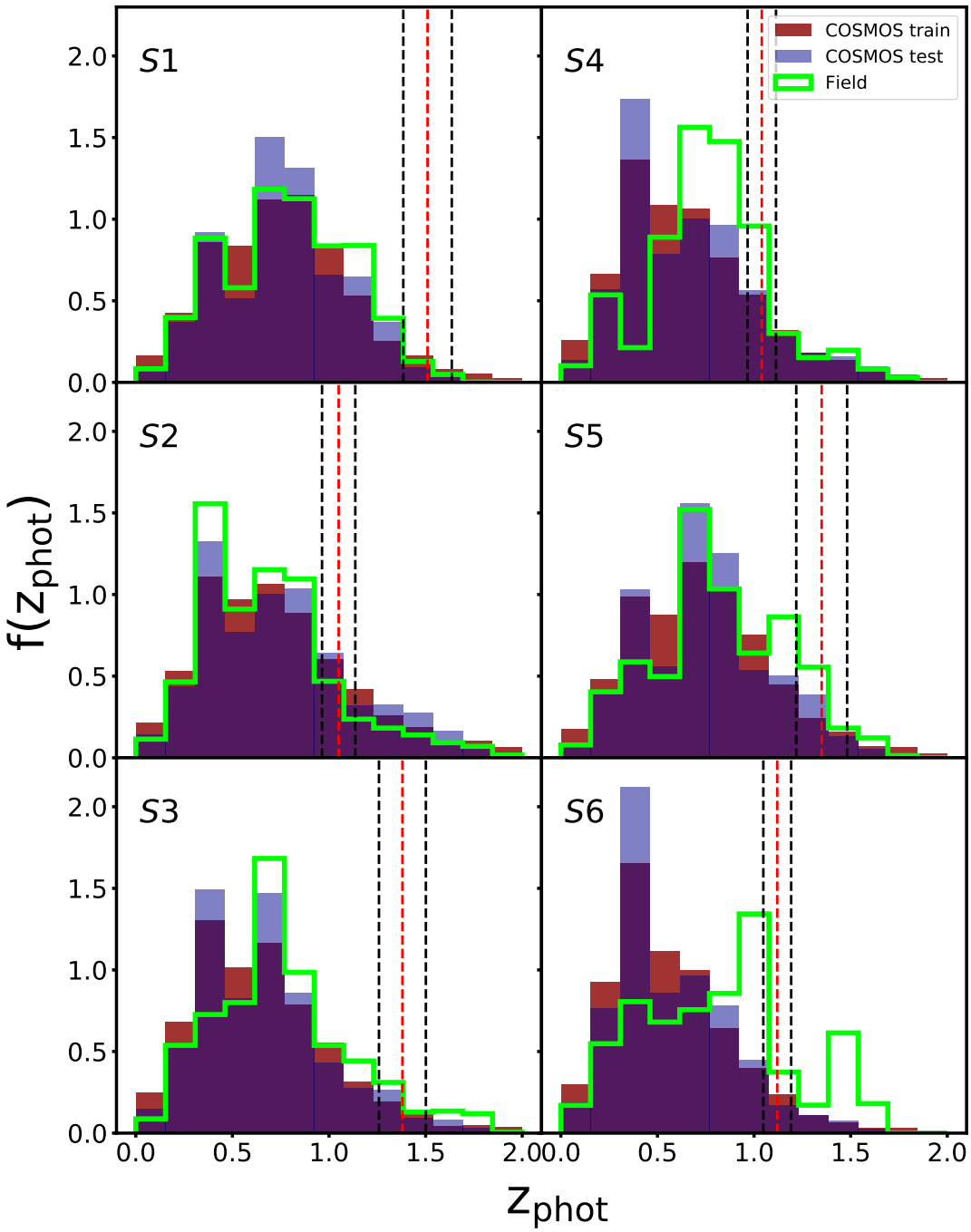

Figure 3.4: Photometric redshift distribution for COSMOS test-set galaxies (blue), COSMOS training-set galaxies (dark red), and for the CFHT fields (lime green), adopting the same magnitude limits. The red and black vertical dashed lines stand for the $\bar{z}$ and $\bar{z}_{\rm slab}$, respectively.

3.2 | The density field

In this section, we will present the methodology adopted for the construction of the maps. First, the concepts and general aspects of the method are presented, and then we show how it was applied to our fields, taking into account the criteria that were defined that we think best suits our case.

3.2.1 | Kernel density estimation

The *Kernel density estimation* (KDE) is a non-parametric way of estimating densities. In astronomy, this technique is widely used since it is not necessary to specify a functional model. This aspect of non-parametric estimators allows us to achieve a description more suited to the real data since these rarely follow simple distributions. Therefore, KDE is an excellent method for achieving density maps taking into account every aspect of how the real data is distributed.

A problem that arises immediately when we analyze a distribution using standard histograms, in one or more dimensions, is to set the position and the limits of the bins, since these values will affect directly affect the counts in each bin. When using KDEs, this problem is overcomed by allowing each point to have its own bin (centered on the point value, and with a certain width to be defined - known as *bandwidth*). The method also allows the bins to overlap, to be represented by different curves (e.g., top-hat, Gaussian, exponential), and, still, a different *weight* can be assigned to each one, depending on the characteristics of the distribution. For this reason, the data is directly responsible for the shape that the density function will assume. Besides all of these practical advantages, according to Ivezić et al. (2014), the KDE is theoretically more robust than the standard histograms.

Figure 3.5 was taken from Ivezić et al. (2014) and exemplifies very well the aspects discussed in the previous paragraph. From top to bottom, the first row of panels shows the problem when defining the position and limits of ordinary histograms bins. The other panels show the distributions from the KDEs approach. The left middle panel show top-hat bins, while the remaining three panels are examples using Gaussian bins with different bandwidths.

Next, we will present the KDE function, its parameters, how we choose the values for each one, and the application on our six fields data.

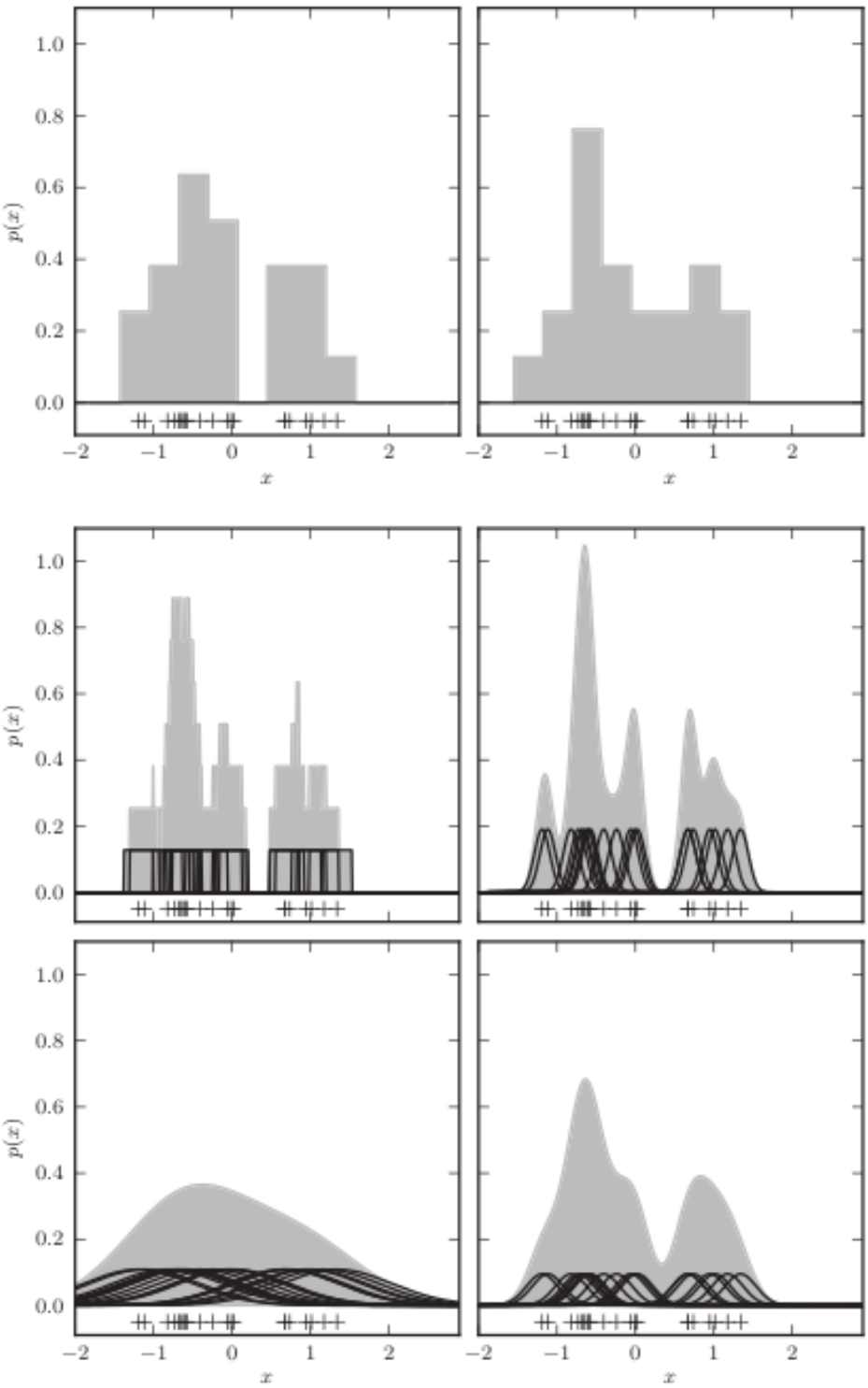

Figure 3.5: Figure reproduced from Ivezić et al. (2014). The top row of panels shows the problem of deciding the exact position and limits of ordinary histograms, and how it can interfere in the analysis. The other panels show examples of how KDEs works. Each black curve represent a unit of the data sample. The middle-panels show examples of top-hat and Gaussian kernels (left and right, respectively), and, in the lower-panels, the Gaussian Kernels are shown with different bandwidths. It is possible to see that they can overlap, and the resultant KDE function (represented by the gray shaded areas) will be the sum of each of these single curves.

3.2.2 | Applying KDE to our data

To verify how galaxies are spatially distributed in a certain field at the triplet redshifts, we computed overdensity significance maps using the *kernel density estimator* (KDE; e.g., Ivezić et al., 2014). The galaxy surface density at a point x is given by

$$\Sigma(x) = \frac{1}{Nh^2} \sum_{\mathrm{i}=1}^{N} K\left(\frac{d(x, x_{\mathrm{i}})}{h}\right), \tag{3.2}$$

where x_i represents the coordinates of one of the N galaxies with photometric redshifts, $K(u)$ is the kernel function (assumed Gaussian), $d(x, x_{\mathrm{i}})$ is the projected (Euclidian) distance between x and x_{i}, and h is the kernel bandwidth.

The bandwidth can be chosen with statistical or physical criteria. In the former case it is common to adopt thumb rules (e.g. *Scott's rule*), *Cross-validation*, or bayesian approaches (e.g., Hyndman et al., 2006). We have noticed, however, that both cases lead to large values of h and, consequently, to too smooth maps of the galaxy distribution. A more physical approach is to adopt a sensible scale for the kernel bandwidth. Chiang et al. (2013) have used the Millennium simulation to estimate the effective radius (R_{e}) of $z = 0$ galaxy clusters progenitors as a function of redshift, finding that, at z $\sim$ 1.3, the typical diameter ranges from 4 to 15 cMpc (lower and upper limit of Fornax-type and Coma-type progenitors, respectively). Motivated by this result, we decided to adopt h = 5 and 10 cMpc, to perform our analysis. The smaller bandwidth is better to probe $\sim$ group scale structures, while the larger one is more appropriate for the more massive objects. For our data, values lower than 5 cMpc are not convenient, since the average distance to the fifth nearest neighbor is $\sim$ 4.5 cMpc. In other words, below this value, the overdensity significance maps are very fragmented into small sets of galaxies, deteriorating the maps information quality due to shot noise.

Another interesting parameter that can be used with the kernel function is a weight ascribed to each galaxy. We use the redshift estimates and their respective errors to calculate the probability that each galaxy is in the QSOs redshift slab. For this, we assume that a photo-z and its correspondent error are the mean and standard deviation of a Gaussian probability density function. Then, we calculate the probability p that a galaxy is in a redshift slab width $\bar{z}_{\mathrm{slab}} = \bar{z} \pm \sigma_{\mathrm{NMAD}}(1 + \bar{z})$ (see Table 2.2). We use p as the weight of the galaxy in the map. That is, the closer, in redshift, the galaxy is to the QSOs system, the greater is its contribution to the density map. The definition of the weight for a given galaxy in our maps is

$$w_{\mathrm{i}}(x_{\mathrm{i}}, y_{\mathrm{i}}) = \frac{p_{\mathrm{i}}}{\sum_{\mathrm{j}=1}^{\mathrm{N}} p_{\mathrm{j}}} \tag{3.3}$$

where x_{i} and y_{i} are the coordinates of a given galaxy on the plane of the sky, and p_{j} is the probability that the j-th galaxy is in the redshift slab of the triplet.

Each field was then covered by a 100 $\times$ 100 grid (where each "pixel" has a width of $\sim$ 0.6 arcmin), the kernel surface density was computed at each node of the grid, and, finally, we

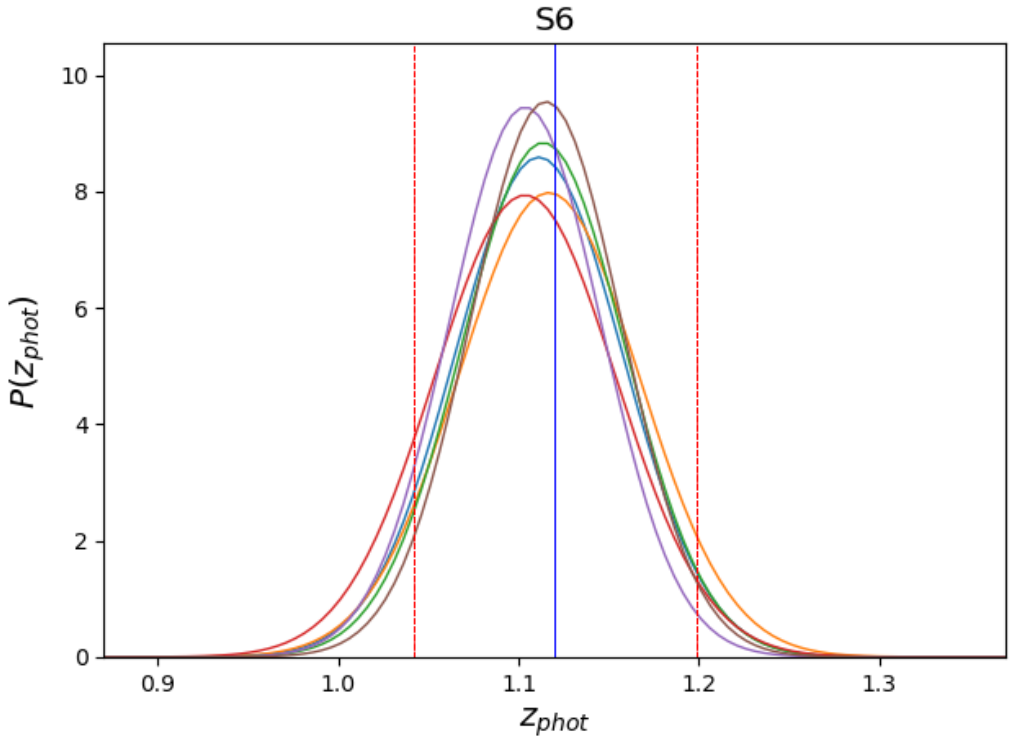

Figure 3.6: The gaussian curves of the galaxies with $p \geqslant 85\%$ of being in the redshift slab of the S6 triplet. The blue solid vertical line is the S6 $\bar{z}$; and the dashed red vertical lines are the limits of the slab according to our criteria.

determined the overdensity significance as

$$\sigma_{\text{gal}} = \frac{\Sigma - \bar{\Sigma}}{\sigma}, \tag{3.4}$$

where $\bar{\Sigma}$ and σ are the mean and standard deviation of the KDE density for the values in the grid.

3.3 | Mocks

This section describes how we reproduce our methods in the simulated data for the comparisons with the observations be as fair as possible. Then, we explain the need to estimate the photometric redshifts for these set of mocks data and how it was done.

3.3.1 | Comparison with Simulations

We have reproduced our method to estimate overdensity significance maps with our mock samples. Both observed and simulated datasets present some differences in the number of galaxies due to the cross-match of our imaging data with the unWISE catalog. Additionally, the magnitude estimation for the PCcones did not consider dust emission, which becomes non-negligible in the MIR, in particular in the *W2* band. Hence, we cannot estimate photometric redshifts directly from the mock magnitudes. However, we address these differences in a statistical way, as explained below.

3.3.2 | Photometric redshifts for mock galaxies

We generated mock photo-zs according to the redshift of the galaxies in the PCcones, following the procedure used by Krefting et al. (2020). The mock photo-z for a galaxy at z_i and reference band magnitude $m_{\mathrm{ref,i}}$ is drawn from a Gaussian with mean z_i and standard deviation equals to the typical error in the photo-z measurements[2] for galaxies with magnitude $m_{\mathrm{ref,i}}$. We have also added catastrophic redshifts as described by Krefting et al. (2020). Assuming a representative outlier fraction (9.5, 7.5, 10.7, 5.7, 11.4, and 4.7% for each field, respectively) these mock catastrophic photo-zs were drawn uniformly between $0 \leqslant z \leqslant 3.0$. With this technique, we achieved a σ_{NMAD} for the simulated photo-zs within $\sim 0.5\%$ from those obtained for each field.

3.4 | Observations results

We present in Table 3.1 overdensity significance values (σ_{gal}) associated to each of the QSOs of our sample, the mean ($\bar{\Sigma}$), and the standard deviation (σ) of the KDE maps for the values in the grid for two different bandwidths chosen for our analysis of 5 and 10 cMpc denoted by the underwritten "5" and "10", respectively. There is also the median of the $r - i$ color of two groups of galaxies divided as: those within an angular distance of 7.5 cMpc of the centroid of the quasars (triplet associated galaxies), and the others (field galaxies).

Figure 3.7 and 3.8 show the resulting overdensity significance maps for the six fields and for the two different bandwidths – as can be seen in the title of each map. In Figure 3.9, the

[2] Krefting et al. (2020) uses as standard deviation the typical width of the photometric redshift probability density distribution, $p(z)$, which are related with the photometric redshift errors.

Table 3.1: Measurements of the triplet members overdensity significance in the density field.

Field	QSOs	$\sigma^a_{\rm gal,5}$	$\bar{\Sigma}^b_5$	σ^c_5	$\sigma^d_{\rm gal,10}$	$\bar{\Sigma}^e_{10}$	σ^f_{10}	Med$(r-i)_{\rm field}$ [g]	Med$(r-i)_{\rm trip}$ [h]
	SDSSJ021612-010519	-0.536			0.469				
S1	SDSSJ021622-010818*	0.033	0.966	0.520	0.529	0.937	0.294	0.479	0.462
	SDSSJ021630-011154*	2.034			0.676				
	SDSSJ022158+000043*	0.179			0.440				
S2	SDSSJ022214-000322	0.254	0.963	0.436	-0.063	0.930	0.285	0.690	0.692
	SDSSJ022223-000745*	-0.187			0.785				
	SDSSJ101603+453316*	0			0.599				
S3	SDSSJ101610+453142	-0.819	0.963	0.567	0.326	0.934	0.349	0.447	0.454
	SDSSJ101620+453257	-0.585			0.335				
	SDSSJ224255-010924*	-0.211			0.624				
S4	SDSSJ224304-010946*	-0.255	0.960	0.403	0.303	0.926	0.263	0.719	0.791
	SDSSJ224313-010552	0.044			0.459				
	SDSSJ232101+001923	-0.600			-0.282				
S5	SDSSJ232119+001826	-0.003	0.961	0.390	-0.624	0.929	0.252	0.545	0.565
	SDSSJ232129+001413*	-0.705			-0.791				
	SDSSJ232812-003238*	2.100			1.799				
S6	SDSSJ232821-003547*	1.367	0.966	0.430	1.599	0.938	0.269	0.604	0.695
	SDSSJ232824.5-003658	2.673			1.786				

[a], [b], [c]: *The overdensity significance, the mean, and standard deviation of the KDE density for a field with 5 cMpc bandwidth.*
[d], [e], [f]: *The same of* [a], [b], [c], *for a field with 10 cMpc bandwidth.*
[g], [h]: *The median of the r-i color of the field and triplet galaxies sample, respectively. The later correspond to the galaxies that are at an angular distance lower than 7.5 cMpc radius from the triplet's centroid.*

yellow fulfilled histograms are the $\sigma_{\rm gal}$ distributions for galaxies with $p \geqslant 50\%$, indicating the densities associated to the triplet members with red dashed vertical lines for radio-loud, and the black solid for radio-quiet quasars. This figure shows that in all but one case (S6), the triplets avoid the densest regions of each field. The triplet in Field S6, on the other hand, seems to be a member of a significant overdensity at $z = 1.12$. The discussion of this results is in the next Section.

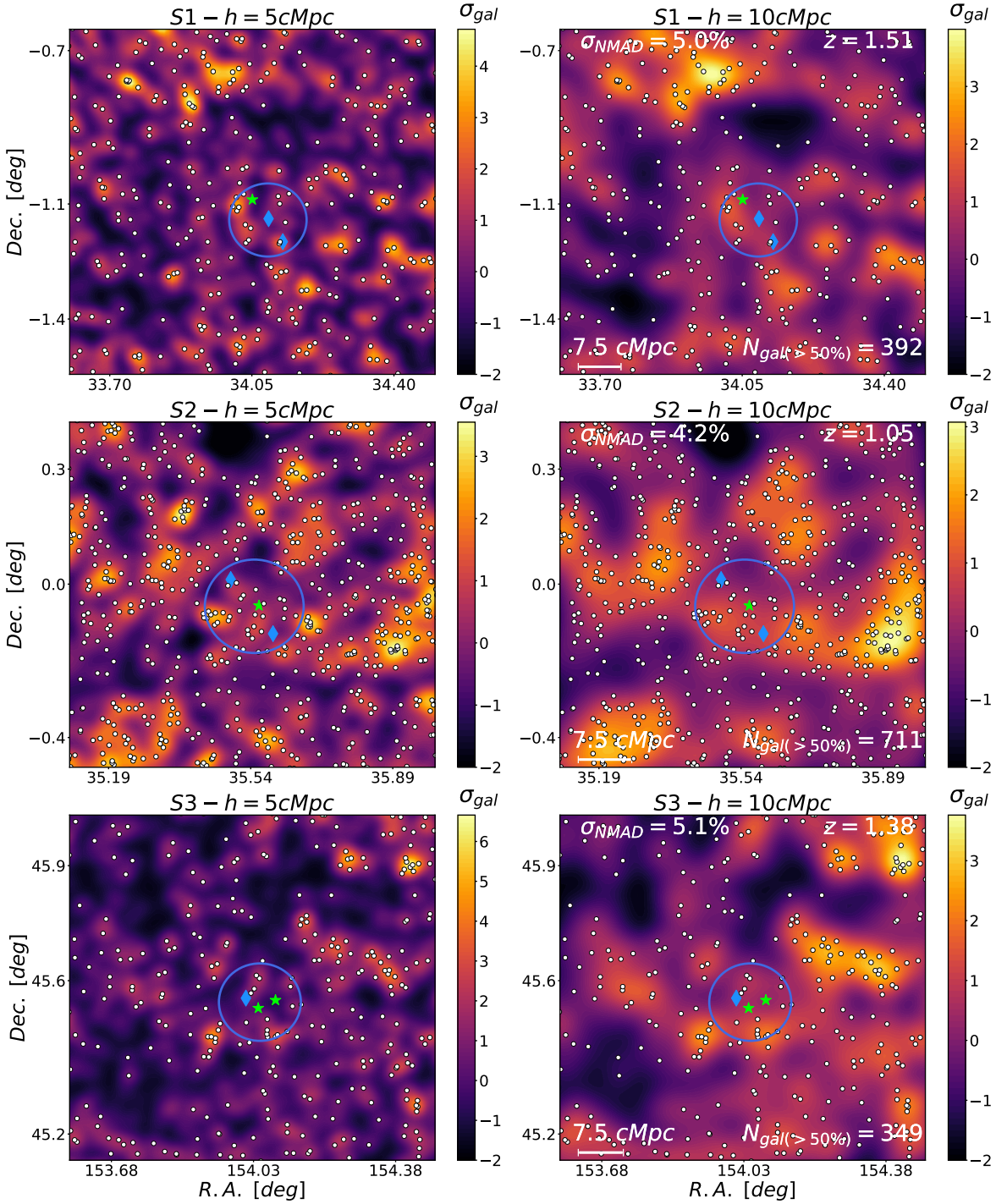

Figure 3.7: Overdensity significance maps for S1, S2, and S3 fields CFHT/Megacam observations. Green stars marks the position of RL QSOs, while blue diamonds the RQs; white dots are the galaxies with $p > 50\%$; blue circles have 7.5 cMpc projected radius, and is centered in the QSOs system's centroid; The title of each map shows the field and the correspondent bandwidth; In each field for $h = 10$ cMpc map, we also present general information as the $\sigma_{\rm NMAD}$ and the number of galaxies with $p \geqslant 50\%$.

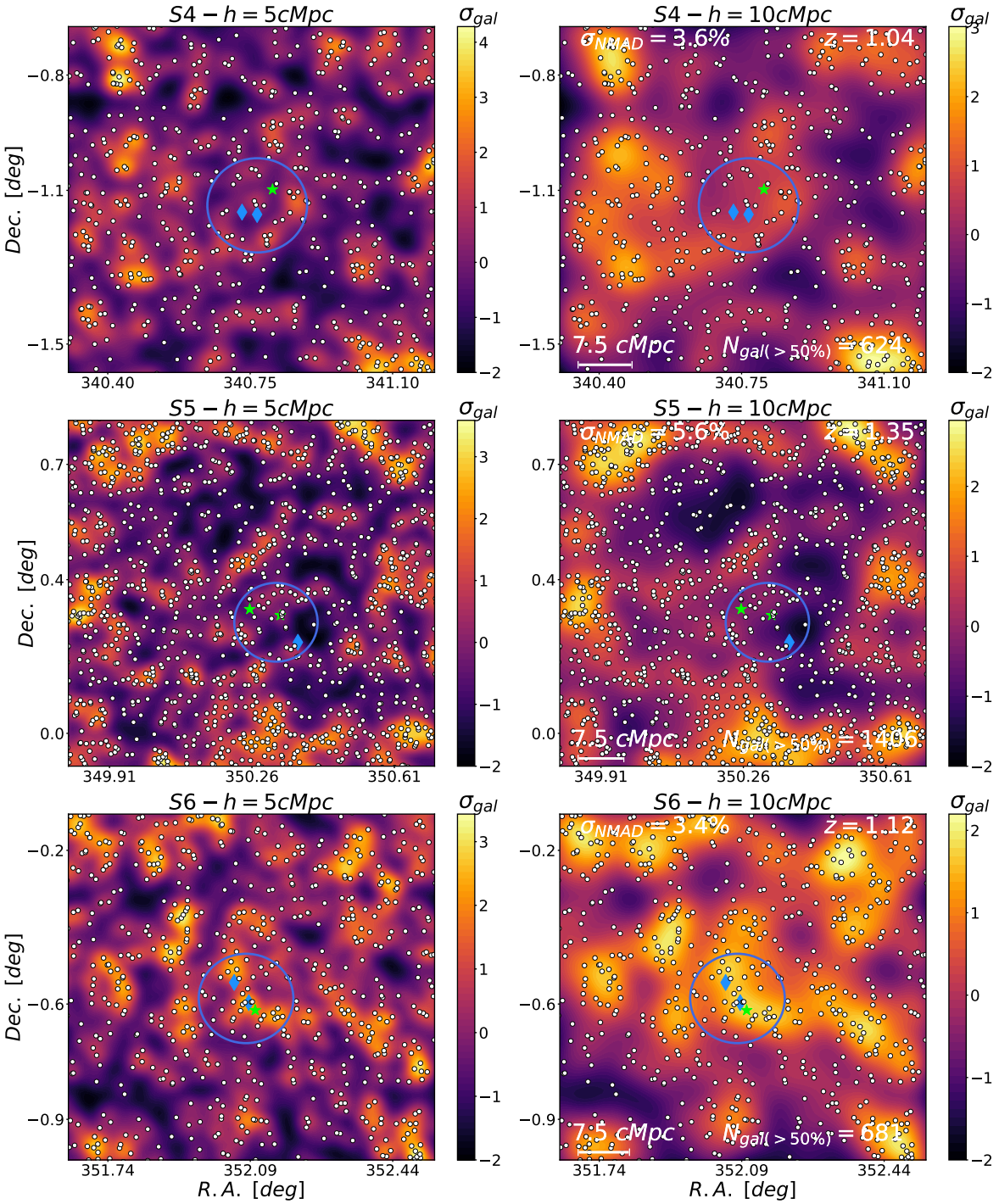

Figure 3.8: The same as Figure 3.7 for S4, S5, and S6 fields.

3.5 | Mocks

The use of simulated data in parallel with observational data is an excellent tool to assess how significant the results are. In this section, we show the calculated overdensity significance for the mocks and an analysis of how efficient our setup is to find protoclusters in the quasar redshifts.

3.5.1 | Overdensity significance maps

Our simulated sample presents an excess of galaxies compared to the observational sample, and we need to remove this excess to perform a reliable comparison. We made it by selecting all galaxies in the mock with photo-z_s within the redshift slab. Then, we compute the overdensity significance maps from a subsample of galaxies according to the effective number of objects within the analyzed redshift interval. For a given slab, we define this quantity as the sum of all probabilities of the galaxies to be within it. These sums (extracted from the observational photo-z_s estimates) are 595, 910, 711, 853, 1594, and 709 for each field, respectively. Notice that at this step all weights are the same. Additionally, we used the same 100×100 pixels grid and the two different kernel bandwidth h, as we described above.

We present in Figure 3.9 the overdensity significance distribution of our mock fields (dark red step line histogram). Despite the applied method for the mock sample be not the same as that performed to observational data, in general, both observed and simulated distributions are similar, with exception of the S6, which shows a peculiar distribution. As we mention in Section 3.1, the S6 field presents a large number of galaxies at higher redshifts.

3.5.2 | Cluster detectability

Following the same procedure described in Araya-Araya et al. (2020), we first identify the overdensity peaks in each PCcone field, by comparing each pixel of these maps with the adjacent ones (in a 3×3 pixels matrix). If the central pixel has the highest value, we select its position as an overdensity peak. After, we link the 20 mock protoclusters with these overdensity peaks within a radius of 5 comoving-Mpc. If there is more than one peak in the linking area, then we attributed the denser one to the protocluster.

We present the overdensity measurements of these structures for each field and the adopted

bandwidth in Table 3.2 and in Figure 3.9.

Table 3.2: Detected overdensities of the 20 protoclusters for our field-like samples at z = 1.04, 1.12, 1.36, and 1.51.

		$\sigma_{gal,5}$ (h = 5 cMpc)						$\sigma_{gal,10}$ (h = 10 cMpc)					
Descendant Type[a]	$\log(M_{z=0}/M_\odot)$[b]	$S1-like$	$S2-like$	$S3-like$	$S4-like$	$S5-like$	$S6-like$	$S1-like$	$S2-like$	$S3-like$	$S4-like$	$S5-like$	$S6-like$
Fornax	14.14	1.52	–	1.84	0.93	0.95	–	2.45	1.48	1.84	0.92	1.16	–
	14.17	2.89	–	2.47	–	0.85	–	2.32	–	2.32	–	1.83	–
	14.2	2.24	1.97	4.13	3.2	0.53	–	1.73	–	1.91	–	0.85	–
	14.23	2.48	1.32	2.56	2.08	2.43	–	0.75	–	0.69	–	1.28	–
	14.27	-0.17	–	1.71	–	2.13	2.96	2.8	–	2.8	1.33	2.62	1.01
	14.31	2.99	1.7	1.57	–	2.28	1.44	2.07	–	1.3	–	2.65	1.87
	14.36	4.25	1.91	3.96	–	3.23	2.04	-0.93	–	3.5	2.45	2.34	–
	14.42	1.17	–	0.74	–	0.64	1.29	1.57	–	3.56	2.34	1.06	1.8
Virgo	14.49	1.91	–	0.69	–	1.83	0.46	0.61	–	2.53	–	4.55	3.49
	14.58	4.01	1.54	2.47	1.41	–	–	1.42	–	2.07	–	1.54	–
	14.73	3.86	2.83	1.92	–	3.08	2.36	1.71	–	0.75	–	2.66	1.07
	14.73	1.37	0.89	2.91	–	2.31	–	4.85	3.06	0.5	–	2.62	–
	14.82	2.06	–	3.22	3.22	5.43	4.28	2.19	1.65	3.21	2.38	3.17	2.97
	14.94	2.46	–	1.1	–	2.88	–	2.51	–	2.07	1.61	1.79	–
Coma	15.03	3.82	2.18	2.86	–	1.7	-0.2	4.19	3.21	2.24	–	1.54	0.63
	15.05	5.4	3.29	2.57	1.59	0.41	–	2.72	2.23	4.06	2.33	5.56	4.8
	15.07	6.22	4.52	2.62	–	1.24	–	0.58	–	1.58	–	2.81	2.64
	15.16	–	–	3.33	1.61	0.87	–	0.85	3.24	1.46	–	1.75	–
	15.26	6.11	4.55	3.48	2.56	3.02	2.72	3.23	3.28	2.06	–	4.86	–
	15.36	1.85	0.5	2.77	–	0.97	–	2.63	1.68	2.63	–	3.04	2.03

[a],[b] *The descendant cluster type, and mass, that the mock protocluster will collapse at z = 0.*
[c],[d] *The overdensity significance of the mock protocluster in a field with bandwidth of 5 and 10 cMpc, and the same setup that we are adopting on the observations.*

The protoclusters that do not appear in the figure are all those without an overdensity peak within a projected distance of 5 cMpc of their centers. In general, using a kernel bandwidth of 5 cMpc, we have detected in our mocks 20, 19, 18, 18, 20, and 20 protoclusters out of the 20 simulated structures, respectively. For a bandwidth of 10 cMpc, the corresponding numbers are 10, 12, 11, 12, 10, and 14, respectively. Therefore, at least 95.8% of the protoclusters are associated to an overdensity peak within a radius of 5 cMpc, if we use $h = 5$ cMpc, whereas this fraction is of only 57.5% when we use $h = 10$ cMpc. This occurs due to projection effects: with the high bandwidth value other structures displaces the overdensity peak away from the position of the simulated protoclusters. Projection effects can also amplify the measured overdensities of structures and of random regions, explaining why we detected some Fornax and Virgo types in denser regions instead of Coma-type protoclusters.

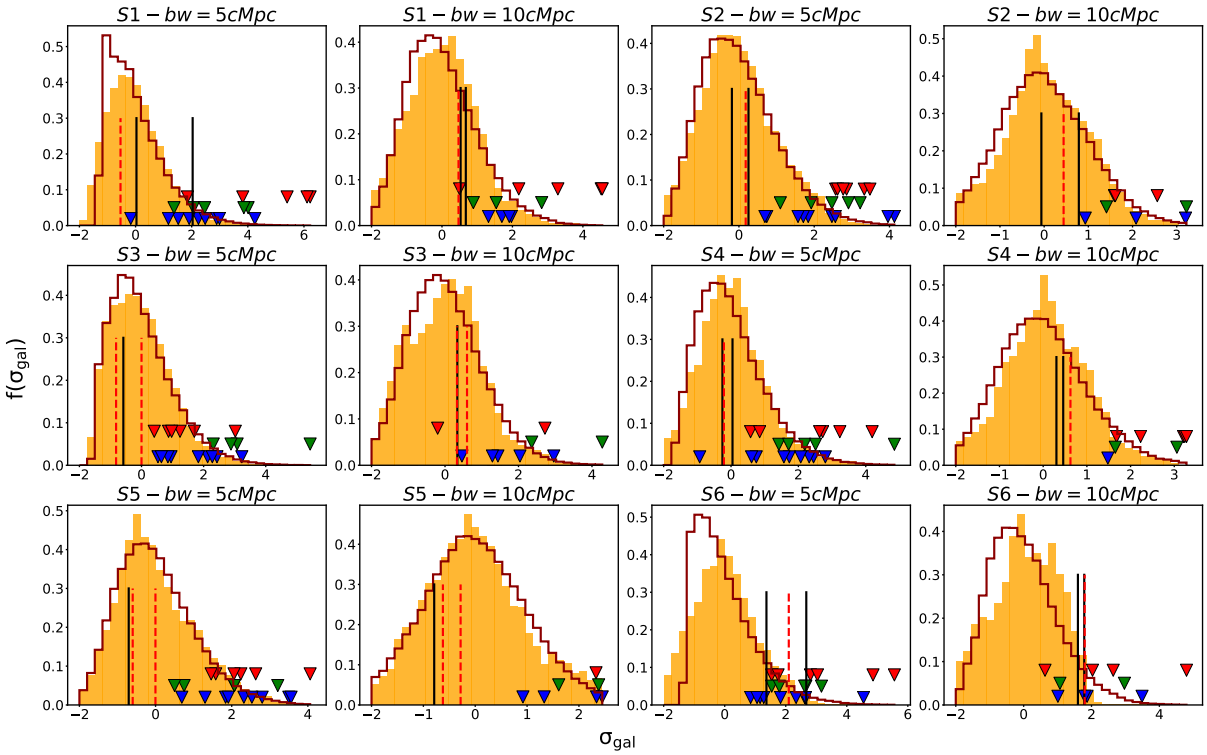

Figure 3.9: The orange filled histograms are the overdensity significance distributions. The red dashed and the solid black vertical lines stand for the measurement of the overdensity significance where RL QSO or RQ QSO members of the triplet are (see Table 3.1), respectively. The red line distributions are from the simulated sample. Red, green, and blue triangles mark the significance overdensity, $\sigma_{\rm gal}$, of Coma, Virgo, and Fornax type protoclusters, respectively.

4 Discussion

Some observations show that AGNs can reside in regions with a high density of galaxies (Seymour et al., 2007), and that is why they are often discussed as potential probes of high redshift structures. There are a number of works along this line, and many structures have already been identified in this way. An example is the CARLA survey, which identified 200 cluster candidates at $1.3 < z < 3.2$ from a survey of 420 radio-loud QSOs or radio-galaxies (Wylezalek et al., 2013). A spectroscopic follow-up with HST for the 20 densest candidates up to $z = 2.8$ confirmed 16 as high redshift clusters (Noirot et al., 2017). However, the follow-up of more systems is required to establish the frequency of real clusters in this sample.

Also, several studies have focused on quasar/AGN systems, such as quasar pairs, with results somewhat contradictory. For instance, Boris et al. (2007), through the analysis of galaxy overdensities, their richness, and the identification of the red sequence in color-magnitude diagrams, reported that 3 out of 4 pairs within the redshift interval $0.9 < z < 1$, presented strong evidence of being part of galaxy clusters. Green et al. (2011) studied the color properties in the visible and possible X-ray emission of a hot IGM from 7 pairs extracted from SDSS-DR6 at $0.4 < z < 1$, finding no evidence of association of these systems with galaxy clusters in their sample. Another interesting example is the work of Onoue et al. (2018), who studied quasar pair environments at $z \sim 1$ and $z > 3$, using optical catalogues from the HSC-SSP survey (Aihara et al., 2018, DRS16A). They found two pairs at $z = 3.3$ and 3.6 associated with regions of high overdensity (5 σ above the mean of the distribution), suggesting that pairs of luminous QSOs seem to be good tracers of protoclusters at these redshifts. At $z \sim 1$, they worked with a total sample of 38 pairs and found that 19% of them are in massive environments ($> 4\sigma$), concluding that pairs of QSOs at this redshift are also good tracers of matter-rich environments.

Here we have extended this type of approach by analyzing QSO triplets at $1 < z < 1.5$, with the requirement that at least one member of the system is a radio-loud quasar. Our main results are depicted in Figure 3.7 and Table 3.1. The blue circles in the figure have a radius of 7.5 cMpc in projected distance. In all but one case, the triplets tend to avoid overdense regions. The exception is the S6 field, where the triplet is at the top of the major overdensity in its redshift slab (see Table 3.1 for the significance overdensity associated with each QSO). It is also interesting to notice that the shape of this overdensity is clearly elongated along the line connecting the three quasars of this field in both maps with different structure scales. The analysis of this field in differents bandwidths shows that even in smaller scales, the quasars are at the peak of an overdensity, which indicates a higher concentration of galaxies close to the triplet. With $h = 10$ cMpc the QSOs remain at the top of this overdensity, revealing that these concentrations should be at the center of a larger structure. Notice that we found the spectrum (Appendix A), the black hole mass, and bolometric luminosity information (Table 2.2) only for the radio-loud member of this field. The other two were observed spectroscopically by Croom

et al. (2009), but we were not able to find this information since the network system for this survey appears to be offline. Also, in Table 2.2, among all the quasars in our sample, the S6 radio-loud has the smallest virial black hole mass and one of the smallest bolometric luminosities. As already mentioned, at high redshifts, the more massive and luminous radio-loud quasars tend to be more associated with denser environments (e.g., Hatch et al., 2014; Stott et al., 2020). However, we found no evidence of this association with the most luminous and massive central black holes quasars of our sample.

Although S4 presents a high concentration of galaxies in the slab (Figure 3.4), there is no significant overdensity associated with any of the triplet members. Therefore, it is possible that larger structures exists in this volume, but with no association with the quasars. Moreover, a detailed work to search for clusters/protoclusters in these images at other redshifts is planned for futuer work.

Our motivation for the adopted kernel bandwidth is related to the typical size of massive protoclusters, as well as to the number of galaxies in each field. From our simulations, we can recover a significant fraction (95.8%) of protoclusters with a bandwidth of 5 cMpc. For $h = 10$ cMpc, many low mass protoclusters do not reside in the densest regions in the field because the higher value of h dilutes the overdensities of compact regions, such as groups and low-mass protoclusters. Additionally, irrespective of the bandwidth, the overdensities are "washed-out" due to the redshift uncertainties. The redshift slabs, which are defined according to the photometric redshift accuracy, are ∼five times wider than the redshift separation between the QSOs, as well ∼seven times wider than the typical 3σ velocity dispersion of Coma-type protoclusters at $1.0 \leqslant z \leqslant 1.5$ (1567 km/s). Nevertheless, we notice from Figure 3.9 that overdensities associated with protoclusters are at the high density end of all distributions for all the field emulations, independently of the kernel bandwidth. It implies that given our observational constraints and photometric redshift uncertainties, we still can separate real structures from random regions.

Is the S6 overdensity real or an artifact produced by projection effects? We present in Figure 4.1 color-magnitude diagrams comparing galaxies "inside" and "outside" the encircled region in the density field[1]. The figure indicates that galaxies closer to the triplet in the S6 field tend to be redder with respect to the $r - i$ color than those in other parts of the field. This trend is less strong for the other triplets. The median colors involved in this analysis are also presented in Table 3.1. To be more quantitative, we have compared through a Kolmogorov-Smirnov (KS) and an Anderson-Darling (AD) test, the color distribution of galaxies within and outside the blue circles of the six fields in Figure 3.7. The resulting KS (AD) p-values that the samples inside and outside the circles were drawn from the same distribution are 0.21 (> 0.25), 0.23 (> 0.25), 0.72 (> 0.25), 0.42 (> 0.25), 0.85 (> 0.25), and 0.03 (0.07), for S1 to S6. For the S6 field, the null hypothesis can be rejected at a confidence level of 97% (93%), contrarily to the other fields. From all these shreds of evidence, we conclude that the S6 triplet seems to be a good candidate for being part of a rich structure, like a galaxy cluster. For the other systems, although some of the overdensities could be associated with galaxy clusters/protoclusters, the triplets are at most at their outskirts.

[1]The circle has a 7.5 cMpc projected radius, while the distance between the centroid of the quasars triplet, and the coordinates of each galaxy, are calculated with the angular distance equation

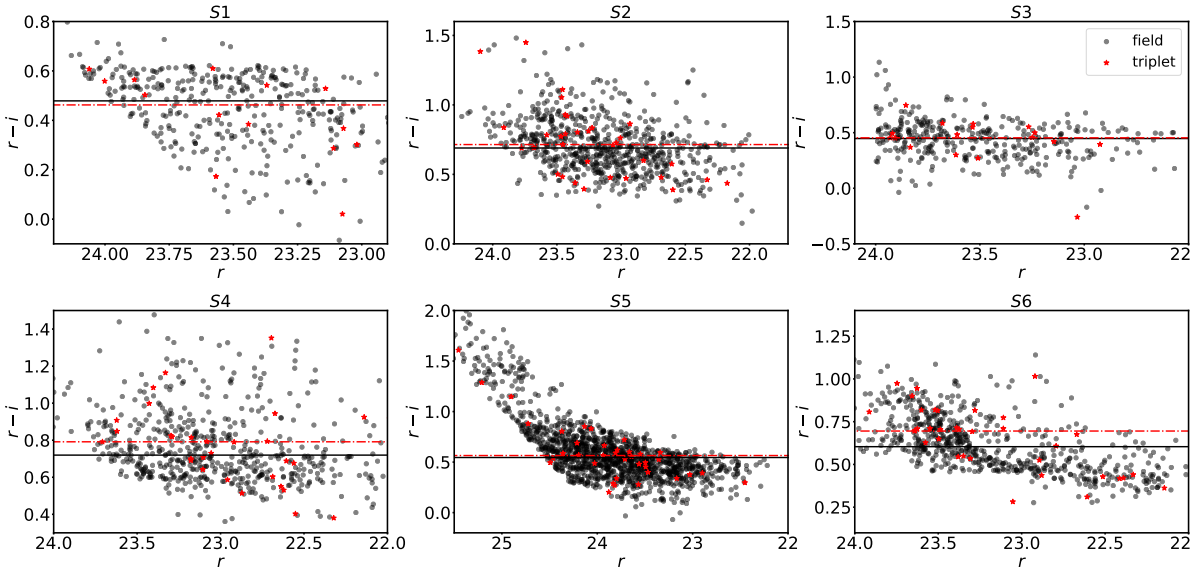

Figure 4.1: The color-magnitude diagrams for our six fields. Red points are the galaxies in the encircled region, whereas black points, the others; red dashed horizontal and black solid lines are the median for the galaxies inside and outside the circle, respectively.

5 Conclusion and Next Steps

In this work, we have investigated the environment of six quasar triplets at $1 \lesssim z \lesssim 1.5$ making use of multiband CFHT/Megacam images, complemented with mid-infrared photometry (3.6 and 4.5μm) from the *unWISE* survey.

We have used photometric redshifts to identify galaxies at the redshifts of the triplets. These photo-zs were obtained with the `ANNz2` software trained with the accurate photometric redshifts of an enlarged version of the *COSMOS2015* catalogue, where we included the W1 and W2 *unWISE* bands. This allowed us to obtain typical accuracies (measured with the $\sigma_{\rm NMAD}$ metrics) of 0.04 when compared with the spectroscopic training set.

The density field in a redshift slab of width $\Delta z = \sigma_{\rm NMAD} \times (1 + z)$ was obtained with a Gaussian Kernel Estimation with a bandwidth of 10 cMpc. The contribution of the galaxies for the map was weighted by the probability of the galaxy be in the redshift slab based on the photometric redshift estimate and its error.

Our analysis shows that one of our six triplets is in an overdense region, possibly a cluster or a protocluster of galaxies at $z = 1.12$. The color-magnitude diagram of the objects in this overdensity also points toward a high concentration of redder galaxies. This structure should be further investigated by spectroscopic follow-up.

Given our sample's size, its Poissonian uncertainty is too large, and, as pointed out in Section 1 and discussed in Section 4, the debate of whether quasar systems are good tracers for high-density regions still should be more investigated with the analysis of other structures of this type throughout different redshifts. Also, it is not trivial to compare results from different works since the sample selection, observational constraints, redshift range, and analysis methods, differ from each other. However, our results help adressing this puzzle by adding 6 systems where 1 ($\sim 17\%$) is consistent with quasars belonging to a region highly populated by galaxies, possibly a protocluster.

Our six images are a source of information that goes far beyond the relationship with the quasars themselves. Therefore, as the next steps, we intend to make a systematic study in search of galaxy clusters, and protoclusters, in all redshifts that our observations are complete. Thus, we will evaluate slab-by-slab if the highest concentrations of galaxies are real candidates for being part of a larger structure through comparisons with the PCcones algorithm. Furthermore, we will prepare spectroscopic follow-ups proposals for the most prominent overdensities in order to confirm if they are real structure.

Bibliography

Infrared Astronomical Satellite (IRAS) Catalogs and Atlases.Volume 1: Explanatory Supplement., vol. 1 (1988).

Abell, G. O., "The Distribution of Rich Clusters of Galaxies.", ApJS, vol. 3, p. 211, 1958. URL http://dx.doi.org/10.1086/190036.

Adami, C. et al., "Galaxy structure searches by photometric redshifts in the CFHTLS", A&A, vol. 509, A81, 2010. 0910.3827, URL http://dx.doi.org/10.1051/0004-6361/200913067.

Adami, C. et al., "The XMM-LSS survey: optical assessment and properties of different X-ray selected cluster classes", A&A, vol. 526, A18, 2011. 1010.6195, URL http://dx.doi.org/10.1051/0004-6361/201015182.

Aihara, H. et al., "The Hyper Suprime-Cam SSP Survey: Overview and survey design", PASJ, vol. 70, S4, 2018. 1704.05858, URL http://dx.doi.org/10.1093/pasj/psx066.

Almosallam, I. A., Jarvis, M. J. and Roberts, S. J., "GPZ: non-stationary sparse Gaussian processes for heteroscedastic uncertainty estimation in photometric redshifts", MNRAS, vol. 462, no. 1, pp. 726–739, 2016. 1604.03593, URL http://dx.doi.org/10.1093/mnras/stw1618.

Araya-Araya, P., Vicentin, M. C., Sodré, J., L., Overzier, R. A. and Cuevas, H., "High Redshift Structure Detection on Photo-z Surveys, Using Protocluster-Lightcones ~1", , 2020. "submmited for publication".

Arnouts, S. et al., "Measuring the redshift evolution of clustering: the Hubble Deep Field South", MNRAS, vol. 329, no. 2, pp. 355–366, 2002. astro-ph/0109453, URL http://dx.doi.org/10.1046/j.1365-8711.2002.04988.x.

Bahcall, N. A., "Clusters and superclusters of galaxies", *arXiv e-prints*, astro-ph/9611148, 1996. astro-ph/9611148.

Bekki, K., "Unequal-Mass Galaxy Mergers and the Creation of Cluster S0 Galaxies", ApJ, vol. 502, no. 2, pp. L133–L137, 1998. astro-ph/9806106, URL http://dx.doi.org/10.1086/311508.

Benítez, N., "Bayesian Photometric Redshift Estimation", ApJ, vol. 536, no. 2, pp. 571–583, 2000. astro-ph/9811189, URL http://dx.doi.org/10.1086/308947.

Bennett, C. L. et al., "Four-Year COBE DMR Cosmic Microwave Background Observations: Maps and Basic Results", ApJ, vol. 464, p. L1, 1996. astro-ph/9601067, URL http://dx.doi.org/10.1086/310075.

Bennett, C. L. et al., "First-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Preliminary Maps and Basic Results", ApJS, vol. 148, no. 1, pp. 1–27, 2003. astro-ph/0302207, URL http://dx.doi.org/10.1086/377253.

Bennett, C. L. et al., "Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Final Maps and Results", ApJS, vol. 208, no. 2, 20, 2013. 1212.5225, URL http://dx.doi.org/10.1088/0067-0049/208/2/20.

Bertin, E., "Automatic Astrometric and Photometric Calibration with SCAMP", in: *Astronomical Data Analysis Software and Systems XV*, edited by C. Gabriel, C. Arviset, D. Ponz and S. Enrique, *Astronomical Society of the Pacific Conference Series*, vol. 351, p. 112 (2006).

Bertin, E., "SWarp: Resampling and Co-adding FITS Images Together", , 2010. 1010.068.

Bertin, E., "Automated Morphometry with SExtractor and PSFEx", in: *Astronomical Data Analysis Software and Systems XX*, edited by I. N. Evans, A. Accomazzi, D. J. Mink and A. H. Rots, *Astronomical Society of the Pacific Conference Series*, vol. 442, p. 435 (2011).

Bertin, E. and Arnouts, S., "SExtractor: Software for source extraction.", A&AS, vol. 117, pp. 393–404, 1996. URL http://dx.doi.org/10.1051/aas:1996164.

Blake, C. et al., "The WiggleZ Dark Energy Survey: the selection function and z = 0.6 galaxy power spectrum", MNRAS, vol. 406, no. 2, pp. 803–821, 2010. 1003.5721, URL http://dx.doi.org/10.1111/j.1365-2966.2010.16747.x.

Bleem, L. E. et al., "Galaxy Clusters Discovered via the Sunyaev-Zel'dovich Effect in the 2500-Square-Degree SPT-SZ Survey", ApJS, vol. 216, no. 2, 27, 2015. 1409.0850, URL http://dx.doi.org/10.1088/0067-0049/216/2/27.

Boris, N. V. et al., "Searching High-Redshift Large-Scale Structures: Photometry of Four Fields around Quasar Pairs at z ~1", ApJ, vol. 666, no. 2, pp. 747–756, 2007. astro-ph/0703562, URL http://dx.doi.org/10.1086/519992.

Bouwens, R. J., Illingworth, G. D., Franx, M. and Ford, H., "z ~7-10 Galaxies in the HUDF and GOODS Fields: UV Luminosity Functions", ApJ, vol. 686, no. 1, pp. 230–250, 2008. 0803.0548, URL http://dx.doi.org/10.1086/590103.

Bower, R. G. and Balogh, M. L., "The Difference Between Clusters and Groups: A Journey from Cluster Cores to Their Outskirts and Beyond", in: *Clusters of Galaxies: Probes of Cosmological Structure and Galaxy Evolution*, edited by J. S. Mulchaey, A. Dressler and A. Oemler, p. 325 (2004). astro-ph/0306342.

Bower, R. G., Lucey, J. R. and Ellis, R. S., "Precision photometry of early-type galaxies in the Coma and Virgo clusters : a test of the universality of the colour-magnitude relation - II. Analysis.", MNRAS, vol. 254, p. 601, 1992. URL http://dx.doi.org/10.1093/mnras/254.4.601.

Boylan-Kolchin, M., Springel, V., White, S. D. M., Jenkins, A. and Lemson, G., "Resolving cosmic structure formation with the Millennium-II Simulation", MNRAS, vol. 398, no. 3, pp. 1150–1164, 2009. 0903.3041, URL http://dx.doi.org/10.1111/j.1365-2966.2009.15191.x.

Brun, R. and Rademakers, F., "ROOT — An object oriented data analysis framework", *Nuclear Instruments and Methods in Physics Research A*, vol. 389, no. 1, pp. 81–86, 1997. URL http://dx.doi.org/10.1016/S0168-9002(97)00048-X.

Bruzual, G. and Charlot, S., "Stellar population synthesis at the resolution of 2003", MNRAS, vol. 344, no. 4, pp. 1000–1028, 2003. astro-ph/0309134, URL http://dx.doi.org/10.1046/j.1365-8711.2003.06897.x.

Bădescu, T. et al., "Discovery of a Protocluster Associated with a Lyα Blob Pair at z = 2.3", ApJ, vol. 845, no. 2, 172, 2017. 1708.00447, URL http://dx.doi.org/10.3847/1538-4357/aa8220.

Cannon, R., Croom, S., Pimbblet, K. and SDSS-2dF Team, "The SDSS-2dF LRG and QSO surveys", *Anglo-Australian Observatory Epping Newsletter*, vol. 103, p. 8, 2003.

Capak, P. L. et al., "A massive protocluster of galaxies at a redshift of z~5.3", Nature, vol. 470, no. 7333, pp. 233–235, 2011. 1101.3586, URL http://dx.doi.org/10.1038/nature09681.

Cardelli, J. A., Clayton, G. C. and Mathis, J. S., "The Relationship between Infrared, Optical, and Ultraviolet Extinction", ApJ, vol. 345, p. 245, 1989. URL http://dx.doi.org/10.1086/167900.

Castignani, G., "Galaxy Clusters Around z 1-2 Low Luminosity Radio Galaxies", in: *Early stages of Galaxy Cluster Formation*, p. 8 (2017). URL http://dx.doi.org/10.5281/zenodo.831235.

Charlor, S. and Bruzual, G., 2007. In press.

Cheng, T. et al., "SCUBA-2 overdensities associated with candidate protoclusters selected from Planck data", MNRAS, vol. 494, no. 4, pp. 5985–5991, 2020. 2004.08560, URL http://dx.doi.org/10.1093/mnras/staa1096.

Chiang, Y.-K., Overzier, R. and Gebhardt, K., "Ancient Light from Young Cosmic Cities: Physical and Observational Signatures of Galaxy Proto-clusters", ApJ, vol. 779, no. 2, 127, 2013. 1310.2938, URL http://dx.doi.org/10.1088/0004-637X/779/2/127.

Chiang, Y.-K., Overzier, R. and Gebhardt, K., "Discovery of a Large Number of Candidate Protoclusters Traced by ~15 Mpc-scale Galaxy Overdensities in COSMOS", ApJ, vol. 782, no. 1, L3, 2014. 1312.4747, URL http://dx.doi.org/10.1088/2041-8205/782/1/L3.

Chiang, Y.-K., Overzier, R. A., Gebhardt, K. and Henriques, B., "Galaxy Protoclusters as Drivers of Cosmic Star Formation History in the First 2 Gyr", ApJ, vol. 844, no. 2, L23, 2017. 1705.01634, URL http://dx.doi.org/10.3847/2041-8213/aa7e7b.

Chiang, Y.-K. et al., "Surveying Galaxy Proto-clusters in Emission: A Large-scale Structure at z = 2.44 and the Outlook for HETDEX", ApJ, vol. 808, no. 1, 37, 2015. 1505.03877, URL http://dx.doi.org/10.1088/0004-637X/808/1/37.

Coleman, G. D., Wu, C. C. and Weedman, D. W., "Colors and magnitudes predicted for high redshift galaxies.", ApJS, vol. 43, pp. 393–416, 1980. URL http://dx.doi.org/10.1086/190674.

Colless, M. et al., "The 2dF Galaxy Redshift Survey: spectra and redshifts", MNRAS, vol. 328, no. 4, pp. 1039–1063, 2001. astro-ph/0106498, URL http://dx.doi.org/10.1046/j.1365-8711.2001.04902.x.

Collister, A. A. and Lahav, O., "ANNz: Estimating Photometric Redshifts Using Artificial Neural Networks", PASP, vol. 116, no. 818, pp. 345–351, 2004. astro-ph/0311058, URL http://dx.doi.org/10.1086/383254.

Cooke, E. A., Hatch, N. A., Muldrew, S. I., Rigby, E. E. and Kurk, J. D., "A z = 2.5 protocluster associated with the radio galaxy MRC 2104-242: star formation and differing mass functions in dense environments", MNRAS, vol. 440, no. 4, pp. 3262–3274, 2014. 1403.4259, URL http://dx.doi.org/10.1093/mnras/stu522.

Cooke, E. A. et al., "A Mature Galaxy Cluster at z=1.58 around the Radio Galaxy 7C1753+6311", ApJ, vol. 816, no. 2, 83, 2016. 1511.05150, URL http://dx.doi.org/10.3847/0004-637X/816/2/83.

Coupon, J. et al., "The galaxy-halo connection from a joint lensing, clustering and abundance analysis in the CFHTLenS/VIPERS field", MNRAS, vol. 449, no. 2, pp. 1352–1379, 2015. 1502.02867, URL http://dx.doi.org/10.1093/mnras/stv276.

Croom, S. M. et al., "The 2dF-SDSS LRG and QSO Survey: the spectroscopic QSO catalogue", MNRAS, vol. 392, no. 1, pp. 19–44, 2009. 0810.4955, URL http://dx.doi.org/10.1111/j.1365-2966.2008.14052.x.

Daddi, E. et al., "Two Bright Submillimeter Galaxies in a z = 4.05 Protocluster in Goods-North, and Accurate Radio-Infrared Photometric Redshifts", ApJ, vol. 694, no. 2, pp. 1517–1538, 2009. 0810.3108, URL http://dx.doi.org/10.1088/0004-637X/694/2/1517.

Danese, L., de Zotti, G. and di Tullio, G., "On velocity dispersions of galaxies in rich clusters.", A&A, vol. 82, pp. 322–327, 1980.

Dannerbauer, H. et al., "An excess of dusty starbursts related to the Spiderweb galaxy", A&A, vol. 570, A55, 2014. 1410.3730, URL http://dx.doi.org/10.1051/0004-6361/201423771.

Dressler, A., "Galaxy morphology in rich clusters: implications for the formation and evolution of galaxies.", ApJ, vol. 236, pp. 351–365, 1980. URL http://dx.doi.org/10.1086/157753.

Eftekharzadeh, S. et al., "Clustering on very small scales from a large sample of confirmed quasar pairs: does quasar clustering track from Mpc to kpc scales?", MNRAS, vol. 468, no. 1, pp. 77–90, 2017. 1702.03491, URL http://dx.doi.org/10.1093/mnras/stx412.

Erben, T. et al., "GaBoDS: The Garching-Bonn Deep Survey. IV. Methods for the image reduction of multi-chip cameras demonstrated on data from the ESO Wide-Field Imager", *Astronomische Nachrichten*, vol. 326, no. 6, pp. 432–464, 2005. astro-ph/0501144, URL http://dx.doi.org/10.1002/asna.200510396.

Farina, E. P., Falomo, R. and Treves, A., "A study of six low-redshift quasar pairs", MNRAS, vol. 415, no. 4, pp. 3163–3167, 2011. 1104.3577, URL http://dx.doi.org/10.1111/j.1365-2966.2011.18931.x.

Finkbeiner, D. P. et al., "Sloan Digital Sky Survey Imaging of Low Galactic Latitude Fields: Technical Summary and Data Release", AJ, vol. 128, no. 5, pp. 2577–2592, 2004. astro-ph/0409700, URL http://dx.doi.org/10.1086/425050.

Fitzpatrick, E. L. and Massa, D., "An Analysis of the Shapes of Ultraviolet Extinction Curves. I. The 2175 Angstrom Bump", ApJ, vol. 307, p. 286, 1986. URL http://dx.doi.org/10.1086/164415.

Fitzpatrick, E. L. and Massa, D., "An Analysis of the Shapes of Ultraviolet Extinction Curves. II. The Far-UV Extinction", ApJ, vol. 328, p. 734, 1988. URL http://dx.doi.org/10.1086/166332.

Fontana, A. et al., "Photometric Redshifts and Selection of High-Redshift Galaxies in the NTT and Hubble Deep Fields", AJ, vol. 120, no. 5, pp. 2206–2219, 2000. astro-ph/0009158, URL http://dx.doi.org/10.1086/316803.

Fukugita, M. et al., "The Sloan Digital Sky Survey Photometric System", AJ, vol. 111, p. 1748, 1996. URL http://dx.doi.org/10.1086/117915.

Galametz, A. et al., "The Mid-infrared Environments of High-redshift Radio Galaxies", ApJ, vol. 749, no. 2, 169, 2012. 1202.4489, URL http://dx.doi.org/10.1088/0004-637X/749/2/169.

Geach, J. E. et al., "The SCUBA-2 Cosmology Legacy Survey: 850 μm maps, catalogues and number counts", MNRAS, vol. 465, no. 2, pp. 1789–1806, 2017. 1607.03904, URL http://dx.doi.org/10.1093/mnras/stw2721.

George, M. R. et al., "Galaxies in X-Ray Groups. I. Robust Membership Assignment and the Impact of Group Environments on Quenching", ApJ, vol. 742, no. 2, 125, 2011. 1109.6040, URL http://dx.doi.org/10.1088/0004-637X/742/2/125.

Gonzalez, A. H. et al., "The Massive and Distant Clusters of WISE Survey. I. Survey Overview and a Catalog of >2000 Galaxy Clusters at z 1", ApJS, vol. 240, no. 2, 33, 2019. 1809.06820, URL http://dx.doi.org/10.3847/1538-4365/aafad2.

Green, P. J. et al., "A Multiwavelength Study of Binary Quasars and Their Environments", ApJ, vol. 743, no. 1, 81, 2011. 1109.1278, URL http://dx.doi.org/10.1088/0004-637X/743/1/81.

Griffith, R. L. and Stern, D., "Morphologies of Radio-, X-ray-, and Mid-infrared-selected Active Galactic Nuclei", AJ, vol. 140, no. 2, pp. 533–545, 2010. 1006.1016, URL http://dx.doi.org/10.1088/0004-6256/140/2/533.

Gu, Y., Fang, G., Yuan, Q. and Lu, S., "Structures, Stellar Population Properties, AGN Fractions, and Environments of Massive Compact Galaxies at 1 < z < 2 in 3D-HST/CANDELS", PASP, vol. 132, no. 1011, 054101, 2020. 2002.11274, URL http://dx.doi.org/10.1088/1538-3873/ab797d.

Gunn, J. E. and Gott, I., J. Richard, "On the Infall of Matter Into Clusters of Galaxies and Some Effects on Their Evolution", ApJ, vol. 176, p. 1, 1972. URL http://dx.doi.org/10.1086/151605.

Halverson, N. W. et al., "Degree Angular Scale Interferometer First Results: A Measurement of the Cosmic Microwave Background Angular Power Spectrum", ApJ, vol. 568, no. 1, pp. 38–45, 2002. astro-ph/0104489, URL http://dx.doi.org/10.1086/338879.

Hatch, N. A. et al., "Galaxy protocluster candidates around z 2.4 radio galaxies", MNRAS, vol. 410, no. 3, pp. 1537–1549, 2011a. 1008.4588, URL http://dx.doi.org/10.1111/j.1365-2966.2010.17538.x.

Hatch, N. A. et al., "Hα emitters in z 2 protoclusters: evidence for faster evolution in dense environments", MNRAS, vol. 415, no. 4, pp. 2993–3005, 2011b. 1103.4364, URL http://dx.doi.org/10.1111/j.1365-2966.2011.18735.x.

Hatch, N. A. et al., "Why z > 1 radio-loud galaxies are commonly located in protoclusters", MNRAS, vol. 445, no. 1, pp. 280–289, 2014. 1409.1218, URL http://dx.doi.org/10.1093/mnras/stu1725.

Hayashi, M., Kodama, T., Tadaki, K.-i., Koyama, Y. and Tanaka, I., "A Starbursting Protocluster in Making Associated with a Radio Galaxy at z = 2.53 Discovered by Hα Imaging", ApJ, vol. 757, no. 1, 15, 2012. 1207.2614, URL http://dx.doi.org/10.1088/0004-637X/757/1/15.

Henriques, B. M. B. et al., "Galaxy formation in the Planck cosmology - I. Matching the observed evolution of star formation rates, colours and stellar masses", MNRAS, vol. 451, no. 3, pp. 2663–2680, 2015. 1410.0365, URL http://dx.doi.org/10.1093/mnras/stv705.

Higuchi, R. et al., "SILVERRUSH. VII. Subaru/HSC Identifications of Protocluster Candidates at z ∼ 67: Implications for Cosmic Reionization", ApJ, vol. 879, no. 1, 28, 2019a. 1801.00531, URL http://dx.doi.org/10.3847/1538-4357/ab2192.

Higuchi, R. et al., "SILVERRUSH. VII. Subaru/HSC Identifications of Protocluster Candidates at z 6-7: Implications for Cosmic Reionization", ApJ, vol. 879, no. 1, 28, 2019b. 1801.00531, URL http://dx.doi.org/10.3847/1538-4357/ab2192.

Hilton, M. et al., "The Atacama Cosmology Telescope: The Two-season ACTPol Sunyaev-Zel'dovich Effect Selected Cluster Catalog", ApJS, vol. 235, no. 1, 20, 2018. 1709.05600, URL http://dx.doi.org/10.3847/1538-4365/aaa6cb.

Hoecker, A. et al., "TMVA - Toolkit for Multivariate Data Analysis", *arXiv e-prints*, physics/0703039, 2007. physics/0703039.

Huang, N. et al., "Galaxy Clusters Selected via the SunyaevZel'dovich Effect in the SPTpol 100-square-degree Survey", AJ, vol. 159, no. 3, 110, 2020. 1907.09621, URL http://dx.doi.org/10.3847/1538-3881/ab6a96.

Hyndman, R., Zhang, X. and King, M., "A Bayesian Approach to Bandwidth Selection for Multivariate Kernel Density Estimation", *Computational Statistics Data Analysis*, vol. 50, pp. 3009–3031, 2006. URL http://dx.doi.org/10.1016/j.csda.2005.06.019.

Ilbert, O. et al., "Accurate photometric redshifts for the CFHT legacy survey calibrated using the VIMOS VLT deep survey", A&A, vol. 457, no. 3, pp. 841–856, 2006. astro-ph/0603217, URL http://dx.doi.org/10.1051/0004-6361:20065138.

Ishigaki, M., Ouchi, M. and Harikane, Y., "A Very Compact Dense Galaxy Overdensity with δ 130 Identified at z 8: Implications for Early Protocluster and Cluster Core Formation", ApJ, vol. 822, no. 1, 5, 2016. 1509.01751, URL http://dx.doi.org/10.3847/0004-637X/822/1/5.

Ivezić, Ž., Connelly, A. J., Vand erPlas, J. T. and Gray, A., *Statistics, Data Mining, and Machine Learning in Astronomy* (2014).

Iye, M., "Subaru studies of the cosmic dawn", *Proceeding of the Japan Academy, Series B*, vol. 87, no. 9, pp. 575–586, 2011. 1110.0195, URL http://dx.doi.org/10.2183/pjab.87.575.

Jester, S. et al., "The Sloan Digital Sky Survey View of the Palomar-Green Bright Quasar Survey", AJ, vol. 130, no. 3, pp. 873–895, 2005. astro-ph/0506022, URL http://dx.doi.org/10.1086/432466.

Jian, H.-Y. et al., "Probability Friends-of-Friends (PFOF) Group Finder: Performance Study and Observational Data Applications on Photometric Surveys", ApJ, vol. 788, no. 2, 109, 2014. 1305.1891, URL http://dx.doi.org/10.1088/0004-637X/788/2/109.

Jian, H.-Y. et al., "Redshift Evolution of Green Valley Galaxies in Different Environments from the Hyper Suprime-Cam Survey", ApJ, vol. 894, no. 2, 125, 2020. 2005.04894, URL http://dx.doi.org/10.3847/1538-4357/ab86a8.

Jones, B. J. T., "A Brief History of Cosmology", in: *From Quantum Fluctuations to Cosmological Structures*, edited by D. Valls-Gabaud, M. A. Hendry, P. Molaro and K. Chamcham, *Astronomical Society of the Pacific Conference Series*, vol. 126, p. 1 (1997).

Jones, D. H. et al., "The 6dF Galaxy Survey: samples, observational techniques and the first data release", MNRAS, vol. 355, no. 3, pp. 747–763, 2004. astro-ph/0403501, URL http://dx.doi.org/10.1111/j.1365-2966.2004.08353.x.

Joshi, G. D. et al., "The fate of disc galaxies in IllustrisTNG clusters", MNRAS, 2020. URL http://dx.doi.org/10.1093/mnras/staa1668.

Kaiser, N., "On the spatial correlations of Abell clusters.", ApJ, vol. 284, pp. L9–L12, 1984. URL http://dx.doi.org/10.1086/184341.

Keisler, R. et al., "A Measurement of the Damping Tail of the Cosmic Microwave Background Power Spectrum with the South Pole Telescope", ApJ, vol. 743, no. 1, 28, 2011. 1105.3182, URL http://dx.doi.org/10.1088/0004-637X/743/1/28.

Kodama, T., Arimoto, N., Barger, A. J. and Arag'on-Salamanca, A., "Evolution of the colour-magnitude relation of early-type galaxies in distant clusters", A&A, vol. 334, pp. 99–109, 1998. astro-ph/9802245.

Koutsouridou, I. and Cattaneo, A., "Bursting and quenching in satellite galaxies", MNRAS, vol. 490, no. 4, pp. 5375–5389, 2019. 1911.01395, URL http://dx.doi.org/10.1093/mnras/stz2916.

Kravtsov, A. V. and Borgani, S., "Formation of Galaxy Clusters", ARA&A, vol. 50, pp. 353–409, 2012. 1205.5556, URL http://dx.doi.org/10.1146/annurev-astro-081811-125502.

Krefting, N. et al., "The Role of Environment in Galaxy Evolution in the SERVS Survey. I. Density Maps and Cluster Candidates", ApJ, vol. 889, no. 2, 185, 2020. 1912.02238, URL http://dx.doi.org/10.3847/1538-4357/ab60a0.

Kuiper, E., Venemans, B. P., Hatch, N. A., Miley, G. K. and Röttgering, H. J. A., "A z $\sim$ 3 radio galaxy and its protocluster: evidence for a superstructure?", MNRAS, vol. 425, no. 2, pp. 801–813, 2012. 1203.2196, URL http://dx.doi.org/10.1111/j.1365-2966.2012.20800.x.

Kurki-Suonio, H., "Physics of the Cosmic Microwave Background and the Planck Mission", *arXiv e-prints*, arXiv:1012.5204, 2010. 1012.5204.

Laigle, C. et al., "The COSMOS2015 Catalog: Exploring the 1 < z < 6 Universe with Half a Million Galaxies", ApJS, vol. 224, no. 2, 24, 2016. 1604.02350, URL http://dx.doi.org/10.3847/0067-0049/224/2/24.

Lang, D., "unWISE: Unblurred Coadds of the WISE Imaging", AJ, vol. 147, no. 5, 108, 2014. 1405.0308, URL http://dx.doi.org/10.1088/0004-6256/147/5/108.

Liu, D. et al., "Automated Mining of the ALMA Archive in the COSMOS Field (A^{3}COSMOS). II. Cold Molecular Gas Evolution out to Redshift 6", ApJ, vol. 887, no. 2, 235, 2019. 1910.12883, URL http://dx.doi.org/10.3847/1538-4357/ab578d.

Lonsdale, C. J. et al., "SWIRE: The SIRTF Wide-Area Infrared Extragalactic Survey", PASP, vol. 115, no. 810, pp. 897–927, 2003. astro-ph/0305375, URL http://dx.doi.org/10.1086/376850.

Lu, S., Cappellari, M., Mao, S., Ge, J. and Li, R., "SDSS-IV MaNGA: stellar population correlates with stellar root-mean-square velocity V_{rms} gradients or total-density-profile slopes at fixed effective velocity dispersion σ_e", MNRAS, vol. 495, no. 4, pp. 4820–4827, 2020. 2003.13625, URL http://dx.doi.org/10.1093/mnras/staa1481.

MacKenzie, T. P. et al., "VizieR Online Data Catalog: Proto-cluster candidate fields (Mackenzie+, 2017)", *VizieR Online Data Catalog*, J/MNRAS/468/4006, 2020.

Madau, P. et al., "High-redshift galaxies in the Hubble Deep Field: colour selection and star formation history to z~4", MNRAS, vol. 283, no. 4, pp. 1388–1404, 1996. astro-ph/9607172, URL http://dx.doi.org/10.1093/mnras/283.4.1388.

Maraston, C., "Evolutionary population synthesis: models, analysis of the ingredients and application to high-z galaxies", MNRAS, vol. 362, no. 3, pp. 799–825, 2005. astro-ph/0410207, URL http://dx.doi.org/10.1111/j.1365-2966.2005.09270.x.

Martinache, C. et al., "Spitzer Planck Herschel Infrared Cluster (SPHerIC) survey: Candidate galaxy clusters at 1.3 < z < 3 selected by high star-formation rate", A&A, vol. 620, A198, 2018. 1810.07330, URL http://dx.doi.org/10.1051/0004-6361/201833198.

Matthews, T. A., Morgan, W. W. and Schmidt, M., "a Discussion of Galaxies Identified with Radio Sources", in: *Quasi-Stellar Sources and Gravitational Collapse*, edited by I. Robinson, A. Schild and E. L. Schucking, p. 105 (1965).

Mayo, J. H. et al., "Overdensities of 24 μm sources in the vicinities of high-redshift radio galaxies", A&A, vol. 539, A33, 2012. 1202.4490, URL http://dx.doi.org/10.1051/0004-6361/201118254.

Mehrtens, N. et al., "The XMM Cluster Survey: optical analysis methodology and the first data release", MNRAS, vol. 423, no. 2, pp. 1024–1052, 2012. 1106.3056, URL http://dx.doi.org/10.1111/j.1365-2966.2012.20931.x.

Mei, S. et al., "Evolution of the Color-Magnitude Relation in High-Redshift Clusters: Blue Early-Type Galaxies and Red Pairs in RDCS J0910+5422", ApJ, vol. 639, no. 1, pp. 81–94, 2006. astro-ph/0601327, URL http://dx.doi.org/10.1086/499259.

Meisner, A. M., Lang, D. and Schlegel, D. J., "Deep Full-sky Coadds from Three Years of WISE and NEOWISE Observations", AJ, vol. 154, no. 4, 161, 2017a. 1705.06746, URL http://dx.doi.org/10.3847/1538-3881/aa894e.

Meisner, A. M., Lang, D. and Schlegel, D. J., "Full-depth Coadds of the WISE and First-year NEOWISE-reactivation Images", AJ, vol. 153, no. 1, 38, 2017b. 1603.05664, URL http://dx.doi.org/10.3847/1538-3881/153/1/38.

Meisner, A. M., Lang, D. and Schlegel, D. J., "Time-resolved WISE/NEOWISE Coadds", AJ, vol. 156, no. 2, 69, 2018. 1710.02526, URL http://dx.doi.org/10.3847/1538-3881/aacbcd.

Molino, A. et al., "CLASH: accurate photometric redshifts with 14 HST bands in massive galaxy cluster cores", MNRAS, vol. 470, no. 1, pp. 95–113, 2017. 1705.02265, URL http://dx.doi.org/10.1093/mnras/stx1243.

Moore, B., Katz, N., Lake, G., Dressler, A. and Oemler, A., "Galaxy harassment and the evolution of clusters of galaxies", Nature, vol. 379, no. 6566, pp. 613–616, 1996. astro-ph/9510034, URL http://dx.doi.org/10.1038/379613a0.

Mörtsell, E., "Calibrating Milky Way dust extinction using cosmological sources", A&A, vol. 550, A80, 2013. 1210.2191, URL http://dx.doi.org/10.1051/0004-6361/201220587.

Mörtsell, E., "Cosmological histories from the Friedmann equation: the Universe as a particle", *European Journal of Physics*, vol. 37, no. 5, p. 055603, 2016. 1606.09556, URL http://dx.doi.org/10.1088/0143-0807/37/5/055603.

Mullis, C. R. et al., "Discovery of an X-Ray-luminous Galaxy Cluster at z=1.4", ApJ, vol. 623, no. 2, pp. L85–L88, 2005. astro-ph/0503004, URL http://dx.doi.org/10.1086/429801.

Noirot, G. et al., "HST Grism Confirmation of Two z 2 Structures from the Clusters around Radio-loud AGN (CARLA) Survey", ApJ, vol. 830, no. 2, 90, 2016. 1609.04162, URL http://dx.doi.org/10.3847/0004-637X/830/2/90.

Noirot, G. et al., "The Carla Survey: Insights From The Densest Carla Structures At 1.4 < Z < 2.8.", in: *Galaxy Evolution Across Time*, p. 72 (2017). URL http://dx.doi.org/10.5281/zenodo.809356.

Onoue, M. et al., "Enhancement of galaxy overdensity around quasar pairs at z < 3.6 based on the Hyper Suprime-Cam Subaru Strategic Program Survey", PASJ, vol. 70, S31, 2018. 1704.06051, URL http://dx.doi.org/10.1093/pasj/psx092.

Overzier, R. A., "The realm of the galaxy protoclusters. A review", A&A Rev., vol. 24, no. 1, 14, 2016. 1610.05201, URL http://dx.doi.org/10.1007/s00159-016-0100-3.

Overzier, R. A. et al., "Clustering of Star-forming Galaxies Near a Radio Galaxy at z=5.2", ApJ, vol. 637, no. 1, pp. 58–73, 2006. astro-ph/0509308, URL http://dx.doi.org/10.1086/498234.

Overzier, R. A. et al., "Lyman Break Galaxies, Lyα Emitters, and a Radio Galaxy in a Protocluster at z = 4.1", ApJ, vol. 673, no. 1, pp. 143–162, 2008. astro-ph/0601223, URL http://dx.doi.org/10.1086/524342.

Oyaizu, H., Lima, M., Cunha, C. E., Lin, H. and Frieman, J., "Photometric Redshift Error Estimators", ApJ, vol. 689, no. 2, pp. 709–720, 2008. 0711.0962, URL http://dx.doi.org/10.1086/592591.

Pentericci, L. et al., "A search for clusters at high redshift. II. A proto cluster around a radio galaxy at z=2.16", A&A, vol. 361, pp. L25–L28, 2000. astro-ph/0008143.

Perlmutter, S. et al., "Measurements of Ω and Λ from 42 High-Redshift Supernovae", ApJ, vol. 517, no. 2, pp. 565–586, 1999. astro-ph/9812133, URL http://dx.doi.org/10.1086/307221.

Pier, J. R. et al., "Astrometric Calibration of the Sloan Digital Sky Survey", AJ, vol. 125, no. 3, pp. 1559–1579, 2003. astro-ph/0211375, URL http://dx.doi.org/10.1086/346138.

Planck Collaboration et al., "Planck 2013 results. XVI. Cosmological parameters", A&A, vol. 571, A16, 2014. 1303.5076, URL http://dx.doi.org/10.1051/0004-6361/201321591.

Planck Collaboration et al., "Planck intermediate results. XXVII. High-redshift infrared galaxy overdensity candidates and lensed sources discovered by Planck and confirmed by

Herschel-SPIRE", A&A, vol. 582, A30, 2015. 1503.08773, URL http://dx.doi.org/10.1051/0004-6361/201424790.

Planck Collaboration et al., "Planck intermediate results. XXXIX. The Planck list of high-redshift source candidates", A&A, vol. 596, A100, 2016. 1508.04171, URL http://dx.doi.org/10.1051/0004-6361/201527206.

Riess, A. G. et al., "Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant", AJ, vol. 116, no. 3, pp. 1009–1038, 1998. astro-ph/9805201, URL http://dx.doi.org/10.1086/300499.

Rigby, E. E. et al., "Searching for large-scale structures around high-redshift radio galaxies with Herschel", MNRAS, vol. 437, no. 2, pp. 1882–1893, 2014. 1310.5710, URL http://dx.doi.org/10.1093/mnras/stt2019.

Rosati, P., Borgani, S. and Norman, C., "The Evolution of X-ray Clusters of Galaxies", ARA&A, vol. 40, pp. 539–577, 2002. astro-ph/0209035, URL http://dx.doi.org/10.1146/annurev.astro.40.120401.150547.

Sadeh, I., Abdalla, F. B. and Lahav, O., "ANNz2: Photometric Redshift and Probability Distribution Function Estimation using Machine Learning", PASP, vol. 128, no. 968, p. 104502, 2016. 1507.00490, URL http://dx.doi.org/10.1088/1538-3873/128/968/104502.

Sako, M. et al., "The Sloan Digital Sky Survey-II Supernova Survey: Search Algorithm and Follow-up Observations", AJ, vol. 135, no. 1, pp. 348–373, 2008. 0708.2750, URL http://dx.doi.org/10.1088/0004-6256/135/1/348.

Salvato, M., Ilbert, O. and Hoyle, B., "The many flavours of photometric redshifts", *Nature Astronomy*, vol. 3, pp. 212–222, 2019. 1805.12574, URL http://dx.doi.org/10.1038/s41550-018-0478-0.

Sánchez, S. F. and González-Serrano, J. I., "Excesses of faint galaxies around seven radio QSOs at 1.0<z<1.6. Clustering at high-z", A&A, vol. 352, pp. 383–394, 1999. astro-ph/9910485.

Sánchez, S. F. and González-Serrano, J. I., "Cluster of galaxies around seven radio-loud QSOs at 1<z<1.6. II. K-band images", A&A, vol. 396, pp. 773–786, 2002. astro-ph/0209355, URL http://dx.doi.org/10.1051/0004-6361:20021455.

Sandrinelli, A., Falomo, R., Treves, A., Farina, E. P. and Uslenghi, M., "The environment of low-redshift quasar pairs", MNRAS, vol. 444, no. 2, pp. 1835–1841, 2014. 1407.7875, URL http://dx.doi.org/10.1093/mnras/stu1526.

Schirmer, M., "THELI: Convenient Reduction of Optical, Near-infrared, and Mid-infrared Imaging Data", ApJS, vol. 209, no. 2, 21, 2013. 1308.4989, URL http://dx.doi.org/10.1088/0067-0049/209/2/21.

Schlafly, E. F., Meisner, A. M. and Green, G. M., "The unWISE Catalog: Two Billion Infrared Sources from Five Years of WISE Imaging", ApJS, vol. 240, no. 2, 30, 2019. 1901.03337, URL http://dx.doi.org/10.3847/1538-4365/aafbea.

Schlafly, E. F. et al., "The DECam Plane Survey: Optical Photometry of Two Billion Objects in the Southern Galactic Plane", ApJS, vol. 234, no. 2, 39, 2018. 1710.01309, URL http://dx.doi.org/10.3847/1538-4365/aaa3e2.

Schlegel, D. J., Finkbeiner, D. P. and Davis, M., "Maps of Dust Infrared Emission for Use in Estimation of Reddening and Cosmic Microwave Background Radiation Foregrounds", ApJ, vol. 500, no. 2, pp. 525–553, 1998. astro-ph/9710327, URL http://dx.doi.org/10.1086/305772.

Schneider, D. P. and SDSS Collaboration, "The SDSS Quasar Survey", in: *American Astronomical Society Meeting Abstracts #198*, *American Astronomical Society Meeting Abstracts*, vol. 198, p. 78.01 (2001).

Schneider, P., *Extragalactic Astronomy and Cosmology* (2006).

Seymour, N., Stern, D. and De Breuck, C., "Spitzer Observations of High Redshift Radio Galaxies", in: *Deepest Astronomical Surveys*, edited by J. Afonso, H. C. Ferguson, B. Mobasher and R. Norris, *Astronomical Society of the Pacific Conference Series*, vol. 380, p. 393 (2007). astro-ph/0604226.

Shen, Y. et al., "A Catalog of Quasar Properties from Sloan Digital Sky Survey Data Release 7", ApJS, vol. 194, no. 2, 45, 2011. 1006.5178, URL http://dx.doi.org/10.1088/0067-0049/194/2/45.

Shi, K., Toshikawa, J., Cai, Z., Lee, K.-S. and Fang, T., "A detailed study of massive galaxies in a protocluster at z=3.13", *arXiv e-prints*, arXiv:2007.07055, 2020. 2007.07055.

Smoot, G. F. et al., "Structure in the COBE Differential Microwave Radiometer First-Year Maps", ApJ, vol. 396, p. L1, 1992. URL http://dx.doi.org/10.1086/186504.

Springel, V. et al., "Simulations of the formation, evolution and clustering of galaxies and quasars", Nature, vol. 435, no. 7042, pp. 629–636, 2005. astro-ph/0504097, URL http://dx.doi.org/10.1038/nature03597.

Stanford, S. A. et al., "The XMM Cluster Survey: A Massive Galaxy Cluster at z = 1.45", ApJ, vol. 646, no. 1, pp. L13–L16, 2006. astro-ph/0606075, URL http://dx.doi.org/10.1086/506449.

Stott, J. P. et al., "Quasar Sightline and Galaxy Evolution (QSAGE) survey – II. Galaxy overdensities around UV luminous quasars at z=1-2", *arXiv e-prints*, arXiv:2006.07384, 2020. 2006.07384.

Sullivan, M. et al., "SNLS3: Constraints on Dark Energy Combining the Supernova Legacy Survey Three-year Data with Other Probes", ApJ, vol. 737, no. 2, 102, 2011. 1104.1444, URL http://dx.doi.org/10.1088/0004-637X/737/2/102.

Tanaka, I. et al., "Discovery of an Excess of Hα Emitters around 4C 23.56 at z = 2.48", PASJ, vol. 63, p. 415, 2011. 1012.1869, URL http://dx.doi.org/10.1093/pasj/63.sp2.S415.

Tanaka, M., "Photometric Redshift with Bayesian Priors on Physical Properties of Galaxies", ApJ, vol. 801, no. 1, 20, 2015. 1501.02047, URL http://dx.doi.org/10.1088/0004-637X/801/1/20.

Taylor, M. B., *TOPCAT & STIL: Starlink Table/VOTable Processing Software*, *Astronomical Society of the Pacific Conference Series*, vol. 347, p. 29 (2005).

Tiley, A. L. et al., "K-CLASH: spatially-resolving star-forming galaxies in field and cluster environments at z 0.2-0.6", MNRAS, 2020. 2005.12471, URL http://dx.doi.org/10.1093/mnras/staa1418.

Toshikawa, J. et al., "A Systematic Survey of Protoclusters at z ~3-6 in the CFHTLS Deep Fields", ApJ, vol. 826, no. 2, 114, 2016. 1605.01439, URL http://dx.doi.org/10.3847/0004-637X/826/2/114.

Toshikawa, J. et al., "GOLDRUSH. III. A systematic search for protoclusters at z 4 based on the >100 deg^2 area", PASJ, vol. 70, S12, 2018. 1708.09421, URL http://dx.doi.org/10.1093/pasj/psx102.

Trussler, J. et al., "The weak imprint of environment on the stellar populations of galaxies", *arXiv e-prints*, arXiv:2006.01154, 2020. 2006.01154.

van den Bosch, F. C. et al., "The importance of satellite quenching for the build-up of the red sequence of present-day galaxies", MNRAS, vol. 387, no. 1, pp. 79–91, 2008. 0710.3164, URL http://dx.doi.org/10.1111/j.1365-2966.2008.13230.x.

van der Burg, R. F. J. et al., "The GOGREEN Survey: A deep stellar mass function of cluster galaxies at 1.0<z<1.4 and the complex nature of satellite quenching", *arXiv e-prints*, arXiv:2004.10757, 2020. 2004.10757.

van Dokkum, P. G., "The Recent and Continuing Assembly of Field Elliptical Galaxies by Red Mergers", AJ, vol. 130, no. 6, pp. 2647–2665, 2005. astro-ph/0506661, URL http://dx.doi.org/10.1086/497593.

Venemans, B. P. et al., "The Most Distant Structure of Galaxies Known: A Protocluster at z=4.1", ApJ, vol. 569, no. 1, pp. L11–L14, 2002. astro-ph/0203249, URL http://dx.doi.org/10.1086/340563.

Venemans, B. P. et al., "Protoclusters associated with z > 2 radio galaxies . I. Characteristics of high redshift protoclusters", A&A, vol. 461, no. 3, pp. 823–845, 2007. astro-ph/0610567, URL http://dx.doi.org/10.1051/0004-6361:20053941.

Véron-Cetty, M. P. and Véron, P., "A catalogue of quasars and active nuclei: 13th edition", A&A, vol. 518, A10, 2010. URL http://dx.doi.org/10.1051/0004-6361/201014188.

Visvanathan, N. and Sandage, A., "The color - absolute magnitude relation for E and S0 galaxies. I. Calibration and tests for universality using Virgo and eight other nearby clusters.", ApJ, vol. 216, pp. 214–226, 1977. URL http://dx.doi.org/10.1086/155464.

Wen, Z. L. and Han, J. L., "Galaxy Clusters at High Redshift and Evolution of Brightest Cluster Galaxies", ApJ, vol. 734, no. 1, 68, 2011. 1104.1667, URL http://dx.doi.org/10.1088/0004-637X/734/1/68.

Wheelock, S. L. et al., "IRAS sky survey atlas: Explanatory supplement", NASA STI/Recon Technical Report N, 1994.

White, S. D. M. and Frenk, C. S., "Galaxy Formation through Hierarchical Clustering", ApJ, vol. 379, p. 52, 1991. URL http://dx.doi.org/10.1086/170483.

White, S. D. M. and Rees, M. J., "Core condensation in heavy halos: a two-stage theory for galaxy formation and clustering.", MNRAS, vol. 183, pp. 341–358, 1978. URL http://dx.doi.org/10.1093/mnras/183.3.341.

Wylezalek, D. et al., "Galaxy Clusters around Radio-loud Active Galactic Nuclei at 1.3 < z < 3.2 as Seen by Spitzer", ApJ, vol. 769, no. 1, 79, 2013. 1304.0770, URL http://dx.doi.org/10.1088/0004-637X/769/1/79.

Yang, X. et al., "Galaxy Groups in the SDSS DR4. I. The Catalog and Basic Properties", ApJ, vol. 671, no. 1, pp. 153–170, 2007. 0707.4640, URL http://dx.doi.org/10.1086/522027.

Zel'dovich, Y. B. and Syunyaev, R. A., "Shock Wave Structure in the Radiation Spectrum During Bose Condensation of Photons", *Soviet Journal of Experimental and Theoretical Physics*, vol. 35, p. 81, 1972.

A QSO available spectra

Here we present SDSS spectra available for our sample of quasars (Figures A.1, A.2, and A.3). There is one case in field S5, and two in S6, for which we do not have this information. The data were downloaded from the SDSS website[1].

[1] http://skyserver.sdss.org/dr15/en/tools/chart/list.aspx

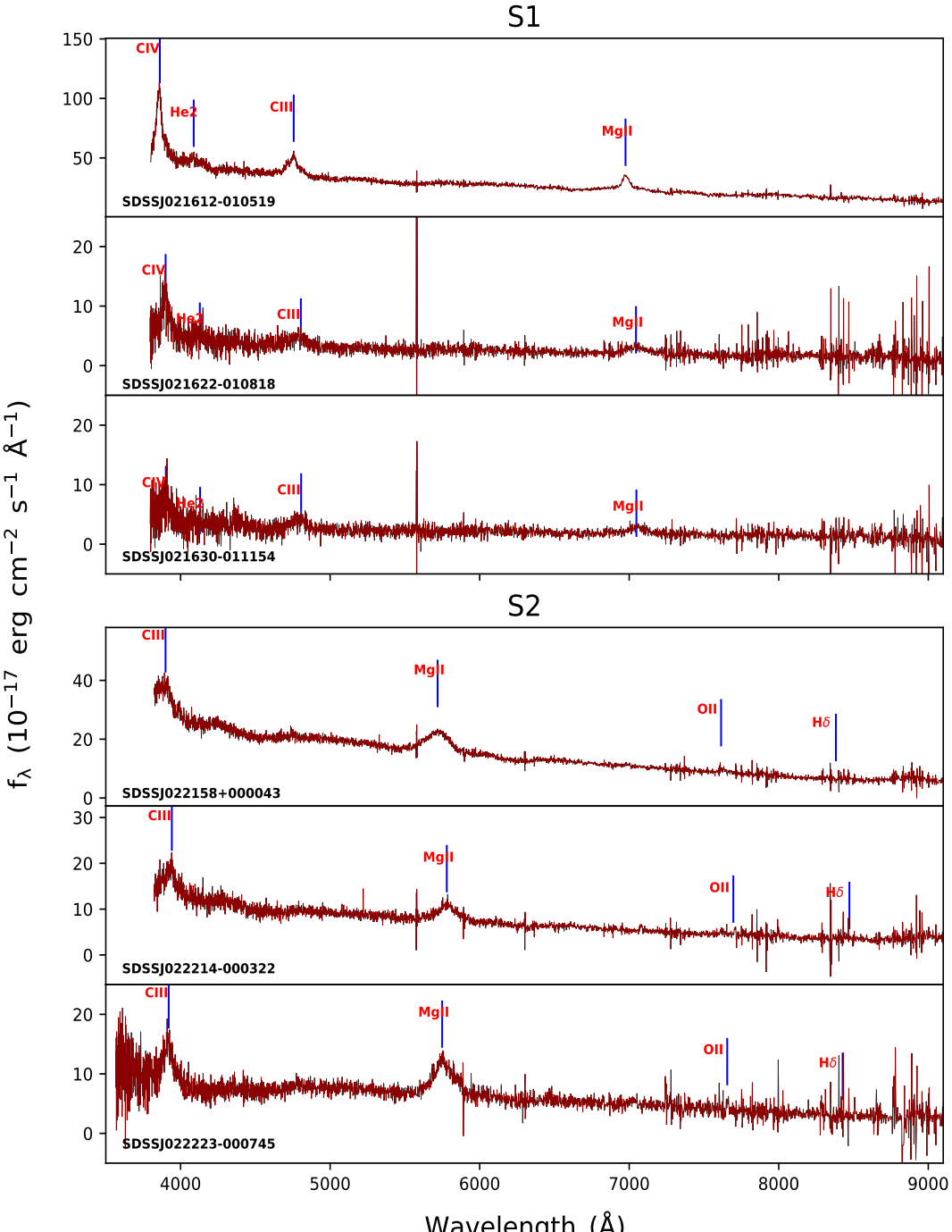

Figure A.1: Available spectra of the QSOs.

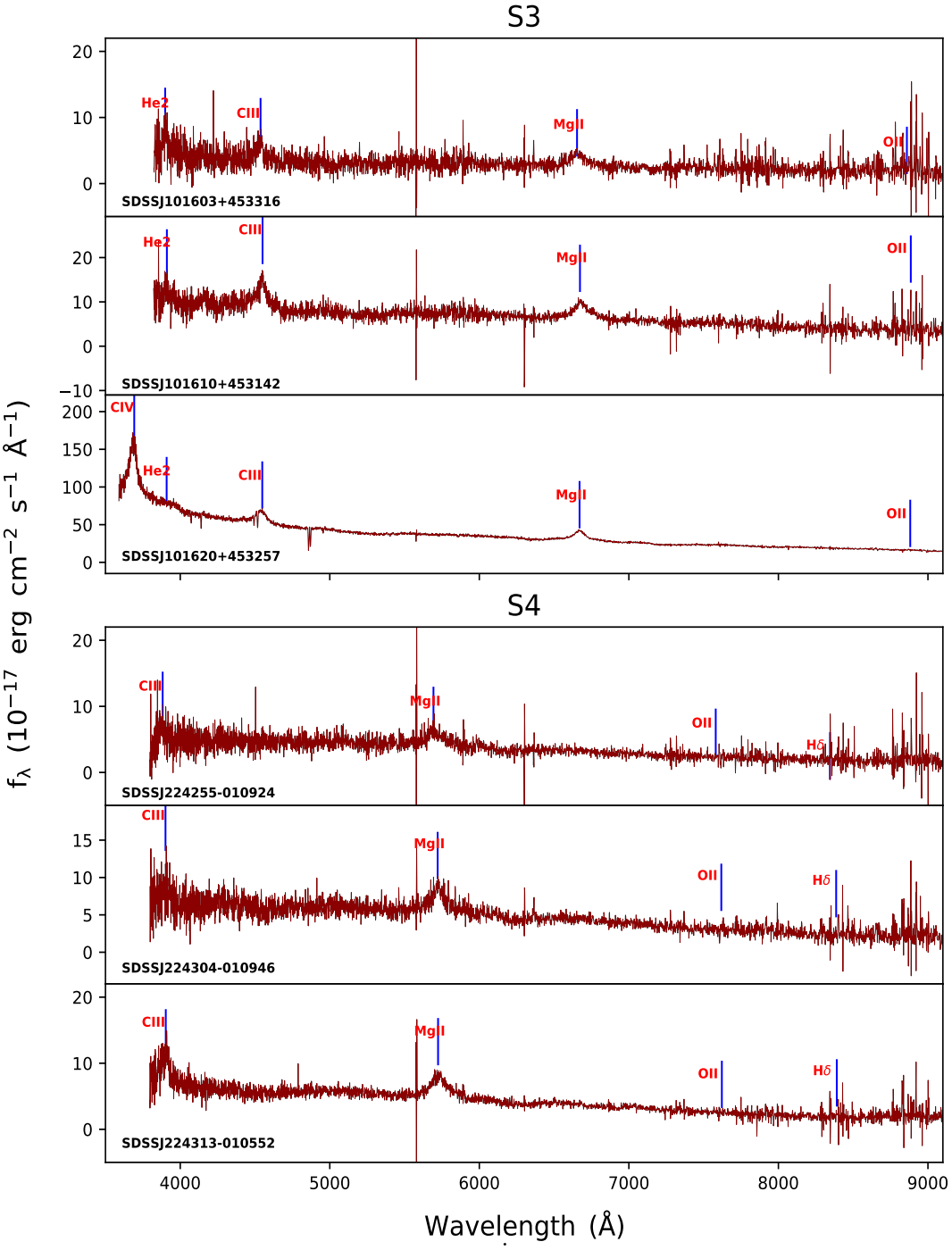

Figure A.2: Continuation of Figure A.1

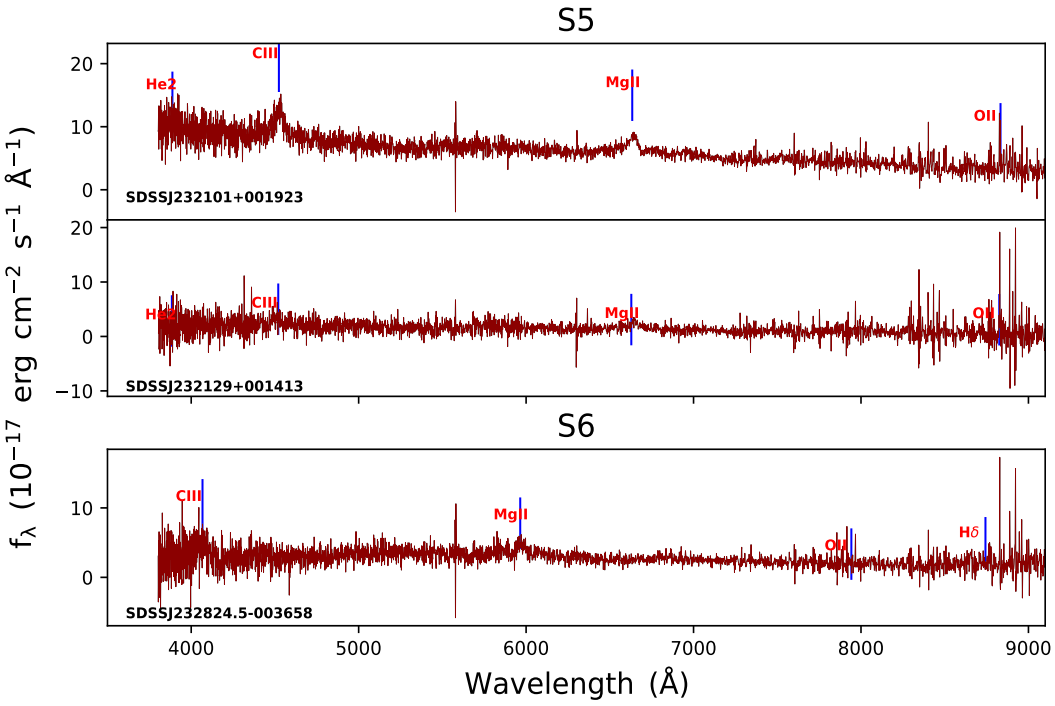

Figure A.3: Continuation of Figure A.2

B Scripts for SExtractor

Here we show an example of the three scripts that we ran in order to detect objects in the image, extract their measurements, and organize them into catalogues, with SExtractor.

```
# execute program
module load SExtractor/2.19.5/gcc
module load plplot/5.9.7

cd /sto/home/laerte/protocluster/marcelo

sex UM416Ih.fits -c pass1.sex -CATALOG_NAME UM416_i.cat -CATALOG_TYPE FITS_LDAC
-PARAMETERS_NAME pass1.param -DETECT_MINAREA 10 -DETECT_THRESH 1.5 -ANALYSIS_THR
ESH 1.5 -FILTER_NAME gauss_4.0_7x7.conv  -WEIGHT_TYPE MAP_WEIGHT -WEIGHT_IMAGE U
M416I.weight.fits -PHOT_APERTURES 5 -MAG_ZEROPOINT 25.718 -GAIN 3013.9 -PIXEL_SC
ALE 0.2 -SEEING_FWHM 0.63 -BACKPHOTO_TYPE LOCAL -BACKPHOTO_THICK 24

sex UM416R.fits -c pass1.sex -CATALOG_NAME UM416_r.cat -CATALOG_TYPE FITS_LDAC -
PARAMETERS_NAME pass1.param -DETECT_MINAREA 10 -DETECT_THRESH 1.5 -ANALYSIS_THRE
SH 1.5 -FILTER_NAME gauss_4.0_7x7.conv  -WEIGHT_TYPE MAP_WEIGHT -WEIGHT_IMAGE UM
416R.weight.fits -PHOT_APERTURES 6 -MAG_ZEROPOINT 25.887  -SEEING_FWHM 0.66 -BAC
KPHOTO_TYPE LOCAL -BACKPHOTO_THICK 24

sex UM416Z.fits -c pass1.sex -CATALOG_NAME UM416_z.cat -CATALOG_TYPE FITS_LDAC -
PARAMETERS_NAME pass1.param -DETECT_MINAREA 10 -DETECT_THRESH 1.5 -ANALYSIS_THRE
SH 1.5 -FILTER_NAME gauss_4.0_7x7.conv  -WEIGHT_TYPE MAP_WEIGHT -WEIGHT_IMAGE UM
416Z.weight.fits -PHOT_APERTURES 8 -MAG_ZEROPOINT 24.679 -GAIN 4861.18 -PIXEL_SC
ALE 0.2 -SEEING_FWHM 0.62 -BACKPHOTO_TYPE LOCAL -BACKPHOTO_THICK 24
```

Figure B.1: The script to obtain the FITS_ LDAC binary format file containing all the identified objects with the parameters information

```
# execute program
#module load SExtractor/2.19.5/gcc
#module load plplot/5.9.7
#module load psfex/3.17.1

#cd /sto/home/laerte/protocluster/marcelo

# psfex for i-filter
psfex UM416_i.cat -c psfconf.c -BASIS_TYPE PIXEL_AUTO -PSF_SIZE 45,45 -SAMPLE_FWH
HMRANGE 3.0,6.5 -SAMPLE_VARIABILITY 0.05 -SAMPLE_MINSN 250 -SAMPLE_MAXELLIP 0.05
 -OUTCAT_TYPE FITS_LDAC -OUTCAT_NAME psfexout_um416_i.cat

# psfex for z-filter
psfex UM416_r.cat -c psfconf.c -BASIS_TYPE PIXEL_AUTO -PSF_SIZE 45,45 -SAMPLE_FWH
HMRANGE 3.0,6.5 -SAMPLE_VARIABILITY 0.05 -SAMPLE_MINSN 250 -SAMPLE_MAXELLIP 0.05
 -OUTCAT_TYPE FITS_LDAC -OUTCAT_NAME psfexout_um416_r.cat
#
# # psfex for r-filter
psfex UM416_z.cat -c psfconf.c -BASIS_TYPE PIXEL_AUTO -PSF_SIZE 45,45 -SAMPLE_FWH
HMRANGE 3.0,6.5 -SAMPLE_VARIABILITY 0.05 -SAMPLE_MINSN 250 -SAMPLE_MAXELLIP 0.05
 -OUTCAT_TYPE FITS_LDAC -OUTCAT_NAME psfexout_um416_z.cat
```

.

Figure B.2: The script for PSFex.

```
# execute program
module load SExtractor/2.19.5
module load plplot/5.9.7

cd /sto/home/laerte/protocluster/marcelo

sex UM416Ih.fits -c sexconf.conf -CATALOG_NAME UM416_i_ascii.cat -CATALOG_TYPE A
SCII_HEAD -PARAMETERS_NAME sexpari.param -DETECT_MINAREA 10 -DETECT_THRESH 1.5 -
ANALYSIS_THRESH 1.5 -FILTER_NAME gauss_4.0_7x7.conv  -WEIGHT_TYPE MAP_WEIGHT -WE
IGHT_IMAGE UM416I.weight.fits -MAG_ZEROPOINT 25.718 -GAIN 3013.9 -PIXEL_SCALE 0.
2 -SEEING_FWHM 0.63 -PHOT_APERTURES 5 -BACKPHOTO_TYPE LOCAL -BACKPHOTO_THICK 24
-PSF_NAME UM416_i.psf -PATTERN_TYPE GAUSS-LAGUERRE

sex UM416Ih.fits,UM416R.fits -c sexconf.conf -CATALOG_NAME UM416_ri_ascii.cat -C
ATALOG_TYPE ASCII_HEAD -PARAMETERS_NAME sexparr.param -DETECT_MINAREA 10 -DETECT
_THRESH 1.5 -ANALYSIS_THRESH 1.5 -FILTER_NAME gauss_4.0_7x7.conv  -WEIGHT_TYPE M
AP_WEIGHT -WEIGHT_IMAGE UM416I.weight.fits,UM416R.weight.fits -PHOT_APERTURES 6
-MAG_ZEROPOINT 25.887 -GAIN 2430.63 -PIXEL_SCALE 0.2 -SEEING_FWHM 0.66 -BACKPHOT
O_TYPE LOCAL -BACKPHOTO_THICK 24 -PSF_NAME UM416_r.psf -PATTERN_TYPE GAUSS-LAGUE
RRE -ASSOC_NAME UM416_i_ascii.cat -ASSOC_DATA 0 -ASSOC_PARAMS 2,3,4 -ASSOCCOORD_
TYPE PIXEL -ASSOC_RADIUS 2.0 -ASSOC_TYPE NEAREST -ASSOCSELEC_TYPE MATCHED

sex UM416Ih.fits,UM416Z.fits -c sexconf.conf -CATALOG_NAME UM416_zi_ascii.cat -C
ATALOG_TYPE ASCII_HEAD -PARAMETERS_NAME sexparz.param -DETECT_MINAREA 10 -DETECT
_THRESH 1.5 -ANALYSIS_THRESH 1.5 -FILTER_NAME gauss_4.0_7x7.conv  -WEIGHT_TYPE M
AP_WEIGHT -WEIGHT_IMAGE UM416I.weight.fits,UM416Z.weight.fits -PHOT_APERTURES 9
-MAG_ZEROPOINT 24.679 -GAIN 4861.18 -PIXEL_SCALE 0.2 -SEEING_FWHM 0.62 -BACKPHOT
O_TYPE LOCAL -BACKPHOTO_THICK 24 -PSF_NAME UM416_z.psf -PATTERN_TYPE GAUSS-LAGUE
RRE -ASSOC_NAME UM416_i_ascii.cat -ASSOC_DATA 0 -ASSOC_PARAMS 2,3,4 -ASSOCCOORD_
TYPE PIXEL -ASSOC_RADIUS 2.0 -ASSOC_TYPE NEAREST -ASSOCSELEC_TYPE MATCHED
```

.

Figure B.3: The final script that will generate the catalogues with all detected objects with the measurements.

C Star/galaxy separation

In this Appendix, we present the star/galaxy separation plots from the S2 to the S6 fields. The criteria are the same described at Figure 2.5.

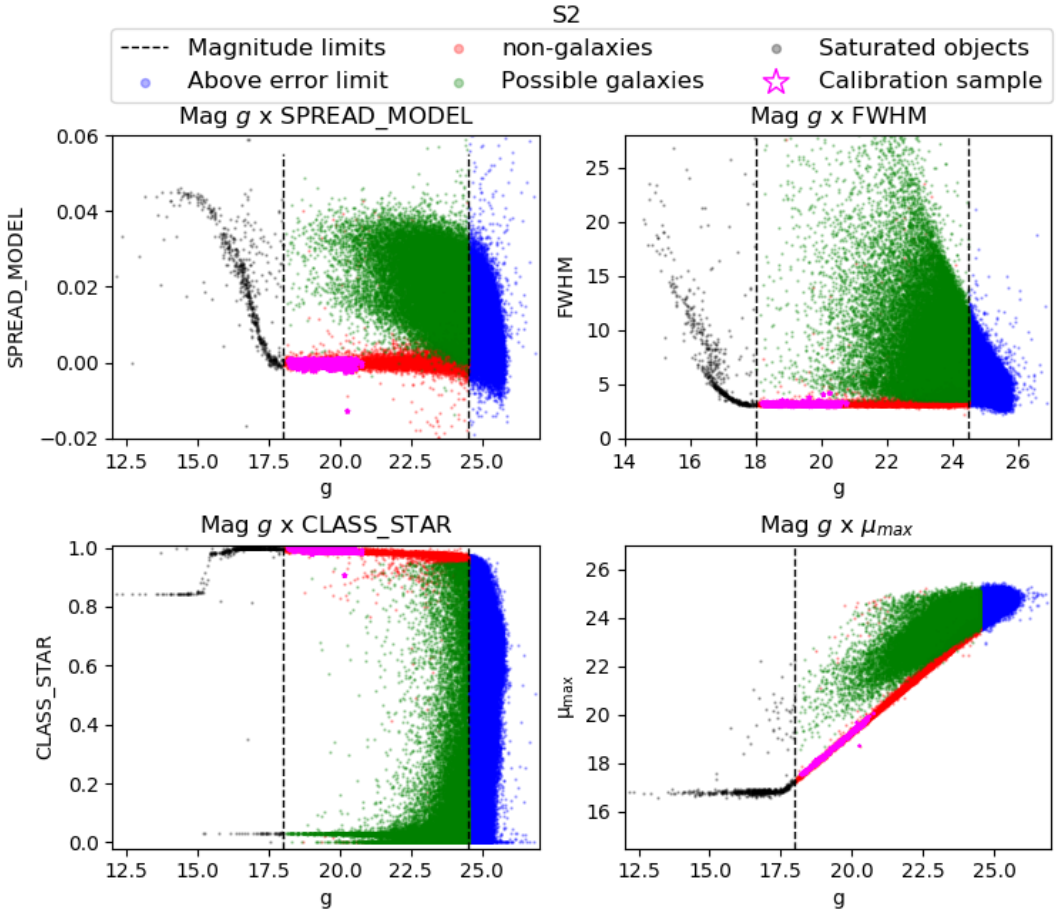

Figure C.1: Star/galaxy separation for S2.

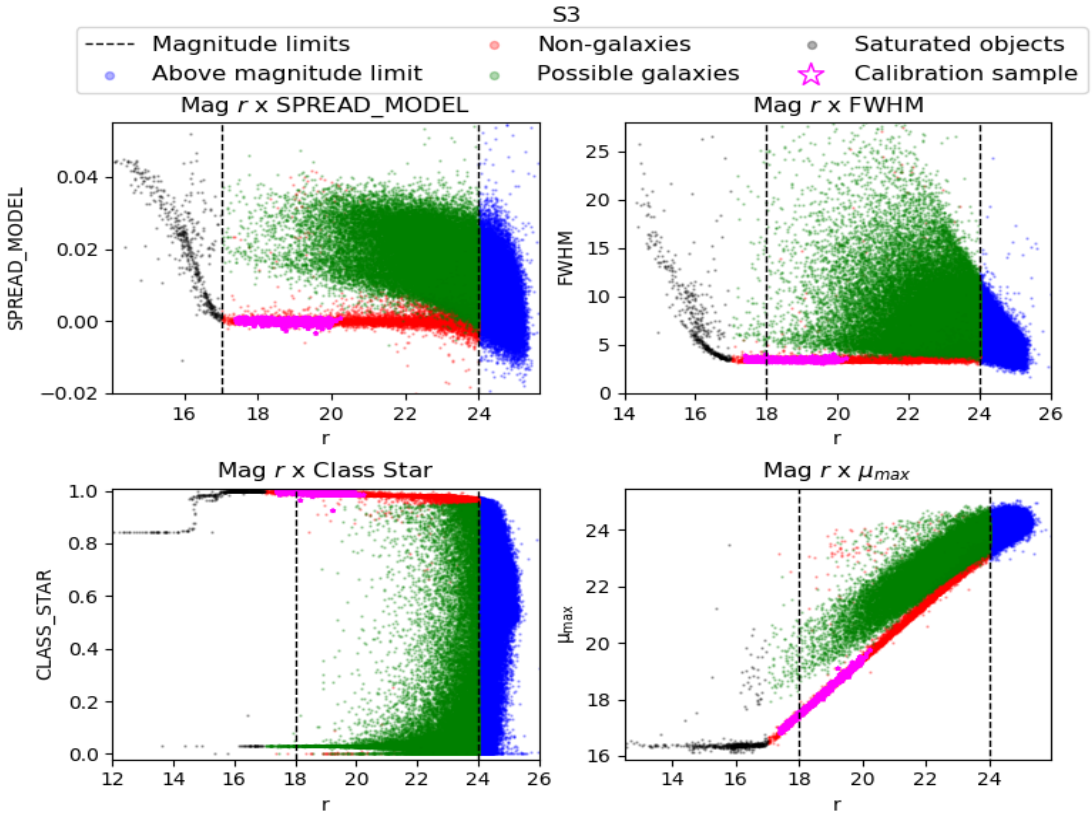

Figure C.2: Star/galaxy separation for S3.

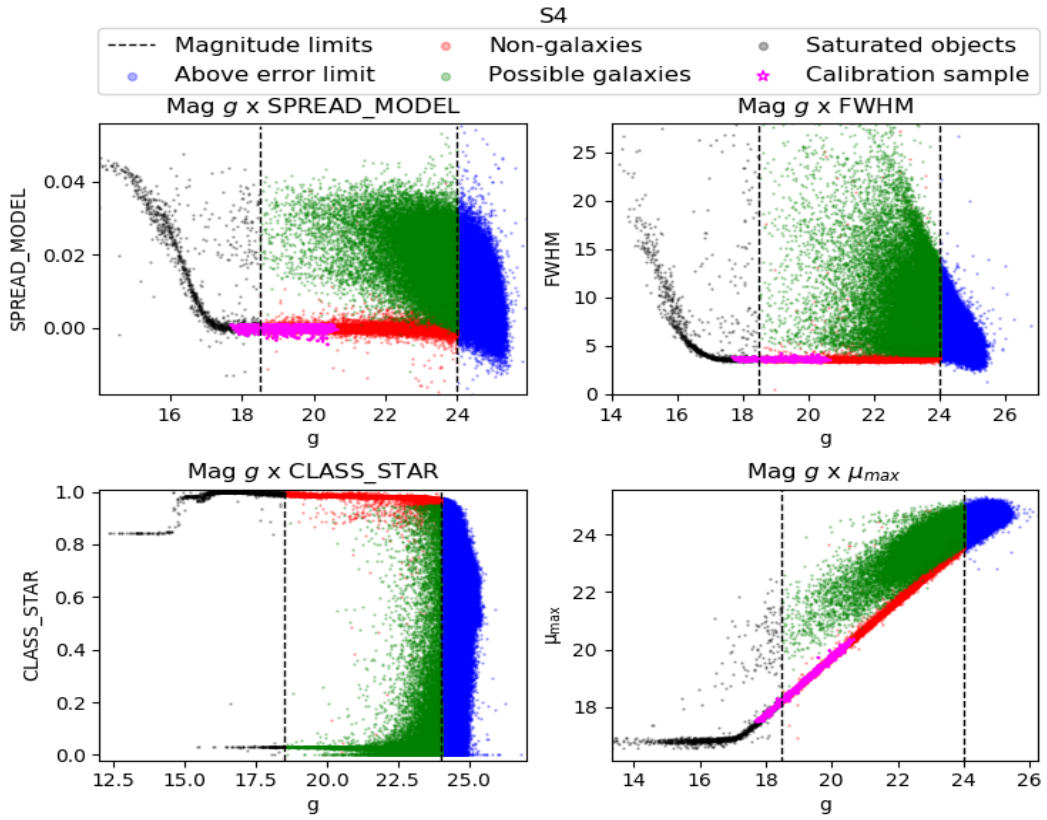

Figure C.3: Star/galaxy separation for S4.

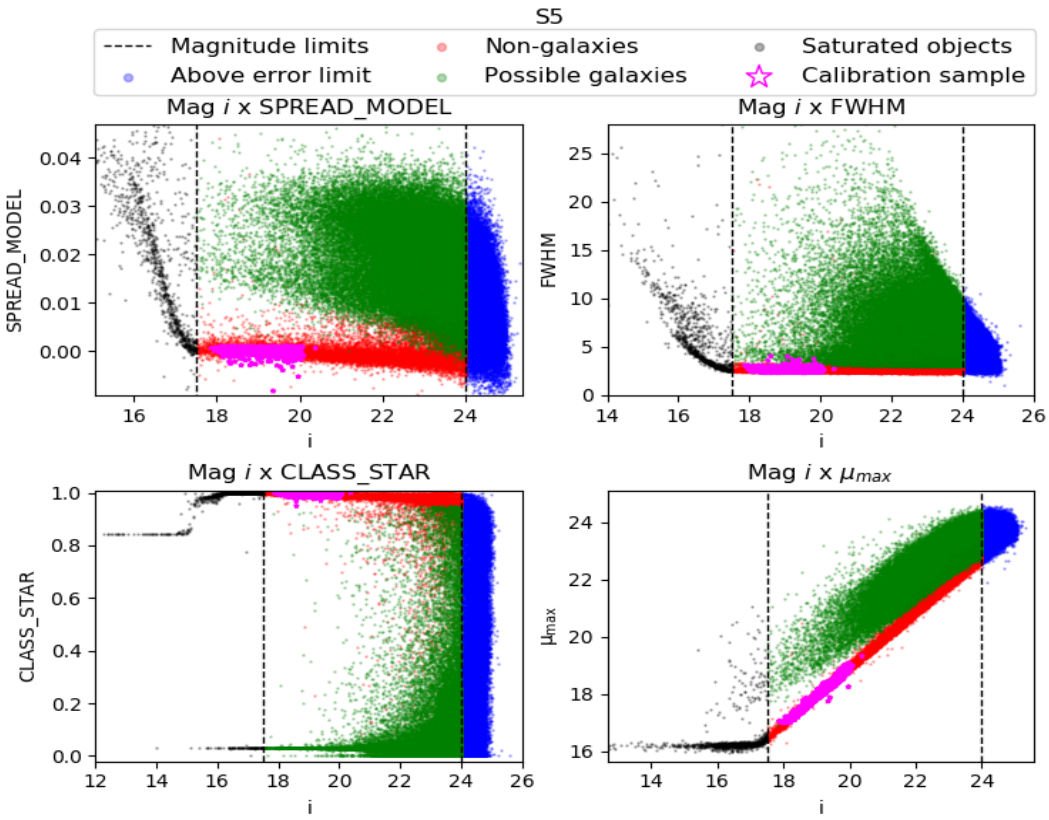

Figure C.4: Star/galaxy separation for S5.

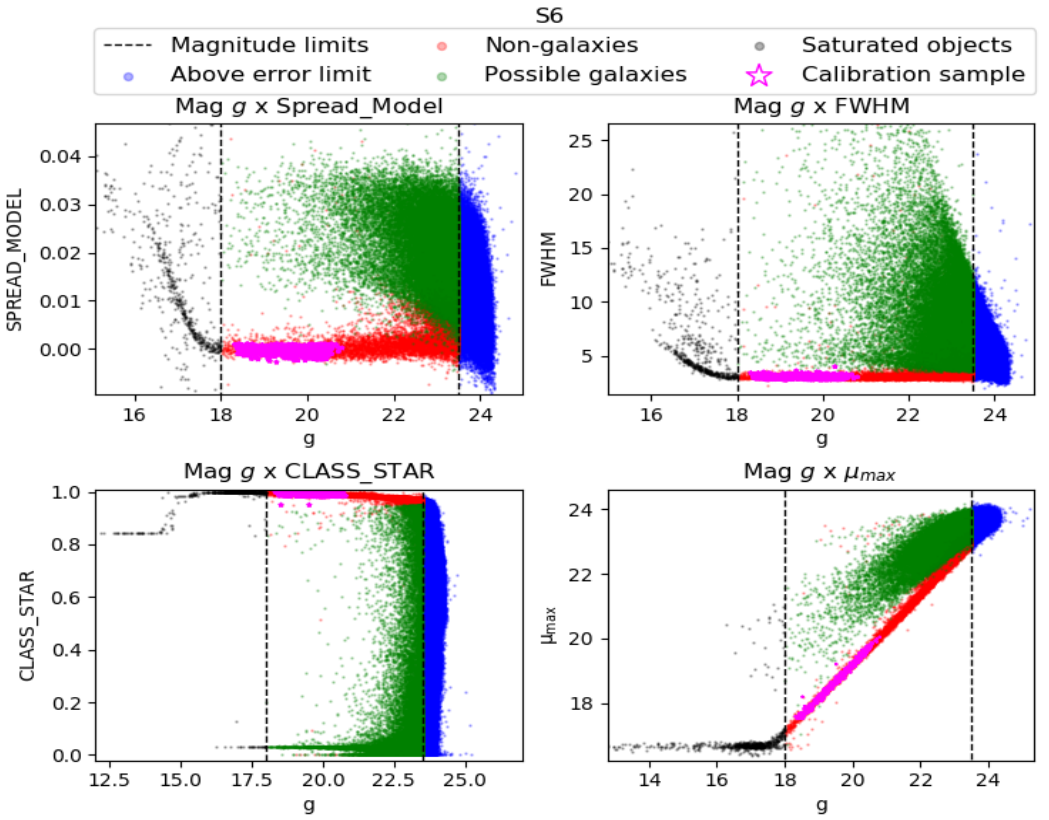

Figure C.5: Star/galaxy separation for S6.

D Galaxy counts

In Figure D.1, we show the galaxy counts distribution from the S2 to the S6 fields. Each row of plots represents one field and the columns are the different photometric bands. The same magnitude limits adopted for each field were applied to the corresponding COSMOS sample.

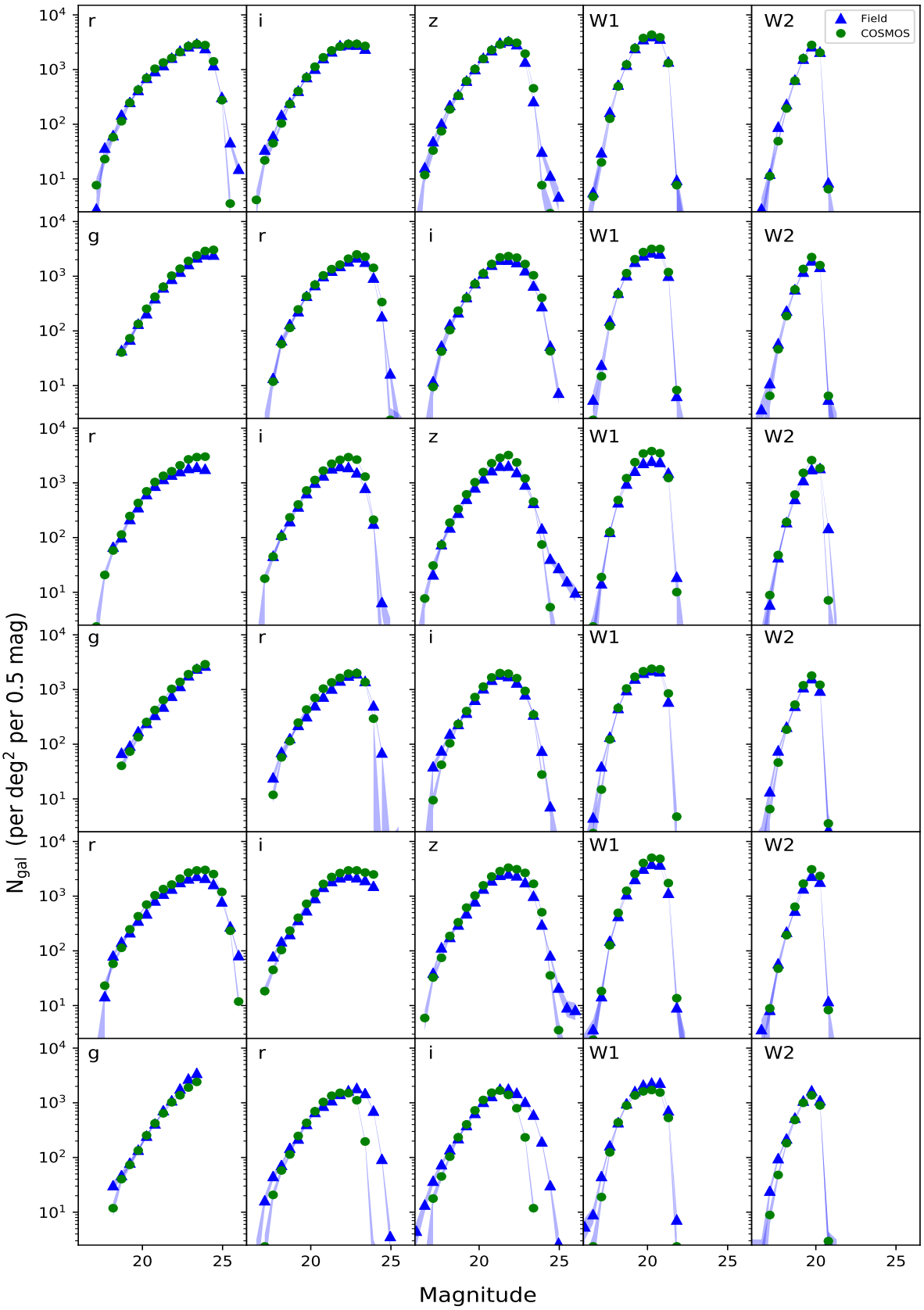

Figure D.1: Galaxy counts for S1, S2, S3, S4, S5, and S6 fields, together with the COSMOS sample with the same magnitude limits (each row, respectively).