

Master thesis

---

---

# Clustering of Lyman-alpha emitters in MUSE surveys

---

---

Carried out at

Leibniz-Institut für Astrophysik Potsdam

Submitted to

Universität Potsdam

Faculty of Mathematics and Natural Sciences

Institute for Physics and Astronomy



By:

**Yohana Herrero Alonso**

Matrikelnummer: 790096

yherreroalonso@aip.de

Supervisor: Prof. Dr. Lutz Wisotzki

APRIL 2019

## Acknowledgements

I would like to thank each single person who made this thesis possible.

In the first place, I want to say thanks to my supervisor **Prof. Dr. Lutz Wisotzki**, who gave me the opportunity of being part of his group and offered me this topic as a possible master thesis work. Thanks to this great chance, I got to know how research groups and scientific conferences work (the MUSE Busy Week was really awesome). As most group leaders and professors, he is a very busy person, but every time I knocked on his door to ask any question he always had a moment, something for which I am very grateful.

In the second place, I would like to thank **Dr. Mirko Krumpe** for the long discussions hold along all these months of work. Right since the beginning, when I barely had any idea on what I was working, he taught me so much about clustering, gave me very helpful feedback, advices, and cleared out all my doubts. Thanks also for letting me know on time that clustering measurements have nothing to do with detecting clusters of galaxies.

In the third place, I would like to thank **Prof. Dr. Christoph Pfrommer**, who agreed to read my master thesis despite having his own students, teaching courses, conferences, etc. Thanks also to **Dr. Tanya Urrutia** who provided me with all necessary MUSE data I asked for and introduced me to the use of MUSE web page. I also really appreciate the email conversations with **Dr. Catrina Diener**, who rapidly replied to my emails the first months when I was trying to reproduce her results and with **Dr. Takamitsu Miyaji**, who let me use his codes and gave me access to his computer in Mexico.

Finally, I would like to thank my family. Special thanks to my mum **Marisol**, who I used to keep on the phone for hours while I was complaining about a line of code on which I was stuck for days. Thanks so much for the moral support every time I felt a bit down or missed being at home with all of you.

# Table of Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Large Scale Structure of the Universe . . . . .	2
1.2	Galaxy Clustering . . . . .	5
1.2.1	Angular clustering method . . . . .	7
1.2.2	Spatial clustering method . . . . .	9
1.3	Lyman Alpha Emitters . . . . .	12
1.4	Detection techniques . . . . .	15
1.4.1	Narrow band selection technique . . . . .	15
1.4.2	Spectrographs . . . . .	16
1.5	MUSE spectrograph . . . . .	18
1.6	Key aims of this work . . . . .	20
<b>2</b>	<b>MUSE samples</b>	<b>22</b>
2.1	MUSE-Wide survey . . . . .	22
2.2	MUSE-Deep survey . . . . .	27
<b>3</b>	<b>Redshift clustering method</b>	<b>29</b>
3.1	MUSE-Wide 24 fields . . . . .	33
3.2	MUSE-Wide 69 fields . . . . .	34
3.3	MUSE-Deep . . . . .	35
3.4	Studying the $K$ estimator in detail . . . . .	36

<i>TABLE OF CONTENTS</i>	iii
<b>4 Analysis of a random sample of galaxies</b>	<b>41</b>
<b>5 Clustering dependence on physical properties</b>	<b>47</b>
<b>6 Spatial clustering method</b>	<b>56</b>
<b>7 Conclusion</b>	<b>66</b>
7.1 Summary . . . . .	66
7.2 Future work . . . . .	68
<b>Bibliography</b>	<b>71</b>
<b>Deutsche Zusammenfassung</b>	<b>80</b>
<b>Eidesstattliche Erklärung</b>	<b>83</b>
<b>APPENDIX</b>	<b>84</b>

# Chapter 1

## Introduction

Galaxies are not randomly distributed in the Universe, they form a "Cosmic Web" that shows a specific structure. This Web was formed when gravitational instabilities produced by primordial density fluctuations grew until they reached a critical density that made them collapse, forming gravitationally bound dark matter halos. These halos, which play an important role in the formation of galaxies, grow hierarchically through accretion or mergers with other halos. Their gravitational interaction with baryonic matter causes gas to fall in, making the gas cool by radiation and collapse to form stars and the first galaxies ( $z \sim 20$ ).

Galaxies thus form as cores inlaid in extensive dark matter halos. The study of galaxy clustering within a single halo provides information on the occupation distribution of the galaxies within that halo. From this halo occupation distribution, gas dynamics, cooling processes, star formation and feedback can be traced. Therefore, an unique opportunity to track the physics behind the galaxy formation is provided, constraining and improving galaxy formation and evolution theories.

Galaxy clustering is also one of the most important diagnostics of observational cosmology, analyzing the clustering of galaxies that belong to different dark matter halos constrains the dark matter halo population and the typical dark matter halo masses. These halo populations tell about the initial conditions of energy density in the Universe and the matter content across the Universe, which make galaxy clustering one of the most significant observational constrains for cosmological simulations, an extraordinary way of testing the cosmological models with observations.

Approaching even larger scales ( $\sim 100 h^{-1}\text{Mpc}$ ), galaxy clustering studies detect an excess which shows evidence for the initial fluctuations in the energy density of the early

Universe. These baryonic acoustic oscillations (BAO) provide length scales and thus a better understanding about the nature of dark matter and the acceleration of the Universe. BAO give the sufficient information to calculate the Hubble and angular diameter distances that illustrate the dark energy behaviour.

The aim of this thesis is to explore how to analyze the spatial distribution of the samples of Lyman Alpha Emitters (LAE) detected by the Multi Unit Spectroscopic Explorer (MUSE) at the Very Large Telescope (VLT). Different methods to study the galaxy clustering will be used and discussed, a random sample of LAEs that mimic our real sample will be created and the galaxy evolution of the samples will be explored by seeking clustering signals depending on distinct galactic properties.

## 1.1 Large Scale Structure of the Universe

The Universe is not homogeneous at small scales ( $< 40$  Mpc) and its inhomogeneity at scales larger than the galactic sizes is known as large scale structure of the Universe. Galaxies are not uniformly distributed, they are allocated in groups or clusters of sizes of few Mpc. Following a well defined structure, they are generally formed along thin filaments and group together into galaxy clusters at the nodes (intersections of filaments of galaxies). However, most of the volume of the universe is occupied by voids, almost empty regions where just a few galaxies lie on. Therefore, the galaxy density in voids is much lower than in filaments or nodes and, comparing it to the baryonic density of a galaxy, the difference is even higher.

It was already in 1934 when Hubble used the Shapley-Ames catalog ([Shapley and Ames, 1932](#)) of bright galaxies to conclude that there was an excess in the number of galaxies compared to the expectation from a random Poisson distribution on angular scales of less than  $\sim 10^\circ$  ([Hubble, 1934](#)). 43 years later, [Seldner et al. \(1977\)](#) published maps showing clearly the inhomogeneous distribution of galaxies across the sky and, it was Peebles the one who analyzed statistically the data. In this way, once the spectra of the galaxies were taken, the first large scale redshift survey was settled in this same year.

From 1978, mapping the three dimensional spatial distribution of galaxies started to become possible and new volumes of the sky commenced to be studied. By studying bigger and separate volumes, it became likely to observe, the so called at present, filaments, nodes and voids. Afterwards, with the improvement of fiber-optics and multi-slit spectrographs, redshift surveys were steeply enhanced. Nowadays, the Sloan Digital Sky Survey I (SDSS) with more than a million redshifts is the most extended survey at low redshifts ( $\bar{z} \approx 0.1$ ),

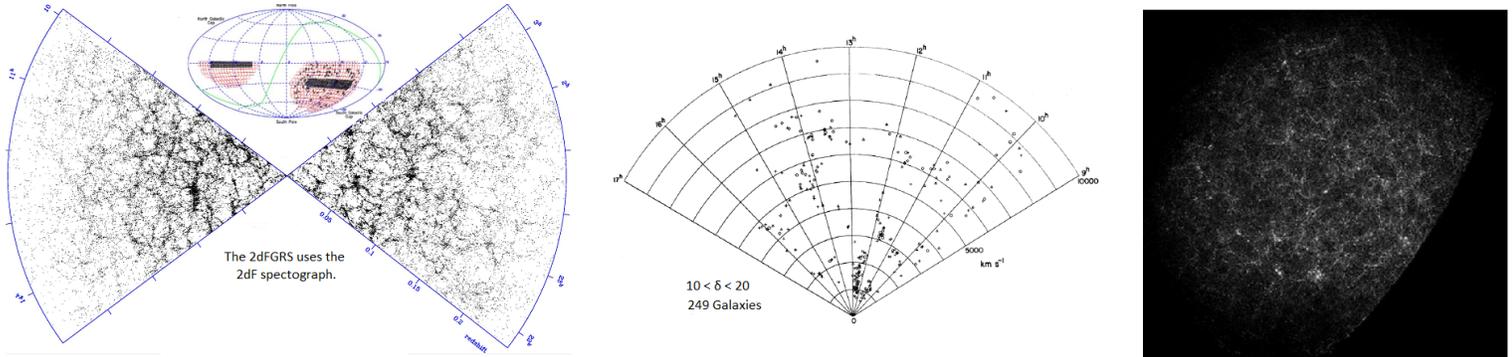


Figure 1.1: Left: Spatial distribution of galaxies as function of redshift and right ascension from the 2dF Survey (Colless et al., 2001). Center: Same as in the left panel but as function of velocity, right ascension and within  $10^\circ < \delta < 20^\circ$  from the CfA survey (Davis and Peebles, 1983). Right: Distribution of galaxies brighter than  $B \approx 19$  in cells of  $10' \times 10'$  in the northern hemisphere from the Seldner catalog (Seldner et al., 1977).

whereas DEEP2 Redshift Survey or the VIMOS-VLT Deep Survey rule at higher redshifts,  $\bar{z} \approx 1$  and  $\bar{z} \approx 2$ , respectively (see Figure 1.1).

Already in the right panel of Figure 1.1 (Seldner et al., 1977), it is possible to appreciate the nonuniform distribution of galaxies on the sky. With the actual catalogs (left panel, Colless et al. (2001)), it is even likely to perceive voids (few tens of Mpc in size), clusters of galaxies and the filamentous web structure of the Universe.

Nowadays, it is even possible to obtain the large scale structure of the Universe that we observe by running simulations. One of the best examples is the Millennium simulation, a N-body run, that is able to show the matter distribution in the Universe as a function of time or redshift (see Figure 1.2 where the difference between theory and observations is negligible). Particularly, it analyzes how the galaxies were formed compared to observations and provides an exceptional tool to compare both observations and predictions. Such is the case, that it can perfectly reproduce the location of the galaxies with tiny differences from the CfA and the 2dF surveys.

Analyzing this large scale structure is essential to have a more accurate understanding of the Universe. The structures seen at present are actually ancient fossils from the conditions of the early Universe. Hence, the history of galaxy and structure evolution and formation is recorded in those galaxies that form the visible structure at present.

The measurements of the large scale structure is also a way to probe conditions before the epoch of recombination, providing information on quantum mechanical effects occurred

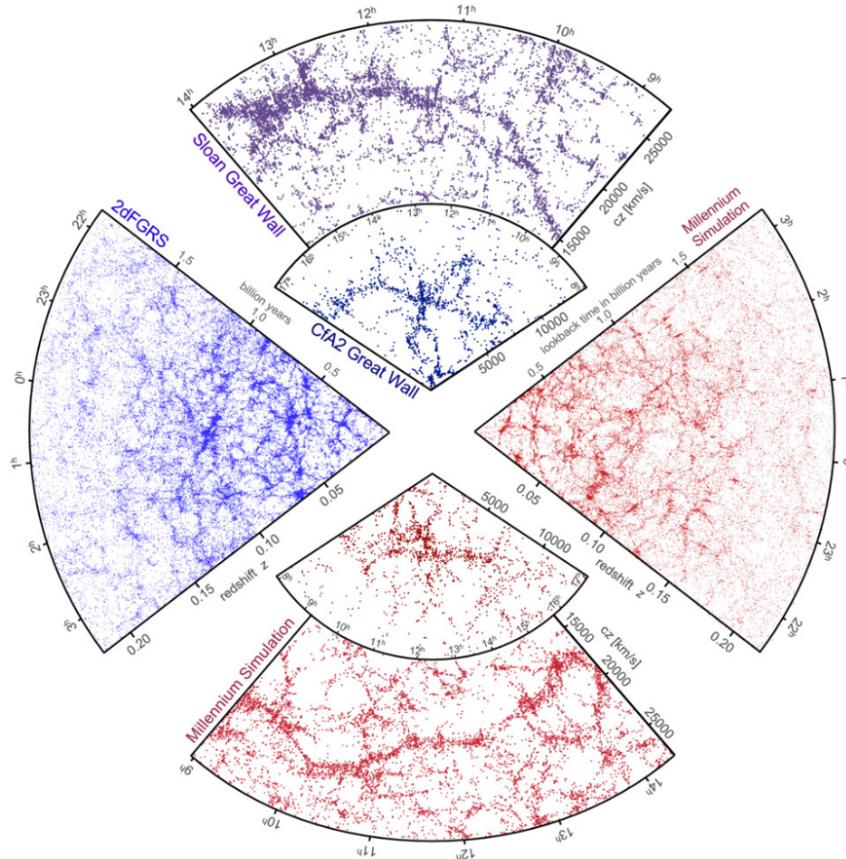


Figure 1.2: From left to right and clockwise: the 2dFGRS, SDSS and CfA2 surveys in bluish colors and the theoretical Millennium galaxy distribution predictions for the same surveys in reddish colors from [Lemson et al. \(2006\)](#).

at that time. As explained by inflationary models, quantum fluctuations expanded into the largest structures that we witness today, explaining the temperature deviations observed in the cosmic microwave background and the anisotropic clustering of galaxies in the Universe. As an example, these fluctuations, which are also called Barionic Acoustic Oscillations (BAO), were detected by [Cole et al. \(2005\)](#) in the 2dFGRS survey by analyzing the power spectrum employing a direct Fourier method. In a different way, BAO were also detected by [Eisenstein et al. \(2005\)](#) in SDSS while studying the spatial clustering of galaxies. The excess in the correlation function was perceived at  $100 h^{-1}\text{Mpc}$  in the redshift range of  $0.16 < z < 0.47$ . However, the meticulousness of the large scale structure depends on matter density, dark energy, galaxy formation and on their evolution as well.

According to the cosmological principle, the Universe is homogeneous and isotropic at large scales. However, if we zoom in different parts of the Universe this homogeneity is over and its clumpiness is visible at all scales, from galaxies (kpc) to very large superclusters (tens of Mpc) passing through clusters (few Mpc). The best way to reveal this structure is by measuring the clustering in precise galaxy redshift surveys.

## 1.2 Galaxy Clustering

Galaxy clustering constrains host dark matter halo masses, host galaxy types, the environments in which galaxies are formed and provide clues on galaxy evolution on which large scale structure depends.

The statistical quantity that is commonly used to describe the clustering is the two point correlation function,  $\xi(r)$ , which is defined as the probability  $dP$  over a random Poisson distribution of finding a pair of galaxies in a volume  $dV$  at a distance  $r$  (Peebles, 1980)

$$dP = n[1 + \xi(r)]dV \quad (1.1)$$

where  $n$  is the number density of the galaxy distribution where we want to study the clustering. This distribution will be compared to a random distribution of galaxies that covers the same 3D area (declination, DEC, right ascension, RA, and redshift,  $z$ ) as the real sample. Due to this comparison, the density of galaxies in the random sample should be considerably higher in order to not introduce an excess in the clustering of the real sample (see Figure 1.3). Moreover, in this way, Poisson errors would be subdominant.

The reason why  $\xi(r)$  is used to illustrate the clustering is because it is the inverse Fourier transform of the power spectrum, which describes the density fluctuations of the Universe. Therefore, a galaxy sample with overdensities with respect to a random sample will have a clustering signal, meaning that  $\xi(r) > 0$  (*i.e.* our galaxy sample clusters stronger than the random sample). On the other hand,  $\xi(r) < 0$  will refer to a sample that cluster less strong than a random distribution of galaxies. The last case is  $\xi(r) = 0$ , which is the correlation function of a sample that clusters like a random Poisson distribution of galaxies and, from now on, it will be referred as "no clustering" or "0 clustering".

However, in practise, probabilities, volumes or correlation functions are not directly measured but  $\xi(r)$  is estimated. To proceed, a random sample of  $R$  number of galaxies is introduced to be analyzed in parallel with a real sample of  $D$  number of galaxies.

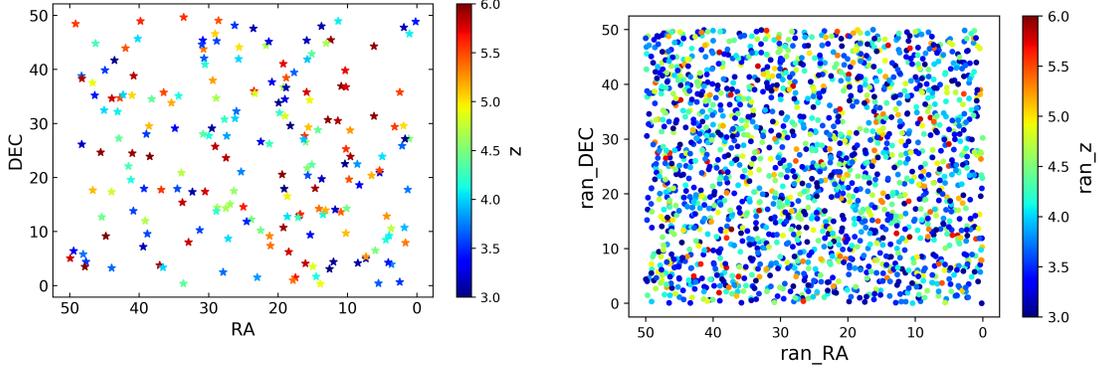


Figure 1.3: Left: real sample of galaxies shown as declination and right ascension distribution. On the right, the same for a much larger random sample than the real survey. In both cases the colorbar shows different depths in redshift and DEC, RA and  $z$  cover the same range of values.

This method is based on pair counting of galaxies within the available samples and the consequent use of an estimator.

The most common estimators used to compute the  $\xi(r)$  discussed in the literature are:

$$- \xi_{PH}(r) = \frac{DD}{RR} - 1 \quad (\text{Peebles and Hauser, 1974})$$

$$- \xi_{Hew}(r) = \frac{DD - DR}{RR} \quad (\text{Hewett, 1982})$$

$$- \xi_{DP}(r) = \frac{DD}{DR} - 1 \quad (\text{Davis and Peebles, 1983})$$

$$- \xi_H(r) = \frac{DD \times RR}{DR^2} - 1 \quad (\text{Hamilton, 1993})$$

$$- \xi_{LS}(r) = \frac{DD - 2DR + RR}{RR} \quad (\text{Landy and Szalay, 1993})$$

In all cases, DD is the number of pairs within our real data sample, RR is the number of pairs counted in the random sample and DR is the number of counted pairs between the real and the random sample. Since the random sample is much larger than the real

sample, the number of pairs DD, DR and RR is normalized. The full process followed to estimate the unknown  $\xi(r)$  will be explained in detail in Chapter 6.

There are several studies that tried to determine which of the above estimators gives the best results. At small scales, their results are almost indistinguishable but at larger scales the differences are obvious. [Pons-Borderia et al. \(1999\)](#) compared the behaviour of six estimators, finding no better performance between the Landy-Szalay and the Hamilton estimator, having these two smaller statistical errors than the rest. So did [Kerscher et al. \(2000\)](#), who showed that the Landy-Szalay and the Hamilton estimators present better properties out of a nine estimators comparison. Though Landy-Szalay estimator requires more computational time, edge corrections are better taken into account and the sensitivity to the random sample size is lower, which otherwise, would alter possible clustering signals at large scales. Due to this deed and to the fact that this estimator reaches minimal variance, it will be one of the selected estimators to carry out the spatial clustering analysis in this thesis (Chapter 6).

With the two point correlation function estimated, the common fit that is performed is a power law,

$$\xi(r) = \left(\frac{r}{r_0}\right)^{-\gamma} \quad (1.2)$$

where  $r_0$  is the correlation length of the galaxy sample, the distance at which the excess probability over a random sample is 1 (*i.e.*  $\xi(r) = 1$ ).

Depending on the kind of available data, clustering analysis can be studied in different ways. In the case that the redshift distribution, RA and DEC are available, the common way to study the clustering is by using the 3D two point correlation function (subsection 1.2.2). On the other hand, if the redshift distribution is not available the angular two point correlation function (subsection 1.2.1) is used instead.

### 1.2.1 Angular clustering method

Obtaining the spectra of a large sample of galaxies is very expensive so the redshift data is not always disposable. Therefore,  $\xi$  is measured in two dimensions as projected onto the plane of the sky and renamed as two-dimensional projected angular correlation function,  $\omega(\theta)$ ,

$$dP = n[1 + \omega(\theta)]d\Omega \quad (1.3)$$

where instead of the radial distance between two galaxies,  $r$ , the angular distance between them,  $\theta$ , is introduced.  $d\Omega$  is the solid angle extended between to galaxies and  $n$  is the mean number of galaxies per steradian.

The two-dimensional clustering on the plane of the sky is a projection of the three-dimensional clustering,  $\xi(r)$ , and it can be fit as

$$\omega(\theta) = A \cdot \theta^{1-\gamma} \quad (1.4)$$

where  $A$  and  $1-\gamma$  are the amplitude and the slope of the fit, respectively.

Normally, the observed angular clustering is estimated with one of the estimators shown in the previous section and  $\omega(\theta)$  is related to  $\xi(r)$  by use of the equation 1.2, the Limber equation (Limber, 1953), and the fit parameters. Since the angular clustering is a projection of  $\xi(r)$ , it can be deprojected by integrating along the redshift direction (*i.e.* the redshift distribution must be known). Exactly because of this reason,  $\omega(\theta)$  might be sometimes hard to interpret. An observed weak signal projected on the sky can be due to a low clustered galaxy sample projected on a small distance or it can be because of the washing out of the clustering signal due to the integration over a large distance. Hence, there is a degeneracy between the real clustering amplitude and the redshift distribution chosen.

As it is already possible to intuit, several issues accompany this way of studying the clustering of galaxies.

- The redshift distribution function must be known accurately. However, due to redshift distortions (see subsection 1.2.2), it is never perfectly assumed. Whereas in the case of photometric surveys the redshift data is often uncertain and results in large uncertainties in the deprojection, in spectroscopic surveys the small number of galaxies with redshift data and the non-random slit allocation may cause artificial clustering.
- Limber equation itself introduce errors that are not taken into account when it comes to calculate the quoted clustering error bars.
- General angular clustering measurements are known to be low by an additive factor, the integral constraint. This issue is due to the use of the data sample itself that might cover small areas on the sky to estimate the mean galaxy density.

Although many attempts, that are beyond this master thesis, are applied to reduce these error sources,  $\xi(r)$  is still uncertain. Narrow band measurements help to constrain

a more accurate redshift distribution but it will not be precise enough unless there is spectroscopy data available. When this situation is present, a much more pinpoint method of clustering measurements is used instead, the spatial clustering method.

### 1.2.2 Spatial clustering method

Having the redshift distribution of a sample of galaxies is of great value in clustering analysis. With redshift data available, spatial clustering analysis can be accurately performed.

Measurements of the  $\xi(r)$  use the redshift of the galaxies to infer the distances between galaxies (see Chapter 3 for details) which introduces two main problems. In the first place, a cosmological model must be assumed as right to calculate those distances and, in the second place, redshift space distortions (RSD) must be considered. RSD are seen as a distortion appeared in the distribution of galaxies while using redshift-space representations. The expansion of the Universe and the peculiar velocities of galaxies are responsible of the Doppler shift that causes the RSD.

1. **Fingers of God (FOG):** The random peculiar velocities (not associated with the Hubble flow) create a Doppler shift that makes the distribution of galaxies in a redshift space representation look elongated and pointing back to the observer along the line of sight (l.o.s.) (red coloured galaxies in the bottom panel of Figure 1.4). At small scales ( $< 1 h^{-1}\text{Mpc}$ ), these velocities deviate the real galaxy velocity from the Hubble flow, stretching the galaxies radially and giving place to the FOG. Furthermore, these velocities are related to the cluster gravity (Virial theorem), varying the observed redshift of the galaxies and, consequently, their calculated distances. Whereas the distance from an observer to the galaxies inside a galaxy cluster can be considered to be the same for all of them, the redshift for each galaxy is different. Those redshift differences are transformed into cluster elongations along the l.o.s. in a redshift space representation (see top part of Figure 1.4).
2. **Kaiser effect:** At larger scales ( $> 1 h^{-1}\text{Mpc}$ ), adjacent galaxies that fall inwards towards structures that are joining up or collapsing have coherent motions. The massive forming structures are the cause of this gravitational infall of galaxies and their coherent motion leads to a contraction of the groups of galaxies when they are represented in a redshift space (see upper panel of Figure 1.5). The effects are smaller than in the FOG case but, looking at the bottom panel of Figure 1.5, the flattening in the  $\xi(r)$  contours is remarkable at large scales. Besides, the elongation at small scales (FOG effect) is obvious.

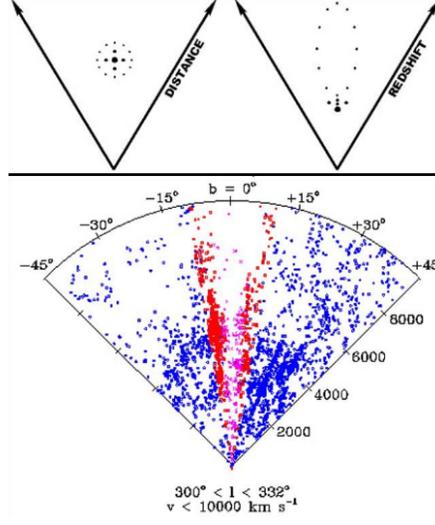


Figure 1.4: Top: group of galaxies in real (left) and redshift (right) space. Bottom: RA-velocity representation of galaxies where the elongation towards the observer is shown in red.

Both kind of RSD, at small and large scales, decrease the amplitude of clustering signals to intermediate scales. Thus, to account for these effects, the separation vector between two galaxies,  $r$ , in equations (Landy and Szalay, 1993) and (Davis and Peebles, 1983) is split into a perpendicular component,  $r_p$ , and a component along the l.o.s.,  $\pi$ , (see Section 6 for a detailed explanation), becoming, respectively

$$\xi(r_p, \pi) = \frac{DD(r_p, \pi) - 2DR(r_p, \pi) + RR(r_p, \pi)}{RR(r_p, \pi)} \quad (1.5)$$

and

$$\xi(r_p, \pi) = \frac{DD(r_p, \pi)}{DR(r_p, \pi)} - 1 \quad (1.6)$$

Then,  $\xi(r)$  is recovered by projecting  $\xi(r_p, \pi)$  along the  $r_p$  axis. This is done by integrating along the  $\pi$  direction (RSD only affect the l.o.s. component) and calculating the projected correlation function,  $\omega_p(r_p)$

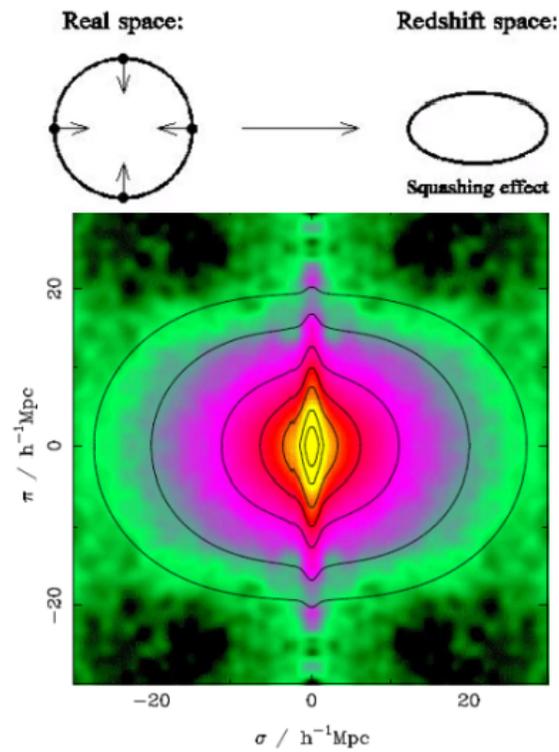


Figure 1.5: Top: group of galaxies in real (left) and redshift (right) space. Bottom:  $\xi(r)$  contours in different colours along the l.o.s. distance,  $\pi$ , varying with the perpendicular distance to the l.o.s.,  $\sigma$ . Yellow colours correspond to large values of  $\xi(r)$  while green colours represent lower values. The image is taken from [John et al. \(2001\)](#).

$$\omega_p(r_p) = \int_{-\infty}^{\infty} \xi(r_p, \pi) d\pi \quad (1.7)$$

where the RSD are considered to be vanished.

### 1.3 Lyman Alpha Emitters

Lyman-alpha emitters (LAE) are galaxies that show Lyman-alpha emission from the neutral hydrogen that surround them. This emission is the most common way to detect galaxies at high redshifts, thus, this is the most used spectral line to trace and study the evolution of galaxies.

Between 1906 and 1914, Dr. Theodore Lyman studied the far ultraviolet region of the electromagnetic spectrum, discovering that when excited hydrogen gas cools down, it emits radiation in this region of the spectrum. When this radiation was viewed through a spectroscope, it seemed to be formed by bright emission lines. He found the brightest of these lines at  $1216 \text{ \AA}$ , the Lyman-alpha line, which lies in the so called Lyman hydrogen series and it is due to the transition of an electron from  $n \geq 2$  to  $n = 1$  (see Figure 1.6).

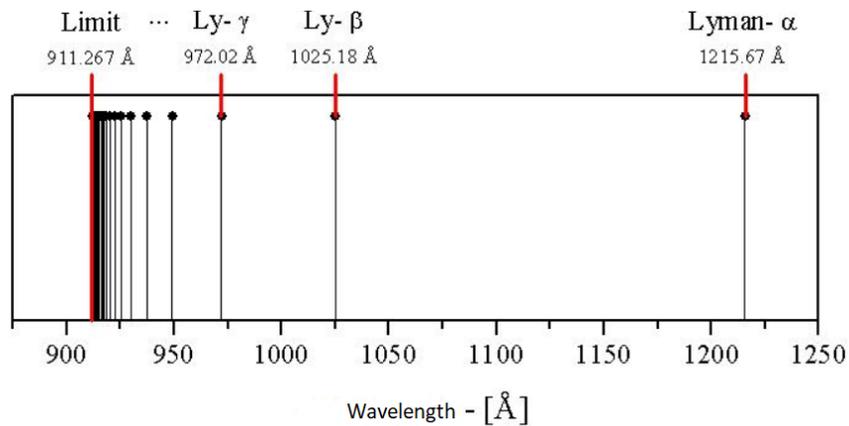


Figure 1.6: Wavelength coverage of the Lyman serie of hydrogen. The Lyman-alpha is shown at  $1215.67 \text{ \AA}$ .

During the same years, the physicist Johannes Rydberg obtained a formula that could predict the wavelengths of these lines. He showed that there would never be an emission line at frequencies less than  $912 \text{ \AA}$ , the Lyman limit. Since the most common element

in the Universe is hydrogen, and stars are related to these emission lines, scientists have found different ways of detecting these emissions from the Earth.

At present time (*i.e.*  $z = 0$ ), the wavelengths of the Lyman lines lie within the ultraviolet range but as the redshift is increased, the lines are shifted to the optical range. Equation 1.8 shows this transition and reason why it is possible to detect this emission with based-ground telescopes.

$$1 + z = \frac{\lambda}{1215.67 \text{ \AA}} \quad (1.8)$$

where  $\lambda$  is the observed wavelength, and  $1215.67 \text{ \AA}$  is the Lyman-alpha emission line wavelength, in the ultraviolet region of the electromagnetic spectrum.

Such Lyman-alpha emission was first suggested as a signature of young galaxies by Partridge and Peebles (1967). They studied the possibility of observing these kind of distant galaxies, specially the newly formed ones. In order to do that, they assumed a specific model of galaxy formation where galaxies formed due to gravitational instabilities of the expanding gas and, in early evolutionary stages, had a phase of high luminosity. The assumptions of this higher luminosity were 1) many of the first stars would be luminous stars of the O-B type and, 2) the fast production of chemical elements in early stages requires high luminosities. They concluded that Milky Way galaxy type luminosities should be around  $3 \cdot 10^{46} \text{ erg s}^{-1}$ ,  $\sim 700$  times higher than the actual one known nowadays.

Moreover, they estimated that 10% of the total luminosity of a young galaxy comes from the Lyman lines, 6-7% from the Lyman-alpha line itself. Figure 1.7 shows the case where all ionizing photons are converted to Lyman photons. The flux was calculated with  $\Delta\nu = 0.002\nu$  for the line, only taking into account the Doppler broadening from the motions of the gas and the stars in young galaxies with velocities of  $\sim 300 \text{ km/s}$ .

Nowadays, it is known that the galaxy formation model that they utilized used too restrictive assumptions which lead them to estimate wrong values for the Lyman-alpha luminosities.

Several decades had to pass until the first discovery of high redshift LAEs was achieved. Following Partridge and Peebles (1967), there were several attempts to discover these kind of galaxies (*e.g.* (Partridge, 1974; Davis and Wilkinson, 1974; Meier, 1976)) but they all failed. It was in the early 90s when the narrow-band technique (see section 1.4.1) started to be successful. The first detections of LAEs at high redshift were done by Lowenthal et al. (1991); Wolfe et al. (1992); Macchetto et al. (1993), who detected one LAE at  $z = 2.3, 2.466$  and  $3.428$ , respectively, using a narrow-band filter. Afterwards, narrow-band techniques

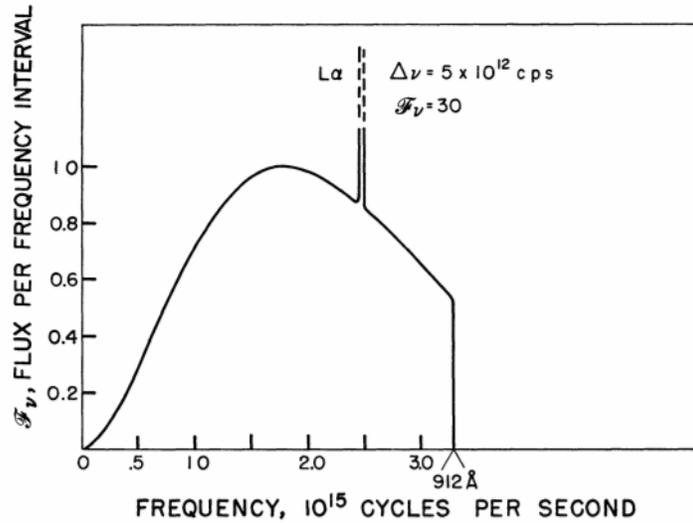


Figure 1.7: Spectra of young galaxies in the specific case that all photons of ionizing radiation become Lyman photons expected by [Partridge and Peebles \(1967\)](#).

started to become more common and effective, detecting more than one LAE at a time. That was the case of [Cowie and Hu \(1998\)](#), who detected a substantial population of LAEs at  $z = 3.4$ . The selected samples included objects with strong Ly $\alpha$ , contributing significantly to the integrated star formation at these epochs.

Up to date, the lines of LAEs are believed to be produced by outbursts of star formation in galaxies, having high importance in the trace of dark matter halos and in the evolution of the distribution of galaxies in the Universe. The formation of stars within galaxies emit photons that excite the hydrogen that surrounds the galaxies. This excited hydrogen is the one that releases the Lyman-alpha emission when it is deexcited.

LAEs are very old nowadays (200-600 million years) and are thought to be progenitors of Milky Way galaxy types. They usually have low masses, between  $10^8 M_{\odot}$  and  $10^{10} M_{\odot}$  and very high specific star formation rates. Furthermore, they are compact galaxies and have low metallicities. However, there are still many unknowns, such as: the photon escape fraction, present outflows, anisotropies in the hydrogen density distribution or their dust content, which is thought to be the leading responsible of the Lyman-alpha escape fraction.

The detection of these types of galaxies can mainly be achieved in two different ways. The first option is the utilization of powerful spectrographs mounted on based-ground telescopes, whereas the second option is accomplished by narrow-band searches.

## 1.4 Detection techniques

### 1.4.1 Narrow band selection technique

An usual detection technique for Lyman alpha emitters is the use of narrow band filters. These filters are designed to only transmit a narrow band of spectral lines from the electromagnetic spectrum (20 nm or less in wavelength range).

The use of narrower bands increases the good performance in detecting galaxies or other celestial objects. For instance, in Figure 1.8 is possible to appreciate how the resolution improves as long as the bands get narrower. The big step from broad to narrow band is quite sharp as it is also shown in Figure 1.10.

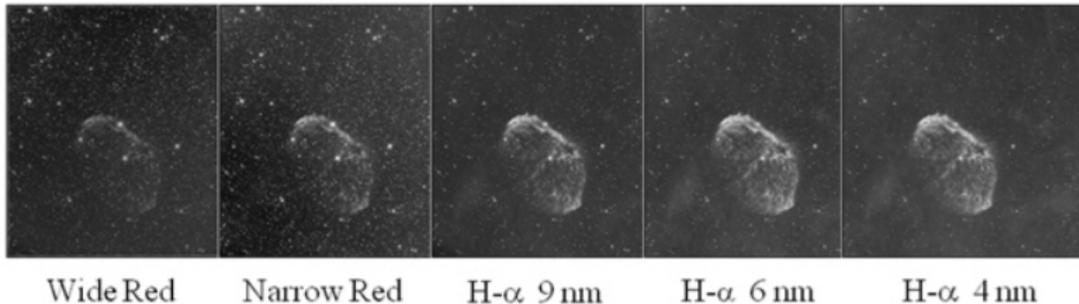


Figure 1.8: Contrast increase in the images of the nebula NGC 6888 by using broad band (WideRed), narrow band in the red (NarrowRed) and narrow bands of 9 nm, 6 nm and 4 nm width, showing a clearer structure while narrowing down the width. The images were taken the same night, with the same system and with equal exposure times.

Narrowband filters are made to resolve a specific emission line. With the wavelength of the emission line known, the suitable width of the filter is employed. These carefully designed filters only let a few wavelengths of light along with the emission line pass, rejecting all other wavelength ranges. An example of Lyman alpha detection is shown in Figure 1.9, where in blue and red two narrow band detections for Lyman alpha at redshifts 5.73 and 6.58 are displayed, respectively. On the background of the image, three broad bands (i, z, and y) are manifested to show the difference between both different kind of bands. While narrow band filters are utilized for single emission lines, broad band filters have a better performance detecting the continuum. This better performance is clearly visible when it comes to detect single galaxies (see Figure 1.10). Whereas the detection using narrow band filters is quite precise, the use of broad band filters does not even get the galaxy in many cases.

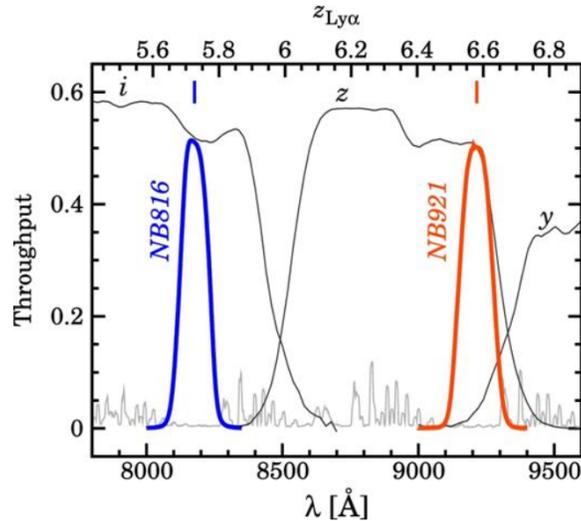


Figure 1.9: Filter transmission curves of narrow band (NB816 in blue and NB921 in red) and broad band (i, z and y) filters in [Shibuya et al. \(2017\)](#). The gray line denotes the OH sky lines and the upper x-axis shows the Lyman alpha redshift.

Even if the use of narrow bands is much more precise than the use of broad band filters, it is still not the best detection technique. Narrow band imaging can only survey a very narrow redshift range at a time (depending on the width of the filter). The large areas of the sky that can be surveyed with narrow band filters are countered with medium precision due to the fact that the width of the filter is wider than the width of the emission line. Therefore, large areas of the sky are covered, which is great to study the clustering, but the redshifts are not perfectly precise since it is an imaging technique, something that is crucial for clustering measurements. Besides, no clustering evolution with redshift is possible with the use of just one filter.

Nevertheless, these uncertainties or cons for clustering analysis are overcome with the use of spectrographs.

### 1.4.2 Spectrographs

Spectrographs are much more expensive instruments than narrow-band filters but, at the same time, much they are more effective. A distribution of very precise redshifts can be provided which offers the opportunity to study the clustering evolution of the sample of galaxies with a high precision.

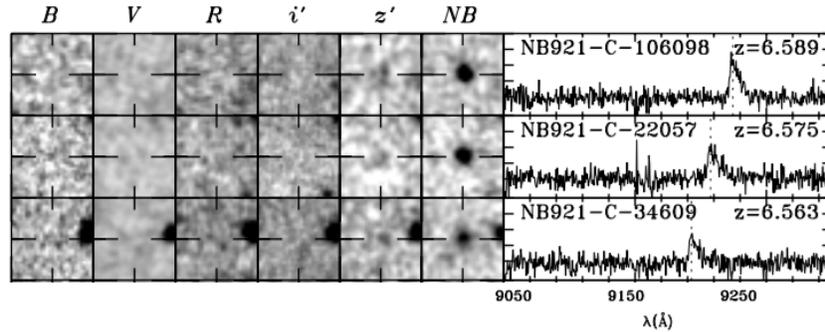


Figure 1.10: Detections of LAEs at redshift 6.6 with broad bands B, V, R, i, z and with the NB921 filter in [Ouchi et al. \(2010\)](#). The spectrum of each galaxy is shown on the right along with the redshift and name of each galaxy. Each detection is presented in a 6''x6'' square.

Spectrographs separate the light into its component wavelengths, revealing information about the chemical content, temperature and motion of planets, comets, stars, interstellar gas and galaxies. These instruments usually cover a wide range of wavelengths, covering several bands of the electromagnetic spectrum. An example of a LAE spectrum is shown in Figure 1.11.

A very common type of spectrographs to detect LAEs are the integral field spectrographs, they equip the spectrograph with an integral field unit, combining spectrography and imaging techniques, obtaining, thus, spatially resolved spectra. The existence of large aperture and high-resolution telescopes requires these kind of spectrographs to analyze the spectra of extended objects as a function of position, which was previously carried out with long-slit spectrographs.

Long-slit spectrographs are rather slow when it comes to get spectra. The spectrum is dispersed perpendicular to the slit of the spectrograph, and, by shifting the position of the slit, the spectrum of points in the imaged field is obtained. Speeding up this process is done by using integral field spectrographs, which obtain spectra in a two-dimensional field.

The most used approaches for integral field units are fibers, image slicers and lenslet arrays.

- **Fibers:** the light coming from LAEs is kept by an array of fibers, which act like the entrance slits plane of the spectrograph. The other side of the fibers are set along a single slit that obtains a spectrum for each fiber.

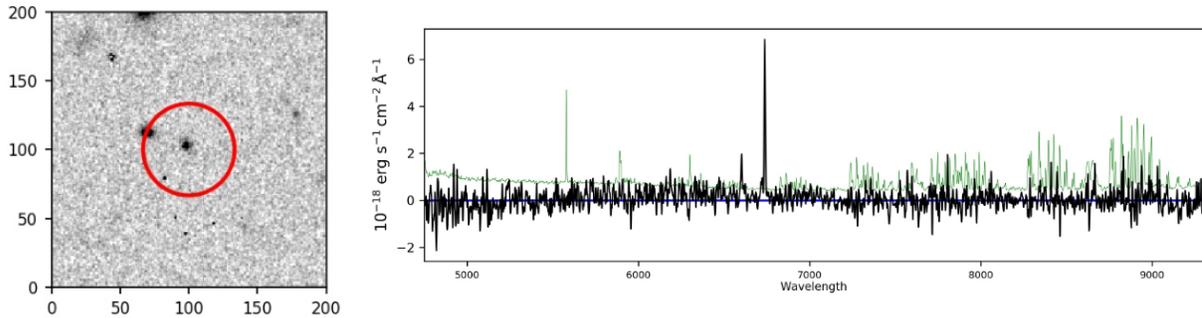


Figure 1.11: Left: LAE 101024044 in candel-cdfs-01 taken by HST ACS with the filter F814W. The red circle is 1'' in radius and the image is 6''x6''. Right: Spectrum of the LAE at  $z = 4.54$  taken by the spectrograph MUSE, which has been smoothed with a 9 pixel Hanning window. The Lyman-alpha emission line is visible at  $\lambda = 6734.81 \text{ \AA}$ .

- **Image slicer:** the image got by the spectrograph is sliced in the image plane and re-arranged in such a way that each different part of the image falls onto a slit and a dispersing element. In that way, a spectrum is obtained for a larger area of interest.
- **Lenslet array:** the entrance slits plane of the spectrograph is formed by a lenslet array, acting as spaxels. All beams generated by the lenslet array are sent through a dispersive element and imaged by a camera. In such a way, the result is a spectrum for each lenslet.

Nowadays, the best functioning spectrograph to detect LAEs is the 3D integral field spectrograph MUSE.

## 1.5 MUSE spectrograph

In order to investigate the structure of the Universe by observing the spatial distribution of galaxies, extensive surveys of thousands of galaxies are demanded to cover big volumes and the large scales of the Universe. Thus, telescopes provided with the best spectrographs are needed and cover the issues present with narrowband techniques.

The VLT<sup>1</sup> is the most productive ground-based telescope, operated by the European Southern Observatory in Chile and working at deep ultraviolet, visible and mid-infrared

<sup>1</sup><https://www.eso.org/public/teles-instr/paranal-observatory/vlt/>

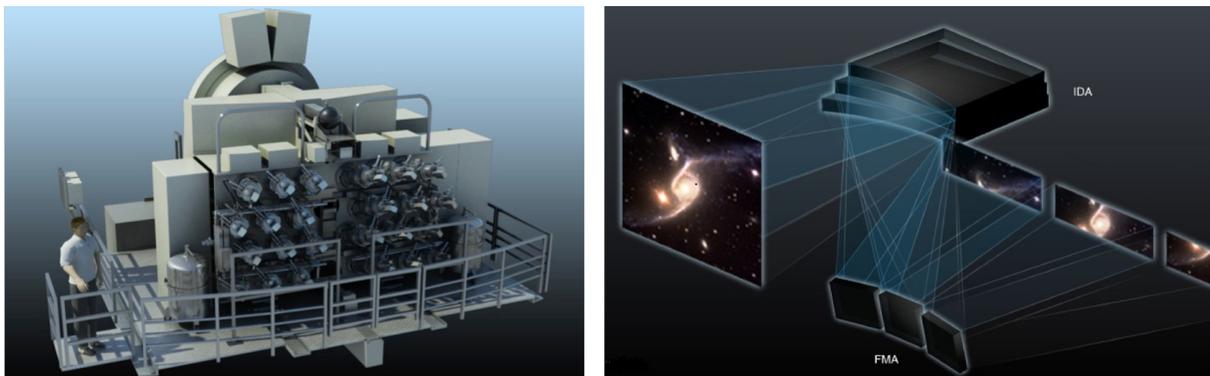


Figure 1.12: Left: Model of the MUSE spectrograph. Right: Sky field slicing by image slicers and the corresponding sending to the spectrograph.

ranges. It consists of four 8.2 m telescopes that can reach angular resolutions of about 0.001 arcseconds when they are all combined, while the single telescopes reach resolutions of 0.05 arcseconds. The VLT is formed by large-field imagers, adaptive optics corrected cameras, spectrographs and high-resolution multi-object spectrographs mounted on four telescopes named Antu, Kueyen, Melipal, and Yepun, where MUSE<sup>2</sup> is also integrated.

MUSE is a 3D integral field spectrograph that splits and slices 1 arcmin<sup>2</sup> field on the sky into 90000 spatial elements, creating a spectrum for each point. Since MUSE obtains a full spectrum for each pixel, the output is a datacube. MUSE was partly built at the Leibniz-Institut für Astrophysik Potsdam (AIP) and, in opposition to most integral field spectrographs, the split of the field of the sky is carried out by image slicers and not by fibers. Its functioning and the complete instrument can be seen in Figure 1.12.

MUSE uses a Wide Field Mode with a field of view that covers 1' x 1'. Therefore, MUSE is ideal to discover distant, high-redshift galaxies, carrying out spectroscopic analyses and, even, studying galaxy mergers. To get the spectra of these galaxies, specially when they are faint, a long integration time, high instrumental throughput and a good instrumental stability is indispensable, which MUSE provides in short exposure times. Furthermore, MUSE also has a Narrow Field Mode with a field of view of 7.5" x 7.5" with spatial resolutions of 10–30 percent Strehl ratio in the I-Z band and a 2D spatial coverage, which is utilized for other purposes.

MUSE was designed to study the formation of galaxies, the fluorescent emission, the cosmic web, reionisation, gravitational lensing, supermassive black holes in nearby galaxies,

<sup>2</sup><https://www.eso.org/public/teles-instr/paranal-observatory/vlt/vlt-instr/muse/>

galaxy kinematics, stellar populations, etc. However, the main goal was to detect the progenitors of nearby galaxies out to redshifts of 6. Since these sources can be extremely faint, they can only be detected through the Lyman-alpha emission line which MUSE gets. MUSE surveys small regions in the sky to very deep redshifts, especially, it searches for low-mass star-forming galaxies at high redshifts, LAEs.

Although these sources can be detected by using narrowband imaging getting just one redshift at a time, MUSE searches for emission lines over a wide range in redshifts,  $2.9 < z < 6.65$ , offering a much higher precision and redshift distribution than narrowband filters.

MUSE reaches Lyman-alpha limiting fluxes of  $3 \cdot 10^{-18}$  erg s<sup>-1</sup> cm<sup>-2</sup> (Herenz et al., 2019) with only 1 h exposure time, being an order of magnitude better than any current narrowband imaging survey.

## 1.6 Key aims of this work

Further studies need to be carried out to understand and know the exact structure that galaxies form in the Universe. It is crucial to analyze the spatial and redshift distribution of samples of galaxies to better constrain large scale structures, dark halo populations and the evolution of Milky Way-like galaxies through different epochs.

A new region in parameter space has recently been opened by the capability of the MUSE instrument to survey small regions in the sky to extreme depths, especially for low-mass star-forming galaxies at high redshifts detectable through their Lyman-alpha emission. In a pilot study by my supervisor's team (Diener et al., 2017) it was shown that MUSE-selected LAEs show a significant clustering signal already in their redshift distribution, even when using a much smaller sample than the one that is now available.

Therefore, the goal of this thesis is to explore how far we can go in measuring the clustering of the special data provided by MUSE, to perform, as far as it is possible, a clustering analysis of the large LAE sample from MUSE surveys. Clustering results will be compared with the different methods applied and the dependence of clustering signals will be studied as several properties are varied. A logical tree of steps will be followed:

1. Learn about the standard methods to measure galaxy clustering and the current state of research in this field.
2. Acquire the data handling know-how (*i.e.* Python programming) needed to perform a clustering analysis based on a given catalogue.

3. Reproduce the results of [Diener et al. \(2017\)](#) using the same input catalogue (the first 24 fields of MUSE-Wide).
4. Extend the analysis to the full MUSE-Wide catalogue and discuss the results in relation to the recent literature and to the results from [Diener et al. \(2017\)](#).
5. Extend the analysis to the MUSE-Deep catalogue using the same method and compare the results from the two catalogues and [Diener et al. \(2017\)](#) results.
6. Develop a code to simulate random samples that properly account for the survey selection and luminosity function, study its clustering and the possible existing bias.
7. Perform the clustering analysis after splitting the samples into subsets selected by certain properties (luminosity, redshift, stellar mass...) and compare the results for different subsets.
8. Implement other clustering statistics than just redshift-space clustering, such as spatial clustering by the use of the two point correlation function (2pcf), the projected correlation function and the random sample generated.

# Chapter 2

## MUSE samples

The short exposure times and the depth accomplished by MUSE make its data special. The wide redshift range covered provides the best opportunity to study the clustering evolution along this range.

MUSE carries out a blind search on the sky but points at the region covered by the Great Observatories Origins Deep Survey (GOODS), with  $320 \text{ arcmin}^2$ , and the Cosmic Evolution Survey ( $2 \text{ deg}^2$ ) comprised of 640 Hubble Space Telescope (HST) orbits (Scoville et al., 2007). Hence, there is photometric data available used to determine the counterparts and the spectral energy distribution of the LAEs in the HST images, providing as important data as stellar masses.

The sky coverage per field of MUSE is small,  $1' \times 1'$  arcmin, but enough to carry out a clustering analysis. This coverage is shown in Figure 2.1 over the footprint in the Chandra Deep/Field South region on the V-Band image from the Garching-Bonn Deep Survey (Hildebrandt et al., 2006).

The main two surveys of MUSE, MUSE-Wide with 100 fields and 1 h exposure time and MUSE-Deep with 9 fields and 10 h exposure time, are represented in Figure 2.1. The 9 fields of the HUDF mosaic are a reduction from MUSE-Deep data, with 1.67 h instead of 10 h exposure time.

### 2.1 MUSE-Wide survey

The MUSE-Wide survey is a GTO programme that focuses on regions with extremely deep HST imaging, specially on the GOODS-South and CANDELS Deep/CDFS (Grogin

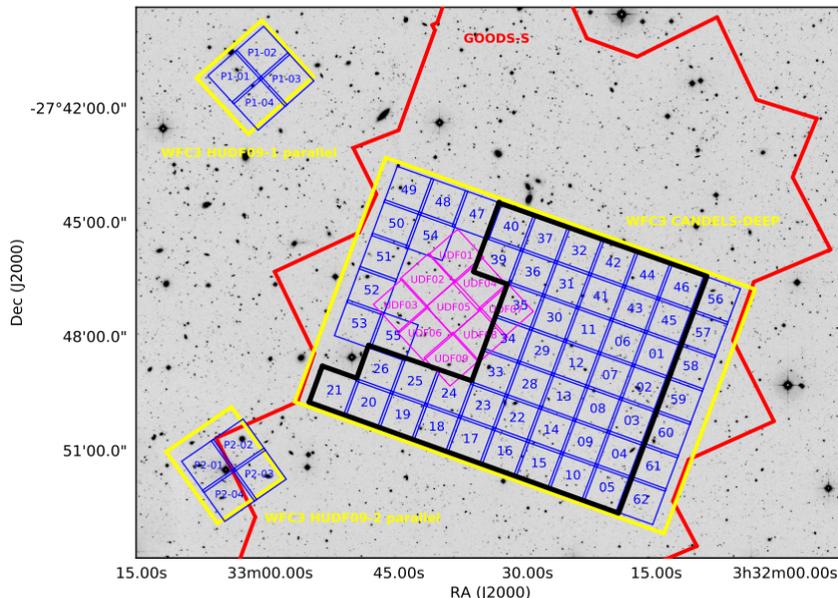


Figure 2.1: Sky coverage of MUSE. The MUSE-Wide survey with 100 fields (blue) and MUSE-Deep with 9 fields (pink). The red lines show the contours of the GOODS-S ACS, in yellow CANDELS-DEEP and HUDF09 parallels WFC3 regions and the black line covers the 44 fields of the last MUSE data release (Urrutia et al., 2019).

et al., 2011) areas that contain a wide range of multiwavelength data. The MUSE GTO programme sets a number of goals such as: the detection of large samples of LAEs that include rare and extreme LAEs and the study of star forming low mass galaxies at intermediate redshifts (Urrutia et al., 2019). Apart from these goals, MUSE has already lead to many new scientific results. For instance, the measurement of the Ly $\alpha$  emitting fraction among  $z > 3$  galaxies (Caruana et al., 2017), the study of spatially resolved properties of galaxies at  $0.1 < z < 0.5$  (Hemmati et al., 2019), the detection of signal clustering already in the first 24 fields of MUSE-Wide (Diener et al., 2017), etc.

Since the main goal of this thesis is to explore different possibilities to analyze the clustering of MUSE surveys, the different samples will be studied in detail and split in various number of fields.

- **24 fields:** The first MUSE data release for the first 24 fields of MUSE-Wide is presented in Herenz et al. (2017). The blind search for LAEs was done in the CDFS region, covering an area of 22.2 arcmin<sup>2</sup> and yielding a number of 234 galaxies in

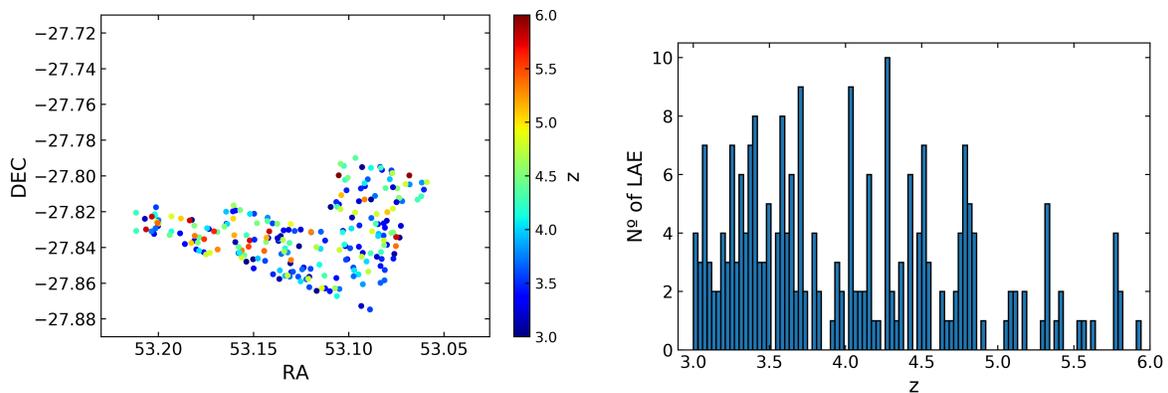


Figure 2.2: Left: sample of galaxies in the first 24 fields of MUSE-Wide shown as DEC, RA and the colorbar showing different depths in redshift. On the right, the histogram represents 234 LAEs as function of redshift.

the range  $3 < z < 6$ . These galaxies, which 50% of them did not have spectroscopic redshift until this catalog was published, were detected with only 1 h exposure time and are shown in Figure 2.2.

- **60 fields:** The next number of fields that will be used are the 60 fields from MUSE-Wide shown in Figure 2.3. They are within the GOODS and WFC3 CANDELS-DEEP regions and, once again, the exposure time is 1 h with  $55.5 \text{ arcmin}^2$  coverage, yielding a total of 681 LAEs at  $3 < z < 6$ .
- **69 fields:** In opposition to the 60 fields, the 69 fields of MUSE-Wide cover an area of  $63.83 \text{ arcmin}^2$  and have 9 deeper fields (1.67 h exposure time) that cover the hole seen in the 60 fields. The number of galaxies detected in this area is 947 and are represented in Figure 2.4.
- **77 fields:** The last sample of galaxies from MUSE-Wide that will be used covers a number of 77 fields with a sky coverage of  $71.23 \text{ arcmin}^2$ . 8 extra fields from WFC3 HUDF09-1 and WFC3 HUDF09-2 parallel are added, giving a total number of LAEs of 1080 at  $3 < z < 6$  shown in Figure 2.5.

If we take a closer look to any of the histograms shown below, it is possible to see spikes in certain parts of the histograms. These spikes show a larger number of LAEs at some redshift ranges which is a clear sign of clustering. It is worth mentioning that the bin sizes play an important role here. In order to clarify this, I have represented with different

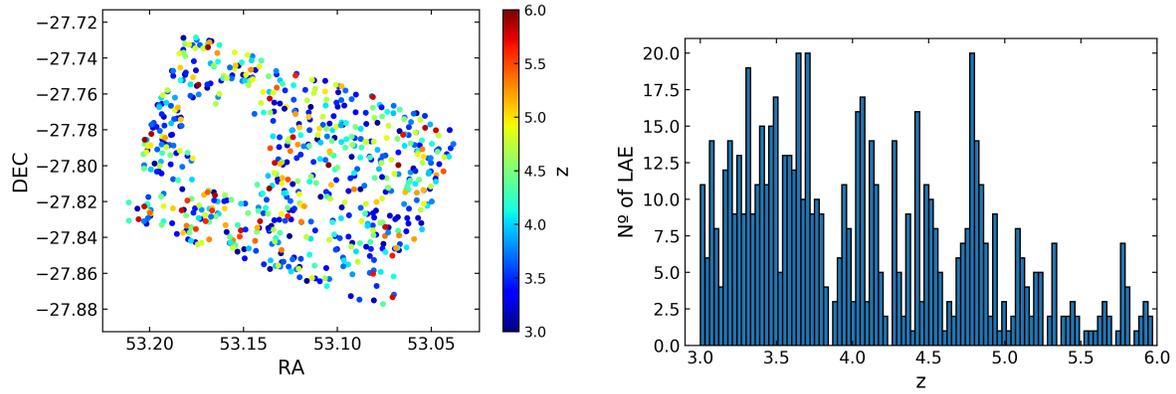


Figure 2.3: Left: sample of galaxies in the 60 fields of MUSE-Wide shown as DEC, RA and the colorbar showing different depths in redshift. On the right, the histogram represents 681 LAEs as function of redshift.

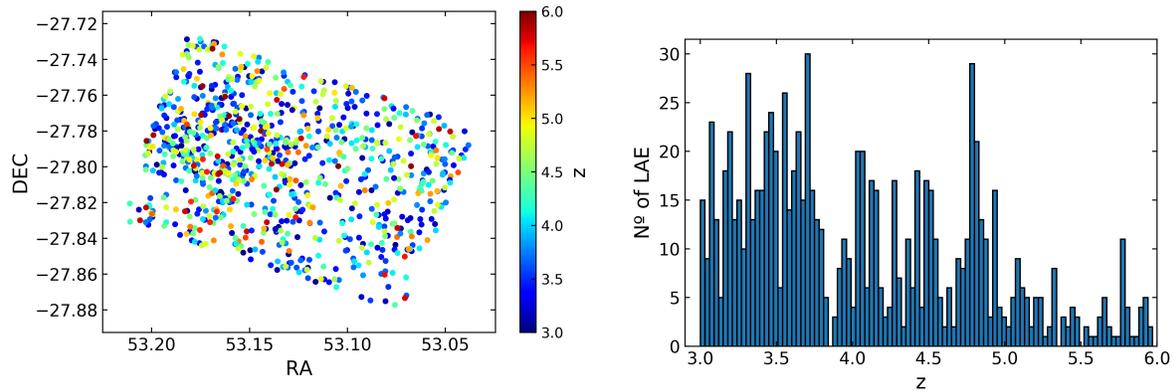


Figure 2.4: Left: sample of galaxies in the 69 fields of MUSE-Wide and the reduction from MUSE-Deep shown as DEC, RA and the colorbar showing different depths in redshift. On the right, the histogram represents 947 LAEs as function of redshift.

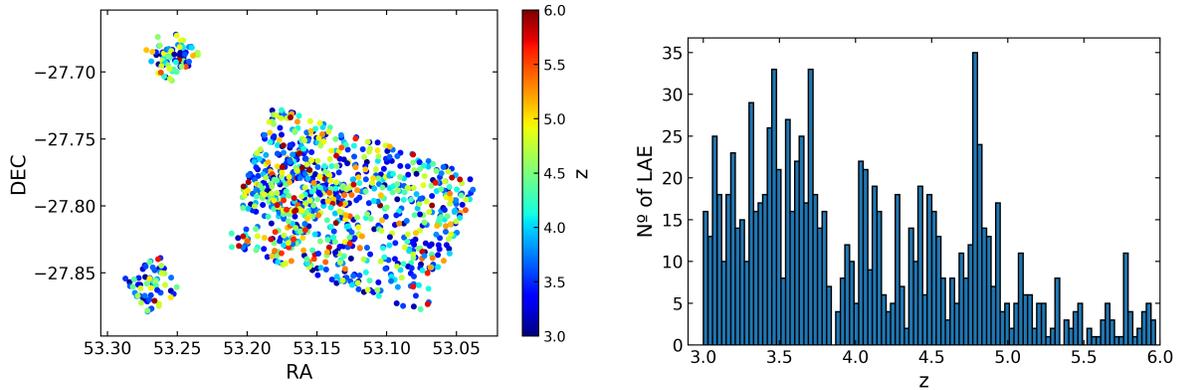


Figure 2.5: Left: sample of galaxies in the 77 fields of MUSE-Wide shown as DEC, RA and the colorbar showing different depths in redshift. On the right, the histogram represents 1080 LAEs as function of redshift.

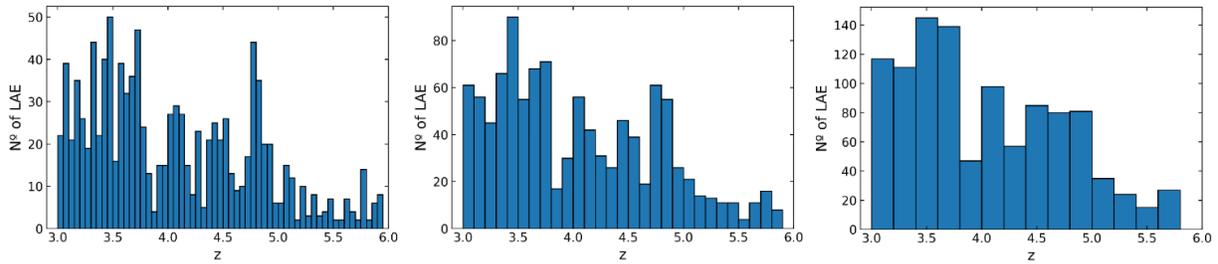


Figure 2.6: Histogram representation of the 1080 LAEs in the 77 fields of MUSE-Wide with different bin widths. From left to right, the bin widths are 0.05, 0.1 and 0.2, respectively.

redshift bin sizes the LAEs detected in the 77 fields of MUSE-Wide (Figure 2.5) in Figure 2.6.

The narrower the redshift bins are, the more visible is the LAEs clustering. By choosing very small bin sizes, the grouping of LAEs is very clear. Such is the case of the left panel of Figure 2.6, around  $z = 3.5$  or  $z = 4.7$ , where the number of LAEs is almost double as the closest bins around these redshifts. This grouping is already seen with bin sizes of 0.1 but it is not so obvious when I choose  $\Delta z = 0.2$ .

## 2.2 MUSE-Deep survey

The MUSE Hubble Ultra Deep Survey contains nine MUSE fields covering 90% of the entire HUDF region,  $3.6' \times 3.6'$ , with an exposure time of 10 h. Besides, there is a deeper field located within the XDF region with 31 h exposure time that covers an area of  $1.15 \text{ arcmin}^2$ . Therefore, many of the LAEs detected by MUSE-Deep are much fainter and less massive than the LAEs detected in MUSE-Wide with 1 h exposure time.

In the first data release (Bacon et al., 2017), 6288 sources were extracted taking the HST body locations as prior. Moreover, a blind search of emission line galaxies was performed, getting 1251 galaxies in the mosaic and 306 in the ultra-deep field.

MUSE-Deep is the deepest spectroscopic survey ever performed, with  $\sim 100$  h of exposure time, it detects almost an order of magnitude more redshift data than the one got on the UDF in the last decade. From those detections, 757 LAEs are within the range  $3 < z < 6$  and are represented in Figure 2.7.

As well as in MUSE-Wide, the large range of redshifts and the high quality of the data, offers the chance of carrying out new studies to test the physical properties of galaxy populations and their environments. For example, Gu erou et al. (2017) studied the stellar kinematics of galaxies at redshift  $0.2 \leq z \leq 0.8$ , Maseda et al. (2017) the properties of CIII emitters, Finley et al. (2017) the Fe II\* emission in star forming galaxies, Ventou et al. (2017) the evolution of galaxy merger fraction since  $z \approx 6$ , Contini et al. (2016) the gas kinematics and Drake et al. (2017b) the faint end of the Lyman-alpha luminosity function.

Studying the clustering of LAE in MUSE-Deep is of huge interest. It will bring the opportunity to compare the clustering of LAEs between the main two surveys and it will be seen that, even with such a small coverage, a clustering signal will already be seen.

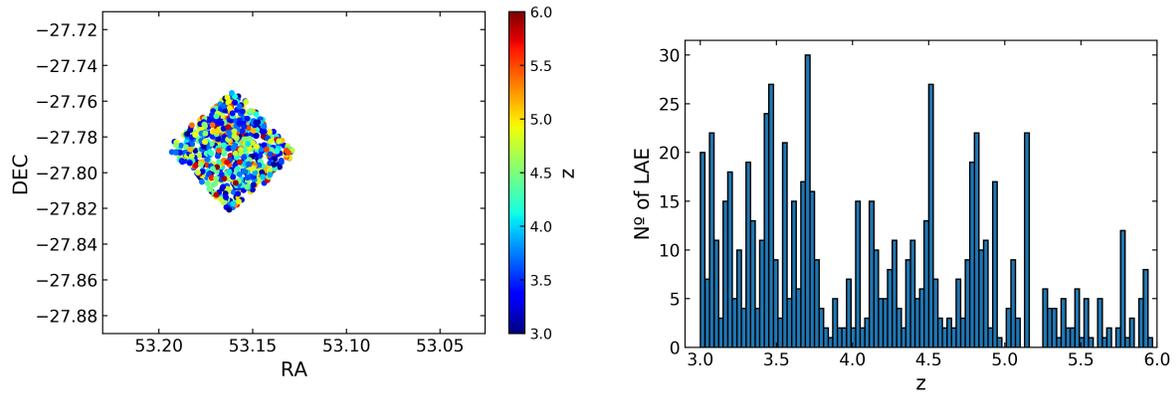


Figure 2.7: Left: sample of galaxies in the 9 fields of MUSE-Deep shown as DEC, RA and the colorbar showing different depths in redshift. On the right, the histogram represents 757 LAEs as function of redshift.

# Chapter 3

## Redshift clustering method

The large redshift range and the small area of the sky covered make MUSE data very special and different from other surveys. Therefore, we should try to find out how we can measure the galaxy clustering in this kind of data. In the first place, due to these characteristics, it is more convenient to study the clustering of the LAEs along the redshift distance instead of performing a spatial analysis with the traditional method. In this way, the redshift clustering method will be the first method that I will apply.

The redshift clustering method was introduced by [Adelberger et al. \(2005\)](#), who studied 28500 photometrically selected galaxies in the redshift range of  $1.4 \leq z \leq 3.5$ , covering an area of  $0.81 \text{ deg}^2$ . First, they studied the clustering with the standard method, the angular clustering, to use the new method afterwards and compare the outputs, getting totally compatible results.

One of the great advantages of this method is the lack of a random sample. As it was already explained in [Chapter 1](#), in the spatial clustering method the creation of a much larger random sample with the same three dimensions as the real sample is crucial. Computationally talking, this increases the computational time by a huge factor (see [Chapter 6](#)). On the other hand, the visualization of the method and the consequent results are not as clear as with the spatial clustering method.

Before explaining and applying the method, one should know the distance between the sampled galaxies. The position on the sky of galaxies can be measured accurately but, measuring the radial distance of these objects is something very different and several distance estimators that depend on the chosen cosmological model should be utilized. In this thesis, [Hogg \(1999\)](#) will be followed.

In the specific case of MUSE samples, after fitting the Lyman-alpha emission line from the detected galaxies, MUSE catalogs provide the three dimensions for each galaxy detection, RA, DEC, and  $z$ . Thus, distances need to be derived from these physical variables.

In the first place, from the cosmological redshift of the galaxies, the comoving distance is derived to calculate, afterwards, the spatial clustering. The comoving distance is defined as the proper distance divided by the scale factor, being the total line of sight comoving distance,  $D_C$ , the integration of all redshift contributions

$$D_C = D_H \cdot \int_0^z \frac{dz'}{E(z')} \quad (3.1)$$

where  $D_H$  is the Hubble distance,  $D_H = \frac{c}{H_0}$ , and  $H_0$  the Hubble constant. The function  $E(z')$  is proportional to the time derivative of the logarithm of the scale factor (*i.e.*,  $\dot{a}(t)/a(t)$ ) and it is defined as

$$E(z') = \sqrt{\Omega_M(1+z')^3 + \Omega_k(1+z')^2 + \Omega_\Lambda} \quad (3.2)$$

where  $\Omega_M$ ,  $\Omega_k$  and  $\Omega_\Lambda$  are the baryonic, dark matter and dark energy or curvature of space density parameters, respectively. The chosen cosmological model is Flat  $\Lambda$ CDM with  $H_0 = 70 \text{ km s}^{-1}\text{Mpc}^{-1}$ ,  $\Omega_{M_0} = 0.3$  and  $\Omega_{\Lambda_0} = 0.7$ . Thus,

$$E(z') = \sqrt{\Omega_M(1+z')^3 + \Omega_\Lambda} \quad (3.3)$$

Taking into account pairs of galaxies, the comoving distance between two galaxies forming a pair is transformed into

$$z_{ij} = D_H \cdot \int_0^{z_1} \frac{dz'}{E(z')} - D_H \cdot \int_0^{z_2} \frac{dz'}{E(z')} \quad (3.4)$$

where  $z_1$  and  $z_2$  are the redshifts of the two galaxies forming each pair. This comoving distance,  $z_{ij}$ , will be also called the l.o.s. distance,  $\pi$  or the radial distance as synonyms along this thesis.

In order to calculate the transverse distance between pairs of galaxies, the angular distance between them,  $\phi$ , will be calculated from RA (or  $\alpha$ ) and DEC (or  $\delta$ ) of the galaxy 1 and 2 of the pair.

$$\phi \approx \sqrt{(\alpha_1 - \alpha_2)^2 + (\delta_1 - \delta_2)^2} \cdot \frac{\pi}{180} \quad (3.5)$$

Afterwards, doing use of the average redshift between pairs of galaxies,  $z_{av}$ , the transverse distance,  $r_{ij}$ , is derived

$$r_{ij} = D_H \cdot \int_0^{z_{av}} \frac{dz'}{E(z')} \quad (3.6)$$

It is worth mentioning again that the distances and, consequently, the statistics describing the clustering depend on the distance definitions and on the cosmological model. For close galaxies, the Hubble distance follows the linear law stated above but, as further the galaxies are, the distance redshift relation is not linear anymore due to the curvature of the space.

Once the distance between all available pairs of galaxies is calculated, the clustering method can be applied. This method is based on the comparison of pair counts,  $N$ , at close redshifts (*i.e.* different l.o.s. distances) for fixed transverse distances. In the same way as for the spatial clustering (see Section 1), an estimator,  $K_{a_1, a_2}^{b_1, b_2}$ , will be introduced to proceed with the pair comparison where  $a_1 < z_{ij} < a_2$  and  $b_1 < z_{ij} < b_2$  bins will be assumed to proceed with the pair counting.

$$K_{a_1, a_2}^{b_1, b_2}(r_{ij}) = \frac{N_{b_1, b_2}(r_{ij})}{N_{b_1, b_2}(r_{ij}) + N_{a_1, a_2}(r_{ij})} \quad (3.7)$$

$N_{a_1, a_2}$  and  $N_{b_1, b_2}$  are the number of galaxy pairs in each bin at a fixed transverse distance  $r_{ij}$ . A visual example of the method is shown in Figure 3.1. At a fixed transverse distance  $r_{ij}$ , the number of galaxy pairs is counted in the first bin (violet) along the l.o.s. corresponding to  $N_{b_1, b_2}$  and in the second bin (blue) corresponding to  $N_{a_1, a_2}$ . Thus, the galaxies represented in the right panel of the figure actually represent pairs of galaxies. Then, the first  $K_{a_1, a_2}^{b_1, b_2}$  value is calculated. Afterwards, we shift to the next transverse distance, count the pairs and estimate  $K_{a_1, a_2}^{b_1, b_2}$  with equation 3.7 following the same procedure until all transverse distances are covered.

As in simultaneous,  $K_{a_1, a_2}^{b_1, b_2}$  is related to the 2pcf through its mean value

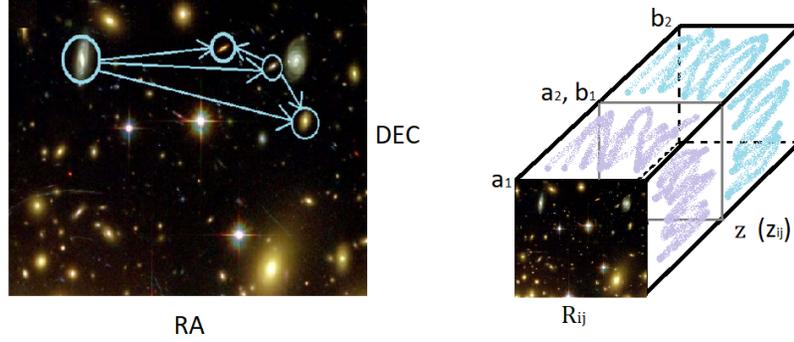


Figure 3.1: Redshift clustering method. Left: galaxy pairs counting in a RA-DEC representation. Right: The colors violet and blue represent the two bins along redshift or l.o.s. distance,  $z_{ij}$ , where the pair comparison is done at fixed selected transverse distances,  $r_{ij}$ . The galaxies represent pairs of galaxies in this  $r_{ij} - z_{ij}$  representation.

$$\langle K_{a_1, a_2}^{b_1, b_2}(r_{ij}) \rangle \approx \frac{(b_2 - b_1) \cdot \sum_{i>j}^{pairs} (1 + \bar{\xi}_{b_1, b_2})}{(b_2 - b_1) \cdot \sum_{i>j}^{pairs} (1 + \bar{\xi}_{b_1, b_2}) + (a_2 - a_1) \cdot \sum_{i>j}^{pairs} (1 + \bar{\xi}_{a_1, a_2})} \quad (3.8)$$

where  $\bar{\xi}_{a_1, a_2}$  is

$$\bar{\xi}_{a_1, a_2} = \frac{1}{a_2 - a_1} \int_{a_1}^{a_2} dz_{ij} \cdot \xi(r_{ij}, z_{ij}) \quad (3.9)$$

and  $\xi(r_{ij}, z_{ij})$  is the Limber equation (eq 1.2) in spatial coordinates (*i.e.*,  $\xi(r_{ij}, z_{ij}) = (r/r_0)^{-\gamma}$ ) with  $r = \sqrt{r_{ij}^2 + z_{ij}^2}$ ,  $r_0$  as the correlation length and the standard value for the slope,  $\gamma = 1.8$  (Diener et al., 2017; Zehavi et al., 2002).

Hence, by using equation 3.7, I calculate the values for  $K$  that represent the clustering and equations 3.8 and 3.9 to fit the estimator. In this way, the correlation length and the  $\gamma$  factor are determined for each sample of galaxies. In opposition to the spatial clustering case where no clustering was given by  $\xi(r) = 0$ , in this method  $K_{a_1, a_2}^{b_1, b_2} = 0.5$  gives the no clustering level.

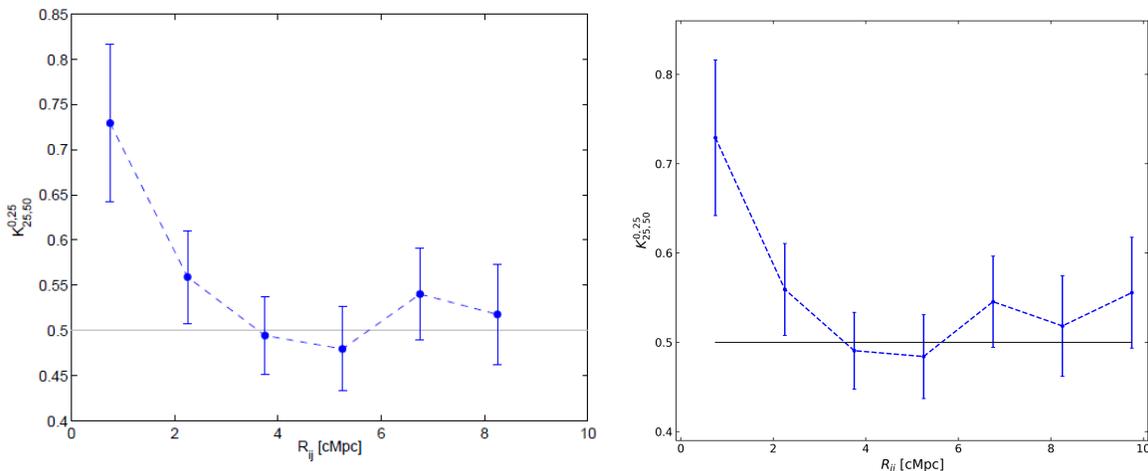


Figure 3.2: Left:  $K_{0,25}^{25,50}(r_{ij})$  estimator variation with the transverse pair distance  $r_{ij}$  for the first 24 fields of MUSE-Wide got by Diener et al. (2017). The gray line represents the expectation for no clustering. Right: the same got during this master work.

### 3.1 MUSE-Wide 24 fields

After Adelberger et al. (2005), this method was used by Diener et al. (2017) in a pilot study of the first 24 MUSE-Wide fields available at that time. Therefore, as a training purpose, the first actual goal of the master thesis was to write the codes and reproduce the results shown in Diener et al. (2017).

Once the comoving distance between pairs of galaxies was derived with equation 3.4, the range of l.o.s. distances chosen was comprehended between 0 and 50 cMpc (comoving Mpc) from which the two bins are created, being  $0 < z_{ij} < 25$  and  $25 < z_{ij} < 50$  cMpc. By binning and fixing the transverse distance calculated with equation 3.6 and counting pairs along the l.o.s. up to 25 and up to 50 cMpc for each fixed  $r_{ij}$  bin, the estimator  $K_{0,25}^{25,50}(r_{ij})$  is calculated and represented in Figure 3.2. The reason why these ranges in  $z_{ij}$  and  $r_{ij}$  were chosen is explained in Section 3.4 and the written code is shown in Appendix 7.2. In this way, the  $K$  estimator is essentially the ratio of LAE pairs within 25 cMpc and the pairs within 50 cMpc along l.o.s. separation between galaxy pairs.

The estimator  $K_{0,25}^{25,50}$  in Diener et al. (2017) and the one calculated in this thesis are basically the same, both showing a clear clustering signal at  $r_{ij} \approx 3$  cMpc, where the

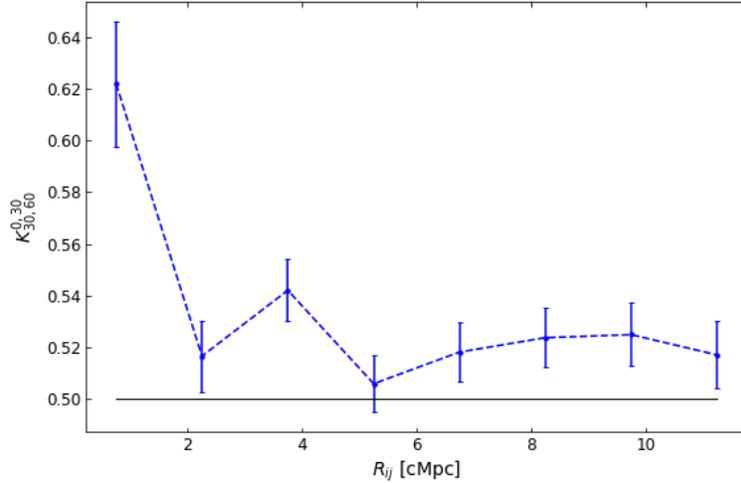


Figure 3.3:  $K_{0,30}^{30,60}$  estimator variation with the transverse pair distance  $r_{ij}$  for the 69 fields of MUSE-Wide survey. The black line represents the expectation for no clustering and the error bars are poissonian.

estimator steeply rises. In both cases, the error bars are calculated in a poissonian way,  $\sqrt{num}/den$ , being  $num$  and  $den$  the numerator and denominator of the estimator  $K_{a_1, a_2}^{b_1, b_2}$  in equation 3.7, respectively. By fitting the increase of the estimator and using equations 3.8 and 3.9, I calculate the correlation length,  $r_0$ , as well as its upper and lower limits, which represent the  $1\sigma$  errors from a  $\chi^2$  analysis. While Diener et al. (2017) got  $r_0 = 2.9_{-1.1}^{+1.0}$  cMpc, I calculated  $r_0 = 3.0_{-1.2}^{+0.9}$  cMpc, being both values completely in agreement.

## 3.2 MUSE-Wide 69 fields

Following the same procedure, the same analysis was extended to the full MUSE-Wide survey. In this case, the range of l.o.s. distance chosen was comprehended between 0 and 60 cMpc from which the two bins are created, being  $0 < z_{ij} < 30$  and  $30 < z_{ij} < 60$  cMpc. The estimator for the full MUSE-Wide sample is shown in Figure 3.3.

The estimator  $K_{0,30}^{30,60}$  shows a clear clustering signal at  $r_{ij} \approx 2$  cMpc, where there is an abruptly increase. By fitting the rise of the estimator and using equations 3.8 and 3.9, I derived a correlation length of  $r_0 = 2.7_{-0.4}^{+0.3}$  cMpc. This correlation length for  $\bar{z} \sim 4.07$

is in agreement with the correlation lengths got by [Gawiser et al. \(2007\)](#) and [Ouchi et al. \(2003\)](#). The former calculated a correlation length of  $r_0 = 3.6_{-1.0}^{+0.8}$  Mpc from the angular clustering of 162 LAEs at  $z \approx 3.1$ , while the latter obtained a correlation length close to mine,  $r_0 = 3.5_{-0.3}^{+0.33}$   $h^{-1}$ Mpc also from the angular clustering of 87 LAEs at  $z = 4.86$ . In both cases the selection technique was by using NB filters. Moreover, [Ouchi et al. \(2010\)](#) calculated the angular correlation function and the correlation length of 207 LAEs at  $z = 3.7$  and  $5.7$  in the Subaru/XMM-Newton Deep Survey field, getting  $r_0 = 2.74_{-0.72}^{+0.58}$   $h^{-1}$ Mpc and  $r_0 = 3.12_{-0.36}^{+0.33}$   $h^{-1}$ Mpc, respectively. Finally [Shioya et al. \(2009\)](#) studied the angular clustering of 79 LAEs at  $z \approx 4.86$  selected by NBs in the COSMOS field. They converted the angular clustering into the spatial clustering by using Limber equation to obtain a two point correlation function with huge error bars in the correlation length,  $\xi(r) = (r/4.4_{-2.9}^{+5.7})^{-1.9 \pm 0.22}$ .

However, despite the agreement of the correlation lengths with the literature and as it is clear in [Figure 3.4](#), both clustering signals for 24 and for 69 fields show very similar increase in the  $K$  estimator but the shapes differ. This is likely due to the larger area covered in the 69 fields case which might decrease the ratio of number of galaxy pairs in  $0 < z_{ij} < 30$  and  $30 < z_{ij} < 60$  cMpc, causing a decrease in the  $K$  estimator and, therefore, a decrease in the clustering signal.

Besides, apart from the different shapes,  $K_{0,30}^{30,60}$  has an excess around 4 cMpc and most of the error bars do not agree with the 0 clustering straight line. This will be studied in more detail with the use of a random sample in [Chapter 4](#). Nevertheless, both correlation lengths agree with each other, being compatible within the error bars.

### 3.3 MUSE-Deep

The same process has been followed for the MUSE-Deep survey. In the same way as I did for the MUSE-Wide survey, both comoving distances are calculated and binned, being  $0 < z_{ij} < 25$  and  $25 < z_{ij} < 50$  cMpc the chosen ranges along the l.o.s. distance. The estimator  $K_{0,25}^{25,50}(r_{ij})$  is calculated and represented in [Figure 3.5](#). Since MUSE-Deep covers a much smaller area on the sky, the transverse distances where I studied the clustering are also smaller than in previous cases.

Taking a closer look at the [Figure 3.5](#), it is possible to appreciate that there seems to be an offset above the 0 clustering line much more clear than for the 69 fields case. This can be explained if we take into account that MUSE-Deep is very small compared to MUSE-Wide, where the clustering signal converged at around  $r_{ij} \sim 3$  Mpc and, in this case, we do not reach those transverse distances.

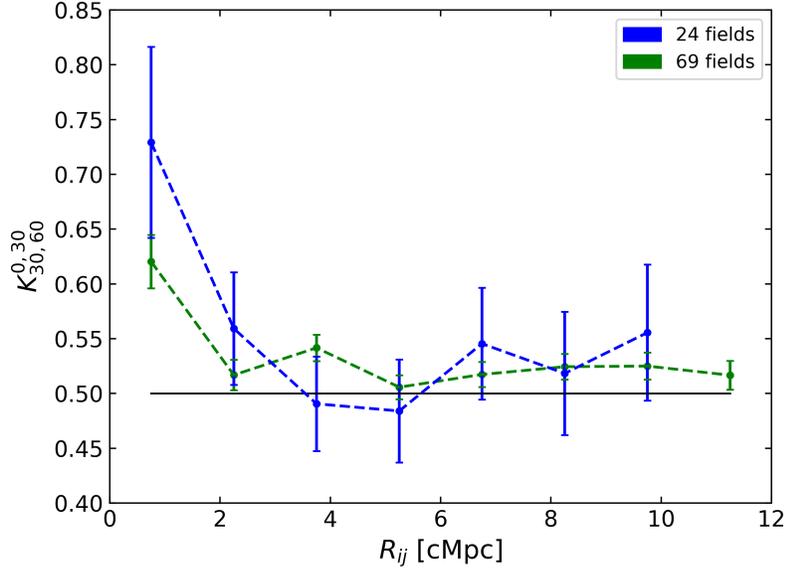


Figure 3.4:  $K$  estimator variations with the transverse pair distance  $r_{ij}$  for the full MUSE-Wide survey (green) and for the first 24 fields (blue). The black line represents the expectation for no clustering and the error bars are poissonian.

However, even in this small area of the sky there is an important increase in the  $K$  estimator, steeper than in the MUSE-Wide survey. This means that the fainter galaxies that are detected in MUSE-Deep but not in MUSE-Wide increase the clustering signal considerably. By fitting the increase of the estimator and using equations 3.8 and 3.9, the correlation length,  $r_0$ , is estimated as well as its upper and lower limits, getting a value of  $r_0 = 2.6_{-0.2}^{+0.2}$  cMpc. In the same way as in the MUSE-Wide survey, this length is compatible with most of the values from the literature mentioned in the previous section that are around  $\bar{z} \approx 4.14$ . Besides, the correlation lengths from MUSE-Wide and MUSE-Deep are in agreement as well.

### 3.4 Studying the $K$ estimator in detail

The values of the  $K$  estimator depend on several parameters. First, the chosen range of  $z_{ij}$  plays an important role as Figure 3.6 shows. Depending on the range selected,  $K$  values vary in important quantities. For instance, by choosing very low  $z_{ij}$  range (black

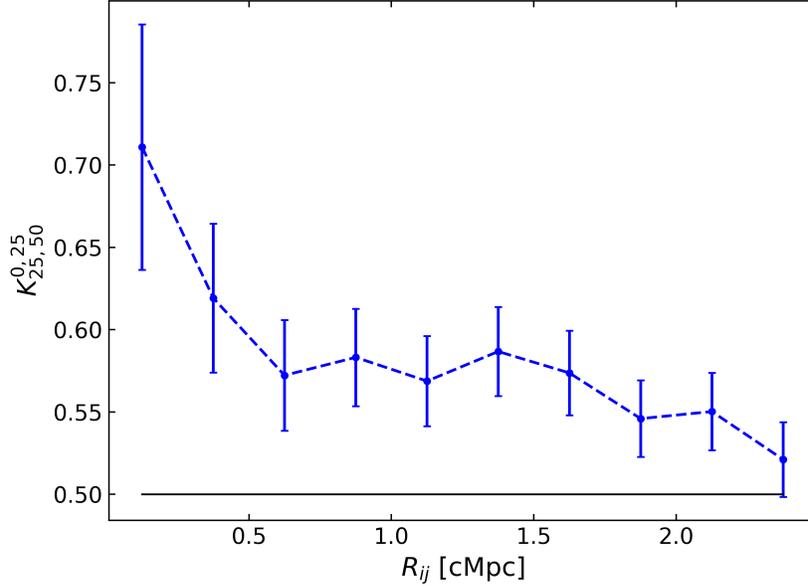


Figure 3.5:  $K_{0,25}^{25,50}$  estimator variation with the transverse pair distance  $r_{ij}$  for the MUSE-Deep survey. The black line represents the expectation for no clustering and the error bars are poissonian.

estimator in the figure), we get a much pronounced clustering signal, while choosing high  $z_{ij}$  ranges makes the signal much lower until its disappearance.

$Z_{ij}$  tells us how far away is one galaxy along the l.o.s. with respect to another one. Since MUSE covers redshifts from 3 to 6, the available range goes to almost 2000 cMpc (*i.e.* 2000 cMpc would be the distance between a pair of galaxies at redshift 3 and 6, respectively). Of course two galaxies separated by 2000 cMpc are not considered as a pair anymore. Following this reasoning,  $K_{0,30}^{30,70}$  drops its signal because the galaxies start to be far away to consider them as pairs. On the other hand,  $K_{0,10}^{10,20}$  does not contain many galaxy pairs (too low distance range that does not cover enough volume) which also augments the error bars and only probes small clustering.

Therefore we need to choose a range in between these low and high ranges where the signal converges, where clustering at small and large scales is accounted and where the minimum possible galaxy pairs is missing. These requirements are accomplished for  $K_{0,25}^{25,50}$  in the first 24 fields of MUSE-Wide and MUSE-Deep and for  $K_{0,30}^{30,60}$  in the 69 fields of MUSE-Wide survey.

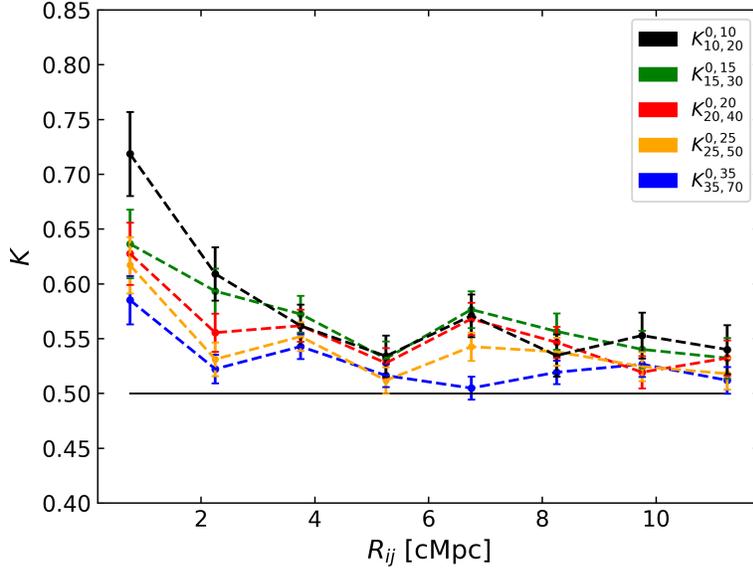


Figure 3.6: Different  $K$  estimators shown in different colors varying with the transverse pair distance  $r_{ij}$  for the full MUSE-Wide survey. The black line represents the expectation for no clustering and the error bars are poissonian.

Secondly, electing different bin sizes in  $r_{ij}$  also modifies the  $K$  estimator shape as it can be seen in Figure 3.7. The use of larger bins smooths completely the throughout of the redshift clustering method, while smaller bins increase all the tiny details of the  $K$  curve.  $R_{ij}$  cannot be larger than the total distance spanned by the total full MUSE-Wide survey because there would be no galaxy pairs further than that limit. As already mentioned, the survey covers an area of  $1' \times 1'$  per field which is translated into maximum values of  $r_{ij} \sim 24$  cMpc (*i.e.* 24 cMpc would be the transverse distance of a pair of galaxies with the minimum and maximum RA and DEC values at a mean redshift of 4.02). Since that distance covers the MUSE edges and there are not so many detected galaxies at those points, the maximum  $r_{ij}$  value is chosen to be 12 cMpc.

Finally, different selection of fields with different exposure times also make a change in the  $K$  estimator, though softer than in previous cases (see Figure 3.8). This can be explained with the decrease of pairs due to the lower number of fields. Moreover, referring to Chapter 2, the 9 udf fields excluded have 1.67 h exposure time against the 1 h exposure of the other 60 fields, which gives a higher number of galaxies per field. The number of

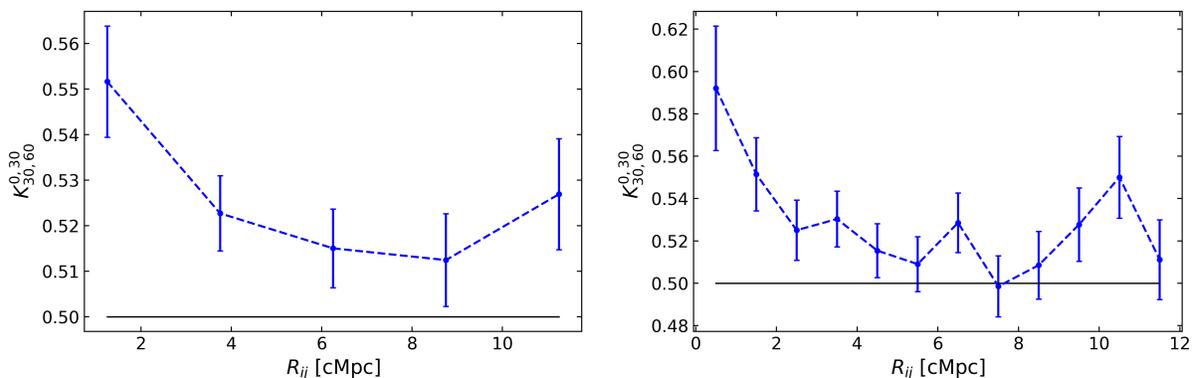


Figure 3.7: Left:  $K_{0,30}^{30,60}$  estimator variation with the transverse pair distance  $r_{ij}$  for the full MUSE-Wide survey, choosing a total of 4 bins. Right: the same with 11 bins. In both plots, the straight line represents the expectation for no clustering.

galaxies in the 9 fields is 266, which is an important quantity compared to the total number of galaxies in the 60 fields survey (947 LAEs).

Despite the changes in the shape of the  $K$  curve or in the maximum value of clustering signal reached, the correlation length barely changes as in the bin sizes case. Including or not the 9 udf deeper fields changes  $r_0$  very slightly, from  $r_0 = 2.7_{-0.4}^{+0.3}$  to  $r_0 = 2.6_{-1.0}^{+0.6}$  cMpc but different  $K$  estimators (*i.e.* different  $z_{ij}$  ranges to count LAEs pairs) change this value in a stronger way. For example, choosing extreme estimators,  $K_{0,10}^{10,20}$  and  $K_{0,35}^{35,70}$ , changes the  $r_0$  from  $r_0 = 3.2_{-0.3}^{+0.3}$  to  $r_0 = 2.9_{-0.4}^{+0.3}$  cMpc.

Due to these different  $r_0$  values, to the different maximum values of the  $K$  estimator using 24 and 69 fields and to the offset above the 0 clustering line, I wanted to test the redshift method. In order to carry out this test, the  $K$  method will be run over a random sample created from our LAEs samples in the next chapter.

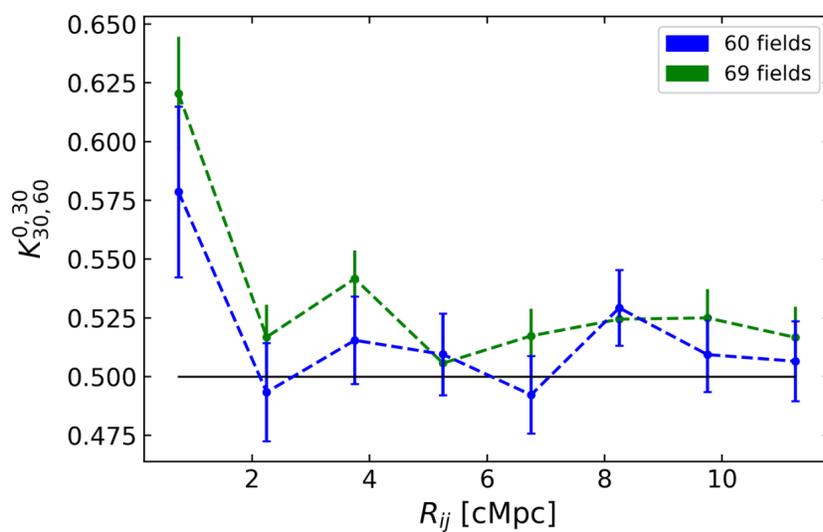


Figure 3.8: Left:  $K_{0,30}^{30,60}$  estimator variation with the transverse pair distance  $r_{ij}$  for the full MUSE-Wide survey (69 fields in green). Right: the same with the 60 fields in blue. In both plots, the straight line represents the expectation for no clustering.

# Chapter 4

## Analysis of a random sample of galaxies

Testing that we are not biased not only gives information about the possibility of having issues inside the method itself but also about the error bars. The best way to test the redshift clustering is by running it over a random sample of galaxies where it is expected to present 0 clustering.

Nevertheless, we cannot take any random sample. The random sample should be realistic, it should account for the LAEs contained in the MUSE samples. In order to achieve this, the survey footprint, the selection function and the luminosity function of our galaxies must be taken into account.

- **Selection function:** As its name indicates, the selection function accounts for the selection effects in the instruments of the telescopes, selection methods, depth, for the way the surveys are built, etc. In our case, the number of galaxies decreases as the distance increases because the surveys are flux-limited. This produces that only bright enough galaxies above the limiting apparent magnitude of the survey are detected. Thus, only the most luminous galaxies are seen at large distances.

In addition to the flux limitation, the sky is not isotropically transparent to the other galaxies' light because of the absorption by dust in our galaxy. The flatness of the galaxy make some regions more obscure than others which affects the brightness of an external galaxy and depends on the direction of the l.o.s.. Besides, the way the samples are built (*i.e.* the masks used for each field, fibre collisions in the spectrographs, taking into account that the luminosity of the galaxies at deep redshifts is

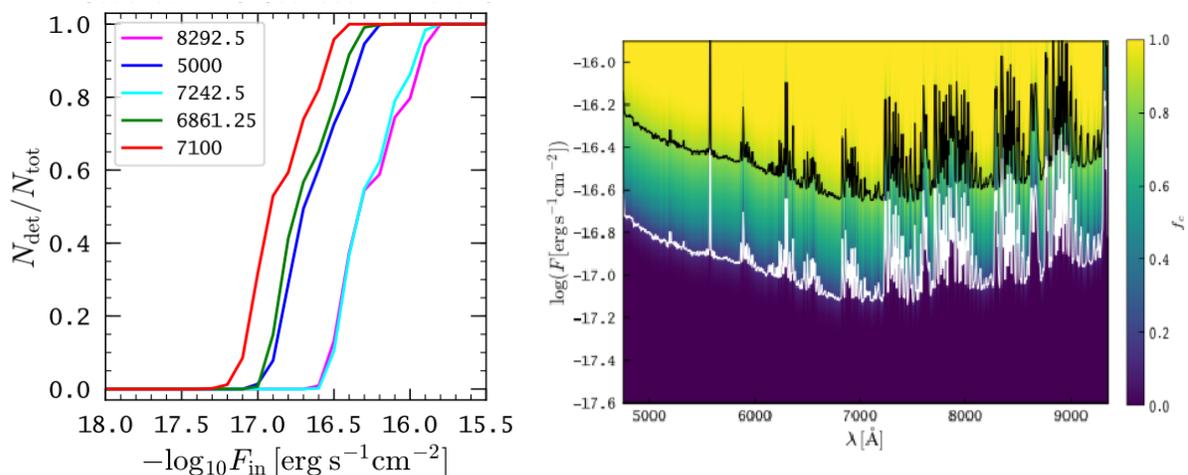


Figure 4.1: Left: Ratio of the stack over the fractions of detected sources,  $N_{\text{det}}$ , and the total number of detections,  $N_{\text{total}}$ , of the ten different MUSE HDFS LAEs used (see (Herenz et al., 2017)). Each colored curve represents the selection function at a different wavelength. Right: Selection function  $f_C(FLy\alpha, \lambda)$  shown as a colorbar for LAEs in the MUSE-Wide survey. The white and black lines indicate the 15% and 85% completeness limits, respectively.

detected at longer wavelength than the one emitted, etc) has also an impact on what type of galaxies are selected, hence, on the selection function.

In order to take into account this incompleteness, the selection function is calculated, providing the probability that a galaxy is detected at a given distance (see Figure 4.1). The selection function is not a step function, there is no fixed flux value above which all galaxies are detected and below which there are no detections. Instead, the selection function is a curve that depends on wavelength.

Traditionally, the selection function is estimated from the luminosity function.

- **Luminosity function:** Given a range of intrinsic luminosities, the luminosity function,  $n(L)$ , is defined by the number density of galaxies in that range. It is a distribution function, a distribution of luminosities within a given range of luminosities.

Luminosity measurements are hard to achieve because the total luminosity requires the exact distance to the object and the integration of the spectrum of the object over all wavelengths. Thereby, instead of the total luminosity, the specific luminosity over a given band is measured, giving the number of galaxies at the frequency range

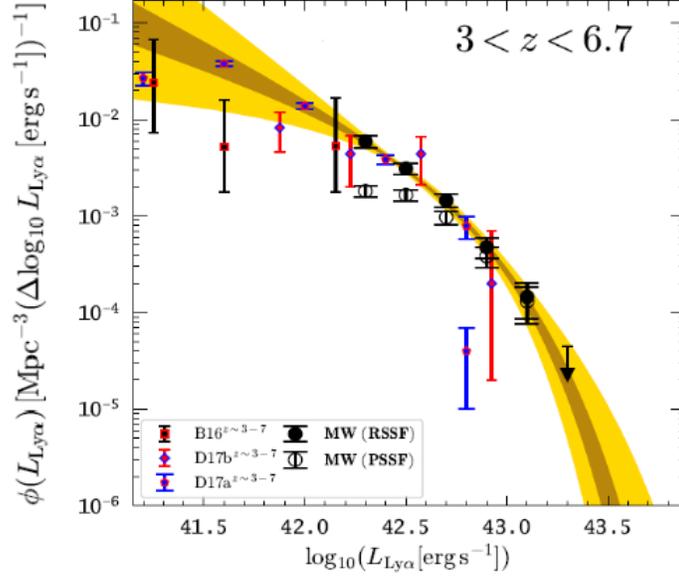


Figure 4.2: Differential MUSE-Wide LAE luminosity function for the total redshift range of MUSE-Wide  $3 < z < 6.7$  from [Herenz et al. \(2019\)](#). Filled black circles show the real source selection function corrected binned, open black circles show the point source selection function corrected binned, blue open circles with red error bars the binned estimates by [Drake et al. \(2017b\)](#), red squares with blue error bars show the binned estimates by [Drake et al. \(2017a\)](#) and the ones by [Bina et al. \(2016\)](#) are the red squares with black error bars. Yellow and dark yellow shaded regions indicate the 68.3% and 95.4% confidence regions for a Schechter parameterisation, respectively.

of the band.

Generally, the Schechter luminosity function (Equation 4.1) is the traditional form to describe the space density of galaxies as a function of their luminosity,  $L$ . This function depends on different parameters that change for different morphologies, environments or redshifts due to galaxy evolution. The luminosity function is formed by an exponential and a power law.

$$n(L)dL = \phi^* \left(\frac{L}{L^*}\right)^\alpha e^{-L/L^*} \frac{dL}{L^*} \quad (4.1)$$

$L^*$  is the characteristic galaxy luminosity that separates faint and bright galaxy ranges in the power law and  $\phi^*$  is the normalization parameter in units of number density.

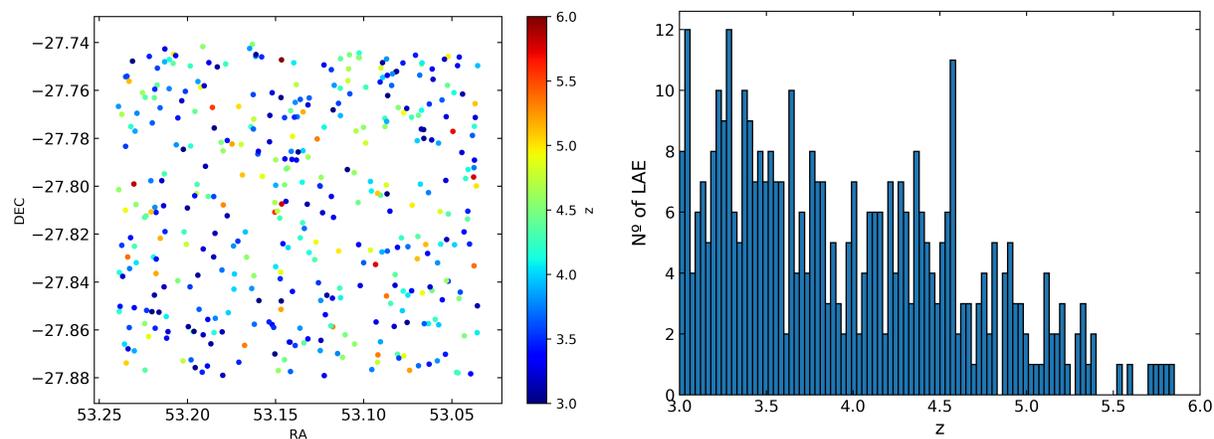


Figure 4.3: Left: random sample of galaxies in DEC and RA distribution. The colorbar shows different depth in redshift. Right: Number of the 400 random LAEs with redshift shown in form of a histogram.

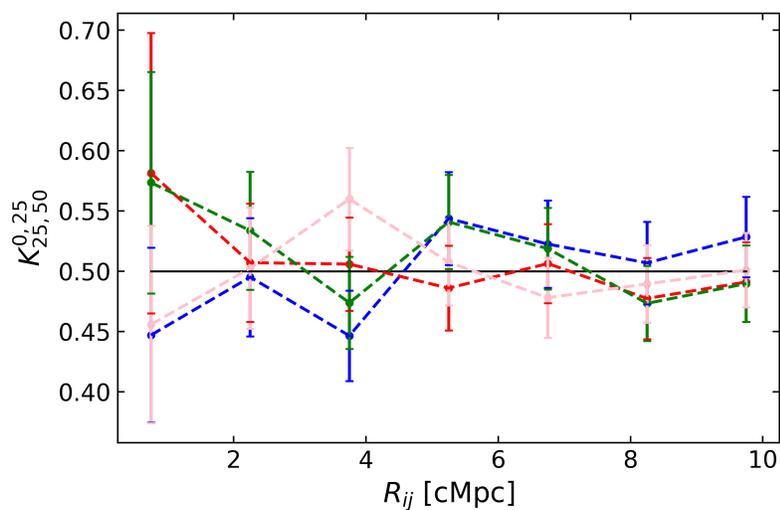


Figure 4.4:  $K_{0,25}^{25,50}$  estimator variation with the transverse pair distance  $r_{ij}$  for four different random samples (different colours) created from the full MUSE-Wide survey. The black straight line shows no clustering.

The luminosity function from LAEs in MUSE-Wide was calculated by [Herenz et al. \(2019\)](#) and it is shown in Figure 4.2. It is visible that the luminosity function decreases with the luminosity of the Lyman-alpha emission line, detecting more galaxies with low luminosities than with large luminosity values.

Taking into account both functions with flat flux assumption, the realistic random sample of LAEs from MUSE surveys is created and shown in Figure 4.3. The  $K$  estimator was afterwards tested on this random sample in the same way as for the real samples and is represented in Figure 4.4.

As we see in Figure 4.4, the random sample does not show any signal clustering, being consistent within the error bars with  $K = 0.5$ . Furthermore, the correlation lengths calculated for random samples have values of the order of  $r_0 = 1.0_{-0.5}^{+1.1}$ ,  $r_0 = 0.1_{-0.1}^{+0.1}$ ,  $r_0 = 0.9_{-0.9}^{+1.5}$  Mpc, etc... which give mean null correlation lengths.

In this way, apart from proving that we do not introduce any excess of clustering signal in our MUSE samples calculations and that we get the expected 0 clustering for a random sample, I can also study how these errors affect our measurements.

In order to estimate the errors that are introduced by using this method and to compare them to the poissonian ones calculated along this thesis, I calculate the  $K$  estimator over 100 different random samples, calculating their corresponding Poisson errors the 100 times. Afterwards, the average of the 100 times of the  $K$  estimator with the average of the 100 times Poisson errors is represented in Figure 4.5.

Having the mean error bars of the 100 random samples gives us an estimation of the excess clustering that we may introduce with this method as a systematic error. The mean values of  $K$  with their corresponding mean error bars above the 0 clustering straight line in Figure 4.5 provide the excess introduced in our measurements. However, the majority of data points with their errors fall on the no clustering line with a small discrepancy in two of the dots. Looking closer at these two points, they have a discrepancy of 0.1% and 0.4% with respect to the 0 clustering line once the error bars are taken into account, a quantity that is not significant in comparison with the whole set of points.

As expected, the errors for the  $K$  values decrease with higher number of galaxies as can be already seen in Figure 3.4. The errors seen in Figure 4.5 are much smaller than the Poisson errors calculated for the clustering of the 24 fields, the full MUSE-Wide and the MUSE-Deep surveys. Therefore, we can conclude that the errors introduced by the method are insignificant and we are not biased when we use the redshift method since we get the expected 0 clustering for a realistic random sample of LAEs that cover the same area as the MUSE-Wide survey.

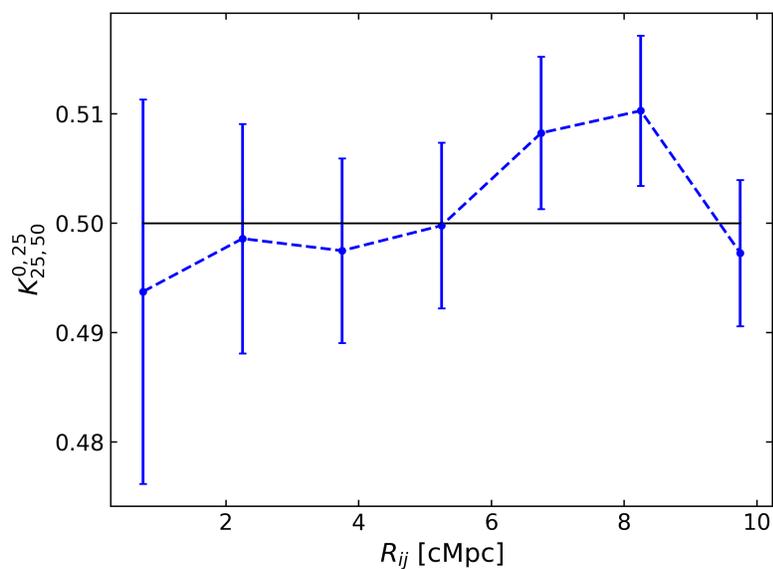


Figure 4.5: Mean  $K_{0,25}^{25,50}$  estimator of the 100 different random samples created as a function of transverse pair distance. The black straight line shows no clustering and the error bars are the errors of the average K values, the mean of the 100 Poisson error calculations for each different random sample divided by  $\sqrt{100}$  (i.e.  $\frac{\sum_{i=1}^{100} \frac{\sqrt{num}}{den}}{100} \cdot \frac{1}{\sqrt{100}}$ ).

# Chapter 5

## Clustering dependence on physical properties

Different physical processes and properties are needed to explain environmental trends in star formation and in galaxy structure. By using clustering signals, one can also trace the dependence of large scale structure on galaxy properties such as luminosity or star formation rate, stellar mass, and track its evolution with redshift. Up to date, the general clustering trend as function of properties that has been found is that more luminous, early type, redder and more stellar massive galaxies are clustered stronger than fainter, late type, bluer and less stellar massive galaxies.

[Zehavi et al. \(2002\)](#) carried out the first clustering analysis in the SDSS survey. With a total of 29.300 galaxies in the range  $0.019 < z < 0.13$  and in an area of  $690 \text{ deg}^2$ . The sample was divided by color, showing that red galaxies exhibit stronger and steeper correlation functions than blue galaxies. For galaxies with absolute magnitudes around  $M^* \sim -1.5$ ,  $M^*$  and  $M^* \sim 1.5$ , they calculated correlation lengths of  $r_0 = 7.4, 6.3$  and  $4.7 h^{-1}\text{Mpc}$  with a slope of  $\gamma = 1.8$ . For red galaxies they calculated  $r_0 \approx 6.78 h^{-1}\text{Mpc}$  with a correlation slope  $\gamma \approx 1.86$ , while for blue galaxies the correlation length and clustering slope were  $r_0 \approx 4.02 h^{-1}\text{Mpc}$  and  $\gamma \approx 1.41$ , respectively.

[Kauffmann et al. \(2004\)](#) found that star formation and nuclear activity depend strongly on local density for a given stellar mass, whereas parameters like size or concentration are nearly independent of it. [Li et al. \(2006\)](#) used a sample of 200000 galaxies from the SDSS in the range  $0.01 < z < 0.03$  to study the evolution of the spatial clustering with stellar mass, colour,  $4000 \text{ \AA}$  break strength, contraction index and stellar surface mass density. They showed that more massive galaxies, redder, with larger  $4000 \text{ \AA}$  break, with larger

concentrations or higher surface mass densities cluster stronger.

On small scales, the clustering amplitude increases steeply for galaxies brighter than  $L^*$ , whereas below that limit it barely varies. On larger scales, clustering amplitude and luminosity are positively correlated (consistent with the studies carried out by Zehavi et al. (2005) in SDSS for galaxies between  $M_r \sim -17.5$  and  $-22.5$  (see Figure 5.1a).

g-r color and D4000 also vary the clustering amplitudes. Redder galaxies show steeper correlation functions and cluster stronger (see Figure 5.1b, c), showing a much stronger dependence than on the concentration or surface mass density. In the case of concentration or surface mass density, there is only clustering differences at small scales when the luminosity and the stellar mass are fixed to a certain value. However, no difference is appreciated between high and low concentrations or surface densities on larger scales. The variation of clustering as a function of concentration was studied by Goto et al. (2003) for galaxies between redshifts 0.05 and 0.1 in SDSS too. According to their study, in intermediate density regions ( $1-6 \text{ Mpc}^{-2}$ ) intermediate-type galaxy fractions increase towards denser regions, while late-disc galaxy types decrease. In dense regions ( $> 6 \text{ Mpc}^{-2}$ ) the opposite effect is seen (see Figure 5.1d).

Norberg et al. (2001) studied the clustering strength of galaxies in the 2dFGRS varying with luminosity. They found an increase in the clustering amplitude with absolute magnitudes. The correlation function rose four times from  $M_{bJ} - 5\log_{10}h = -18$  to  $-22.5$  and the most luminous galaxies clustered 3 times stronger than  $L^*$  galaxies. Same results were predicted by the galaxy formation models of Benson et al. (2000) and the studies carried out by Zehavi et al. (2011).

Norberg et al. (2002) and Madgwick et al. (2003) studied the clustering as function of the galaxy spectral type in 2dFGRS. Norberg et al. (2002) divided the sample into late type galaxies (with strong emission lines) and early type galaxies (more quiescent galaxies), finding that the slope of the correlation function is steeper for early types than for late types, having almost the same clustering strength dependence on luminosity. The clustering amplitude increased 2.5 times from  $L^*$  to  $4L^*$  and, at all luminosities, the correlation function amplitude for early types was double than that of late types. In this way, they showed that it is the luminosity and not the galaxy type the dominant factor that determines the clustering strength variation with luminosity. Following the same line, Madgwick et al. (2003) proved that, at small separations, passive galaxies cluster stronger than more active star forming galaxies, having slopes of 1.93 and 1.5, respectively.

The same study was carried out by Budavari et al. (2003) in the SDSS, finding that  $r_0$  increases 1.6 times with luminosity over the sampled luminosity range. In agreement with the studies mentioned above, they found that galaxies with spectral types similar to

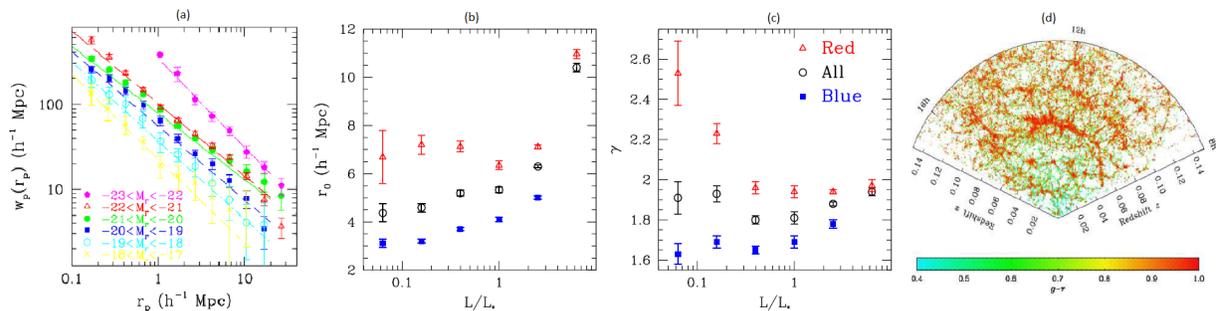


Figure 5.1: (a) Luminosity dependence of galaxy clustering in SDSS where different colors represent different absolute magnitudes from [Zehavi et al. \(2005\)](#). (b, c) Correlation length and slope as a function of luminosity. Blue, red and black represent blue, red and total number of galaxies in SDSS. (d) Distribution of galaxies in a redshift space where the different colors represent the  $g-r$  color shown in the colorbar. Last three panels were taken from [Zehavi et al. \(2011\)](#)

elliptical galaxies have a correlation length of  $\sim 6.59 h^{-1}\text{Mpc}$  and a slope of 0.96 (for the angular correlation function), while blue galaxies have a clustering length of  $4.51 h^{-1}\text{Mpc}$  and a slope of  $\gamma \approx 0.68$ .

[Kauffmann et al. \(1999\)](#) focused on the differences in clustering for galaxies with different luminosity, colour, morphology and star formation rate to learn about the galaxy formation process. The clustering evolution of galaxies selected in the rest frame B band (442 nm, blue) was compared to the clustering of galaxies selected with the I band (806 nm, red), getting at least  $0.5 h^{-1}\text{Mpc}$  higher correlation lengths for redder galaxies. In both cases,  $r_0$  increases from  $z = 1.5$  to  $z = 5$ . Then, the clustering variation with magnitudes was analyzed, obtaining that more negative magnitudes cluster stronger than higher ones. Again, in both cases the  $r_0$  curve increases over redshift. Moreover, they found that galaxies with higher stellar formation rates (SFR) cluster stronger and both  $r_0$  trends (for low and high SFR) increase after redshifts of  $\sim 2$ . Lastly, the clustering of early type galaxies at fixed stellar mass shows very similar  $r_0$  values over the redshift range  $0 < z < 3$ , getting minimum values of 6.5 and maximum values of 7.1 Mpc.

Since galaxy clustering measurements trace the underlying dark matter (DM) density field and vice versa, the use of N-body simulations to study the clustering of haloes provides insights of galaxy clustering. For a given cosmology, the galaxy clustering amplitude depends on their DM halo masses and its evolution depends on the change of the halo mass distributions over time. Therefore, a combination of N-body simulations and semi-analytic modelling of galaxy formation was carried out by [Kauffmann et al. \(1999\)](#) to analyze the

clustering evolution with redshift assuming the  $\Lambda$ CDM model ( $\Omega = 0.3$ ,  $\Delta = 0.7$  and  $H_0 = 70 \text{ km s}^{-1}\text{Mpc}^{-1}$ ) and the  $\tau$ CDM model ( $\Omega = 1$ ,  $\tau = 0.2$  and  $H_0 = 50 \text{ km s}^{-1}\text{Mpc}^{-1}$ ).

For halos of masses  $\sim 11 - 11.5M_\odot$  they found correlation lengths of  $\sim 4.0$  Mpc at  $z = 0$ ,  $\sim 3.0$  Mpc at  $z = 1$ ,  $\sim 2.8$  Mpc at  $z = 2$  and  $\sim 3.0$  Mpc at  $z = 3$ . For halos of masses  $\sim 11.5 - 12M_\odot$  they found correlation lengths of  $\sim 3.0$  Mpc at  $z = 0$ ,  $\sim 2.6$  Mpc at  $z = 1$ ,  $\sim 2.8$  Mpc at  $z = 2$  and  $\sim 3.0$  Mpc at  $z = 3$ . For halos of masses  $\sim 12 - 12.5M_\odot$  they found correlation lengths of  $\sim 3.2$  Mpc at  $z = 0$ ,  $\sim 3.4$  Mpc at  $z = 1$ ,  $\sim 3.8$  Mpc at  $z = 2$  and  $\sim 4.6$  Mpc at  $z = 3$ . In the last place, halos of masses  $\sim 12.5 - 13M_\odot$  they found correlation lengths of  $\sim 4.4$  Mpc at  $z = 0$ ,  $\sim 4.6$  Mpc at  $z = 1$  and  $\sim 5.9$  Mpc at  $z = 2$ . However, all these values are lower than using  $\tau$ CDM model due to the later structure formation assumed in this model.

Up to date, several attempts to study the clustering evolution of LAEs with redshift were done and shown in Figure 5.2 (see Chapter 6 for an explanation on how they obtained these values). It is seen that LAEs occupy denser regions of the universe at higher redshifts, shown by the increase of the clustering strength.

In the specific case of LAEs, [Ouchi et al. \(2003\)](#) studied the angular clustering of 87 LAEs selected by a combination of two broad bands and one NB in the Subaru survey ( $z \approx 4.86$ ) as a function of Lyman-alpha luminosity, UV continuum luminosity and Lyman-alpha equivalent width. They found that brighter LAEs cluster stronger and no difference between UV continuum luminosity or equivalent width was appreciated, indicating that galaxies with bright Lyman-alpha emission may be biased against the underlying dark matter halos in a stronger way than those with bright UV continuum.

In order to interpret clustering dependencies in the field of galaxy formation models, galaxy clustering is preferentially studied in terms of galaxy mass, size or mean stellar age instead of luminosity or color which are the traditional physical properties for comparison. This is due to the fact that luminosity and stellar mass do not always necessarily correlate. Depending on the dust content and on the fraction of young stars, the clustering comparison gets tricky, specially between galaxies at low and high redshifts due to the strong evolution of stellar formation rates with  $z$ .

However, in our case we only have certain properties available. In the first place, MUSE provides the largest redshift range up to date so the evolution of the LAEs in the samples can be tracked with redshift. Due to the growth of structure over different epochs caused by gravity, clustering is expected to follow the same trend. Nevertheless, its evolution depends on  $\Lambda$  and  $\Omega_M$ . A larger  $\Lambda$  gives larger voids and increases density limits between overdense and underdense areas. Studying the galaxy clustering at high redshift is essential to break these relations constrained by low redshift measurements. Thereby, determining

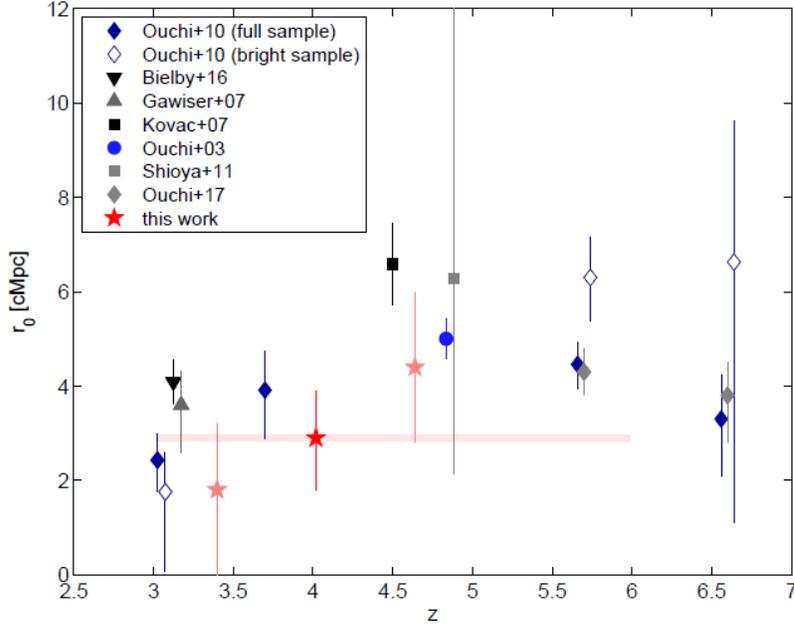


Figure 5.2: Literature values of  $r_0$  as function of  $z$ . Red stars show the calculated  $r_0$  values from the first 24 fields of MUSE-Wide split by redshift (Diener et al., 2017) while the rest of the values are calculated from NB surveys in our redshift range (red bar) shown in the legend.

the clustering evolution with redshift constrain both cosmological parameters and galaxy evolution theories.

Since overdense regions increase their density as gravity forces galaxies to move towards groups or clusters, an increase of  $r_0$  over time is expected. On the other hand, the expansion of the Universe (*i.e.* the  $\Lambda$  parameter) and the formation of new galaxies, while others grow in mass and luminosity, have the opposite effect.

In order to test these ideas, I divide the MUSE-Wide sample in low ( $3 < z < 4$ ), medium ( $4 < z < 5$ ) and high ( $5 < z < 6$ ) redshifts and I calculate the  $K$  estimator in the three ranges. The ranges selection has been done in such a way that we have enough number of galaxies (491 for low, 347 for medium and 111 for high redshifts) to be able to use the redshift clustering method. Since the range of high redshift galaxies have many less galaxies, the error bars will be larger than in the other two redshift ranges. This result is shown in Figure 5.3.

Despite the large error bars, specially for the range  $5 < z < 6$ , the tendency of the  $K$

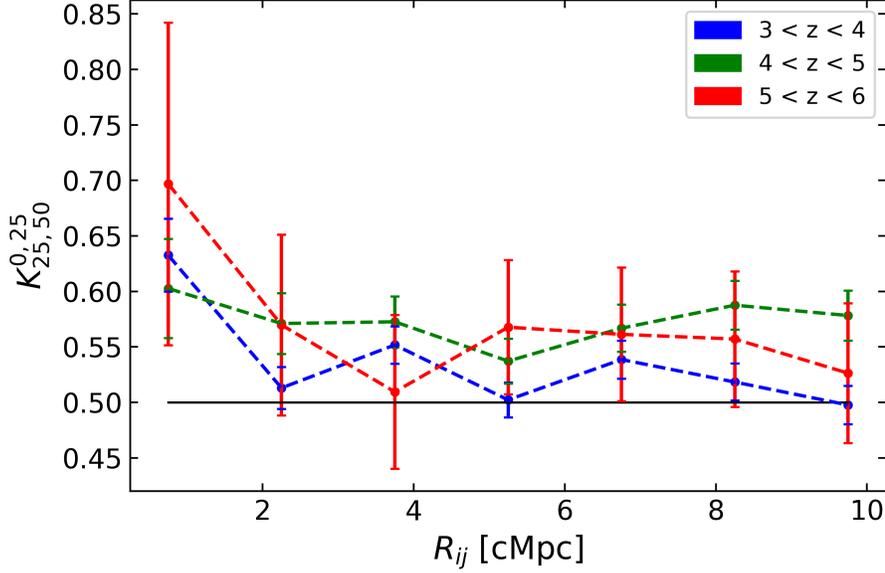


Figure 5.3:  $K_{0.25}^{0.25}$  estimator variation with the transverse pair distance  $r_{ij}$  for three different ranges of redshift of the LAEs of MUSE-Wide. The blue line shows the clustering at  $3 < z < 4$ , the green at  $4 < z < 5$  and the red at  $5 < z < 6$ . The black straight line shows no clustering and the error bars are poissonian.

estimator seems to show higher clustering signals as we increase the redshift. However, due to the errors, all redshifts seem to cluster in a similar way. The correlation length for the three cases was calculated, getting values of  $r_0 = 2.7_{-0.5}^{+0.4}$ ,  $r_0 = 3.8_{-0.6}^{+0.5}$ , and  $r_0 = 3.5_{-1.9}^{+1.4}$  for low, medium and high redshifts, respectively.

Since large samples that are so deep in redshift are not common, there are barely any study to which we can compare my results to. However, NB clustering studies can combine several NBs to study the clustering evolution with redshift. This is the case of [Ouchi et al. \(2010\)](#) whose results are in agreement with the possible trend appreciated in Figure 5.3.

In the second place, around 50% of the LAEs in MUSE-Wide present a catalogued counterpart so the spectral energy distribution (SED) of the available LAE was determined with FAST ([Kriek et al., 2009](#)). From those best SED fittings of the photometric points, parameters such as stellar ages, star formation rates and extinctions are calculated. Nevertheless, the most trustworthy parameters are the stellar masses, even if they could be several factors off since it is really hard to get the right stellar mass at high redshifts

with the available data. The LAEs stellar masses are available for a total of 44 fields from MUSE-Wide.

Thereby, the 44 fields were split in low ( $\log M < 8.5$ ) and high stellar masses ( $\log M > 8.5$ ), with a total number of galaxies of 123 and 133 for each mass range. Following the same procedure as previous times, the clustering evolution with the stellar masses is studied in Figure 5.4. Unfortunately, once again, the low number of galaxies produce large error bars that do not allow us to differentiate between the way that high and low stellar masses cluster.

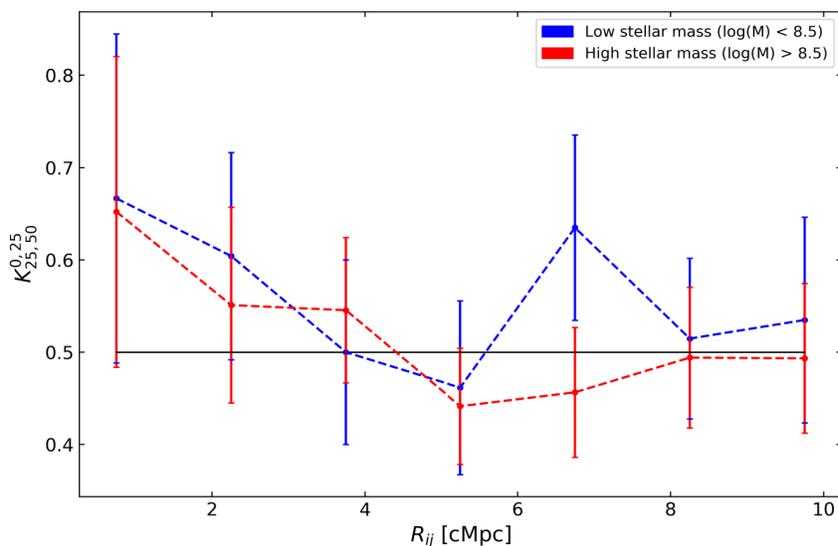


Figure 5.4:  $K_{0,25}^{25,50}$  estimator variation with the transverse pair distance  $r_{ij}$  for the LAEs with low (blue) and high stellar masses (red) in the first 44 fields of MUSE-Wide. The black straight line shows no clustering and the error bars are poissonian.

Calculating the correlation length for this case does not provide any difference between low,  $r_0 = 3.5_{-3.0}^{+2.2}$ , and high stellar masses,  $r_0 = 3.1_{-2.5}^{+1.8}$ , the errors are so large that no clustering signal difference can be appreciated.

In the third place, the variation of clustering with the equivalent width (EW) of the Ly $\alpha$  lines of the 60 fields from MUSE-Wide was studied since it is believed that the EW of the Ly $\alpha$  line and the dust content, the clumpiness of the interstellar medium and the metallicity of the galaxy are related (Finkelstein et al., 2008). In this case, I divided the MUSE-Wide survey in low (EW < 120) and high (EW > 120) EW values, with 242 and 234 LAEs in each EW division. Carrying out the same clustering analysis, the clustering variation of LAEs as function of their EW is shown in Figure 5.5.

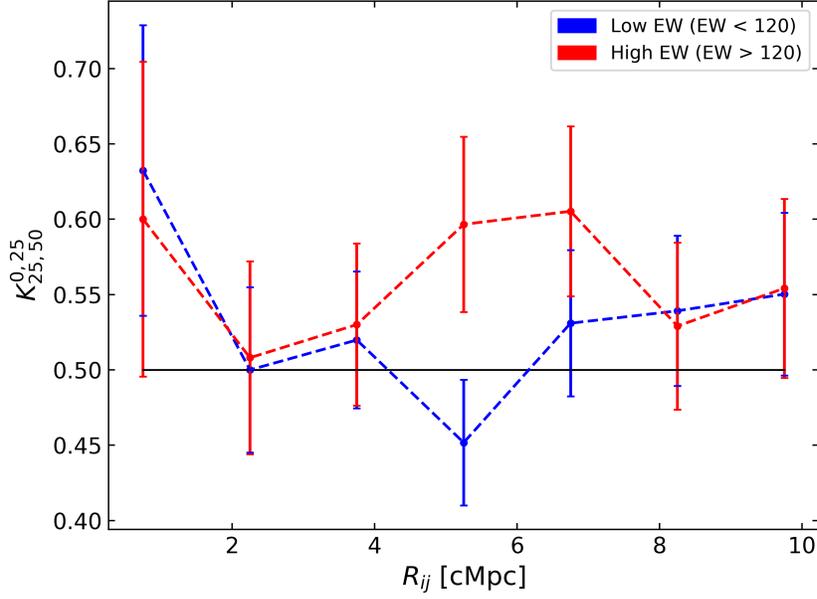


Figure 5.5:  $K_{0,25}^{0,25}$  estimator variation with the transverse pair distance  $r_{ij}$  for the LAEs with low (blue) and high (red) EW of the Lyman $\alpha$  values in the 60 fields of MUSE-Wide. The black straight line shows no clustering and the error bars are poissonian.

Once again, the low number of sampled galaxies gives large error bars that make us not distinguish between the clustering of galaxies with low and high EW. Correlation lengths were calculated for low EW,  $r_0 = 1.2^{+1.4}_{-1.2}$ , and for high EW,  $r_0 = 2.7^{+1.3}_{-1.7}$ , confirming the plot results of similar clustering within the errors for both samples. This no difference in clustering as function of EW was also obtained by [Ouchi et al. \(2003\)](#).

In the fourth place, the variation with Ly $\alpha$  luminosity of how LAEs cluster was studied. In this case, I divided the MUSE-Wide survey in low ( $\log L < 42.4$ ) and high ( $\log L > 42.4$ ) Ly $\alpha$  luminosities, with 254 and 222 LAEs in each luminosity division. Carrying out the same clustering analysis, the clustering variation of LAEs as function of their Ly $\alpha$  luminosity is shown in Figure 5.6.

Due to the smaller error bars than in the previous cases, it is possible to start seeing a suggestive trend of LAEs with higher Ly $\alpha$  luminosities clustering in a stronger way. Anyway, it is not possible to claim that more luminous LAEs cluster stronger than fainter LAEs yet but a small difference can already start to be seen. The calculated correlation length for fainter LAEs is  $r_0 = 2.3^{+1.0}_{-1.3}$ , while for high luminosities it is  $r_0 = 3.2^{+1.3}_{-1.6}$ .

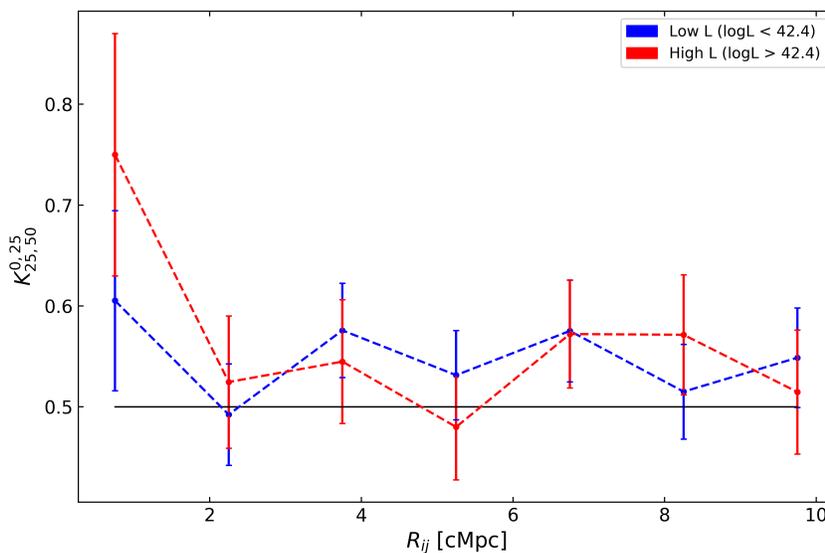


Figure 5.6:  $K_{0,25}^{0,25}$  estimator variation with the transverse pair distance  $r_{ij}$  for the LAEs with low (blue) and high luminosities (red) in the 60 fields of MUSE-Wide. The black straight line shows no clustering and the error bars are poissonian.

Hypothetically, if the error bars would be smaller, these results would be meaningful and completely compatible with the results obtained by [Norberg et al. \(2001\)](#); [Zehavi et al. \(2011\)](#); [Budavari et al. \(2003\)](#), who used the total luminosity of the galaxies instead of the  $\text{Ly}\alpha$  luminosities.

Finally, the star formation rate,  $SFR$  of the LAEs of MUSE-Wide was calculated from the  $\text{Ly}\alpha$  luminosity,  $L_{\text{Ly}\alpha}$ , using equation 5.1 ([Dijkstra and Westra, 2010](#))

$$SFR(\text{Ly}\alpha) = 9.1 \cdot 10^{-43} L_{\text{Ly}\alpha} \quad (5.1)$$

Since both luminosity and  $SFR$  are proportional, the clustering plots and correlation lengths are the same, thus, the clustering graph has been omitted.

# Chapter 6

## Spatial clustering method

For the first time, MUSE data brings the opportunity to study the spatial clustering of Ly $\alpha$  selected LAEs. The precise spectroscopic redshifts and its large depth covered get rid of the main issues of the last decades. Nowadays, the redshift distribution does not have to be extracted from photometric surveys or from NB filters, with the errors that it would entail. Thus, the angular clustering method and the Limber equation are not needed anymore, eliminatting the main error sources of recent times.

As part of the exploration on how to perform clustering analysis of this special data (small coverage but very deep in redshift), expanding the initial goals of my thesis to carry out a comparison between the redshift clustering results and the ones obtained by the traditional and reliable method should be done to check if both methods agree. The spatial clustering method has been used since the 70s in very different kind of surveys, providing results in agreement with cosmological simulations.

[Zehavi et al. \(2002\)](#) carried out the first clustering analysis in the SDSS survey with a total of 29.300 galaxies in the range  $0.019 < z < 0.13$ , covering an area of  $690 \text{ deg}^2$ . They calculated the  $\xi(r_p, \pi)$  of the sample, obtaining a correlation length of  $r_0 \approx 8.0 h^{-1}\text{Mpc}$  and a spatial clustering slope of  $\gamma \approx 1.75$ . The correlation function showed clear signs of RSD caused by the peculiar velocity of the galaxies and the coherent motions of infalling structures. They derived galaxy pairwise velocity dispersions of the order of  $600 \pm 100 \text{ km/s}$ . Similar results were obtained by [Tucker et al. \(1997\)](#) in Las Campanas Redshift Survey.

[Durkalec et al. \(2015\)](#) studied the evolution of clustering using the VIMOS Ultra Deep Survey in the range  $2.0 < z < 5.0$ . They measured the  $\omega_p(r_p)$  of 3022 galaxies with spectroscopic redshifts in a  $0.8 \text{ deg}^2$  area and found  $r_0 = 3.95 \pm 0.54 h^{-1}\text{Mpc}$  at  $z \approx 2.5$

and  $r_0 = 4.35 \pm 0.60 h^{-1}\text{Mpc}$  at  $z \approx 3.5$  after using mock samples to evaluate and correct for the survey selection function.

However, for the specific case of LAEs, since they were always NB-selected with no redshift depth, angular clustering was everything that it could be done. Gawiser et al. (2007) studied the angular clustering of a sample of 162 LAEs at  $z \approx 3.1$  selected by deep NB MUSYC imaging of the Extended Chandra Deep Field-South. The angular clustering was converted to spatial clustering by using Limber equation and the correlation lengths were calculated. LAEs were found to have a correlation length of  $r_0 = 3.6_{-1.0}^{+0.8} h^{-1}\text{Mpc}$ . A very similar length was obtained by Ouchi et al. (2003), who studied the clustering of 87 LAEs selected with the NB technique at  $z = 4.86$  in the Subaru Deep Field, 543 arcmin<sup>2</sup>. They also calculated the angular clustering of the sample and translated it into the spatial clustering, getting  $r_0 = 3.5_{-0.3}^{+0.3} h^{-1}\text{Mpc}$ .

Kovač et al. (2007) analyzed the angular clustering of 151 LAE at higher redshifts ( $z \approx 4.5$ ) from the Large Area Lyman Alpha survey. The LAEs were within an area of 36' x 36' that was covered with five narrowbands. From the power law parameters of the angular clustering, they derive the correlation length of the real-space two-point correlation function,  $r_0 = 3.20 \pm 0.42 h^{-1}\text{Mpc}$ . Due to a possible contamination by randomly distributed sources, the correlation length was recalculated to  $r_0 = 4.61 \pm 0.6 h^{-1}\text{Mpc}$ . Comparing to the expectation from simulations, they found that LAEs reside in halos with a typical minimum mass of  $1-2 \cdot 10^{11} h^{-1}\text{M}_\odot$ .

Following the same process, Khostovan et al. (2018) utilized 5000 LAEs from Slicing COSMOS 4K and from archival NB497 imaging of SA22. From the angular correlation function they calculated  $r_0 = 3 - 6 h^{-1}\text{Mpc}$  at  $z \sim 2.5-6$ .

Similar analysis were carried out by Ouchi et al. (2017) for different purposes. They studied the clustering of 2000 LAEs at  $z = 5.7$  and  $6.6$  selected with NB filters from the Hyper Suprime-Cam Subaru Strategic Program survey (14–21 deg<sup>2</sup> area). They derived the angular correlation function to study the clustering evolution with redshift and conclude, after an extensive halo modelling work, that the LAEs probably evolve into massive super- $L^*$  galaxies in the present day universe.

In our case, I have a sample of  $\sim 1000$  LAEs so I will perform the first spatial clustering analysis of Ly $\alpha$  selected LAEs by means of the spatial correlation function. As already explained in Chapter 1,  $\xi(r_p, \pi)$  is estimated by counting pairs on a 2D grid of separation  $r_p$  and  $\pi$  and the use of one of the estimators previously mentioned. In this master thesis, I will use the Landy-Szalay estimator (Equation 1.5) and the Davis estimator (Equation 1.6).

In the first place, once the l.o.s. and the radial distances between all possible pairs

of galaxies have been calculated in the same way as in Chapter 3.1 (see Figure 6.1 for a scheme to count the LAEs pairs),  $r_p$  is binned in 11 equally spaced logarithmic bins, from 0.04 to  $32.02 h^{-1}\text{Mpc}$  and  $\pi$  is computed in steps of 10 from 10 to  $200 h^{-1}\text{Mpc}$  in order to cover all possible correlated pairs within MUSE fields.

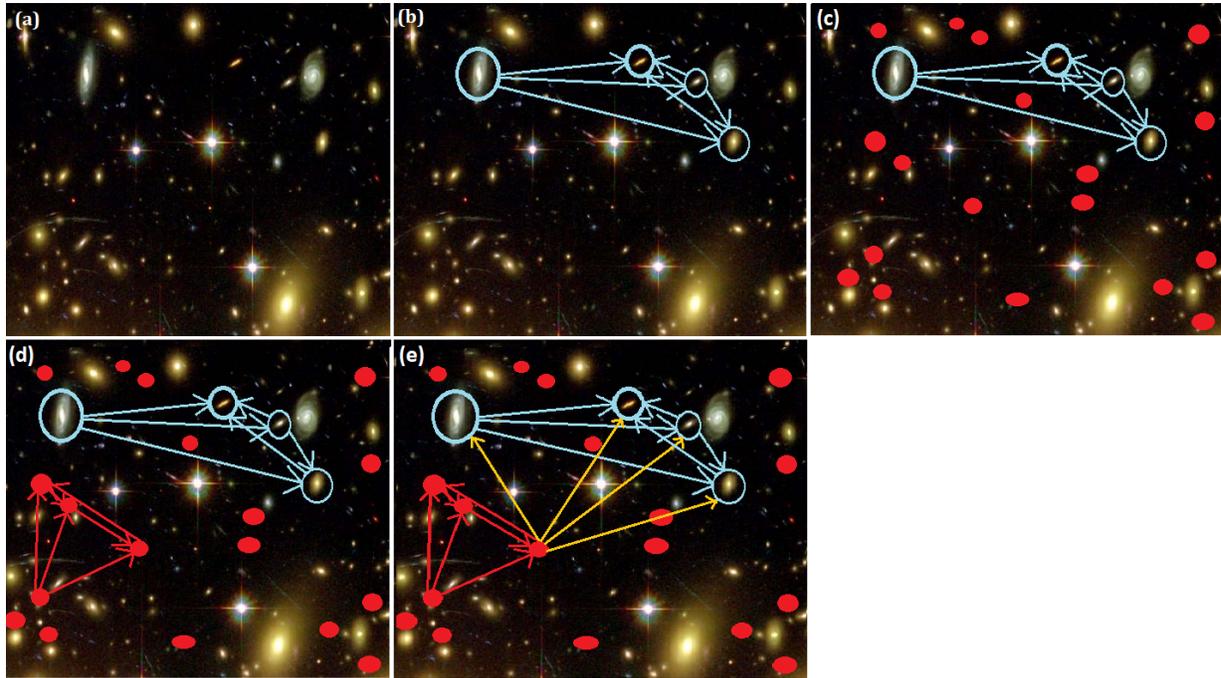


Figure 6.1: (a) Real galaxies in RA (x axis) and DEC (y axis) representation. (b) Real galaxies pair counting represented by blue arrows from galaxy to galaxy (DD pairs). (c) Introduction of the random sample of galaxies shown as red ovals. (d) Random galaxies pair counting represented by red arrows from random to random galaxy (RR pairs). (e) Random-real galaxies pair counting shown by orange arrows between real and random galaxies (DR pairs).

Second, during the first run, pairs of real, random and real-random galaxies are counted as shown in Figure 6.1 with separations from  $r_p = 0.04$  to  $r_p = 0.06 h^{-1}\text{Mpc}$  (first  $r_p$  bin) and fixed  $\pi = 10 h^{-1}\text{Mpc}$ . The next run gives us the pairs with separations comprehended in the nex  $r_p$  bin but same fixed  $\pi$ , and so on until all  $r_p$  bins are covered at the first fixed  $\pi$  value. Once all  $r_p$  bins are filled, the process starts again with the second  $\pi$  value,  $\pi = 20 h^{-1}\text{Mpc}$ . In this way, I collapse  $\xi(r_p, \pi)$  for all  $\pi$  values as a functon of  $r_p$ , calculating as many correlation function curves as bins we have in  $\pi$ .

Third, the 2 point correlation function  $\xi(r_p, \pi)$  is calculated with Equation 1.6 or 1.5 and deprojected with Equation 1.7 to get rid of the redshift space distortions (see Chapter 1). In practice, Equation 1.7 is not integrated to infinite separations between pairs of galaxies. Instead,

$$\omega_p(r_p) = \int_{-\infty}^{\infty} \xi(r_p, \pi) d\pi = 2 \int_0^{\infty} \xi(r_p, \pi) d\pi \approx 2 \int_0^{\pi_{max}} \xi(r_p, \pi) d\pi \quad (6.1)$$

where  $\pi_{max}$  is the l.o.s. distance between pairs of LAEs up to where the integration is performed. It must be chosen in such a way that most correlated pairs DD, RR and DR are taken into account. However, a too large value would only increase the noise instead of increasing  $\omega_p(r_p)$  (see the black areas of Figure 1.5 that show the noise). Very low values of  $\pi_{max}$  do not include most correlated pairs so  $\omega_p(r_p)$  would be underestimated. On the other hand, very high values of  $\pi_{max}$  account for all the pairs but increase the noise in  $\omega_p(r_p)$ . This is due to the fact that at very large l.o.s. distances there are not many correlated galaxy pairs. Therefore, the number of pairs would be that low that  $\omega_p(r_p)$  would barely increase but the noise would due to the correlation of random pairs of galaxies and not of real galaxies. Typical  $\pi_{max}$  values are within the range 20–80  $h^{-1}$ Mpc (see Zehavi et al. (2002, 2011); Durkalec et al. (2015); Coil et al. (2007)), which are large enough to include most correlated pairs and give stable results without including too much noise from distant, physically uncorrelated pairs.

In order to choose the right  $\pi_{max}$  value,  $\xi(r_p, \pi)$  is integrated up to different  $\pi_{max}$  values from which one of them will be chosen when  $\omega_p(r_p)$  saturates. I have studied  $\omega_p(r_p)$  for 20 equally separated values in the range  $10 < \pi_{max} < 200 h^{-1}$ Mpc. The  $\omega_p(r_p)$  calculated with the Landy-Szalay estimator and with the Davis estimator are represented on the left panel of Figures 6.2 and 6.3 for the full  $r_p$  and  $\pi$  ranges, respectively.

Due to the insufficient number of real LAEs pairs at high  $r_p$  separations,  $\omega_p(r_p)$  does not show a smooth curve in this region, thus I have chosen the  $r_p$  range where the curve does not show so much variation. From the 20 different  $\pi_{max}$  values, 6 of them have been chosen to be represented, for simplicity, in the right panel of Figure 6.2 for the Landy-Szalay estimator and in the right panel of Figure 6.3 for the Davis estimator. In this way, I will be able to fit  $\omega_p(r_p)$  to calculate  $r_0$  and  $\gamma$ .

Due to the larger coverage of the 77 fields (see Chapter 2), I will use this sample to study the spatial clustering. However, the inner 9 fields of the UDF will be excluded because of the lack of the selection function in this region (without which we cannot create a realistic random sample of the LAEs there included and we would always have an excess of pairs due to the longer exposure time which detects more galaxies in this area). Thereby, I will

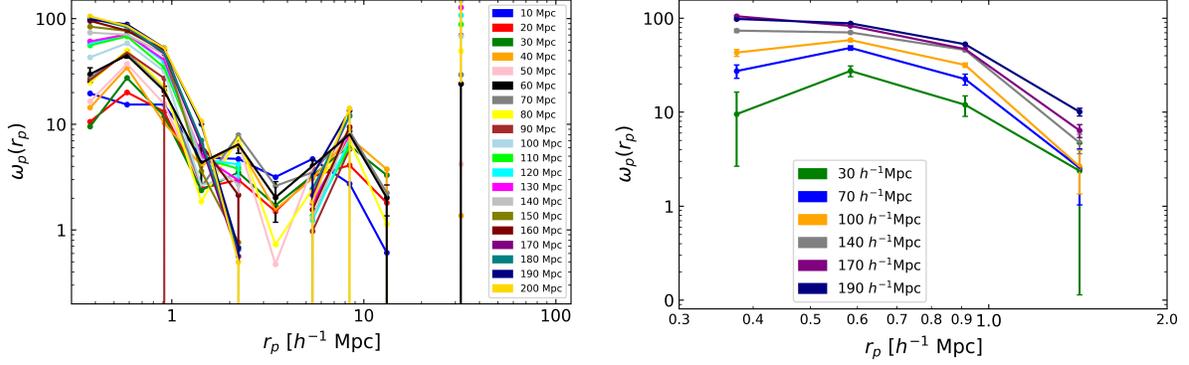


Figure 6.2: Left: Projected 2 point correlation function calculated with the Landy-Szalay estimator as a function of the full transverse distance range between pairs of galaxies,  $r_p$ . Right: the same but with the shorter selected  $r_p$  range. In both cases, the various colors represent different  $\omega_p$  curves for each different upper integration limit value,  $\pi_{max}$  were chosen.

employ a total of 68 fields of MUSE-Wide (77 MUSE-Wide fields - 9 UDF fields) between  $3.3 < z < 6$  that contain 814 LAEs in an area of  $62.9 \text{ arcmin}^2$ .

In the same way as for the redshift clustering, I used the  $\Lambda$ CDM cosmological model ( $\Omega_M = 0.3, \Omega_\Lambda = 0.7$  and  $H_0 = 100 \text{ km s}^{-1} \text{Mpc}^{-1}$ ) and, in this case, I considered  $h = H_0/100 \text{ km s}^{-1} \text{Mpc}^{-1}$ .

In both cases for the selected  $r_p$  range, the error bars are Poissonian and were calculated with an error propagation analysis. For the case of the Landy-Szalay estimator,  $\omega_{pL}$ ,

$$\delta\omega_{pL} = \sqrt{\left(\frac{\partial\xi}{\partial DD} \cdot \delta DD\right)^2 + \left(\frac{\partial\xi}{\partial DR} \cdot \delta DR\right)^2 + \left(\frac{\partial\xi}{\partial RR} \cdot \delta RR\right)^2} = \sqrt{\left(\frac{\delta DD}{RR}\right)^2 + \left(\frac{-2 \cdot \delta DR}{RR}\right)^2 + \frac{(-DD + 2DR) \cdot \delta DD}{RR^2}} \quad (6.2)$$

where  $\delta DD$ ,  $\delta RR$  and  $\delta DR$  are the square root of the number of objects in each bin,  $\sqrt{DD}$ ,  $\sqrt{RR}$  and  $\sqrt{DR}$ , respectively.

In the same way, for the case of the Davis estimator,  $\omega_{pD}$ ,

$$\delta\omega_{pD} = \sqrt{\left(\frac{\partial\xi}{\partial DD} \cdot \delta DD\right)^2 + \left(\frac{\partial\xi}{\partial DR} \cdot \delta DR\right)^2} = \sqrt{\left(\frac{\delta DD}{DR}\right)^2 + \left(\frac{-DD \cdot \delta DR}{DR^2}\right)^2} \quad (6.3)$$

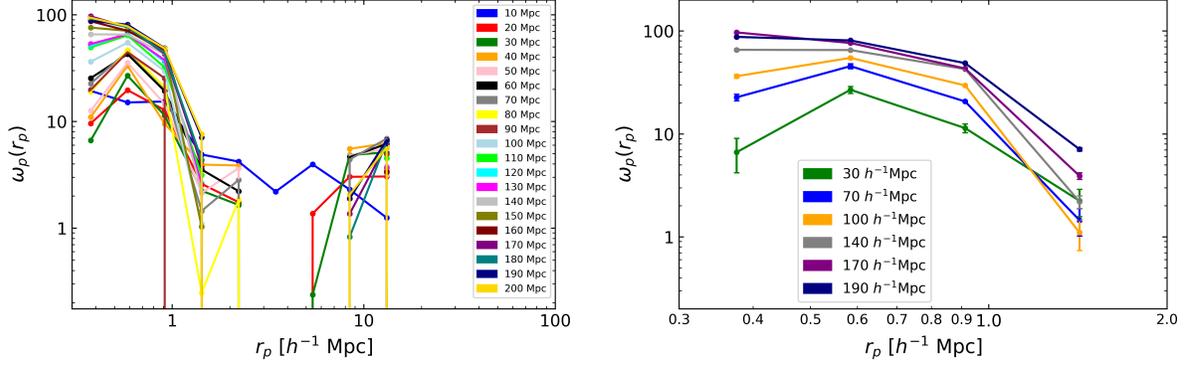


Figure 6.3: Left: Projected 2 point correlation function calculated with the Davis estimator as a function of the full transverse distance range between pairs of galaxies,  $r_p$ . Right: the same but with the shorter selected  $r_p$  range. In both cases, the various colors represent different  $\omega_p$  curves for each different upper integration limit value,  $\pi_{max}$  were chosen.

For both estimators, the  $\omega_p(r_p)$  has been fitted with a power law, where the equation 1.2 would be its equivalent, to obtain  $r_0$  and  $\gamma$ . By using the analytical solution,

$$\omega_p(r_p) = r_p \cdot \left(\frac{r_0}{r_p}\right)^{-\gamma} \cdot \frac{\Gamma(1/2)\Gamma((\gamma - 1)/2)}{\Gamma(\gamma/2)} \quad (6.4)$$

where  $\Gamma(x)$  is the Gamma function, I have calculated values of  $r_0 = 3.84 \pm 0.06 h^{-1}\text{Mpc}$  when the Landy-Szalay estimator was used and  $r_0 = 3.67 \pm 0.06$  when the Davis estimator was applied in the right panels of Figures 6.2 and 6.3. Both  $r_0$  values have been obtained by fixing  $\gamma$  at 1.8 and 1.9 since they are the standard values used in the literature. Afterwards, a  $\chi^2$  analysis has shown that  $\gamma = 1.9$  gives a better fit of  $\omega_p(r_p)$ . The errors in the  $r_0$  are  $1\sigma$  errors and are calculated with the minimum squared method from the  $\omega_p(r_p)$  fitting. Both values are in completely agreement with the values calculated with the  $K$  estimator method even if the selected  $r_p$  coverage in both methods differ.

As it is possible to see in both figures, there is again a clear clustering signal as it was already shown in the redshift clustering method. In this case, the signal is the excess in the  $\omega_p$  curves that rise up to 100 (remember that the random clustering is given by  $\omega_p = 0$  or  $\xi = 0$ . Thereby, in a logarithmic scale the random clustering would be at  $-\infty$ , while in a linear scale it would be at 0).

The point where the saturation of the projected correlation function occurs is hard to tell because of the lack of realistic error bars. As far as greater values of  $\pi_{max}$  are

considered, the larger the error bars would become. In our case, the error calculation and, thus, the point at which  $\omega_p$  saturates can only be obtained by hydrodynamical simulations (see Section 7.2). Therefore, an intermediate  $\pi_{max}$  value will be selected for now,  $\pi_{max} = 70 h^{-1}\text{Mpc}$ , and justified afterwards. It has also been tested that increasing the number of random galaxies or using another random sample makes basically no difference to the results.

Despite the short transverse distances selected,  $\omega_p$  already gives an excess clustering signal, covering  $r_p$  values of  $\sim 2 h^{-1}\text{Mpc}$ . Typical sizes of DM halos are  $\sim 1-2 h^{-1}\text{Mpc}$  which means that our  $\omega_p$  tests LAEs that belong to the same DM halo, something that was never done before.

Moreover, a drop of  $\omega_p$  is seen, which is due to the decrease of correlated galaxy pairs while  $r_p$  distances increase. Furthermore, at small  $r_p$  distances ( $r_p < 0.6 h^{-1}\text{Mpc}$ ), the flattening caused by the Kaiser effect is visible (same effect seen in Figure 1.5 at small  $r_p$  values).

However, an intriguing question is why I have chose to integrate up to values of  $\pi_{max} = 70 h^{-1}\text{Mpc}$ . Does it make sense that we consider as a correlated pair two galaxies that are separated by  $70 h^{-1}\text{Mpc}$ ? Yes, and the physical answer is given by the peculiar velocities of the galaxies at these high redshifts (strong FOG effect).

As it was introduced in Chapter 1, when we derive comoving distances from redshifts we have to take into account the RSD. As it can be seen in Figure 6.4, an overdensity of galaxies in a real space/distance representation (*e.g.* blue circle in the left panel of Figure 6.4a) is translated into a much more elongated structure in redshift space (*e.g.* blue curve in the right panel of Figure 6.4a) due to the peculiar velocities. This same effect is observed in Figure 1.5 at small  $r_p$  distances, where the contours of the two point correlation function are also elongated along the l.o.s. distance.

Let's consider the case where we have two galaxies in a  $z = 4$  environment. One galaxy is moving towards us at  $z_2 = 4.0$  and another one is going further away from us at  $z_2 = 4.15$  (redshifted), being both going apart from each other with velocities of 1000 km/s (see Figure 6.4b).

The l.o.s. separation  $\pi$  between both is calculated with equation 3.4, getting  $\pi \approx 100 \text{ Mpc} = 70 h^{-1}\text{Mpc}$ . However, to know the real distance between the pair of galaxies, the shift in the l.o.s. distance due to random motions of galaxies must be considered. This shift in distance is calculated through the shift in the wavelength of the emission lines due to the peculiar velocities

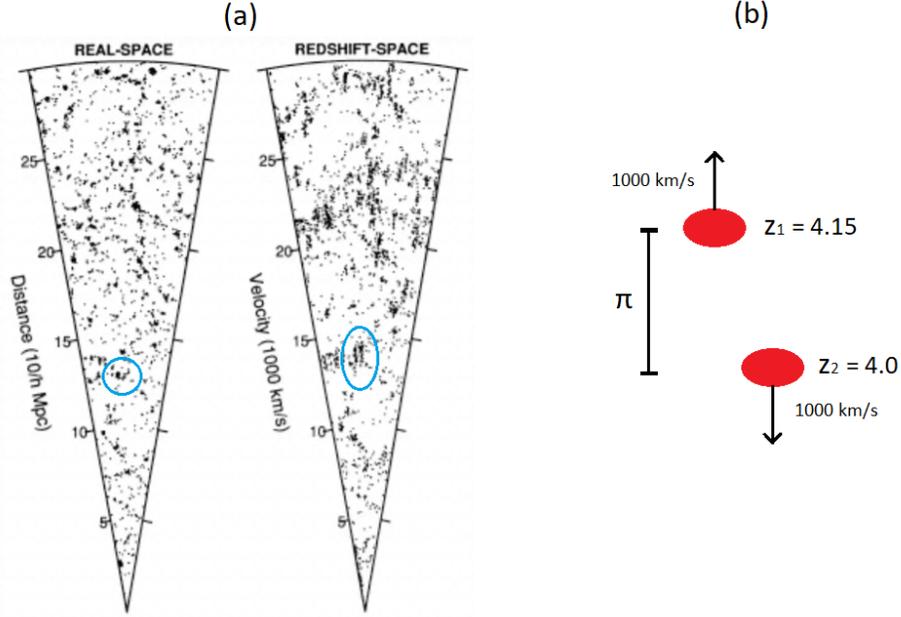


Figure 6.4: (a). Left panel: representation of galaxies in real space where the blue circle shows one of the galaxy overdensities. Right panel: representation of galaxies in a redshift space where the elongation of the overdensities are clear. The blue curve shows the elongation of our chosen overdensity. (b) Pair of galaxies moving apart from each other at velocities of 1000 km/s, at a l.o.s. separation  $\pi$  and at redshifts 4.0 and 4.15, respectively.

$$\frac{\lambda'}{\lambda_0} = \sqrt{\frac{1 + \beta}{1 - \beta}} \quad (6.5)$$

where  $\beta = v_p/c$  is the fraction of peculiar velocity and the speed of light and  $\lambda_0$  and  $\lambda'$  are the emitted and the observed wavelengths, respectively. For our example,  $\frac{\lambda'}{\lambda_0} = 1.01$ . This same relation can also be expressed as a function of redshift

$$\frac{\lambda'}{\lambda_0} = \frac{1 + z'}{1 + z_0} = \frac{1 + z_0 + \Delta z}{1 + z_0} \quad (6.6)$$

By using  $z_0 = 4$ , I calculate  $\Delta z = 0.05$ . Therefore, the shift in redshift is translated into shifts in comoving distances of  $\sim 30 h^{-1} \text{Mpc}$  (eq 3.4).

Converting comoving distances to real distances,  $d$ , is done by

$$d = \frac{\pi}{1+z} \quad (6.7)$$

which, for our example, gives real distances of few Mpc. This result is in agreement with the size of one DM halo (few Mpc), which is also the range of  $r_p$  that we are able to cover (see Figures 6.3 and 6.2), and with the correlation length obtained for the sample.

Therefore, the considered  $\pi_{max}$  values in Figures 6.3 and 6.2 are actually real space separations of the order of few Mpc, after taking into account the shift in distance caused by the random motions of the LAEs.

A factor that has not been mentioned or taken into account is the fact that peculiar velocity dispersions are based on gaussian distributions, on  $1\sigma$  measurements, which does not always have to be the case.

As a curiosity exercise since both clustering graphs show very similar curves and maximum values of  $\omega_p$ , both estimators have been compared in Figure 6.5.

As it was possible to observe in Figures 6.2 and 6.3, both estimators give very similar results at intermediate  $r_p$  values ( $0.6 < r_p < 1.0$ ), obtaining always a slight shift to higher values in the case of the Landy-Szalay estimator. This is in agreement with the results got by Zehavi et al. (2002), who showed that both estimators gave very similar results and are completely consistent within the error bars.

Nevertheless, once  $r_p$  become too small and specially too large, we approach the limit of the method, the point at which pairs of galaxies start to be not sufficient to continue carrying out this method. This is why the more advanced estimator is still able to get higher values of  $\omega_p$ , while the simpler estimator does not, giving lower values of  $\omega_p$  and steeply increasing the error bars.

Once again, the errors in Figure 6.5 are calculated by the same error propagation analysis as above,

$$\delta\xi = \sqrt{\left(\frac{\partial r}{\partial\omega_{pL}} \cdot \delta\omega_{pL}\right)^2 + \left(\frac{\partial r}{\partial\omega_{pD}} \cdot \delta\omega_{pD}\right)^2} = \sqrt{\left(\frac{\delta\omega_{pL}}{\omega_{pD}}\right)^2 + \left(\frac{-\omega_{pL} \cdot \delta\omega_{pD}}{\omega_{pD}^2}\right)^2} \quad (6.8)$$

where the fraction of both estimators,  $\frac{\omega_{pL}}{\omega_{pD}}$ , has been called  $r$  for simplicity.

In comparison with other clustering studies, the number of galaxies in our samples is not so large. Thereby, the computational time between the Landy-Szalay and the Davis estimator differs in less than one minute. This is why for further studies, I will always use the Landy-Szalay estimator.

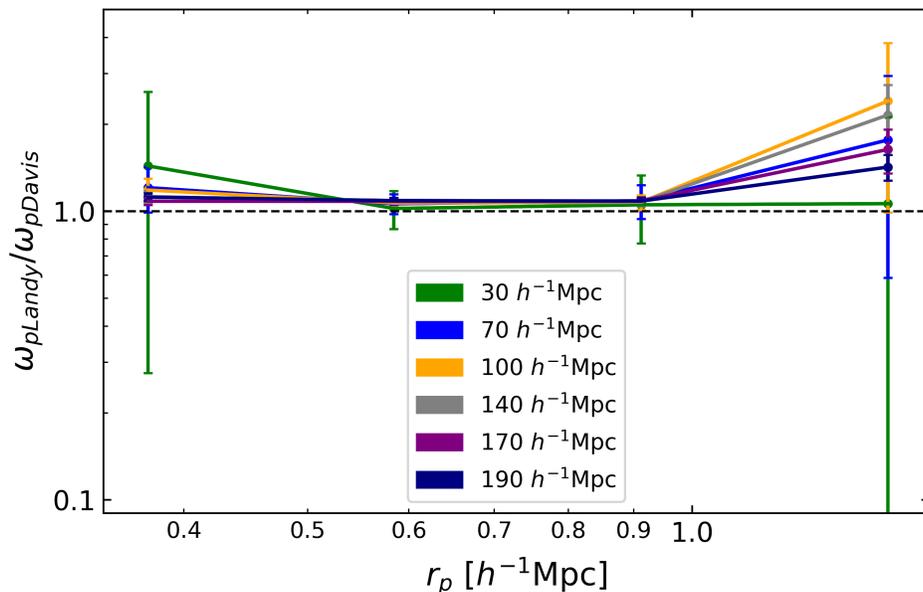


Figure 6.5: Ratio of the projected 2 point correlation function calculated with the Landy-Szalay and the Davis estimator as a function of transverse distance between pairs of galaxies,  $r_p$ . Various colors represent different  $\omega_p$  curves for each different upper integration limit values,  $\pi_{max}$ .

However, where I observe the greatest computational time saving is by using the redshift clustering method where no random sample is needed. Therefore, the computational time is abruptly shorter.

In all cases, this chapter will be used for future goals (see section 7.2). Since the spatial clustering method was successfully applied for decades, it was used as a comparison and test of the new redshift clustering method, which is the one on which I have mainly focused along this thesis. I have shown that it is possible to measure the clustering of MUSE LAEs, that with both methods I calculate  $r_0$  values completely in agreement and that both  $\omega_p$  tests LAEs that are within the one-halo term, something that is completely new in the high redshift galaxy domain.

# Chapter 7

## Conclusion

### 7.1 Summary

Studying the distribution of LAEs in the Universe reveals how the structure of the Universe looked at  $3 < z < 6$ . Galaxy clustering analysis constrain the size and population of the dark matter (DM) halos, tests the cosmological models and shows evidence of the fluctuations in the early Universe. By studying the clustering evolution of Lyman Alpha Emitters (LAEs) through different epochs, the galaxy formation and evolution theories are rehearsed.

Since the first goal of this master thesis was to reproduce the results shown in [Diener et al. \(2017\)](#), I have analyzed the first 24 fields of MUSE-Wide and afterwards, I have extended the study to the full samples of MUSE-Wide and MUSE-Deep. The number of LAEs contained in each survey is 234, 947 (814 in 68 fields of MUSE Wide) and 757, covering areas of  $22.2 \text{ arcmin}^2$ ,  $63.83 \text{ arcmin}^2$  ( $62.9 \text{ arcmin}^2$  for the 68 fields) and  $8.32 \text{ arcmin}^2$ , respectively. The redshift coverage of all samples comprehends  $3 < z < 6$ , being  $z = 4.04$ ,  $z = 4.07$  and  $z = 4.14$  the mean redshift of each survey, respectively.

Due to the special MUSE data, the large redshift range and the limited angular coverage, I explored different possibilities to measure the clustering of the LAEs. Because of the type of data, the samples are ideal to perform the radial clustering method. I have studied the redshift clustering through the  $K_{0,30}^{30,60}$  and  $K_{0,25}^{25,50}$  estimators and obtained correlation lengths of  $r_0 = 2.7_{-0.4}^{+0.3}$  and  $r_0 = 2.6_{-0.2}^{+0.2}$  in the MUSE-Wide and MUSE-Deep samples, respectively. The slope of the  $K$  estimator was  $\gamma = 1.8$  in both cases and it yields consistent results in both regions of the sky. The uncertainties in the estimated correlation lengths

are  $\pm 13\%$  and  $\pm 8\%$ , respectively, being both lengths in agreement with the values in the literature at similar redshifts.

Afterwards, the method has been studied in detail by creating a realistic random sample of the LAEs contained in the MUSE surveys. The selection and luminosity functions and the footprints of MUSE-Wide survey have been taken into account to develop a realistic random sample of LAEs on which the redshift clustering can be tested. As expected, the  $K$  estimator provided values consistent with no clustering after the method was ran 100 times and averaged. This analysis, apart from ensuring that we are not biased, also offered a better understanding on how realistic Poisson errors are. In all cases, the errors introduced by the method itself are much smaller than the Poisson errors calculated.

I also studied the clustering dependence with several physical properties. Such was the case of redshift, Ly $\alpha$  luminosity, stellar mass contained in the galaxy, and equivalent width of the Ly $\alpha$  emission line. Despite that the errors bars were quite large to be able to say precisely which subsample cluster stronger, clear tendencies were seen in various cases. The  $K$  estimator seems to give higher clustering trends for LAEs as the redshift increases. In this way, we can connect different observed populations of LAEs at different epochs which, combining it with theoretical models for the DM halos growth, one can say that galaxies formed early on in more overdense regions of the Universe.

Even more suggestive trends are seen when the clustering is studied as a function of luminosity. More luminous LAEs seem to cluster stronger than fainter LAEs, residing in more massive halos. For the rest of the cases (stellar mass and equivalent width) the error bars are too large to be able to distinguish between the clustering of LAEs at low and high values of both properties. However, I am hopeful that future error calculations (see Section 7.2) will bring light to help to solve this issue.

The results obtained in this thesis are consistent with the idea that LAEs are the progenitors of the galaxies seen in the local Universe, including Milky Way type galaxies. Since I found that luminous LAEs very likely cluster stronger than fainter LAEs, luminous LAEs may be the protoclusters, the densest regions in the early Universe that evolve to the massive galaxy clusters that we observe today.

In order to assure the reality of the clustering signal obtained with the radial clustering method, I have extended the thesis to use the traditional spatial clustering method in the MUSE-Wide sample. The two most common estimators, Landy-Szalay and Davis estimators, were applied to calculate the projected correlation function, giving both basically the same results. Apart from the appreciation of a clear clustering signal and a total coverage of the one-halo term ( $\omega_p$  gives a smoothed clustering signal up to  $r_p = 1.4 h^{-1}\text{Mpc}$ ), the correlation lengths calculated from both estimators,  $r_0 = 3.84 \pm 0.06 h^{-1}\text{Mpc}$

and  $r_0 = 3.67 \pm 0.06 h^{-1}\text{Mpc}$ , agree completely with the values obtained with the redshift clustering method. In this case, the best fit for the slope of the correlation function was  $\gamma = 1.9$ .

In conclusion, LAEs provide a great insight into the formation and evolution of the galaxies in our local Universe. By studying their clustering in  $3 < z < 6$ , we were able to study for the first time the one-halo term of our LAEs samples by means of the projected two point correlation function, to test different and new methods of galaxy clustering, obtaining their correlation lengths and slopes, to study the dependency of LAEs clustering as function of different physical properties, to connect galaxies from the end of the epoch of reionization to close to the peak of cosmic star formation epoch and to reproduce realistic random samples of the LAEs detected by MUSE. On the process, we are able to trace the large scale structure of the Universe, test galaxy formation and evolution theories, constrain the parameters of cosmological models, determine the DM halo populations, their distribution, and the DM halo masses that contain the LAEs of our samples.

## 7.2 Future work

Galaxy clustering is a field that is experiencing a great development since the last few decades. The increase of the precision in the observations drive this progress. The larger quantities of data that nowadays is possible to collect provide the best framework to study galaxy clustering and, therefore, the formation and evolution of the large-scale structure of the Universe.

One of the first things that needs to be done as a continuation of this work is related to the error calculations. As discussed in previous sections, all error bars along this thesis were poissonian. This should not be the case because all data points in the plots are correlated. Thereby, the errors will be calculated from a comparison with simulations. The Universe will be represented in a box of certain size and will be populated with galaxies in the same redshift range as our LAEs. Hundreds of light cones will scan the Universe box and, from each of them,  $\omega_p$  will be calculated. From the comparison of our clustering results and the mean  $\omega_p$  calculated from those simulations, more representative error bars will be achieved. Hopefully, this improvement in the errors will allow us to precisely study how LAEs cluster as a function of different physical properties, to determine exactly where  $\omega_p$  saturates and to know the exact  $r_p$  range that we can use to study the clustering in our samples.

The next step is one of the most important results provided by clustering measurements, calculating the typical mass of the DM haloes where the galaxies reside. In order to calculate them, a halo occupation distribution (HOD) analysis must be carried out.

The clustering of galaxies does not necessarily have to follow the same path as the clustering of the DM where they are contained. The bias factor constrains the fact that the spatial distribution of baryons differ from that of DM, giving a relation between the spatial distribution of galaxies and the underlying DM.

The bias factor can be calculated by comparison to cosmological simulations. Using a certain cosmological model, the evolution of the clustering of DM is predicted with cosmological N-body simulations. Performing a HOD modelling for the one-halo and two-halo terms (pairs of galaxies belonging to the same parent halo,  $1-2 h^{-1}\text{Mpc}$ , or to different ones) and adding both terms together, results in a correlation function that mainly follows a power law. This power law is then compared to the clustering power law of our observed LAEs through which I will acquire the bias.

After this comparison, the galaxies will be assumed to be in halos of a given mass with the same bias factor as in the simulations, which will allow me to calculate the typical DM halo masses.

These proceedings depend on the number of galaxies in each halo (the halo occupation) and on the location on the galaxies in the halo. In general, it is assumed that one galaxy is at the center of the halo (central galaxy), while the rest are distributed around it (satellite galaxies). By carrying out a HOD analysis, the probability that a halo contains  $x$  number of galaxies will be obtained.

Performing this kind of analysis will provide very interesting results. The mass of the DM halo will be calculated, simulations will provide a good comparison to check our previous results and, by combining the bias with the galaxy number density, I will constrain the fraction of halos with certain galaxy types. Therefore, I will know the distribution of central and satellite galaxies within the DM haloes of our LAEs.

On the other hand, studying the clustering of other type of galaxies is of huge interest. Data already exists (provided by MUSE) to perform this analysis at lower redshifts. Such is the case of the O[II] emitters present in MUSE-Wide, whose clustering will be studied.

Nevertheless, there are several improvements that do not depend on me, they will come hand in hand with new observational data. Extensive spectroscopic redshift surveys will bring light to the clarification of the nature and the properties of the large-scale structure. The shape and type of structures in the galaxy distribution will be able to be revealed.

- MUSE is the deepest spectroscopic survey ever performed, with 10 and 30 hours depth provides more than 1600 redshifts. However, an extension of the project is approaching and new areas of the sky will be covered. Moreover, in a farther future,

MUSE Blue will cover the blue part of the spectrum (370-600 nm), will detect many more galaxies, and will provide a larger area per field ( $1.4 \times 1.4 \text{ arcmin}^2$ ) than the coverage available today ( $1 \times 1 \text{ arcmin}^2$ ).

- The Hobby-Eberly Dark Energy Experiment (HEDTEX) will detect more than one million LAEs in the redshift range of  $1.7 < z < 3.5$ . It will allow us to study the clustering at lower redshifts and have the complete evolution of LAEs through all epochs. The spectrograph (VIRUS) will scan more than  $200 \text{ deg}^2$  in 110 observing nights with a wavelength range of 340 – 565 nm.
- The upcoming James Webb Space Telescope will detect galaxies primarily in the infrared region of the electromagnetic spectrum by detecting the Ly $\alpha$ , H $\alpha$  and He II lines among others. It will cover redshifts in the range  $1 < z < 6$  with both imaging and spectroscopy, providing a great opportunity to study the clustering evolution of LAEs.
- The Taipan survey (multi-object spectroscopic survey scanning the wavelength range of  $370 < \lambda < 870 \text{ nm}$ ) will detect millions of galaxies at much lower redshifts ( $z < 0.3$ ) in the whole southern sky. This survey will be able to measure  $H_0$  with 0.9% precision and the growth rate of structure with 5% which will, therefore, improve the precision of clustering measurements. This survey will provide the most extensive map of mass distribution and motions from peculiar velocities of the southern hemisphere.

In this way, these large samples of low to high redshift galaxies will trace the Universe structure and its evolution with time, as well as, its galaxy formation. Currently, guessing which of the current models will success with the new data cannot be predicted.

Therefore, at the end of this future work I will be able to show how LAEs cluster in redshift space (redshift clustering), how they cluster spatially (spatial clustering) in larger and more representative samples of the Universe with proper error bars, how much stronger the LAEs cluster with respect to the DM (bias factor) and how much is the typical mass of the DM haloes of LAEs. From all these various results, apart from constraining cosmological parameters, galaxy evolution models and connecting galaxy populations at different epochs, the way LAEs cluster and the formation and evolution of LAEs within their parent DM haloes will be better understood.

# Bibliography

- Adelberger, K. L., Steidel, C. C., Pettini, M., Shapley, A. E., Reddy, N. A., and Erb, D. K. (2005). The spatial clustering of star-forming galaxies at redshifts  $1.4 \leq z \leq 3.5$ . *ApJ*, 619(2):697.
- Bacon, R., Conseil, S., Mary, D., Brinchmann, J., Shepherd, M., Akhlaghi, M., Weilbacher, P. M., Piqueras, L., Wisotzki, L., Lagattuta, D., et al. (2017). The muse hubble ultra deep field survey-i. survey description, data reduction, and source detection. *A&A*, 608:A1.
- Benson, A., Cole, S., Frenk, C., Baugh, C., and Lacey, C. G. (2000). The nature of galaxy bias and clustering. *MNRAS*, 311(4):793–808.
- Bina, D., Pelló, R., Richard, J., Lewis, J., Patrício, V., Cantalupo, S., Herenz, E., Soto, K., Weilbacher, P., Bacon, R., et al. (2016). Muse observations of the lensing cluster abell 1689. *A&A*, 590:A14.
- Budavari, T., Connolly, A. J., Szalay, A. S., Szapudi, I., Csabai, I., Scranton, R., Bahcall, N. A., Brinkmann, J., Eisenstein, D. J., Frieman, J. A., et al. (2003). Angular clustering with photometric redshifts in the sloan digital sky survey: Bimodality in the clustering properties of galaxies. *ApJ*, 595(1):59.
- Caruana, J., Wisotzki, L., Herenz, E. C., Kerutt, J., Urrutia, T., Schmidt, K. B., Bouwens, R., Brinchmann, J., Cantalupo, S., Carollo, M., et al. (2017). The muse-wide survey: a measurement of the  $\text{Ly}\alpha$  emitting fraction among  $z > 3$  galaxies. *Monthly Notices of the Royal Astronomical Society*, 473(1):30–37.
- Coil, A. L., Hennawi, J. F., Newman, J. A., Cooper, M. C., and Davis, M. (2007). The deep2 galaxy redshift survey: clustering of quasars and galaxies at  $z = 1$ . *ApJ*, 654(1):115.

- Cole, S., Percival, W. J., Peacock, J. A., Norberg, P., Baugh, C. M., Frenk, C. S., Baldry, I., Bland-Hawthorn, J., Bridges, T., Cannon, R., Colless, M., Collins, C., Couch, W., Cross, N. J. G., Dalton, G., Eke, V. R., Propris, R. D., Driver, S. P., Efstathiou, G., Ellis, R. S., Glazebrook, K., Jackson, C., Jenkins, A., Lahav, O., Lewis, I., Lumsden, S., Maddox, S., Madgwick, D., Peterson, B. A., Sutherland, W., and and, K. T. (2005). The 2df galaxy redshift survey: power-spectrum analysis of the final data set and cosmological implications. *MNRAS*, 362(2):505–534.
- Colless, M., Dalton, G., Maddox, S., Sutherland, W., Norberg, P., Cole, S., Bland-Hawthorn, J., Bridges, T., Cannon, R., Collins, C., Couch, W., Cross, N., Deeley, K., Propris, R. D., Driver, S. P., Efstathiou, G., Ellis, R. S., Frenk, C. S., Glazebrook, K., Jackson, C., Lahav, O., Lewis, I., Lumsden, S., Madgwick, D., Peacock, J. A., Peterson, B. A., Price, I., Seaborne, M., and and, K. T. (2001). The 2df galaxy redshift survey: spectra and redshifts. *MNRAS*, 328(4):1039–1063.
- Contini, T., Epinat, B., Bouché, N., Brinchmann, J., Boogaard, L., Ventou, E., Bacon, R., Richard, J., Weilbacher, P., Wisotzki, L., et al. (2016). Deep muse observations in the hdfs-morpho-kinematics of distant star-forming galaxies down to  $10^8 m$ . *A&A*, 591:A49.
- Cowie, L. L. and Hu, E. M. (1998). High- $z$   $ly\alpha$  emitters. i. a blank-field search for objects near redshift  $z = 3.4$  in and around the hubble deep field and the hawaii deep field ssa 22. *The Astronomical Journal*, 115(4):1319.
- Davis, M. and Peebles, P. J. E. (1983). A survey of galaxy redshifts. v - the two-point position and velocity correlations. *ApJ*, 267:465.
- Davis, M. and Wilkinson, D. T. (1974). Search for primeval galaxies. *The Astrophysical Journal*, 192:251–260.
- Diener, C., Wisotzki, L., Schmidt, K. B., Herenz, E. C., Urrutia, T., Garel, T., Kerutt, J., Saust, R. L., Bacon, R., Cantalupo, S., Contini, T., Guiderdoni, B., Marino, R. A., Richard, J., Schaye, J., Soucail, G., and Weilbacher, P. M. (2017). The MUSE-wide survey: detection of a clustering signal from  $lyman\alpha$  emitters in the range  $3 < z < 6$ . *MNRAS*, 471(3):3186–3192.
- Dijkstra, M. and Westra, E. (2010). Star formation indicators and line equivalent width in  $ly\alpha$  galaxies. *MNRAS*, 401(4):2343–2348.
- Drake, A.-B., Garel, T., Wisotzki, L., Leclercq, F., Hashimoto, T., Richard, J., Bacon, R., Blaizot, J., Caruana, J., Conseil, S., et al. (2017a). The muse hubble ultra deep field

- survey-vi. the faint-end of the  $\text{Ly}\alpha$  luminosity function at  $2.91 < z < 6.64$  and implications for reionisation. *A&A*, 608:A6.
- Drake, A. B., Guiderdoni, B., Blaizot, J., Wisotzki, L., Herenz, E. C., Garel, T., Richard, J., Bacon, R., Bina, D., Cantalupo, S., et al. (2017b). Muse deep-fields: the  $\text{Ly}\alpha$  luminosity function in the hubble deep field-south at  $2.91 < z < 6.64$ . *MNRAS*, 471(1):267–278.
- Durkalec, A., Le Fèvre, O., Pollo, A., De La Torre, S., Cassata, P., Garilli, B., Le Brun, V., Lemaux, B., Maccagni, D., Pentericci, L., et al. (2015). Evolution of clustering length, large-scale bias, and host halo mass at  $2 < z < 5$  in the vimos ultra deep survey (vuds). *A&A*, 583:A128.
- Eisenstein, D. J., Zehavi, I., Hogg, D. W., Scoccimarro, R., Blanton, M. R., Nichol, R. C., Scranton, R., Seo, H.-J., Tegmark, M., Zheng, Z., Anderson, S. F., Annis, J., Bahcall, N., Brinkmann, J., Burles, S., Castander, F. J., Connolly, A., Csabai, I., Doi, M., Fukugita, M., Frieman, J. A., Glazebrook, K., Gunn, J. E., Hendry, J. S., Hennessy, G., Ivezić, Z., Kent, S., Knapp, G. R., Lin, H., Loh, Y.-S., Lupton, R. H., Margon, B., McKay, T. A., Meiksin, A., Munn, J. A., Pope, A., Richmond, M. W., Schlegel, D., Schneider, D. P., Shimasaku, K., Stoughton, C., Strauss, M. A., SubbaRao, M., Szalay, A. S., Szapudi, I., Tucker, D. L., Yanny, B., and York, D. G. (2005). Detection of the baryon acoustic peak in the large-scale correlation function of SDSS luminous red galaxies. *ApJ*, 633(2):560–574.
- Finkelstein, S. L., Rhoads, J. E., Malhotra, S., Grogin, N., and Wang, J. (2008). Effects of dust geometry in  $\text{Ly}\alpha$  galaxies at  $z = 4.4$ . *The Astrophysical Journal*, 678(2):655.
- Finley, H., Bouché, N., Contini, T., Paalvast, M., Boogaard, L., Maseda, M., Bacon, R., Blaizot, J., Brinchmann, J., Epinat, B., et al. (2017). The muse hubble ultra deep field survey-vii.  $\text{Fe II}^*$  emission in star-forming galaxies. *A&A*, 608:A7.
- Gawiser, E., Francke, H., Lai, K., Schawinski, K., Gronwall, C., Ciardullo, R., Quadri, R., Orsi, A., Barrientos, L. F., Blanc, G. A., et al. (2007).  $\text{Ly}\alpha$ -emitting galaxies at  $z = 3.1$ :  $\text{L}^*$  progenitors experiencing rapid star formation. *ApJ*, 671(1):278.
- Giavalisco, M., Ferguson H C, K. A. M., Dickinson, M. Alexander, D. M. B., et al. (2004). The great observatories origins deep survey: Initial results from optical and near-infrared imaging. *The Astrophysics Journal*, 600:L93.
- Goto, T., Yamauchi, C., Fujita, Y., Okamura, S., Sekiguchi, M., Smail, I., Bernardi, M., and Gomez, P. L. (2003). The morphology-density relation in the sloan digital sky survey. *MNRAS*, 346(2):601–614.

- Grogin, N. A., Kocevski, D. D., Faber, S., Ferguson, H. C., Koekemoer, A. M., Riess, A. G., Acquaviva, V., Alexander, D. M., Almaini, O., Ashby, M. L., et al. (2011). Candels: the cosmic assembly near-infrared deep extragalactic legacy survey. *ApJS*, 197(2):35.
- Guérou, A., Krajnović, D., Epinat, B., Contini, T., Emsellem, E., Bouché, N., Bacon, R., Michel-Dansac, L., Richard, J., Weilbacher, P. M., et al. (2017). The muse hubble ultra deep field survey-v. spatially resolved stellar kinematics of galaxies at redshift  $0.2 \leq z \leq 0.8$ . *A&A*, 608:A5.
- Guo, H., Zheng, Z., Zehavi, I., Dawson, K., Skibba, R. A., Tinker, J. L., Weinberg, D. H., White, M., and Schneider, D. P. (2014). Velocity bias from the small-scale clustering of sdss-iii boss galaxies. *MNRAS*, 446(1):578–594.
- Hamilton, A. J. S. (1993). Toward better ways to measure the galaxy correlation function. *ApJ*, 417:19.
- Hemmati, S., Mobasher, B., and Jafariyazani, M. (2019). Spatially resolved properties of  $0.1 < z < 0.5$  galaxies from the muse-wide survey and candels survey. In *American Astronomical Society Meeting Abstracts# 233*, volume 233.
- Henault, F., Bacon, R., Bonneville, C., Boudon, D., Davies, R. L., Ferruit, P., Gilmore, G. F., LeFevre, O., Lemonnier, J.-P., Lilly, S., Morris, S. L., Prieto, E., Steinmetz, M., and de Zeeuw, P. T. (2003). MUSE: a second-generation integral-field spectrograph for the VLT. In Iye, M. and Moorwood, A. F. M., editors, *Instrument Design and Performance for Optical/Infrared Ground-based Telescopes*. SPIE.
- Herenz, E. C., Urrutia, T., Wisotzki, L., Kerutt, J., Saust, R., Werhahn, M., Schmidt, K. B., Caruana, J., Diener, C., Bacon, R., et al. (2017). The muse-wide survey: A first catalogue of 831 emission line galaxies. *A&A*, 606:A12.
- Herenz, E. C., Wisotzki, L., Saust, R., Kerutt, J., Urrutia, T., Diener, C., Schmidt, K. B., Marino, R. A., De La Vieuville, G., Boogaard, L., et al. (2019). The muse-wide survey: A determination of the lyman  $\alpha$  emitter luminosity function at  $3 < z < 6$ . *Astronomy & Astrophysics*, 621:A107.
- Hewett, P. C. (1982). The estimation of galaxy angular correlation functions. *MNRAS*, 201(4):867–883.
- Hildebrandt, H., Erben, T., Dietrich, J. P., Cordes, O., Habertzettl, L., Hetterscheidt, M., Schirmer, M., Schmithuesen, O., Schneider, P., Simon, P., and Trachternach, C. (2006). GaBoDS: The garching-bonn deep survey. *A&A*, 452(3):1121–1128.

- Hogg, D. W. (1999). Distance measures in cosmology. *arXiv preprint astro-ph/9905116*.
- Hubble, E. (1934). The distribution of extra-galactic nebulae. *ApJ*, 79:8.
- John, P., Cole Shaun, N. P., Baugh Carlton M., B.-H. J., Bridges Terry, C. R., D. Colless, M. C., Couch, C., et al. (2001). A measurement of the cosmological mass density from clustering in the 2df galaxy redshift survey. *Nature*, 410:169.
- Kauffmann, G., Colberg, J. M., Diaferio, A., and White, S. D. (1999). Clustering of galaxies in a hierarchical universe ii. evolution to high redshift. *MNRAS*, 307(3):529–536.
- Kauffmann, G., White, S. D., Heckman, T. M., Ménard, B., Brinchmann, J., Charlot, S., Tremonti, C., and Brinkmann, J. (2004). The environmental dependence of the relations between stellar mass, structure, star formation and nuclear activity in galaxies. *MNRAS*, 353(3):713–731.
- Kerscher, M., Szapudi, I., and Szalay, A. S. (2000). A comparison of estimators for the two-point correlation function. *ApJ*, 535(1):L13–L16.
- Khostovan, A. A., Sobral, D., Mobasher, B., Matthee, J., Cochrane, R. K., Soltani, N. C., Jafariyazani, M., Paulino-Afonso, A., Santos, S., and Calhau, J. (2018). The clustering of typical  $\text{Ly}\alpha$  emitters from  $z \sim 2.5$ -6: host halo masses depend on  $\text{Ly}\alpha$  and uv luminosities. *arXiv preprint arXiv:1811.00556*.
- Kovač, K., Somerville, R. S., Rhoads, J. E., Malhotra, S., and Wang, J. (2007). Clustering of  $\text{Ly}\alpha$  emitters at  $z \approx 4.5$ . *ApJ*, 668(1):15.
- Kriek, M., Van Dokkum, P. G., Labbé, I., Franx, M., Illingworth, G. D., Marchesini, D., and Quadri, R. F. (2009). An ultra-deep near-infrared spectrum of a compact quiescent galaxy at  $z = 2.2$ . *ApJ*, 700(1):221.
- Krumpe, M., Miyaji, T., and Coil, A. L. (2010). The spatial clustering of rosat all-sky survey agns. i. the cross-correlation function with sdss luminous red galaxies. *ApJ*, 713(1):558.
- Landy, S. D. and Szalay, A. S. (1993). Bias and variance of angular correlation functions. *ApJ*, 412:64.
- Lemson, G. et al. (2006). Halo and galaxy formation histories from the millennium simulation: Public release of a vo-oriented and sql-queryable database for studying the evolution of galaxies in the  $\Lambda$ cdm cosmogony. *arXiv preprint astro-ph/0608019*.

- Li, C., Kauffmann, G., Jing, Y., White, S. D., Börner, G., and Cheng, F. (2006). The dependence of clustering on galaxy properties. *MNRAS*, 368(1):21–36.
- Limber, D. N. (1953). The analysis of counts of the extragalactic nebulae in terms of a fluctuating density field. *ApJ*, 117:134.
- Lowenthal, J. D., Hogan, C. J., Green, R. F., Caulet, A., Woodgate, B. E., Brown, L., and Foltz, C. B. (1991). Discovery of a ly-alpha galaxy near a damped ly-alpha absorber at  $z = 2.3$ . *The Astrophysical Journal*, 377:L73–L77.
- Macchetto, F., Lipari, S., Giavalisco, M., Turnshek, D., and Sparks, W. (1993). Identification of a lyman-alpha radio-quiet galaxy at redshift  $z = 3.428$ —a primeval galaxy? *The Astrophysical Journal*, 404:511–520.
- Madgwick, D. S., Hawkins, E., Lahav, O., Maddox, S., Norberg, P., Peacock, J. A., Baldry, I. K., Baugh, C. M., Bland-Hawthorn, J., Bridges, T., et al. (2003). The 2df galaxy redshift survey: galaxy clustering per spectral type. *MNRAS*, 344(3):847–856.
- Maseda, M. V., Brinchmann, J., Franx, M., Bacon, R., Bouwens, R. J., Schmidt, K. B., Boogaard, L. A., Contini, T., Feltre, A., Inami, H., et al. (2017). The muse hubble ultra deep field survey-iv. global properties of c iii] emitters. *A&A*, 608:A4.
- Meier, D. L. (1976). The optical appearance of model primeval galaxies. *The Astrophysical Journal*, 207:343–350.
- Monaco, P., Møller, P., Fynbo, J., Weidinger, M., Ledoux, C., and Theuns, T. (2005). Tracing large-scale structure at high redshift with lyman- $\alpha$  emitters: the effect of peculiar velocities. *A&A*, 440(3):799–808.
- Norberg, P., Baugh, C. M., Hawkins, E., Maddox, S., Madgwick, D., Lahav, O., Cole, S., Frenk, C. S., Baldry, I., Bland-Hawthorn, J., et al. (2002). The 2df galaxy redshift survey: the dependence of galaxy clustering on luminosity and spectral type. *MNRAS*, 332(4):827–838.
- Norberg, P., Baugh, C. M., Hawkins, E., Maddox, S., Peacock, J. A., Cole, S., Frenk, C. S., Bland-Hawthorn, J., Bridges, T., Cannon, R., et al. (2001). The 2df galaxy redshift survey: luminosity dependence of galaxy clustering. *MNRAS*, 328(1):64–70.
- Ouchi, M., Harikane, Y., Shibuya, T., Shimasaku, K., Taniguchi, Y., Konno, A., Kobayashi, M., Kajisawa, M., Nagao, T., Ono, Y., et al. (2017). Systematic identification of laes for visible exploration and reionization research using subaru hsc (silverrush). i. program strategy and clustering properties of 2000 ly $\alpha$  emitters at  $z = 6-7$

- over the 0.3–0.5 gpc<sup>2</sup> survey area. *Publications of the Astronomical Society of Japan*, 70(SP1):S13.
- Ouchi, M., Shimasaku, K., Furusawa, H., Miyazaki, M., Doi, M., Hamabe, M., Hayashino, T., Kimura, M., Kodaira, K., Komiyama, Y., et al. (2003). Subaru deep survey. ii. luminosity functions and clustering properties of ly $\alpha$  emitters at  $z = 4.86$  in the subaru deep field. *ApJ*, 582(1):60.
- Ouchi, M., Shimasaku, K., Furusawa, H., Saito, T., Yoshida, M., Akiyama, M., Ono, Y., Yamada, T., Ota, K., Kashikawa, N., Iye, M., Kodama, T., Okamura, S., Simpson, C., and Yoshida, M. (2010). Statistics of 207 ly $\alpha$  emitters at a redshift near 7: Constraints on reionization and galaxy formation models. *ApJ*, 723(1):869–894.
- Partridge, B. (1974). A search for primeval galaxies at high redshifts.
- Partridge, R. B. and Peebles, P. J. E. (1967). Are young galaxies visible? *ApJ*, 147:868.
- Peebles (1980). The large-scale structure of the universe. *Princeton University Press*, 196:435.
- Peebles, P. J. E. and Hauser, M. G. (1974). Statistical analysis of catalogs of extragalactic objects. III. the shane-wirtanen and zwicky catalogs. *ApJS*, 28:19.
- Pezzotta, A., De La Torre, S., Bel, J., Granett, B., Guzzo, L., Peacock, J., Garilli, B., Scodreggio, M., Bolzonella, M., Abbas, U., et al. (2017). The vimos public extragalactic redshift survey (vipers)-the growth of structure at  $0.5 < z < 1.2$  from redshift-space distortions in the clustering of the pdr-2 final sample. *A&A*, 604:A33.
- Pons-Borderia, M.-J., Martinez, V. J., Stoyan, D., Stoyan, H., and Saar, E. (1999). Comparing estimators of the galaxy correlation function. *ApJ*, 523(2):480–491.
- Scoville, N., Aussel, H., Brusa, M., Capak, P., Carollo, C. M., Elvis, M., Giavalisco, M., Guzzo, L., Hasinger, G., Impey, C., Kneib, J.-P., LeFevre, O., Lilly, S. J., Mobasher, B., Renzini, A., Rich, R. M., Sanders, D. B., Schinnerer, E., Schminovich, D., Shopbell, P., Taniguchi, Y., and Tyson, N. D. (2007). The cosmic evolution survey (COSMOS): Overview. *ApJS*, 172(1):1–8.
- Seldner, M., Siebers, B., Groth, E. J., and Peebles, P. J. E. (1977). New reduction of the lick catalog of galaxies. *ApJ*, 82:249.
- Shapley and Ames (1932). A survey of the external galaxies brighter than the thirteenth magnitude. *Annals of Harvard College Observatory*, 88:41–76.

- Shibuya, T., Ouchi, M., Konno, A., Higuchi, R., Harikane, Y., Ono, Y., Shimasaku, K., Taniguchi, Y., Kobayashi, M. A. R., Kajisawa, M., Nagao, T., Furusawa, H., Goto, T., Kashikawa, N., Komiyama, Y., Kusakabe, H., Lee, C.-H., Momose, R., Nakajima, K., Tanaka, M., Wang, S.-Y., and Yuma, S. (2017). Silverrush. ii. first catalogs and properties of 2000  $\text{Ly}\alpha$  emitters and blobs at  $z = 6 - 7$  identified over the 14-21  $\text{deg}^2$  sky. *Publications of the Astronomical Society of Japan*.
- Shioya, Y., Taniguchi, Y., Sasaki, S., Nagao, T., Murayama, T., Saito, T., Ideue, Y., Nakajima, A., Matsuoka, K., Trump, J., et al. (2009). Photometric properties of  $\text{Ly}\alpha$  emitters at  $z \approx 4.86$  in the cosmos 2 square degree field. *ApJ*, 696(1):546.
- Tucker, D., Oemler Jr, A., Kirshner, R., Lin, H., Shectman, S., Landy, S., Schechter, P., Müller, V., Gottlöber, S., and Einasto, J. (1997). The las campanas redshift survey galaxy—galaxy autocorrelation function. *MNRAS*, 285(1):L5–L9.
- Urrutia, T., Wisotzki, L., Kerutt, J., Schmidt, K. B., Herenz, E. C., Klar, J., Saust, R., Werhahn, M., Diener, C., Caruana, J., Krajnović, D., Bacon, R., Boogaard, L., Brinchman, J., Enke, H., Maseda, M., Nanayakkara, T., Steinmetz, M., and Weilbacher, P. M. (2019). The muse-wide survey: Survey description and first data release. *A&A*.
- Ventou, E., Contini, T., Bouché, N., Epinat, B., Brinchmann, J., Bacon, R., Inami, H., Lam, D., Drake, A., Garel, T., et al. (2017). The muse hubble ultra deep field survey-ix. evolution of galaxy merger fraction since  $z \approx 6$ . *A&A*, 608:A9.
- Wisotzki, L., Bacon, R., Brinchmann, J., Cantalupo, S., Richter, P., Schaye, J., Schmidt, K., Urrutia, T., Weilbacher, P., Akhlaghi, M., et al. (2018). Nearly all the sky is covered by lyman- $\alpha$  emission around high-redshift galaxies. *Nature*, 562(7726):229.
- Wolfe, A., Turnshek, D., Lanzetta, K., and Oke, J. (1992). Lyman-alpha emission from the damped lyman-alpha system toward h0836+ 113. *The Astrophysical Journal*, 385:151–172.
- Zehavi, I., Blanton, M. R., Frieman, J. A., Weinberg, D. H., Mo, H. J., Strauss, M. A., Anderson, S. F., Annis, J., Bahcall, N. A., Bernardi, M., et al. (2002). Galaxy clustering in early sloan digital sky survey redshift data. *ApJ*, 571(1):172.
- Zehavi, I., Zheng, Z., Weinberg, D. H., Blanton, M. R., Bahcall, N. A., Berlind, A. A., Brinkmann, J., Frieman, J. A., Gunn, J. E., Lupton, R. H., et al. (2011). Galaxy clustering in the completed sdss redshift survey: The dependence on color and luminosity. *ApJ*, 736(1):59.

Zehavi, I., Zheng, Z., Weinberg, D. H., Frieman, J. A., Berlind, A. A., Blanton, M. R., Scoccimarro, R., Sheth, R. K., Strauss, M. A., Kayo, I., et al. (2005). The luminosity and color dependence of the galaxy correlation function. *ApJ*, 630(1):1.

# Deutsche Zusammenfassung

Das Studium der Verteilung von LAEs im Universum offenbart uns wie das Universum von  $3 < z < 6$  strukturiert ist. Durch die Analyse der Galaxiencluster/anhäufungen kann man ermitteln, wo Materieüberdichten liegen, die Größe und Population der Dunkle-Materie-Halos einschränken, kosmologischen Modelle testen und Schwankungen im frühen Universum aufzeigen. Außerdem gibt uns die Untersuchung der Cluster-Entwicklung in verschiedenen kosmologischen Epochen die Möglichkeit, gängige Galaxienbildungs- und Evolutionstheorien zu überprüfen.

Das erste Ziel dieser Masterarbeit war es, die in [Diener et al. \(2017\)](#) gezeigten Ergebnisse zu reproduzieren, analysierte ich die ersten 24 Felder von MUSE-Wide und nachdem ich die Studie auf die vollständige MUSE-Wide-Survey und MUSE-Deep-Survey ausgedehnt habe. Die Anzahl der LAEs beträgt 234, 947 (814 in 68 Felder) und 757 LAEs (Lyman Alpha Emitter). Dabei wurden Bereiche von  $22.2 \text{ arcmin}^2$ ,  $63.83 \text{ arcmin}^2$  ( $62.9 \text{ arcmin}^2$  für die 68 Felder) und  $8.32 \text{ arcmin}^2$ . Die Rotverschiebungsabdeckung aller Stichprobe umfasst  $3 < z < 6$ , mit  $z = 4.04$ ,  $z = 4.07$  und  $z = 4.14$  die mittlere Rotverschiebung jeder Stichprobe.

Aufgrund der speziellen MUSE-Daten, des großen Rotverschiebungsbereichs und der kleinen Winkelabdeckung habe ich verschiedene Möglichkeiten zur Messung der Clustering der LAEs ausprobiert. Aufgrund des Datentyps sind die Proben ideal für die radiale Clustering-Methode. Ich studierte das Rotverschiebungs-Clustering anhand des Schätzers  $K_{0,30}^{30,60}$  und ermittelte die Korrelationslängen von  $r_0 = 2.7_{-0.4}^{+0.3}$  und  $r_0 = 2.6_{-0.2}^{+0.2}$  in den MUSE-Wide und MUSE-Deep Stichproben. Die Steigung des K-Schätzers war  $\gamma = 1.8$  und liefert in beiden Himmelsbereichen konsistente Ergebnisse. Die Unsicherheiten in den geschätzten Korrelationslängen sind jeweils  $\pm 13\%$  und  $\pm 8\%$ , wobei beide Längen mit den Literaturwerten bei ähnlichen Rotverschiebungen übereinstimmen.

Anschließend wurde die Methode detailliert untersucht, indem eine realistische Zufallsstichprobe der in den MUSE-Umfragen enthaltenen LAEs erstellt wurde. Die Auswahl und Helligkeitsfunktionen und Fußabdruck der MUSE-Wide wurden berücksichtigt, um

eine realistische Zufallsstichprobe von LAEs zu entwickeln, an denen das Rotverschiebungs Clustering getestet werden konnte. Wie erwartet lieferte der K-Schätzer Werte, die mit null Clustering übereinstimmen, nachdem die Methode 100 mal ausgeführt und gemittelt wurde. Diese Analyse bot nicht nur die Gewähr, dass unsere Messerergebnisse nicht verfälscht sind, sondern auch ein besseres Verständnis darüber, wie realistisch Poisson-Fehler sind. In allen Fällen sind die durch die Methode selbst eingeführten Fehler viel kleiner als die berechneten Poisson-Fehler.

Ich untersuchte auch die Clusterabhängigkeit mit verschiedenen physikalischen Eigenschaften: die Rotverschiebung, der Sichtbare Galaxienmasse, der Leuchtkraft der  $Ly\alpha$  und der Equivalentbreite der Emissionslinie von  $Ly\alpha$ . Obwohl die Fehlerbalken zu groß waren um eine klare Aussage über die Clusteringstärke einzelner Unterstichproben zu machen, konnten in verschiedenen Fällen deutliche Tendenzen beobachtet werden. Der K-Schätzer scheint mit zunehmender Rotverschiebung höhere Cluster-Trends für LAEs hinzuweisen. Auf diese Weise können wir verschiedene beobachtete Populationen von LAEs in verschiedenen Epochen miteinander verbinden. In Kombination mit theoretischen Modellen des Halowachstum der dunklen Materie (DM) kommen wir zu dem Schluss, dass sich Galaxien früh in sehr überdichten Regionen des Universums gebildet haben.

Noch deutlichere Trends werden beobachtet, wenn wir das Rotverschiebungs-Clustering als Funktion der Helligkeit untersuchen. Hellere LAEs, die massiveren Halos wohnen, scheinen sich stärker gruppieren als licht LAEs. Für den Rest der Fälle (Galaxienmasse und Equivalentbreite) sind die Fehlerbalken zu groß, um zwischen dem Clustering von LAEs bei niedrigen und hohen Werten beider Eigenschaften unterscheiden zu können. Ich hoffe jedoch, dass die neuen Fehlerberechnungen (siehe Abschnitt 7.2) zur Lösung dieses Problems beitragen können.

Die Ergebnisse dieser Arbeit zeigen, dass LAEs Vorgänger der verschiedenen Arten von Galaxien sein können, die im nahen Universum zu sehen sind, einschließlich Galaxien ähnlich der Milchstraße. Da, wie herausfand, helle LAEs sich stärker als licht LAEs gruppieren, könnten helle LAEs Protocluster sein, die dichtesten Regionen im frühen Universum, die sich zu den massiven Galaxienhaufen entwickeln, die wir heute beobachten.

Um die Richtigkeit des mit der Radialen-Clustering-Methode erhaltenen Clustersignals sicherzustellen, habe ich die Masterarbeit erweitert, um die traditionelle räumliche Clustering Methode in der gesamten MUSE-Wide-Survey zu verwenden. Die beiden am häufigsten verwendeten Schätzer, die Schätzer von Landy-Szalay und Davis, wurden verwendet, um die projizierte Korrelationsfunktion zu berechnen, wobei beide grundsätzlich die selben Ergebnisse liefern. Abgesehen von der Einschätzung eines eindeutigen Clustersignals und einer vollständigen Abdeckung des Einhalo-Terms ( $\omega_p$  gibt ein positives

Signal bis zu  $r_p = 1.4h^{-1}\text{Mpc}$ ), die von beiden Schätzern berechneten Korrelationslängen stimmten vollständig mit den Werten, die mit der Rotverschiebungs-Clustering-Methode erhalten wurden. In diesem Fall war die beste Anpassung für die Steigung der Korrelationsfunktion  $\gamma = 1.9$ .

Zusammenfassend bieten LAEs einen großen Einblick in die Entstehung und Entwicklung der Galaxien im nahen und fernen Universum. Indem wir deren Clustering in einer Rotverschiebung von  $3 < z < 6$  untersuchen, konnten wir zum ersten Mal den Ein-Halo-Term unserer LAEs anhand der projizierten Korrelationsfunktion untersuchen, verschiedene und neue Methoden des Clustering testen, ihre Korrelationslängen und Steigungen ermitteln, die Abhängigkeit von Galaxien Clustering als Funktion verschiedener physikalischer Eigenschaften untersuchen, Galaxien vom Ende der Epoche der Reionisation bis zum Höhepunkt der kosmischen Sternentstehung miteinander verbinden und realistische MUSE LAEs Stichproben erstellen. Auf diesem Weg sind wir in der Lage, die großräumige Struktur des Universums zu verfolgen, Galaxienbildungs- und Evolutionstheorien zu testen und die Parameter kosmologischer Modelle einzuschränken. Insbesondere können wir die DM-Halopopulationen, ihre Verteilung und die DM-Halomassen bestimmen, die die LAEs unserer Stichprobe enthalten.

# Eidesstattliche Erklärung

Hier versichere ich, dass ich die vorliegende Masterarbeit selbstständig und ohne fremde Hilfe verfasst habe. Alle verwendeten Quellen und Materialien sind in der Literaturliste aufgeführt, weitere wurden nicht benutzt. Wortwörtlich und inhaltlich entnommene Stellen sind als solche kenntlich gemacht. Die Masterarbeit hat keiner anderen Prüfungsbehörde vorgelegen.

Potsdam, den

Yohana Herrero

# APPENDIX

```
#Zf, RA and DECf are the redshifts, right ascensions and declinations,  
#directly read from MUSE catalogues.
```

```
from __future__ import division  
from astropy.utils.data import get_pkg_data_filename  
from astropy.table import Table  
from astropy.io import fits  
import matplotlib.pyplot as plt  
import numpy as np  
import math  
from astropy.cosmology import FlatLambdaCDM  
cosmo = FlatLambdaCDM(H0=70, Om0=0.3)  
import itertools  
import matplotlib  
  
#Right Ascension correction  
RA_offset = (RA - np.mean(RA))*np.cos(DECf*math.pi/180)  
RAf = np.mean(RA) + RA_offset  
  
#Radial distance calculation  
zij = np.array([])  
co = np.array([])  
cos = np.array([])  
for k, zk in enumerate(Zf):  
    for l, zl in enumerate(Zf[k+1:]):  
        co = np.append(co, cosmo.comoving_distance(zk).value)  
        cos = np.append(cos, cosmo.comoving_distance(zl).value)  
        zij = np.append(zij, abs(co-cos))
```

```

#Angular separation calculation
phi = np.array([])
d = np.array([])
r = np.array([])
for i, deci in enumerate(DECf):
    d = np.append(d, deci-DECf[i+1:])
for j, rai in enumerate(RAf[i+1:]):
    r = np.append(r, RAf[i]-rai)
phi = np.append(phi, np.sqrt((r)**2+(d)**2)*math.pi/180)
#Transverse distance calculation
rij = np.array([])
pairs = itertools.combinations(Zf, 2)
rzav = np.array([])

for pair in pairs:
    rzav= np.append(rzav, cosmo.comoving_distance((pair[0] + pair[1])/2).value)
    rij = np.append(rij, rzav * phi )

#K estimator calculation, the l.o.s. distance selection is from 30 to 60
kab = np.array([])
bins = np.arange(0,12,1.5) #MUSE-Wide limit of the field of view
err = np.array([])
binp = np.array([])
for k, bini in enumerate(bins):
    idxtrans = (rij >= bini) & (rij < (bini+1.5))
    idxl.o.s.1 = (zij > 0) & (zij < 30)
    idxl.o.s.2 = (zij > 0) & (zij < 60)
    kab = np.append(kab, sum(idxtrans & idxlos1)/sum(idxtrans & idxlos2))
    err = np.append(err, math.sqrt(sum(idxtrans & idxlos1))/sum(idxtrans & idxlos2))
    binp = np.append(binp, bini + 1.5/2)

#K fit
r0val = np.arange(1., 5., 0.1)
xi_ij_a = np.array([])
xi_ij_b = np.array([])
err1 = np.array([])
k1 = np.array([])

```

```

idx = (rij < 5.)
transin = rij[idx]
for j, r0valj in enumerate(r0val):
for h, transinh in enumerate(transin):
integ1 = integrate.quad(lambda Z: (math.sqrt(transinh**2 + Z**2) /
r0valj)**(-1.8),30,60)[0]
integ2 = integrate.quad(lambda Z: (math.sqrt(transinh**2 + Z**2) /
r0valj)**(-1.8),0,30)[0]
xi_ij_a = np.append(xi_ij_a, (1/(60-30)) * integ1)
xi_ij_b = np.append(xi_ij_b, (1/(30-0)) * integ2)
k1 = np.append(k1, (((30-0)*sum(1+xi_ij_b[j*len(transin):])))/(((30-0)*
sum(1+xi_ij_b[j*len(transin):])+(60-30)*sum(1+xi_ij_a[j*len(transin):])))

kmeasured = sum(idx & idxlos1)/sum(idx & idxlos2)
err1 = np.append(err1, math.sqrt(sum(idx & idxlos1))/sum(idx & idxlos2))
difference = np.abs(k1-kmeasured)
differencelow = np.abs(k1-(kmeasured-err1))
differenceup = np.abs(k1-(kmeasured+err1))
min_dif = np.argmin(difference)
r0opt = r0val[min_dif]
min_dif_up = np.argmin(differenceup)
r0up = r0val[min_dif_up]
min_dif_low = np.argmin(differencelow)
r0low = r0val[min_dif_low]

print('The ideal r0', r0opt)
print('Upper limit', r0up)
print('Lower limit', r0low)

```