The relationship between the intergalactic medium and galaxies

TEJOS, NICOLAS,ANDRES

The relationship between the intergalactic medium and galaxies

Nicolas Tejos

Abstract

In this thesis we study the relationship between the intergalactic medium (IGM) and galaxies at $z \lesssim 1$, in a statistical manner. Galaxies are mostly surveyed in emission using optical spectroscopy, while the IGM is mostly surveyed in absorption in the ultra-violet (UV) spectra of background quasi-stellar objects (QSOs). We present observational results investigating the connection between the IGM and galaxies using two complementary methods:

- We use galaxy voids as tracers of both underdense and overdense regions. We use archival data to study the properties of H I absorption line systems within and around galaxy voids at $z < 0.1$. Typical galaxy voids have sizes $\gtrsim 10$ Mpc and so our results constrain the very large-scale association. This sample contains 106 H I absorption systems and 1054 galaxy voids.

- We use a sample of H I absorption line systems and galaxies from pencil beam surveys to measure the H I–galaxy cross-correlation at $z \lesssim 1$. Our sample is composed of a combination of archival and new data taken by the author and collaborators. This survey covers transverse separations between H I and galaxies from ~ 100 kpc (proper) up to ~ 10 Mpc, filling the gap between the very large scales and those associated with the so-called circumgalactic medium (CGM). This sample contains 654 H I absorption systems and 17509 galaxies.

Our results hint towards a picture in which there are *at least* three types of association between the diffuse gas in the Universe and galaxies at $z \lesssim 1$:

- *One-to-one direct association* because galaxies do contain diffuse gas.

- *Indirect association* because both the IGM and galaxies trace the same overdense underlying dark matter distribution. We provide quantitative evidence for this association. Moreover, we show that not all galaxies are related to the diffuse gas in the same way. In particular, a non negligible fraction of 'non-star-forming' galaxies might reside in environments devoid of diffuse H I.

- *No association* because there are regions in the Universe that contain a significant amount of diffuse gas but that are devoid of galaxies. In these regions, only the IGM follows the underdense underlying dark matter distribution because galaxies are not present. We provide quantitative evidence for this scenario.

The relationship between the intergalactic medium and galaxies

Nicolas Andres Tejos Salgado

A Thesis presented for the degree of
Doctor of Philosophy

Extragalactic Astronomy and Cosmology
Institute for Computational Cosmology
Department of Physics
University of Durham
United Kingdom

July 2013

Contents

List of Tables

List of Figures

Declaration

The work described in this thesis was undertaken between 2009 and 2013 while the author was a research student under the supervision of Professor Simon L. Morris and Dr. Tom Theuns at the Extragalactic Astronomy and Cosmology Group and the Institute for Computational Cosmology in the Department of Physics at the University of Durham. This work has not been submitted for any other degree at the University of Durham or any other university.

Chapter 2 has been published in the following paper:

- *Large scale structure in absorption: Gas within and around galaxy voids* Tejos, N.; Morris, S. L.; Crighton, N. H. M.; Theuns, T., Altay, G.; Finn, C. W. 2012, MNRAS, 425, 245-260, (Tejos et al., 2012).

Chapter 4 has been recently accepted for publication in the form of a paper:

- *On the connection between the intergalactic medium and galaxies: The H* I*–galaxy cross-correlation at* $z \lesssim 1$ Tejos, N.; Morris, S. L.; Finn, C. W.; Crighton, N. H. M.; Bechtold, J.; Jannuzi, B. T.; Schaye, J.; Theuns, T.; Altay, G.; Le Fèvre, O.; Ryan-Weber, E.; Dave, R., 2013, MNRAS, accepted (Tejos et al., 2013).

During this time, the author has also contributed to the following published work:

- *A high molecular fraction in a sub-damped absorber at* $z = 0.56$ Crighton, N. H. M.; Bechtold, J.; Carswell, R. F.; Dave, R.; Foltz, C. B.; Jannuzi, B. T.; Morris, S. L.; O'Meara, J. M.; Prochaska, J. X.; Schaye, J.; Tejos, N. 2013, MNRAS, 433, 178-193, (Crighton et al., 2013).

Other published papers by the author are:

- *Casting light on the 'anomalous' statistics of Mg* II *absorbers toward gamma-ray burst afterglows: The incidence of weak systems* Tejos, N.; Lopez, S.; Prochaska, J. X.; Bloom, J. S.; Chen, H.-W.; Dessauges-Zavadsky, M; Maureira, M. J. 2009, ApJ, 706, 1309-1315, (Tejos et al., 2009).

- *Galaxy clusters in the line of sight to background quasars. II. Environmental effects on the sizes of baryonic halo sizes* Padilla, N.; Lacerna, I.; Lopez, S.; Barrientos, L. F.; Lira, P.; Andrews, H; Tejos, N. 2009, MNRAS, 395, 1135-1145, (Padilla et al., 2009).

- *Galaxy clusters in the line of sight to background quasars. I. Survey design and incidence of Mg* II *absorbers at cluster redshifts* Lopez, S.; Barrientos, L. F.; Lira, P.; Padilla, N.; Gilbank, D. G.; Gladders, M. D.; Maza, J.; Tejos, N.; Vidal, M.; Yee, H. K. C. 2008, ApJ, 679, 1144-1161, (Lopez et al., 2008).

- *On the incidence of C* IV *absorbers along the sight lines to gamma-ray bursts* Tejos, N.; Lopez, S.; Prochaska, J. X.; Chen, H.-W.; Dessauges-Zavadsky, M. 2007, ApJ, 671, 622-627, (Tejos et al., 2007).

Most of the work in this thesis was completed by the author. Those parts not done by him are properly acknowledged in the text or in the following list:

- In Chapter 4, the acquisition, reduction and galaxy redshift determination of DEIMOS data were mostly done by Neil H. M. Crighton, Jill Bechtold, Buell T. Jannuzi and Allison L. Strom. The author used these data to create and characterize the final heterogeneous catalog.
- In Chapter 4, the acquisition, reduction and galaxy redshift determination of GMOS data were mostly done by Neil H. M. Crighton. The author used these data to create and characterize the final heterogeneous catalog.
- In Chapter 4, the reduction of COS data and creation of the H I line list were mostly done by Charles W. Finn and Neil H. M. Crighton. The author helped to assess the uncertainty of the process by performing an independent search in a subsample of the data. The characterization of the H I catalog was mostly done by the author.
- In Chapter 4, the reduction of FOS data was done by Neil H. M. Crighton. The characterization of the H I catalog was mostly done by the author.

Acknowledgements

First of all, I am deeply grateful to my parents, Pedro and Loreto, and my siblings, Pedro, Juan Pablo and Loreto. Their unconditional love and constant support have been fundamental to my development as a person. My sincere and heartfelt gratitude to my beloved wife Macarena; thanks for being always there for me, for sharing your life with mine, and for making me the happiest man in the world.

Many thanks to Simon Morris for having been an excellent mentor and supervisor. I thank him for his trust and support in all my scientific endeavours, for his insightful comments and suggestions, and for teaching me to be always skeptical (even of my own conclusions). I also thank Tom Theuns for his support and for always suggesting new and interesting ideas to explore. Special thanks to Neil Crighton for his major contribution to the analysis presented in this Thesis, in particular for sharing with me his programming skills and software, and for always having time to discuss results and problems encountered. Many thanks to Charles Finn for being an excellent collaborator and friend.

I am also grateful to Tom Shanks and Philipp Richter for having kindly agreed to be the reviewers of this Thesis. I thank them for their comments and suggestions that improved the quality of the present work.

My time in Durham would not have been so enjoyable if it weren't for all the good times shared with my fellow astronomers. I would like to thank my office mates and all who regularly participated in our journal clubs; thank you all for sharing knowledge, for having helped to create a nice working atmosphere, and for being such good friends! Special thanks to: Alvaro, Elise, Violeta, Gabriel, Claudia, Yetli, Juan, Nikos[2], Alex[2], Peter[2], Sarah, Mark, Charles, Mike, Rachel, Matthieu, Michelle, Tamsyn, Neil, Rich, Pablo, Cai, Wojciech, Yuichi, Yusei, Lilian, Nok, George, Flora, Agnese, Poshak, Manolis, Rob, Will and Till.

Finally, I thank Eugenio Candia for having helped me to reconnect with Maths and Physics in a crucial moment of my vocational formation. I will always remember his teaching as an important milestone along the path I am travelling.

Dedicated to

My ever growing family

"Believe nothing on the faith of traditions,
even though they have been held in honor
for many generations and in diverse places.
Do not believe a thing because many people speak of it.
Do not believe on the faith of the sages of the past.
Do not believe what you yourself have imagined.
Believe nothing on the sole authority of your masters and priests.
After examination, believe what you yourself have tested
and found to be reasonable, and conform your conduct thereto."

— Siddhãrtha Gautama

Chapter 1 | *Introduction*

Given that the aim of this thesis is to address the relationship between the intergalactic medium (IGM) and galaxies, in this chapter we explicitly define what we mean when referring to these two concepts. To do so, we briefly review the key observational facts that has lead us to our current galaxy formation paradigm, as well as the main observables of these extragalactic objects.

1.1 Dark matter, dark energy and ΛCDM

Our current understanding of the Universe relies on the presence of a significant fraction of 'exotic'[1] forms of matter and energy, commonly referred by the use of the prefix 'dark'. Optically invisible matter has been invoked to explain several observations, spanning from the velocity dispersion of galaxies in galaxy clusters (e.g. Zwicky, 1933, 1937), rotational curves of spiral galaxies (e.g. Babcock, 1939; Rubin et al., 1980), gravitational lensing of galaxy clusters (e.g. Clowe et al., 2006) and the power spectrum of the cosmic microwave background (CMB) (e.g. Komatsu et al., 2011; Planck Collaboration et al., 2013). Some of this 'invisible' matter can still be non-exotic (normal), but just not emitting detectable optical light (e.g. gas, dust, planets, faint stars, etc.). However, these observations—interpreted under our accepted theoretical models—indicate that $\sim 80\%$ of the total matter in the Universe is indeed exotic 'dark matter' (e.g. Komatsu et al., 2011; Planck Collaboration et al., 2013). Direct experiments to detect such a dark matter particle in the laboratory have not been successful yet (e.g. Aprile et al., 2012, although see Bernabei et al. 2010 for a recent claim), hinting towards an extremely weak or

[1] With respect to what is normally found on the Earth.

null electromagnetic interaction of dark matter with baryons[2].

Observations of galaxies in the Local Universe have revealed that their radial velocities are proportional to their distances (e.g. Hubble, 1929). Such a behaviour is naturally explained under the framework of an expanding isotropic universe, giving important support to Einstein's General Theory of Relativity (GR, Einstein, 1916) applied to the Universe as a whole (e.g. Friedman, 1922; Lemaître, 1927; Robertson, 1935; Walker, 1935). If the Universe is expanding at the present, it should have been more compact, denser and hotter in the past, starting from a state in which all the constituents of the Universe may have been coupled (e.g. Lemaître, 1931). Observational evidence for such an state comes from (i) the CMB, interpreted as the radiation from the moment in which photons and matter decoupled after the Big Bang[3] (e.g. Dicke et al., 1965; Penzias and Wilson, 1965); and (ii) the abundances of light elements (in particular Deuterium, ^{2}H; e.g. Pettini and Cooke 2012), interpreted as being produced by the Big Bang nucleosynthesis (e.g. Gamow, 1948).

Under the framework of GR, the Universe can have a geometrical curvature. Current observations however—mainly from the CMB—favor no curvature at all (e.g. Komatsu et al., 2011; Planck Collaboration et al., 2013). In a flat, matter dominated universe, the expansion will eventually stop because gravity is an attractive force. Observational experiments to measure such a deceleration revealed an unexpected result: the Universe is not decelerating but rather accelerating (Riess et al., 1998; Perlmutter et al., 1999). Such a behaviour requires an extra component to the energy density of the Universe acting against gravity, referred as 'dark energy'. This dark energy can be described by a 'cosmological constant', Λ, in Einstein's field equations. Given the expansion of the Universe, the matter

[2]Note that astronomers usually include electrons and neutrinos (fermions) in the baryonic component census to make an explicit distinction between baryons ('normal' matter) and dark matter ('exotic' matter). Also note that fermions are considerably lighter than baryons and so, they do not contribute significantly to the total matter density.

[3]We define Big Bang as the 'origin' of the Universe, i.e. the extrapolated singularity in which the Universe's energy density and size tend to infinity and zero, respectively.

and radiation density energies decrease with time. In contrast, the density energy of Λ remains constant, which leads to a natural equivalence with the energy of space itself (vacuum). We note however that there is a significant discrepancy between the observed value of the cosmological dark energy density with that of the vacuum energy density predicted by the Quantum Field Theory (by ~ 120 orders of magnitude!; e.g. Weinberg, 1989; Carroll, 2001). Solving the tension between such observations and the theoretical predictions will probably lead to a major breakthrough in our understanding of the Universe. We note that some efforts to explain current observations without the need of 'dark matter' or 'dark energy' have been indeed presented (e.g. Milgrom, 1983; Moffat, 2006; Alfonso-Faus, 2008; Hajdukovic, 2011, 2012, among many others), all of which deserving proper scientific scrutiny. For the purposes of this thesis however, we will assume that we live in an accelerated expanding flat Universe, in which both dark matter and dark energy dominate the current energy densities (e.g. Komatsu et al., 2011; Planck Collaboration et al., 2013).

This cosmological paradigm is commonly referred as ΛCDM, which stands for 'Λ cold dark matter' (CDM). As mentioned, Λ corresponds to the cosmological constant used to explain 'dark energy', and CDM corresponds to the type of matter used to explain 'dark matter' (see Section 1.2 for further details). The term 'cold' in CDM is used to account for the massive and non-relativistic nature of these particles, which is needed to explain the observed large-scale structure (traced by galaxies) as being related to the temperature fluctuations observed in the CMB. If dark matter were relativistic ('hot'; e.g. neutrinos), then the resulting clustering of galaxies would be inconsistent with observations (e.g. White et al., 1983). Given that CDM is presureless and disipationeless, its evolution is only governed by gravity (e.g. Springel et al., 2005). Primordial baryonic material[4] will therefore cool (by radiation), condense, and eventually form stars and galaxies at the peaks of the underlying dark matter distribution (e.g. Press and Schechter,

[4]Primordial material is defined as that produced by the Big Bang nucleosynthesis.

1974; Rees and Ostriker, 1977; White and Rees, 1978; White and Frenk, 1991; Kereš et al., 2005). For an extended and exhaustive description of galaxy formation under the ΛCDM paradigm, we refer the reader to Mo et al. (2010).

1.2 Cosmology

Independently of the real nature of dark matter and dark energy, our current cosmological paradigm allows us to place extragalactic observations in a useful framework. For an isotropic and homogeneous universe, the most general space-time metric, $\mathrm{d}s$, has the form:

$$\mathrm{d}s^2 = -c^2\mathrm{d}t^2 + a^2(t)\left\{\frac{\mathrm{d}r^2}{1-kr^2} + r^2\mathrm{d}\theta^2 + r^2\sin^2(\theta)\mathrm{d}\phi^2\right\} , \tag{1.1}$$

(e.g. Friedman, 1922; Lemaître, 1927; Robertson, 1935; Walker, 1935), where c is the speed of light (constant), t is the cosmological time, r, θ and ϕ are the 'co-moving' spatial spherical coordinates (radial distance, polar angle and azimuthal angle respectively), k is the sign of the spatial curvature of the Universe ($k \in \{-1, 0, 1\}$ corresponding to open, flat and close respectively), and $a(t)$ is the cosmological scale factor. Basically, this is a metric of a universe which expands or contracts by a factor $a(t)$, which is a function of t only.

From such a metric with $k = 0$, the proper distance between two objects separated by r in co-moving coordinates is,

$$d_\mathrm{p}(t) = \int_0^r a(t)\mathrm{d}r = a(t)r , \tag{1.2}$$

which leads to a proper apparent velocity of,

$$\dot{d}_\mathrm{p}(t) = \dot{a}(t)r = \left(\frac{\dot{a}}{a}\right) d_\mathrm{p}(t) , \tag{1.3}$$

given by the expansion (or contraction) of the Universe. This is a generalized expression of Hubble's findings (Hubble, 1929):

$$v = H_0 d \,, \tag{1.4}$$

where v is the observed velocity, d is the observed distance, and H_0 is commonly referred as the Hubble constant. In fact, Equation 1.3 defines the Hubble parameter,

$$H(t) \equiv \left(\frac{\dot{a}}{a}\right) \,, \tag{1.5}$$

which is a function of cosmological time. In the Local Universe, we have that $H_0 = H(t_0)$, where t_0 is the present time.

One of the principles of Einstein's Special Theory of Relativity (SR, Einstein, 1905) is that the light in vacuum propagates at a constant (finite) speed (e.g. see Michelson et al., 1935, for observational evidence), which is independent of the reference frame in which it is emitted or observed (e.g. see Michelson and Morley, 1887; Michelson, 1925; Michelson and Gale, 1925, for observational evidence). This means that the further the distance to an observed object is, the further in the past its photons were emitted. In GR, because of the expansion of the Universe, photons will also get their wavelengths dilated,

$$\frac{\lambda_0}{\lambda_\mathrm{e}} = \frac{a(t_0)}{a(t_\mathrm{e})} \,, \tag{1.6}$$

where λ_0 is the observed wavelength at t_0, and λ_e is the emitted wavelength at t_e, due to the time dilation,

$$\frac{\Delta t_0}{a(t_0)} = \frac{\Delta t_\mathrm{e}}{a(t_e)} \,. \tag{1.7}$$

This means that knowing the rest-frame wavelength of a photon, λ_rest, we can infer the scale factor at which a photon was emitted, $a(t_\mathrm{e})$, by looking at the difference between the observed and rest-frame wavelength,

$$z_\mathrm{cos} \equiv \frac{\lambda_\mathrm{obs} - \lambda_\mathrm{rest}}{\lambda_\mathrm{rest}} = \frac{a(t_0)}{a(t_\mathrm{e})} - 1 \,, \tag{1.8}$$

where $z_{\rm cos}$ is defined as the cosmological redshift. Extragalactic objects will have other components contributing to the observed redshift (non-cosmological), like Doppler redshift ($z_{\rm pec}$; due to the line-of-sight component of peculiar velocities) and gravitational redshift ($z_{\rm grav}$; due to the time dilation in gravitational potentials). In such a case the observed redshift will be given by,

$$1 + z_{\rm obs} \approx (1 + z_{\rm cos})(1 + z_{\rm pec})(1 + z_{\rm grav}) \, . \tag{1.9}$$

For the typical galaxies and IGM clouds considered in this thesis, the gravitational redshift can be neglected however. Assuming locally flat space-time, the relativistic Doppler redshift given by the peculiar velocity along the line-of-sight, $v_{\rm pec}$, is

$$1 + z_{\rm pec} = \sqrt{\frac{1 + v_{\rm pec}/c}{1 - v_{\rm pec}/c}} \, . \tag{1.10}$$

which for $v_{\rm pec} \ll c$ can be approximated by $1 + z_{\rm pec} \approx 1 + v_{\rm pec}/c$. It follows that the peculiar velocity difference (along the line-of-sight), Δv, between two objects having a small observed redshift difference, Δz, at redshift $z_{\rm cos}$ is

$$\frac{\Delta v}{c} = \frac{\Delta z}{(1 + z_{\rm cos})} \, . \tag{1.11}$$

In order to know the cosmological time at which the photons of an extragalactic object were emitted (or absorbed) one needs to know the functional form of $a(t)$. Using the metric presented in Equation 1.1 with $k = 0$ in Einstein's field equations, we obtain the so-called Friedman equation,

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3c^2}\epsilon + \frac{\Lambda}{3} \, , \tag{1.12}$$

where G is the gravitational constant and ϵ is the energy density of the Universe, and Λ is the cosmological constant (for a pedagogical derivation see Ryden, 2003). Assuming isotropy and homogeneity, ϵ is a function of t only. If we model the Universe as a perfect fluid, by the First Law of Thermodynamics (conservation of

energy) we obtain the so-called fluid equation,

$$\dot{\epsilon} + 3\left(\frac{\dot{a}}{a}\right)(\epsilon + p) = 0 \,, \tag{1.13}$$

where p is the pressure (e.g. Ryden, 2003). From these equations Λ is equivalent to having $\epsilon_\Lambda = \frac{\Lambda}{8\pi G}$ and $p_\Lambda = -\epsilon_\Lambda$.

In the general case, Equations 1.12, 1.13 provide 2 independent equations and 3 unknown variables: a, ϵ and p. A third independent equation is given by the relation between p and ϵ, the equation of state. For individual fluids such equation is of the form

$$p(\epsilon) = w\epsilon \,, \tag{1.14}$$

where w is assumed to be independent of time.[5] For a fluid composed of photons $w_\mathrm{r} = 1/3$; for a fluid composed of presureless non-relativistic matter $w_\mathrm{m} = 0$; and for a fluid composed of dark energy coming from a cosmological constant $w_\Lambda = -1$. Considering an Universe containing only those components, from Equation 1.13 we have that

$$\epsilon_i(a) = \epsilon_{i,0} a^{-3(1+w_i)} \,, \tag{1.15}$$

where the sub-index i denotes an individual component. $\epsilon_{i,0}$ is the value when $a = 1$, which we arbitrarily define as the present, $a(t_0) \equiv 1$. As expected, ϵ_m decreases in proportion to the increase in volume as $\epsilon_\mathrm{m} \propto a^{-3}$; $\epsilon_\mathrm{r} \propto a^{-4}$ decreases faster than an increase in volume by a factor of a^{-1}, which is consistent with the extra energy lost due to the wavelength dilatation (e.g. see Equation 1.6); and ϵ_Λ is constant.

The total density energy is usually written as,

$$\Omega = \Omega_\mathrm{r} + \Omega_\mathrm{m} + \Omega_\Lambda \,, \tag{1.16}$$

[5] This might not always be the case though.

where $\Omega_i \equiv \epsilon_i/\epsilon_{c,0}$, and

$$\epsilon_{c,0} \equiv \frac{3c^2 H_0^2}{8\pi G} \,, \tag{1.17}$$

is the critical density energy at the present, defined as that of a flat Universe.[6] From Equation 1.15,

$$\Omega(a) = \Omega_{\mathrm{r},0} a^{-4} + \Omega_{\mathrm{m},0} a^{-3} + \Omega_{\Lambda,0} \tag{1.18}$$

where as usual, the sub-index 0 denotes that of the present value. Replacing our new definitions into Equation 1.12 we have

$$\frac{\dot{a}}{a} = H_0 (\Omega_{\mathrm{r},0} a^{-4} + \Omega_{\mathrm{m},0} a^{-3} + \Omega_{\Lambda,0})^{1/2} \tag{1.19}$$

in which H_0, $\Omega_{\mathrm{r},0}$, $\Omega_{\mathrm{m},0}$ and $\Omega_{\Lambda,0}$ are all observables.

In this thesis we will use the observed redshift of an extragalactic object to obtain their cosmological radial (along the line-of-sight) co-moving distance. From the metric presented in Equation 1.1 and our previous findings, the radial co-moving distance is given by,

$$\begin{aligned} r &= c \int_{t_\mathrm{e}}^{t_0} \frac{\mathrm{d}t}{a(t)} \\ &= c \int_{a(t_\mathrm{e})}^{a(t_0)\equiv 1} \frac{\mathrm{d}a}{a(t)\dot{a}(t)} \\ &= \frac{c}{H_0} \int_{a(t_\mathrm{e})}^{1} \frac{\mathrm{d}a}{a^2 (\Omega_{\mathrm{r},0} a^{-4} + \Omega_{\mathrm{m},0} a^{-3} + \Omega_{\Lambda,0})^{1/2}} \,, \end{aligned} \tag{1.20}$$

which in terms of cosmological redshift (observable) is finally given by

$$r = \frac{c}{H_0} \int_0^{z_\mathrm{cos}} \frac{\mathrm{d}z}{\{\Omega_{\mathrm{r},0}(1+z)^4 + \Omega_{\mathrm{m},0}(1+z)^3 + \Omega_{\Lambda,0}\}^{1/2}} \,. \tag{1.21}$$

[6] Note that for an actually flat Universe the critical density energy is indeed the total density energy.

The latest observational constraints for these cosmological parameters are close to $H_0 \approx 70$ km s^{-1} Mpc^{-1}, $\Omega_{\mathrm{r},0} \approx 0$, $\Omega_{\mathrm{m},0} \approx 0.3$ and $\Omega_{\Lambda,0} \approx 0.7$, with a curvature parameter $k \approx 0$ (e.g. Komatsu et al., 2011; Planck Collaboration et al., 2013). The highest redshifts involved in this thesis are $z \sim 1$; hence, the energy contribution of photons will be ignored, i.e. $\Omega_{\mathrm{r},0}(1+z)^4 = 0$.[7]

In this thesis we focus on the baryonic component of the universe at $z \lesssim 1$, which corresponds to half of the history of the Universe for this adopted cosmological model.

We also measure distances associated with a given projected angle in the sky, $\Delta\theta$, at an observed $z_{\cos}$. From the metric presented in Equation 1.1 and our previous findings, the angular proper distance is given by,

$$d_{\mathrm{A,p}}(t) = \int_0^{\Delta\theta} a(t) r \mathrm{d}\theta = a(t) r \Delta\theta \ . \tag{1.22}$$

where r is the radial co-moving distance given by Equation 1.21. Consequently, the angular co-moving distance is given by,

$$d_{\mathrm{A}}(t) = r\Delta\theta \ . \tag{1.23}$$

The luminosity distance to an object of luminosity L is defined as,

$$d_{\mathrm{L}} \equiv \sqrt{\frac{L}{4\pi F}} \ , \tag{1.24}$$

where F is the flux received by the observer. In GR, the flux of an object of luminosity L will get dimmer not only because of the co-moving distance ($\propto r^{-2}$), but also because of the energy lost due to the expansion of space ($\propto a$; see Equation 1.6) and time dilation ($\propto a$; see Equation 1.7). (Flux is a measure of energy per unit time, per unit area.) Consequently,

$$F = \frac{L}{4\pi}\left(\frac{a}{r}\right)^2 \ . \tag{1.25}$$

[7]Note that at sufficiently large $z_{\cos}$, the contribution of photons will become dominant and should not be neglected.

Replacing into 1.24 and using redshift instead,

$$d_{\mathrm{L}}(z) = (1+z)r \ . \tag{1.26}$$

1.3 Basic rest-frame observables

In this section we briefly describe the main observables for a rest-frame galaxy and intergalactic cloud that are relevant for this thesis.

1.3.1 Galaxy observables

Galaxies have been surveyed using optical spectroscopy. Thus, the spectrum of the galaxy is the main observable. The continuum in a galaxy spectrum is dominated by the emission of the stars, while the emission lines are commonly produced in gas regions around young stars. Therefore, these emission lines are good tracers of star-formation activity: the more intense the line, the larger the current star-formation rate. The most common of these optical emission lines are the hydrogen Hα λ6563 Å and Hβ λ4861 Å (Balmer series), and the oxygen [O II] $\lambda\lambda$ 3726,3729 Å and [O III] $\lambda\lambda$ 4959,5007 Å. Some emission lines can also be produced by active galactic nuclei (AGN) activity (including the previous ones), but these are typically broader and much more highly ionized.

In contrast, absorption features in the optical are typically produced by stellar atmospheres. The most common of these are the D4000 break—produced by a combination of stellar Ca II H λ3969 Å and Ca II K λ3934 Å absorption and the Balmer break λ3646 Å—, higher order hydrogen Balmer series (e.g. Hγ λ4341 Å and Hδ λ4101 Å), the G band of the CH molecule λ4303 Å, and Mg λ5175 Å. The intensity of these absorption features—especially the D4000 break—are good indicators of the star-formation history of the galaxy: the weaker the break, the more recent the star-formation activity. As an example, Figure 1.1 shows the rest-frame spectrum of a typical star-forming (top panel) and non-star-forming (bottom panel) galaxy.

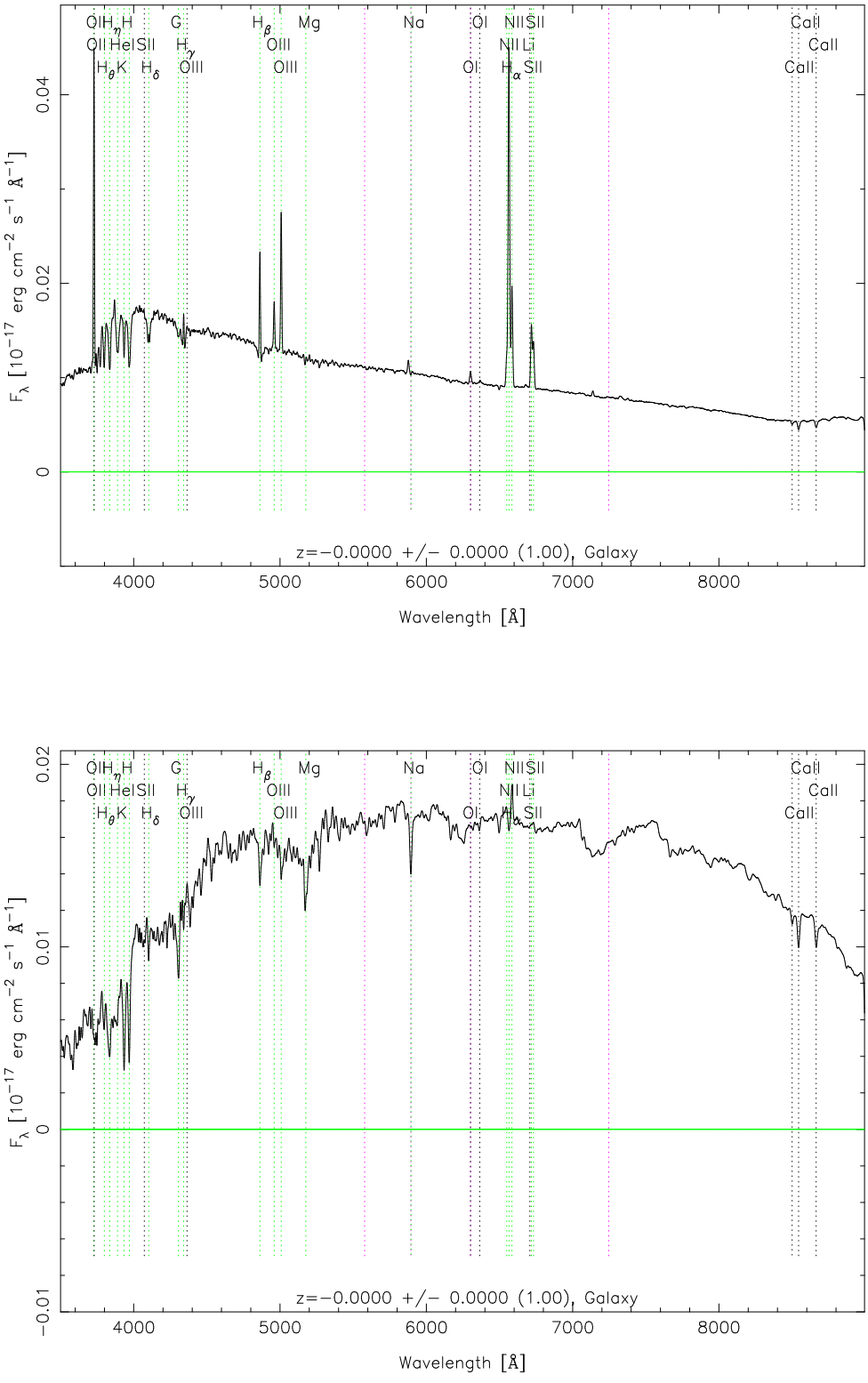

Figure 1.1: Composite rest-frame spectrum of a typical star-forming (top panel) and non-star-forming (bottom panel) galaxy obtained from the Sloan Digital Sky Survey (SDSS; Abazajian et al., 2009). Vertical dashed lines show some of the most important spectral features. Copyright by SDSS.

Dust in the galaxy will modify the spectrum by absorbing and scattering photons. The net effect is a 'reddening', as dust will preferentially absorb and scatter shorter wavelengths. This reddening will consequently affect the estimation of the star-formation history based on emission/absorption features. This effect can be corrected for by assessing the amount of dust or by using observables that are less affected by dust extinction (e.g. ratios of emission lines at similar wavelengths).

1.3.2 Intergalactic medium observables

Because of the extremely low densities of the ionized IGM, its observation is difficult and limited. Currently, the only feasible way to observe such a diffuse gas is through intervening absorption line systems in the spectra of bright background sources. The idea is to use a source with a known and simple spectrum, and look for absorption that is inconsistent with being part of the source itself (e.g. Gunn and Peterson, 1965; Greenstein and Schmidt, 1967; Burbidge et al., 1968). These features are interpreted as being due to absorption of intervening material, whose redshifts can be determined from the observed wavelength of the identified transition (e.g. see Rauch, 1998, for a review). This technique limits the IGM characterization to being one-dimensional, but allows the tracing of extremely weak column densities of neutral hydrogen (e.g. $N_{HI} \gtrsim 10^{13}$ cm^{-2}) in a fairly unbiased way.[8] By combining multiple lines-of-sight and galaxy surveys, an averaged three dimensional picture can still be obtained.

In this thesis, we will survey the IGM using ultra-violet (UV) spectroscopy of a backgroung QSO, by means of the H I Lyα λ1216 Å transition. Depending on the redshift and H I column density of the cloud, we will also observe the H I Lyβ λ1026 Å transition. Contrary to what is observed at high redshifts ($z \gtrsim 2$), the Lyα forest at $z \lesssim 1$ is much more sparse. This fact allows us to characterize individual lines by fitting an appropriate profile. Figure 1.2 shows a schematic of

[8]Note that the selection function is dominated by the background source, and so it is, in principle, independent of the intervening clouds themselves. But also note that intervening galaxies can indeed introduce a bias because of the presence of dust and/or gravitational lensing.

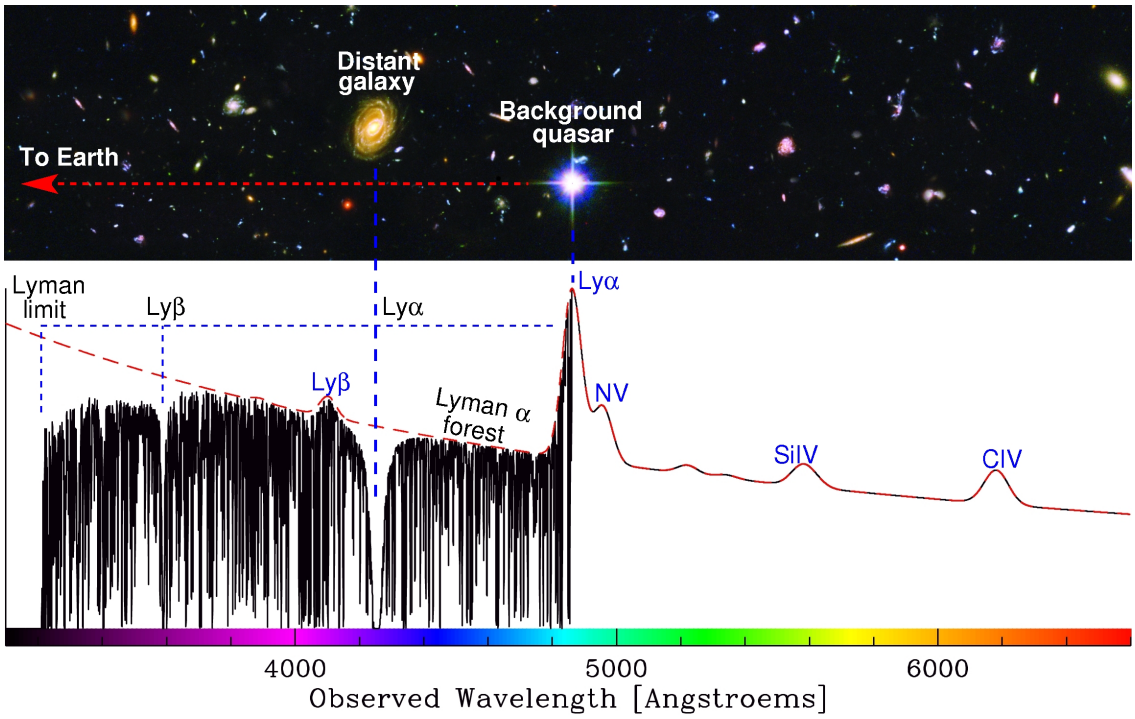

Figure 1.2: Schematic of the QSO absorption line technique. The observed QSO spectrum is represented by the black lines, which is the result of invervening absorption superposed to the original (unabsorbed) QSO continuum and broad emission lines (red lines) shifted by a given redshift. Vertical dashed lines show some of the most important spectral features. Copyright by Michael Murphy.

the QSO absorption line technique.

The observed absorption profile will be the results of many processes. First, the natural profile of the line is given by quantum mechanics. Because of the uncertainty principle, the transition wavelength is not perfectly constrained, and so there is a probability of absorption at a different wavelength. This probability function is well approximated by a Lorentzian profile centred at the transition energy, whose width can be determined in the laboratory. Second, thermal motions of the gas will broaden the intrinsic profile because of the Doppler effect. This thermal broadening is well characterized by a Gaussian profile, whose width is $\propto \sqrt{T/m}$, where T is the temperature and m the mass of the absorbing ion. Third, collisions between particles will also introduce broadening of the line profile. In this case the profile will be Lorentzian, whose width is $\propto n\sigma_{\rm col}\sqrt{T/m}$, where T is the temperature and m the mass of the absorbing ion, n is the density

and $\sigma_{\rm col}$ is the collisional cross-section. Fourth, the presence of turbulence in the gas will introduce an extra broadening, usually *assumed* to be Gaussian, whose width is proportional to the bulk velocity dispersion along the line-of-sight. Fifth, the expansion of the Universe will also cause an apparent broadening, with a profile that can be asymmetric. This effect is only important for extremely underdense and uncollapsed material; at a first approximation, it can be included as a turbulence term (Gaussian). Sixth, the spectrograph from which the line is observed will introduce an instrumental broadening, whose profile is called the line-spread-function (LSF) and is usually well known. The LSF is usually well approximated by Gaussian profile, but this is not always the case. Finally, the observed absorption line profile will be the convolution of all the aforementioned profiles.

Physical processes will lead then to either Gaussian or Lorentzian profiles. The convolution of these two is known as a Voigt profile, and is widely used in astronomy to fit these lines (e.g. van de Hulst and Reesinck, 1947). In this way, the Voigt profile of the line contains the relevant information on the physical conditions of the IGM and will serve as the main observable (after deconvolution with the known LSF).

A useful observable quantity is the 'equivalent width', defined as,

$$W \equiv \int_{\lambda_1}^{\lambda_2} (1 - \frac{F(\lambda)}{F_{\rm c}(\lambda)}) {\rm d}\lambda \, , \tag{1.27}$$

where λ_1 and λ_2 are the wavelength limits of the absorption line, $F(\lambda)$ is the observed flux at a given wavelength, and $F_{\rm c}(\lambda)$ is the flux of the background source in the absence of absorption (i.e. the 'continuum'). In the absence of emission from the absorbing material (which is usually the case), W is equivalent to,

$$W = \int_{\lambda_1}^{\lambda_2} (1 - e^{-\tau(\lambda)}) {\rm d}\lambda \, , \tag{1.28}$$

where $\tau(\lambda)$ is the 'optical depth' of the absorbing cloud for that transition. The

optical depth is $\tau(\lambda) = \sigma_{\rm abs} N(\lambda)$, where $\sigma_{\rm abs}$ is the absorption cross-section and $N(\lambda)$ is the column density per unit wavelength. The total column density is defined as

$$N \equiv \int n(\vec{r})\, \mathrm{d}l \,, \tag{1.29}$$

which corresponds to the integrated volumetric density, $n(\vec{r})$, along the line-of-sight path. The absorption cross-section is $\sigma_{\rm abs} \propto f_{\rm osc}\phi(\lambda)$, where $f_{\rm osc}$ is the so-called 'oscillator strength' and $\phi(\lambda)$ is the line profile described above (Voigt). The oscillator strength is a measure of the probability of having that transition, which can be measured in the laboratory. The relation between W and N, $W(N)$, is known as the 'curve of growth' (COG).

For optically thin transitions ($\tau \ll 1$), we have that $(1 - e^{-\tau}) \approx \tau$ and so $W \propto N$. The profile of $\tau \ll 1$ lines is usually dominated by thermal broadening (Gaussian), and so it is common to parametrize its width as a 'Doppler parameter', $b \equiv \sqrt{2}\sigma_{\rm Gauss}$, where $\sigma_{\rm Gauss}$ is the standard deviation of the Gaussian profile. For purely thermal broadening,

$$b = \sqrt{\frac{2k_{\rm B}T}{m}} \,, \tag{1.30}$$

where $k_{\rm B}$ is the Boltzmann constant and T and m take the previous meaning. In the presence of turbulence,

$$b = \sqrt{\frac{2k_{\rm B}T}{m} + b_{\rm turb}^2} \,, \tag{1.31}$$

where $b_{\rm turb}$ is the turbulence broadening (assumed Gaussian). Collisional broadening can be usually neglected in studies of the IGM, as the densities are very small.

For optically thick lines ($\tau \sim 1$) the Gaussian profile is no longer a good representation. Moreover, $W \propto b\sqrt{\ln(N/b)}$ (e.g. see Draine, 2011), and so there is a degeneracy between N and b for a fixed W, making it difficult to (confidently) recover these physical quantities from W. One can overcome this problem by

looking at another transition of the same ion, for which the line appears optically thin. Due to observational limitations this is not always possible however.

For extremely saturated lines ($\tau \gg 1$), the damping wings of the Lorentzian profile start to take over, breaking the degeneracy between N and b. In this regime $W \propto \sqrt{N}$. We refer the reader to Draine (2011) for further description of the physics of absorption (and emission) lines.

In this thesis we focus on H I absorption line systems in the optically thin regime. As mentioned, these profiles are well described by Gaussians and both column densities and Doppler parameters can be measured confidently.

1.4 Galaxy definition

Even though is widely used in astronomy, the concept of 'galaxy' is not well defined (e.g. see Forbes and Kroupa, 2011, for a recent discussion about this). From a theoretical perspective—under the ΛCDM paradigm—galaxies are often defined as the gravitationally bounded baryonic component, within the potential well of a (cold)[9] dark matter halo, that has formed stars. Despite the undoubtedly success of ΛCDM, it is still worrying to rely on the presence of exotic dark matter for the galaxy definition; as mentioned in previous sections, the real nature of dark matter remains a mystery. Willman and Strader (2012) have recently proposed a more conservative definition as follows: "*a galaxy is a gravitationally bound collection of stars, whose properties cannot be explained by a combination of baryons and Newton's laws of gravity*". Because of observational limitations however, it is not always easy (or even possible) to assess the total mass and exact dynamics of extragalactic objects, making this last proposed definition unpractical. It seems more convenient then to define a galaxy in terms of its *observed* properties. Given that the work presented in this thesis is mainly based on spectroscopy, we define a galaxy as:

- a gravitationally bound system,

[9] Hereafter we will omit the term cold when referring to dark matter.

- whose observed spectrum is consistent with that of a single or multiple stellar populations, and/or with that of diffuse gas in emission, shifted by a fixed velocity,
- with an integrated luminosity greater than that of a globular cluster ($\gtrsim 10^{6-7} L_{\odot}$)

In this way, we are not considering the so called 'dark galaxies'—those having cold gas (e.g. H I 21 cm emission) but no stars (e.g. Doyle et al., 2005)—as galaxies. Similarly, tidal streams are not considered galaxies either, as they often lack significant amount of stars. Our definition will also exclude AGN from being galaxies, as long as its continuum and emission lines do not require an observed stellar component.

The main observables of galaxies that are relevant for this thesis are:

- redshift,
- position in the sky,
- spectral type, and
- luminosity.

The redshift and position in the sky will be used to constrain their co-moving coordinates, while spectral type and luminosity will be used to constrain their star-formation history. We refer the reader to Mo et al. (2010) and Draine (2011) for extensive and comprehensive description of current interpretation of the physics of galaxies.

1.5 Intergalactic medium definition

The intergalactic medium (IGM) is loosely defined as the baryonic material that is not part of a galaxy. Consequently, the IGM definition is intimately related to that of the galaxy. A relevant quantity to assess then is where a galaxy ends or

in other words, how the 'edge' of a galaxy is defined. This is a tricky question because there might not be sharp transitions in terms of the gas or matter density of galaxies.

According to theoretical models, the density profiles of dark matter are smooth, with no sharp edges at large distances. The most commonly adopted density profiles are $\propto r^{-2}$ (isothermal), or $\propto r^{-3}$ (e.g. Navarro et al., 1996, 1997), where r is the distance to the center of the halo. Baryons might presumably follow the same underlying dark matter smooth distribution, but because of observational limitations, this hypothesis has been difficult to test. Even though the surface brightness of stars in galaxies seem to provide much steeper profiles, e.g. $\propto e^{-kr^{1/n}}$, with k and n constants (e.g. Sérsic, 1963), a considerably fraction of baryons in galaxies are in diffuse gas that might not follow the distribution of stars (stars are decoupled from the bulk of gas).

At present, observations of the H I 21 cm transition provide a viable way to map gas in emission, but this is limited to the neutral component only, and at relatively large column densities ($N_{\rm HI} \gtrsim 10^{18}$ cm^{-2}; e.g. Doyle et al., 2005). Even though these maps do show sharp transitions at larger radii, these apparent edges are likely driven by a sudden change in the ionization state of the gas rather than its density (e.g. Zheng and Miralda-Escudé, 2002; Altay and Theuns, 2013).

The UV background (UVB) is an important ingredient needed to understand the physical condition of the gas in the Universe. In our current picture of galaxy formation, galaxies form from the cooling and condensation of primordial material. These galaxies start to produce photons with energies capable of ionizing neutral hydrogen (> 13.6 eV), via stellar emission and/or AGN activity (e.g. Haardt and Madau, 1996). A fraction of these photons escape from the galaxies and start to ionize the IGM. Observational constraints indicate that the hydrogen component of the IGM is already mostly ionized at $z \sim 6$ (e.g. Fan et al., 2001, 2006). For a sufficiently large H I column density however, the gas clouds can be 'self-shielded' from the UVB, and so their inner parts can remain mostly neutral. The typical H I column densities at which self-shielding is important are

$N_{\rm HI} \gtrsim 10^{18}$ cm^{-2} (e.g. Altay et al., 2011; Altay and Theuns, 2013).

From a theoretical point of view, the galaxy edge is usually defined in terms of the so-called r_{200} or 'virial radius', i.e. the radius of an spherical region centred in a galaxy that has an averaged density $200\times$ the critical mass density of the Universe (e.g. Cole and Lacey, 1996; Navarro et al., 1997). However, such a virial radius does not imply any intrinsic physical difference between what is inside and outside it. As an example, the Milky Way virial radius is about ~ 200 kpc (proper), but there is not evidence for a change in the physical condition of the medium at that particular scale. Again, it seems more practical to aim for a definition of the IGM, based on its *observed* properties instead. Therefore, for the purposes of this thesis we define IGM as:

- a baryonic system that emits or absorbs light,
- at a proper distance larger than 300 kpc from the closest galaxy.

The 300 kpc limit is somewhat arbitrary from the point of view of H I but it is the value usually adopted in the literature. Motivation for such a limit arises from the presence of highly ionized metal line systems that seem to be fairly common around galaxies within such a distance (e.g. C IV, O VI; Tumlinson et al., 2011; Stocke et al., 2013; Prochaska et al., 2011b).

In this way, 'dark galaxies' and tidal streams will be considered as part of the IGM as long as they do not lie close to a galaxy. If they lie close to a galaxy these systems are left undefined (they are neither galaxies nor IGM). In order to account for all the processes that might or might not be directly linked to galaxies but happen close to galaxies, an intermediate definition has been usually adopted in the literature: the circumgalactic medium (CGM). Characterization of the CGM is very important for understanding galaxy evolution for obvious reasons: the CGM is the interface where enriched material from the galaxies is expelled and the new material from the IGM is accreted. Even though in this thesis we do not cover scales around galaxies much smaller than this limit, for the sake of completeness

we (arbitrarily) define the interface between the CGM and galaxies to be at 50 kpc (but see Richter, 2012, for a potential motivation of this value).

Absorption line systems in the spectra of background objects provide a means to observe the physical conditions of both the CGM and the IGM. In particular, H I absorption systems are commonly classified into three main categories, according to their H I column densities: Damped Lyman Alpha Systems (DLAs, $N_{\rm HI} \geq 10^{20.3}$ cm^{-2}; e.g. Wolfe et al. 2005), Lyman Limit Systems (LLSs, $10^{17} \leq N_{\rm HI} < 10^{20.3}$ cm^{-2}; O'Meara et al. 2007) and the Lyα forest ($N_{\rm HI} < 10^{17}$ cm^{-2}; e.g. Rauch 1998). Due to their column densities, DLAs are mostly neutral, whereas the Lyα forest is mostly ionized. LLSs are in the transition of these two regimes. This classification is independent of the presence of a close galaxy and so we can explore the statistical connection between galaxies and absorption line systems as a function of H I column density.

The main observables of such absorption lines that are relevant for this thesis are:

- redshift,
- position in the sky,
- column density, and
- Doppler parameter.

The redshift and position in the sky will be used to constrain their co-moving coordinates, while column density and Doppler parameters will be used to constrain their typical densities and temperature (and turbulence), respectively. For a recent review on the physics of the IGM and its wide applications in astrophysics and cosmology, we refer the reader to Meiksin (2009).

1.6 Motivation and structure of the thesis

The motivation of the present work is simple: understand the relationship between the IGM and galaxies in a statistical manner. To do so, we will use observational

samples of galaxies and H I absorption line systems found in the same volumes at $z \lesssim 1$. We will cover a wide range of scales and environments, especially towards the somewhat unexplored underdense regions. Our results range from scales of ~ 100 kpc (proper) to $\gtrsim 10$ Mpc, covering both the CGM and IGM. We aim to provide meaningful observational results, that can serve as references for current theoretical models of the evolution of baryonic matter in the Universe.

The structure of the thesis is the following:

- In Chapter 2 we explore the statistical connection of H I absorption line systems and the large-scale structure of galaxies, as traced by galaxy voids. This work is based on archival data and represents a new, simple and reliable method for identifying IGM associated to different environments. We present results on the statistical properties of H I absorption systems within and around galaxy voids at $z < 0.1$ and compare them with the prediction from a state-of-the-art hydrodynamical simulation. We finally discuss the implications of our results on the IGM–galaxy connection.

- In Chapter 3 we summarize the basic definitions of the two-point correlation function, which will be used as the main tool in the analysis presented in Chapter 4. We also derive basic analytical properties of the two-point correlation, which will be used to interpret our observational results obtained in Chapter 4.

- In Chapter 4 we address the statistical connection between H I absorption systems by means of the two-point correlation function. This work is based on a combination of archival data and new data taken by the author and collaborators. The final dataset corresponds to the largest sample of H I and galaxies in the same volume, available to date. We present results for the H I–galaxy cross-correlation as well as the galaxy–galaxy and H I–H I auto-correlations, measured independently along and transverse to the line-of-sight. This is the first time that these three quantities have been measured from the same dataset, and the first time that the H I–H I auto-correlation

has been measured transverse to the line-of-sight at $z \lesssim 1$. We have also introduced a new quantity based on the Cauchy-Schwarz inequality, which assesses the degree of linear dependence between the samples and hence helps with the interpretation of the results. We discuss our results an provide a simplistic model that can account for them.

- In Chapter 5 we finally summarize our main results and conclusions, and briefly review broad future projects that are natural continuations of the present work.

Chapter 2

IGM within and around galaxy voids

2.1 Overview

We investigate the properties of the H I Lyα absorption systems (Lyα forest) *within* and *around* galaxy voids at $z \lesssim 0.1$. We find a significant excess (> 99 per cent confidence level, c.l.) of Lyα systems at the edges of galaxy voids with respect to a random distribution, on ~ 5 Mpc scales. We find no significant difference in the number of systems inside voids with respect to the random expectation. We report differences between both column density ($N_{\rm HI}$) and Doppler parameter ($b_{\rm HI}$) distributions of Lyα systems found inside and at the edge of galaxy voids at the $\gtrsim 98$ and $\gtrsim 90$ per cent c.l. respectively. Low density environments (voids) have smaller values for both $N_{\rm HI}$ and $b_{\rm HI}$ than higher density ones (edges of voids). These trends are theoretically expected and also found in GIMIC, a state-of-the-art hydrodynamical simulation. Our findings are consistent with a scenario of *at least* three types of Lyα systems: (1) containing embedded galaxies and so directly correlated with galaxies (referred as 'halo-like'), (2) correlated with galaxies only because they lie in the same over-dense LSS, and (3) associated with under-dense LSS with a very low auto-correlation amplitude ($\approx$ random) that are not correlated with luminous galaxies. We argue the latter arise in structures still growing linearly from the primordial density fluctuations inside galaxy voids that have not formed galaxies because of their low densities. We estimate that these under-dense LSS absorbers account for $25 - 30 \pm 6$ per cent of the current Lyα population ($N_{\rm HI} \gtrsim 10^{12.5}$ cm^{-2}) while the other two types account for the

remaining $70-75\pm12$ per cent. Assuming that *only* $N_{\rm HI} \geq 10^{14}$ cm^{-2} systems have embedded galaxies nearby, we have estimated the contribution of the 'halo-like' Lyα population to be $\approx 12-15\pm4$ per cent and consequently $\approx 55-60\pm13$ per cent of the Lyα systems to be associated with the over-dense LSS.

2.2 Introduction

The inter-galactic medium (IGM) hosts the main reservoirs of baryons at all epochs (see Prochaska and Tumlinson 2009 for a review). This is supported by both observations (e.g. Fukugita et al., 1998; Fukugita and Peebles, 2004; Shull et al., 2012) and simulations (e.g. Cen and Ostriker, 1999; Theuns et al., 1999; Davé et al., 2010). Efficient feedback mechanisms that expel material from galaxies to the IGM are required to explain the statistical properties of the observed galaxies (e.g. Baugh et al., 2005; Bower et al., 2006; Schaye et al., 2010). Given that galaxies are formed by accreting gas from the IGM, a continuous interplay between the IGM and galaxies is then in place. Consequently, understanding the relationship between the IGM and galaxies is key to understanding galaxy formation and evolution. This has been recognized since the earliest *Hubble Space Telescope* (HST) spectroscopy of QSOs, where the association between low-z IGM absorption systems and galaxies was investigated for the first time (e.g. Spinrad et al., 1993; Morris et al., 1993; Morris and van den Bergh, 1994; Stocke et al., 1995; Lanzetta et al., 1995).

The large scale environment in which matter resides is also important, as it is predicted (e.g. Borgani et al., 2002; Padilla et al., 2009) and observed (e.g. Lewis et al., 2002; Lopez et al., 2008; Padilla et al., 2010) to have non negligible effects on the gas and galaxy properties. Given that baryonic matter is expected to fall into the considerably deeper gravitational potentials of dark matter, the IGM gas and galaxies are expected to be predominantly found at such locations forming the so called 'cosmic web' (Bond et al., 1996). Identification of large scale structures (LSS) like galaxy clusters, filaments or voids and their influence over the IGM and

galaxies is then fundamental to a complete picture of the IGM/galaxy connection and its evolution over cosmic time.

With the advent of big galaxy surveys such as the *2dF* (Colless et al., 2001) or the *Sloan Digital Sky Survey* (SDSS, Abazajian et al., 2009) it has been possible to directly observe the nature and extent of the distribution of stellar matter in the local universe. Galaxies tend to lie in the filamentary structure which simulations predict, however, very little is known about the *actual* gas distribution at low-z. In this work we focus on the study of H I Lyα (hereafter referred simply as Lyα) absorption systems found *within* and *around* galaxy voids at $z \lesssim 0.1$.

Galaxy voids are the best candidates to start our statistical study of LSS in absorption. Voids account for up to $60 - 80\%$ of the volume of the universe at $z = 0$ (e.g. Aragón-Calvo et al., 2010; Pan et al., 2012). Some studies have suggested that when a minimum density threshold is reached, voids grow in a spherically symmetric way (e.g. Regos and Geller, 1991; van de Weygaert and van Kampen, 1993). This suggests that voids have a relatively simple geometry, which makes them comparatively easy to define and identify from current galaxy surveys (although see Colberg et al., 2008, for a discussion on different void finder algorithms). Galaxy voids are a unique environment in which to look for evidence of early (or even primordial) enrichment of the IGM (e.g. Stocke et al., 2007). It is interesting that galaxy voids are present even in the distribution of low mass galaxies (e.g. Peebles, 2001; Tikhonov and Klypin, 2009) and so there must be mechanisms that prevent galaxies from forming in such low density environments.

Previous studies of Lyα absorption systems associated with voids at low-z have relied on a 'nearest galaxy distance' (NGD) definition. (e.g. Penton et al., 2002; Stocke et al., 2007; Wakker and Savage, 2009). In order to have a clean definition of void absorbers the NGD must be large, leading to small samples. For instance, Penton et al. (2002) found only 8 void absorbers (from a total of 46 systems) defined as being located at $> 3\ h_{70}^{-1}$ Mpc from the nearest $\geq L^*$ galaxy.

Wakker and Savage (2009) found 17 void absorbers (from a total of 102) based on the same definition. Stocke et al. (2007) had to relax the previous limit to > 1.4 h_{70}^{-1} Mpc in order to find 61 void absorbers (from a total of 651 systems), although only 12 were used in their study on void metallicities. Note that a low NGD limit (of 1.4 h_{70}^{-1} Mpc) could introduce some contamination of not-void absorbers. This is because filaments in the 'cosmic web' are expected to be a couple of Mpc in radius (González and Padilla, 2010; Aragón-Calvo et al., 2010; Bond et al., 2010). Considering the Local Group as an example, being 1.4 h_{70}^{-1} Mpc away from either the Milky Way or Andromeda cannot be considered as being in a galaxy void. On the other hand, given that there is a population of galaxies inside voids (e.g. Rojas et al., 2005; Park et al., 2007; Kreckel et al., 2011), the NGD definition could also miss some 'real' void absorbers relatively close to bright isolated galaxies. In fact, Wakker and Savage (2009) found that there may be no void absorbers in their sample (based on the NGD definition) if the luminosity limit to the closest galaxy is reduced to $0.1L^*$. Note however that their sample is very local ($z \leq 0.017$ or $\lesssim 70$ h_{70}^{-1} Mpc away), and it might be biased because of the local overdensity to which our Local Group belongs.

In this work we use a different approach to define void absorption systems. We based our definition on current galaxy void catalogues (typical radius of > 14 h_{70}^{-1} Mpc), defining void absorbers as those located inside such galaxy voids. This leads to larger samples of well identified void absorbers compared to previous studies. Moreover, this approach allows us to define a sample of absorbers located at the very edges of voids, that can be associated with walls, filaments and nodes, allowing us to get some insights in the distribution of gas in the 'cosmic web' itself. This definition is different from the NGD based ones and it focuses on the 'large scale' ($\gtrsim 10$ Mpc) relationship between Lyα forest systems and galaxies. The results from this work will offer a good complement to previous studies based on 'local' scales ($\lesssim 2$ Mpc).

This Chapter is structured as follows. The catalogues of both Lyα systems and galaxy voids that we used in this work are described in §2.3. Definition of our

LSS in absorption samples and the observational results are presented in §2.4. We compare our observational results with a recent cosmological hydrodynamical simulation in §2.5. We discuss our findings in §2.6 and summarize them in §4.10. All distances are in co-moving coordinates assuming $H_0 = 100\,h$ km s^{-1} Mpc^{-1}, $h = 0.71$, $\Omega_{\rm m} = 0.27$, $\Omega_\Lambda = 0.73$, $k = 0$ cosmology unless otherwise stated. This cosmology was chosen to match the one adopted by Pan et al. (2012) (D. Pan, private communication; see §2.3.2).

2.3 Data

2.3.1 IGM in absorption

We use QSO absorption line data from the Danforth and Shull (2008, hereafter DS08) catalogue, which is one of the largest high-resolution ($R \equiv \frac{\lambda}{\Delta\lambda} \approx 30\,000 - 46\,000$), low-$z$ IGM sample to date[1]. Briefly, the catalogue lists 651 Lyα absorption systems at $z_{\rm abs} \leq 0.4$, with associated metal lines (O VI, N V, C IV, C III, Si IV, Si III and Fe III; when the spectral coverage and signal-to-noise allowed their observation), taken from 28 AGN observed with both the *Space Telescope Imaging Spectrograph* (STIS, Woodgate et al., 1998) on the HST, and the *Far Ultraviolet Spectroscopic Explorer* (FUSE, Moos et al., 2000). The systems are characterized by their rest-frame equivalent widths (W_r), or upper limits on W_r, for each individual transition. Column densities ($N_{\rm HI}$) and Doppler parameters ($b_{\rm HI}$) were inferred using the apparent optical depth method (AODM, Savage and Sembach, 1991) and/or Voigt profile line fitting. In particular for the Lyα transition, a curve-of-growth (COG) solution was used when other Lyman series lines were available (see also §2.4.4). We refer the reader to DS08 (and references therein) for further description and discussion.

In order to identify absorbing gas associated with LSS (drawn from the SDSS

[1]We note that after this Chapter was accepted as a paper in 2012, a new pre-print by Tilton et al. (2012) appeared with an updated version of the DS08 catalogue. We checked that our results were not considerably affected using this new catalogue instead.

Table 2.1: IGM sightlines from DS08 that intersect the SDSS survey.

Sight Line	RA (J2000)	Dec (J2000)	$z_{\rm AGN}$	S/N^a
PG 0953+414	09 56 52.4	+41 15 22	0.23410	14
Ton 28	10 04 02.5	+28 55 35	0.32970	9
PG 1116+215	11 19 08.6	+21 19 18	0.17650	18
PG 1211+143	12 14 17.7	+14 03 13	0.08090	30
PG 1216+069	12 19 20.9	+06 38 38	0.33130	3
3C 273	12 29 06.7	+02 03 09	0.15834	35
Q 1230+0115	12 30 50.0	+01 15 23	0.11700	12
PG 1259+593	13 01 12.9	+59 02 07	0.47780	12
NGC 5548	14 17 59.5	+25 08 12	0.01718	13
Mrk 1383	14 29 06.6	+01 17 06	0.08647	16
PG 1444+407	14 46 45.9	+40 35 06	0.26730	10

[a] Median HST/STIS signal-to-noise ratio per two-pixel resolution element in the $1215 - 1340$ Å range (C. Danforth, private communication). The expected minimum equivalent width, $W_{\rm min}$, at a c.l. of cl corresponding to a given S/N can be estimated from $W_{\rm min} = \frac{cl\lambda}{R(S/N)}$, where R is the spectral resolution (e.g., see DS08).

DR7), we use a subsample of the DS08 AGN sightlines that intersect the SDSS volume (PG 0953+414, Ton 28, PG 1116+215, PG 1211+143, PG 1216+069, 3C 273, Q 1230+0115, PG 1259+593, NGC 5548, Mrk 1383 and PG 1444+407; see Table 2.1). Despite the fact that PG 1216+069 spectrum has a poor quality, it is still possible to find strong systems in it, and so we decided to do not exclude it from the sample (this inclusion does not affect our results; see §2.4.1)[2]. We use the rest of the sightlines in the DS08 catalogue to derive the general properties of the average absorber for comparison (see §2.4.4).

In our analysis we focus on statistical comparisons of the H I properties in different LSS environments. Metal systems have smaller redshift coverage and lower number densities than Lyα absorbers. Consequently, we do not aim to draw statistical conclusions from them. We intend to pursue metallicity studies in

[2]We note that Chen and Mulchaey (2009) have presented a Lyα absorption system list along PG 1216+069 sightline at a better sensitivity than that of DS08. In order to have an homogeneous sample, we did not include this new data in our analysis however.

future work.

2.3.2 Galaxy voids

We use a recently released galaxy-void catalogue from SDSS DR7 galaxies (Pan et al., 2012, hereafter P12), which is the largest galaxy-void sample to date. Hereafter we will use the term void to mean galaxy-void unless otherwise stated. P12 identified $\gtrsim 1000$ cosmic voids using the `VoidFinder` algorithm described by Hoyle and Vogeley (2002), with redshifts between $0.01 \lesssim z \lesssim 0.102$. To summarize, it first uses a nearest neighbor algorithm on a volume limited galaxy survey. Galaxies whose third nearest neighbor distance is greater than $6.3\ h^{-1}$Mpc are classified as potential void galaxies, whereas the rest are classified as *wall* galaxies. Void regions are identified by looking for maximal empty spheres embedded in the *wall* galaxy sample. These individual void spheres have radii between $10 < R_{\rm void} \lesssim 25\ h^{-1}$Mpc, with mean radius $\langle R_{\rm void} \rangle \approx 13\ h^{-1}$Mpc. The minimum radius of $10\ h^{-1}$Mpc for the void spheres was imposed. Only galaxies with spectroscopic redshifts were used and therefore we expect the uncertainties in the void centres and radii to be small ($\lesssim 1$ Mpc; we will discuss the effects of peculiar velocities in §2.4.1). Independent void regions are defined by combining all the adjacent spheres that share more than 10% of their volume with another. Void galaxies are defined as those galaxies that lie within a void region. We refer the reader to P12 for further description and discussion.

In our analysis, for simplicity, we use the individual spheres as separate voids instead of using the different independent void regions. This choice has the following advantages. First, it allows us to use a perfectly spherical geometry, making it possible to characterize each void by just one number: its radius. Thus, we can straightforwardly scale voids with different sizes for comparison. Secondly, this approach allows us to identify regions at the very edges of the voids. P12 found that the number density of galaxies has a sharp peak at a distance $\approx R_{\rm void}$ from the centre of the void spheres, a clear signature that walls are well defined (at least from the point of view of bright galaxies at low redshifts). This is

also consistent with the predictions of linear gravitation theory (e.g. Icke, 1984; Sheth and van de Weygaert, 2004) and dark matter simulations (e.g. Benson et al., 2003; Colberg et al., 2005; Ceccarelli et al., 2006, P12). Therefore, by looking for absorption systems very close to the edge of voids, we expect to trace a different cosmic environment. Thirdly, using the individual void spheres securely identifies void regions. The void-edge sample on the other hand, could be contaminated by void regions associated with the intersections of two void spheres. We checked that this is not the case though (see §2.4.2). This contamination should only reduce any possible difference between the two samples rather than enhance them. We also note that systematic uncertainties produced by assuming voids to be perfect spheres should also act to reduce any detected difference.

2.4 Data analysis and results

2.4.1 Number density of absorption systems around voids

We have cross-matched the IGM absorption line catalogue from DS08 (see §2.3.1) with the void catalogue from P12 (see §2.3.2). A total of 106 Lyα absorption systems were found in the 11 sightlines that intersect the void sample volume (i.e., those with $0.01 \leq z_{\rm abs} \leq 0.102$).

We first look for a possible difference in the number density of Lyα systems as a function of the distance to voids. We take two approaches. First, we define X as the three-dimensional distance between an absorption system and the closest void centre in $R_{\rm void}$ units, so

$$X \equiv \min_{\rm sample} \frac{s}{R_{\rm void}} \, , \tag{2.1}$$

where s is the co-moving distance between the centre of a given void and the absorber. Thus, $0 \leq X < 1$ corresponds to absorption systems inside voids and $X > 1$ corresponds to absorption systems outside voids. A value of $X \approx 1$ corresponds to absorption systems around void edges as defined by the galaxy

distribution.

The second approach defines D as the three-dimensional distance between an absorption system and the closest void edge in co-moving h^{-1}Mpc, so

$$D \equiv \min_{\text{sample}} (s - R_{\text{void}}) \, [h^{-1}\text{Mpc}]. \tag{2.2}$$

Negative D values correspond to absorption systems inside voids while positive values correspond to absorbers outside voids. Values of $D \approx 0$ h^{-1}Mpc are associated with absorption systems around void edges as defined by the galaxy distribution.

Distances were calculated assuming the absorption systems to have no peculiar velocities with respect to the centre of the voids. Although this assumption might be realistic for gas inside voids, it might not be the case for gas residing in denser environments, where gas outflows from galaxies might dominate. However, some studies have suggested that the bulk of Lyα forest lines have little velocity offset with respect to galaxies (e.g. Theuns et al., 2002; Wilman et al., 2007, and also §4). As an example, a velocity difference of ~ 200 km s^{-1} at redshift $z_{\text{abs}} \lesssim 0.1$ would give an apparent distance shift of the order of ~ 2 h^{-1}Mpc, which is somewhat higher than, but comparable to the systematic error of the void centre determination (given that voids regions are not perfectly spherical as assumed here). Note that the uncertainty in the void centre is smaller than the uncertainty of a single galaxy because the void is defined by an average over many galaxies. As previously mentioned, such an uncertainty should not artificially create a false signal but rather should dilute any real difference.

Although the X and D coordinates are not independent, we decided to show our results using both. This has the advantage of testing the consistency of our results using two slightly differently motivated definitions. X is a scaled coordinate, good for stacking voids of different radius. It is also good for comparisons with some of the P12 results. D gives a direct measure of the actual distances involved, while still using R_{void}. For convenience, results associated with the X definition will be shown normally in the text while results associated with the D definition

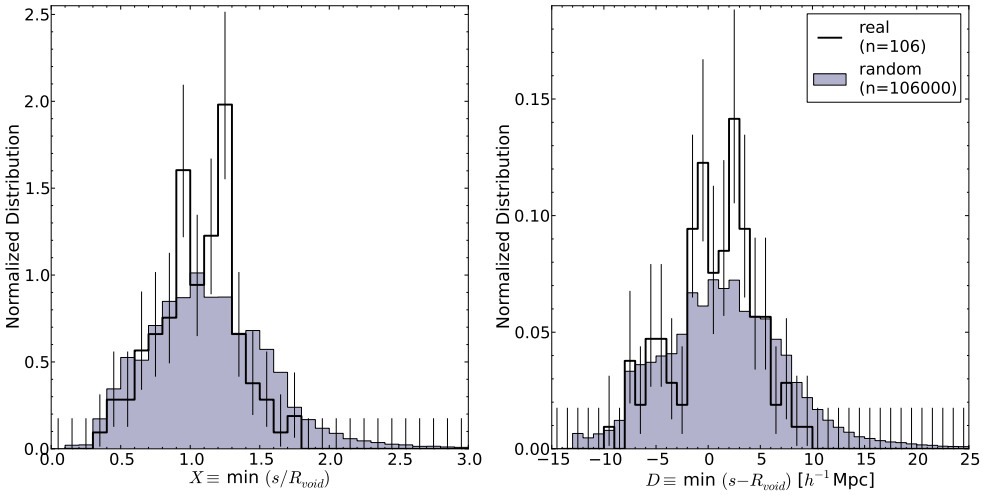

Figure 2.1: Normalized (in area) distribution of H I absorption systems as a function of X (left panel; 0.1 binning) and D (right panel; $1\ h^{-1}$Mpc binning) for both real and random samples. Error bars correspond to the Poissonian uncertainty from the analytical approximation $\sigma_n^+ \approx \sqrt{n + \frac{3}{4}} + 1$ and $\sigma_n^- \approx \sqrt{n - \frac{1}{4}}$ (Gehrels, 1986). Real and random distributions are different at a $\gtrsim 99.5\%$ confidence level (see §2.4.1 for further details).

will be shown in parenthesis: X (D) format.

Figure 2.1 shows the histogram of absorption systems as a function of X and D (left and right panel respectively). In order to show the effects of the geometry of the survey, the random expectations are also shown (shaded distributions). To generate the random samples, we placed 1000 random absorption systems per real one, uniformly between $z_{\rm lim} < z_{\rm abs} \leq 0.102$ for each sightline, where $z_{\rm lim}$ corresponds to the maximum of 0.01 and the minimum observed redshift for a Lyα in that sightline. In this calculation we have masked out spectral regions over a velocity window of ± 200 km s^{-1} around the position where strong Galactic absorption could have been detected (namely: C I, C II, N V, O I, Si II, P III, S I, S II, Fe II and Ni II) *before* the random redshifts are assigned. A total of $106\,000$ random absorbers were generated. We observe a relative excess of absorption systems compared to the random expectation between $X \simeq 0.9$–1.3 and/or $D \simeq -2$–4 h^{-1}Mpc. Assuming Poisson uncertainty, there were 61 ± 8 (65 ± 8)[3] observed,

[3]Results regarding distances from the centre of voids are presented in a X (D) format (see §2.4.1). Reference to this footnote will be omitted hereafter.

while $\approx 38.3 \pm 0.2$ (42.4 ± 0.2) were expected from the random distribution. This corresponds to an $\approx 3\sigma$ excess. Similarly, there is a significant ($\approx 3\sigma$) deficit of absorption systems at $X \gtrsim 1.3$ and/or $D \gtrsim 4$ h^{-1}Mpc, for which 17 ± 4 (19 ± 4) systems were observed compared to the 33.9 ± 0.2 (34.1 ± 0.2) randomly expected. We also checked that such an excess and deficit did not appear by chance in 1000 realizations, consistent with the $< 0.1\%$ probability of occurrence. No significant difference is found for systems at $X \lesssim 0.9$ and/or $D \lesssim -2$ h^{-1}Mpc, for which the 28 ± 5 (22 ± 5) found are consistent with the random expectation of 33.4 ± 0.2 (29.0 ± 0.2). The Kolmogorov-Smirnov (KS) test between the full unbinned samples gives a $\approx 0.3\%$ (0.5%) probability that both the random and the real data come from the same parent distribution. We checked that no single sightline dominates the signal by removing each individual one and repeating the previous calculation. We also checked that masking out the spectral regions associated to possible Galactic absorption does not have an impact on our results as the same numbers (within the errors) are recovered when these regions are not excluded. These results hint at a well defined gas structure around voids, possibly analogous to that seen in galaxies. The current data are not sufficient to confirm (at a high confidence level) the reality of the apparent two-peaked shape seen in the real distributions however.

2.4.2 Definition of large scale structure in absorption

We define three LSS samples observed in absorption:

- *Void absorbers:* those absorption systems with $X < 1$ and/or $D < 0$ h^{-1}Mpc. A total of 45 (45) void absorbers were found.

- *Void-edge absorbers:* those absorption systems with $1 \leq X < 1.3$ and/or $0 \leq D < 4$ h^{-1}Mpc. A total of 44 (42) void-edge absorbers were found.

- *Unclassified absorbers:* those absorption systems with $X \geq 1.3$ and/or $D \geq 4$ h^{-1}Mpc. A total of 17 (19) unclassified absorbers were found.

To be consistent with the galaxy-void definition, we use $X = 1$ and/or $D = 0$ h^{-1}Mpc as the limits between void and void-edge absorbers. The division between void-edge and unclassified absorbers was chosen to match the transition between the overdensity to underdensity of observed absorbers compared to the random expectation at $X > 1$ and/or $D > 0$ h^{-1}Mpc (see Figure 2.1).

We have assumed here that the centre of galaxy voids will roughly correspond to the centre of gas voids, however that does not necessarily imply that gas voids and galaxy voids have the same geometry. In fact, as we do not find a significant underdensity in the number of void absorbers with respect to the random expectation, it is not clear that such voids are actually present within the Lyα forest population. Of course, the fact that we do not detect this underdensity, does not imply that the gas voids are not there. A better way to look at these definitions is by considering void absorbers as those found in galaxy underdensities (galaxy voids) and void-edge absorbers as those found in regions with a typical density of galaxies. We do not have a clear picture of what the unclassified absorbers correspond to. Unclassified absorbers are those lying at the largest distances from the *catalogued* voids, but this does not necessarily imply that they are associated with the highest density environments only. In fact, there could be high density regions also located close to void-edges, at the intersection of the cosmic web filaments. Given that voids of radius $\lesssim 10$ h^{-1}Mpc are not present in the current catalogue it is also likely that some of the unclassified absorbers are associated with low density environments. Therefore, one interpretation of unclassified-absorbers could be as being a mixture of all kind of environments, including voids, void-edges and high density regions.

We checked the robustness of these definitions by looking at the number of voids and void-edges which can be associated with a given absorber. In other words, for a given absorption system, we counted how many voids or void-edges could have been associated with it by taking simply $X \equiv s/R_{\rm void}$ or $D \equiv s - R_{\rm void}$ (in contrast to having taken the minimum values). Out of the 45 void absorbers, 41 are associated with only one void and 4 are associated with 2 voids, independently

Table 2.2: General properties of our LSS samples.[a]

Sample	$\log(N_{HI}[\mathrm{cm}^{-2}])$		b_{HI} [$\mathrm{km\,s^{-1}}$]	
	mean	median	mean	median
Void	13.21 ± 0.67 (13.21 ± 0.67)	13.05 (13.05)	28 ± 15 (28 ± 15)	25 (25)
Edge	13.50 ± 0.70 (13.52 ± 0.69)	13.38 (13.38)	33 ± 17 (34 ± 17)	28 (28)
Unclassified	13.20 ± 0.45 (13.17 ± 0.48)	13.36 (13.36)	33 ± 11 (31 ± 11)	32 (31)

[a] Results are presented in a X (D) format (see §2.4.1).

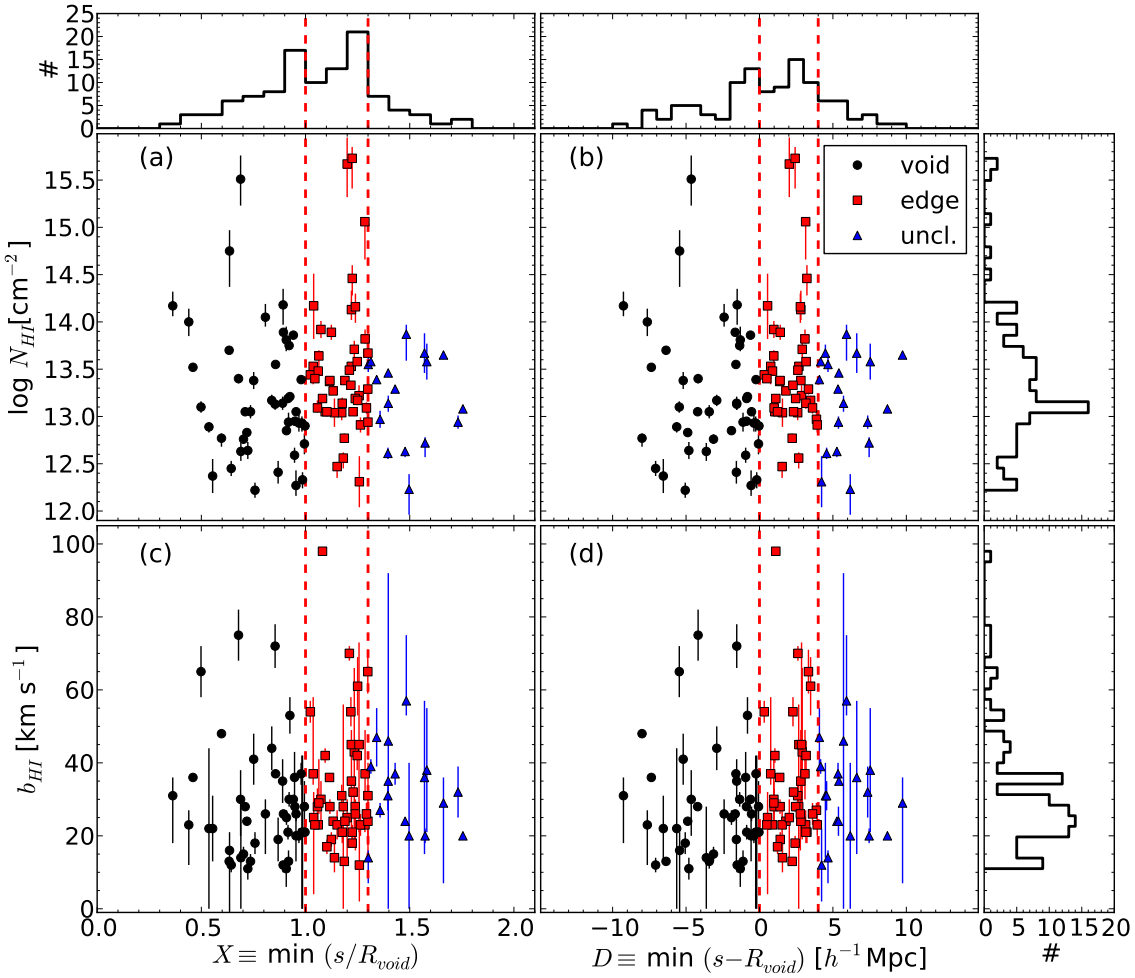

Figure 2.2: Panels (a) and (b) show the distribution of column densities of H I as a function of X and D respectively. Panels (c) and (d) show the distribution of Doppler parameters as a function of X and D respectively. Our LSS samples are shown by different color/symbols: void (black circles), void-edge (red squares) and unclassified (blue triangles). Histograms are also shown around the main panels. Vertical red dashed lines show the limits of our LSS definitions (see §2.4.2).

of the definition used (either X or D). Likewise, out of the 44 (42) void-edge absorbers, 31 (28) are associated with just one void-edge, 12 (13) are associated with two void-edges and 1 (1) is associated with three void-edges. This last system is located at $X = 1.04$ ($D = 0.55\ h^{-1}$Mpc) and has $N_{\rm HI} = 10^{14.17\pm0.35}$ cm^{-2} and $b_{\rm HI} = 25^{+21}_{-7}$ km s^{-1} at a redshift of $z_{\rm abs} = 0.01533$. From these values the system does not seem to be particularly peculiar. Finding an association with more than two void-edges is not surprising as long as the filling factor of voids is not small[4]. Void absorbers have on average 1.1 ± 0.3 voids associated with them, with a median of 1. Void-edge absorbers have in average 1.3 ± 0.5 (1.4 ± 0.5) void-edges associated with them, with a median of $1(1)$. These values give a median one-to-one association. Therefore, we conclude then that the LSS definitions used here are robust.

2.4.3 Properties of absorption systems in different large scale structure regions

Figure 2.2 shows the distribution of column densities and Doppler parameters as a function of both X and D. At first sight, no correlation is seen between $N_{\rm HI}$ or $b_{\rm HI}$ and distance to the centre of voids. Table 2.2 gives the mean and median values of $\log(N_{HI}[{\rm cm}^{-2}])$ and $b_{\rm HI}$ for our void, void-edge and unclassified absorption systems. These results show consistency within 1σ between the three LSS samples.

A closer look at the problem can be taken by investigating the possible differences in the full $N_{\rm HI}$ and $b_{\rm HI}$ distributions of void, void-edge and unclassified absorbers.

Column density distributions

Figure 2.3 shows the distribution of column density for the three different LSS defined above (see §2.4.2). The top panels show the normalized fraction of systems as a function of $N_{\rm HI}$ (arbitrary binning), whilst the bottom panels show

[4]For reference, voids found by P12 have a filling factor of 62%.

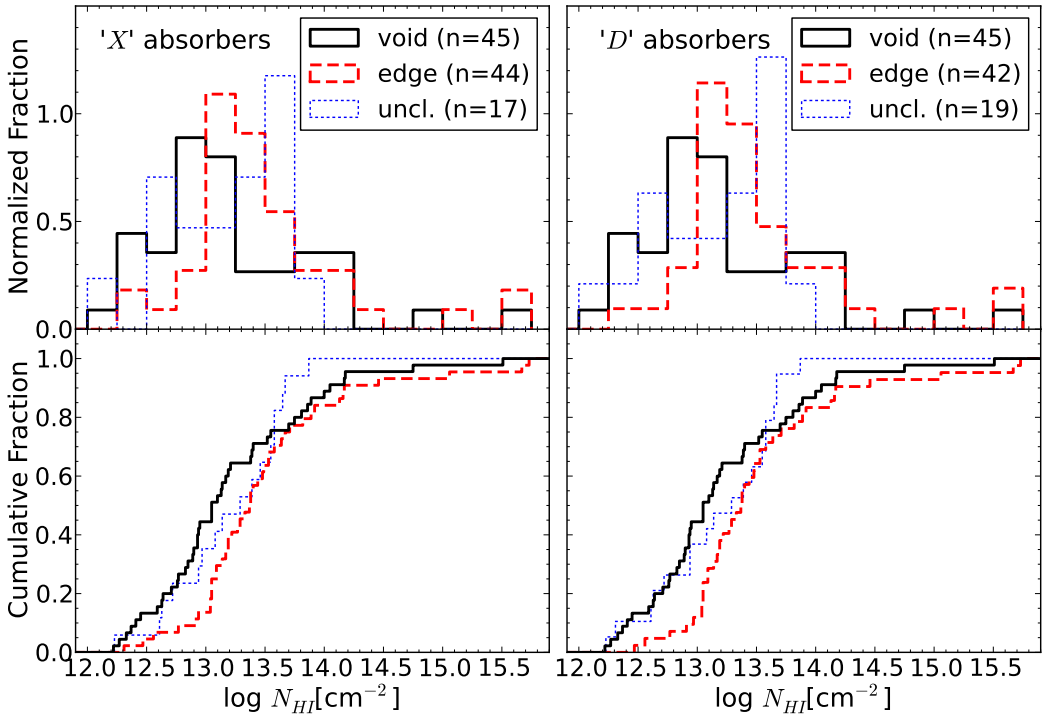

Figure 2.3: H I column density distribution for the three different LSS defined in this work (see §2.4.2): void absorbers (solid-black lines), void-edge absorbers (red-dashed lines) and unclassified absorbers (blue-dotted lines). Top panels show the normalized distribution using arbitrary binning of 0.5 dex. Bottom panels show the cumulative distributions for the unbinned samples. Left and right panels correspond to absorbers defined using X and D coordinates respectively.

the cumulative distributions (unbinned). We see from the top panels that this distribution seems to peak systematically at higher $N_{\rm HI}$ from void to void-edge and from void-edge to unclassified absorbers. We also observe a suggestion of a relative excess of weak systems ($N_{\rm HI} \lesssim 10^{13}$ cm^{-2}) in voids compared to those found in void-edges. This can also be seen directly in Figure 2.2 (see panels (a) and (b)). The KS test gives a probability $P^{\log N}_{\rm void/edge} \approx 2\%$ (0.7%) that void and void-edge absorbers come from the same parent distribution. This implies a $> 2\sigma$ difference between these samples. No significant difference is found between voids or void-edges with unclassified absorbers, for which the KS test gives probabilities of $P^{\log N}_{\rm void/uncl.} \approx 74\%$ (66%) and $P^{\log N}_{\rm edge/uncl.} \approx 56\%$ (24%) respectively. These results can be understood by looking at the bottom panels of Figure 2.3, as we see that the

maximum difference between the void and void-edge absorbers distributions is at $N_{\rm HI} \lesssim 10^{13.8}$ cm^{-2}. On the other hand, no big differences are observed at $N_{\rm HI} \gtrsim 10^{13.8}$ cm^{-2}. In fact, by considering just the systems at $N_{\rm HI} < 10^{13.8}$ cm^{-2}, the significance of the difference between void and void-edge absorbers is increased, with $P^{\log N}_{\rm void/edge} \approx 0.9\%$ (0.2%). Likewise, at $N_{\rm HI} \geq 10^{13.8}$ cm^{-2}, void and void-edge absorber distributions agree at the $\approx 86\%$ (86%) confidence level. We note however that there were ≤ 10 systems per sample for this last comparison and therefore, it is likely to be strongly affected by low number statistics.

We also investigated possible differences between void, void-edge and unclassified absorbers and their complements (i.e., all the systems that were not classified as these: not-void, not-void-edge, not-unclassified). Not-voids correspond to the combination of void-edge and unclassified absorbers and so on. The KS gives probabilities of $P^{\log N}_{\rm void/not-void} \approx 4\%$ (4%), $P^{\log N}_{\rm edge/not-edge} \approx 3\%$ (0.6%) implying that void and void-edge absorbers are somewhat inconsistent with their complements. On the other hand, the distribution of unclassified absorbers is consistent with the distribution of their complements with a KS probability of $P^{\log N}_{\rm uncl./not-uncl.} \approx 64\%$ (54%). These results are summarized in Table 2.3.

Doppler parameter distributions

Figure 2.4 shows the distribution of Doppler parameter for the three different LSS defined above (see §2.4.2). The top panels show the normalized fraction of systems as a function of $b_{\rm HI}$ (arbitrary binning), whilst bottom panels show the cumulative distributions (unbinned). This figure suggests a relative excess of low-$b_{\rm HI}$ systems ($b_{\rm HI} \lesssim 20$ km s^{-1}) in voids compared to those from void-edge and unclassified samples. A relative excess of unclassified absorbers compared to that of voids or void-edges at high-$b_{\rm HI}$ values ($b_{\rm HI} \gtrsim 35$ km s^{-1}) is also suggested by the figure. The KS test gives a probability $P^{b}_{\rm void/edge} \approx 8\%$ (6%) that void and void-edge absorbers come from the same parent distribution. This implies no detected difference between void and void-edge absorbers. Likewise, no significant difference is found between voids or void-edges with unclassified

Table 2.3: Kolmogorov-Smirnov (KS) test probabilities between different samples.[a]

	void/edge	void/uncl.	edge/uncl.	void/not-void	edge/not-edge	uncl./not-uncl.
KS-Prob($\log N_{\rm HI}$)	2% (0.7%)	74% (66%)	56% (24%)	4% (4%)	3% (0.6%)	64% (54%)
KS-Prob($b_{\rm HI}$)	8% (6%)	18% (17%)	71% (75%)	7% (7%)	20% (14%)	32% (32%)

[a] Results are presented in a X (D) format (see §2.4.1).

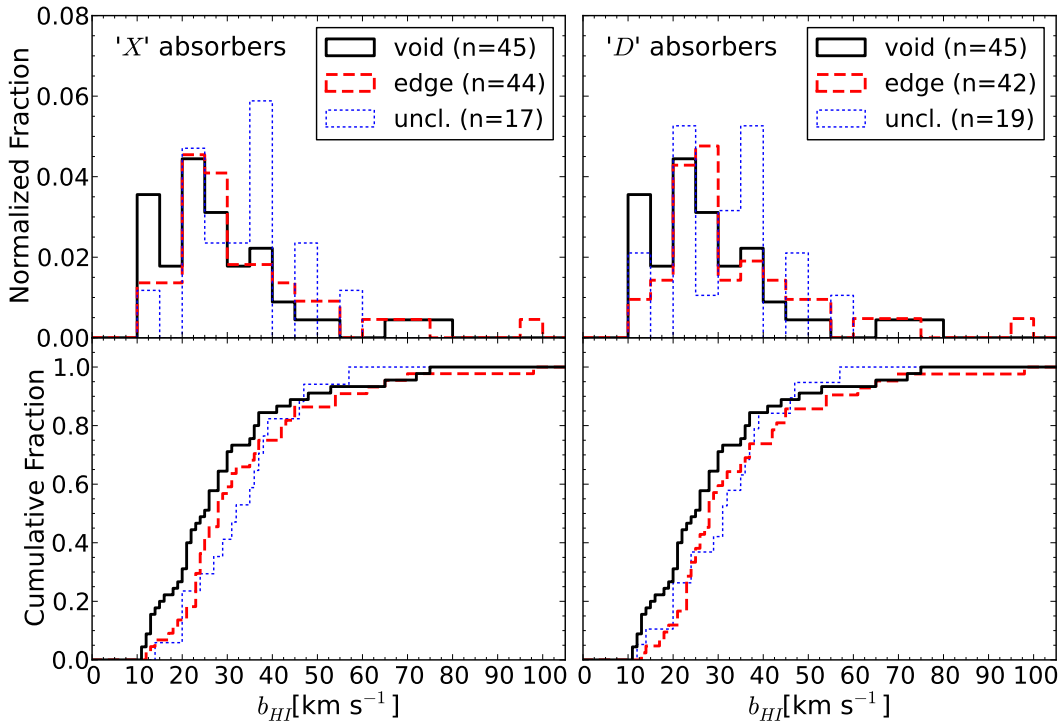

Figure 2.4: H I Doppler parameter distribution for the three different LSS defined in this work (see §2.4.2): void absorbers (black-solid lines), void-edge absorbers (red-dashed lines) and unclassified absorbers (blue-dotted lines). Top panels show the normalized distribution using arbitrary binning of 5 km s^{-1}. Bottom panels show the cumulative distributions for the unbinned samples. Left and right panels correspond to absorbers defined using X and D coordinates respectively.

absorbers, for which the KS test gives probabilities of $P^b_{\rm void/uncl.} \approx 18\%$ (17%) and $P^b_{\rm edge/uncl.} \approx 71\%$ (75%) respectively.

As before, we also investigated possible difference between LSS and their complements. In this case, neither void, void-edge or unclassified absorbers are significantly different than their complements with KS probabilities of $P^b_{\rm void/not-void} \approx 7\%$ (7%), $P^b_{\rm edge/not-edge} \approx 20\%$ (14%) and $P^b_{\rm uncl./not-uncl.} \approx 32\%$ (32%). These results are also summarized in Table 2.3.

2.4.4 Check for systematic effects

Comparison between our subsample and the whole DS08 sample

In this section we explore whether the subsample of H I systems used here is statistically different from the rest of the H I population found in the other sightlines of the DS08 catalogue. For this, we compared the column density ($N_{\rm HI}$) and Doppler parameter ($b_{\rm HI}$) distributions between the 106 absorption systems found inside the void catalogue volume ($0.01 \leq z_{\rm abs} \leq 0.102$ in 11 sightlines) with those 545 outside this volume. The KS test gives a probability of $\approx 5\%$ and $\approx 90\%$ that both $N_{\rm HI}$ and $b_{\rm HI}$ distributions, inside and outside the void catalogue volume, are drawn from the same parent distribution respectively. This shows there is no significant difference in the $b_{\rm HI}$ distribution and an $\approx 2\sigma$ difference between the $N_{\rm HI}$ distributions.

One explanation for such a difference could be due to an intrinsic evolution of the Lyα forest between $z_{\rm abs} \lesssim 0.102$ and $0.102 \lesssim z_{\rm abs} < 0.4$. To check this, we performed a KS comparison between the 286 absorbers at $z_{\rm abs} < 0.102$ (regardless of whether they are inside the void catalogue volume or not) with the 365 systems at higher redshifts, to look for a possible difference in both $N_{\rm HI}$ and $b_{\rm HI}$ distributions. The distribution of $b_{\rm HI}$ does not show any difference (KS Prob. $\approx 87\%$). On the other hand, the KS probability for the two $N_{\rm HI}$ distributions is $\approx 8\%$, hinting that such evolution could be present. We note, however, that an observational bias between low and high-z systems (e.g. due to different selection functions) could also explain the observed difference. We did not further explore this matter.

Finally, we repeated the previous comparisons between the 106 systems inside the void catalogue volume with those 155 outside it having $0.01 \leq z_{\rm abs} \leq 0.102$. The KS test gives this time a probability of $\approx 23\%$ and $\approx 81\%$ of both $N_{\rm HI}$ and $b_{\rm HI}$ inside and outside the void catalogue volume are drawn from the same parent distribution, respectively. No significant ($> 2\sigma$) differences are found for these samples. Thus, we conclude that the properties of the systems in the sightlines used in this work are not statistically different from the properties of the systems

in other sightlines when we restrict ourselves to the same redshift range.

We note that for our comparisons we use systems with $N_{\rm HI}$ below the completeness of the column density distribution itself ($N_{\rm HI} \lesssim 10^{13.4}$ cm^{-2}). The classification of absorber environment does not depend on column density but rather corresponds to a geometrical association of absorbers with known galaxy voids. Therefore, this should not affect any of our results. We also checked that none of the AGN used here were observed in particular for having sightlines intersecting a known void region.

Systematics in the DS08 characterization method

A particular source of concern is the fact that DS08 used different methods for obtaining the H I column densities and Doppler parameters in the catalogue. They used curve-of-growth (COG) solutions when other Lyman series lines were available other than Lyα. For the rest of the systems they used either a single Voigt profile fit (preferentially for strong or blended lines) and/or the apparent optical depth method (AODM, Savage and Sembach, 1991, preferentially for weak, asymmetric or noisy lines). We will refer to these as Lyα-only methods. Given that we have been using systems without distinction between these different methods we will explore any possible bias that this could produce.

Figure 2.5 shows H I column densities versus H I Doppler parameters for the 106 systems in our sample, where the different characterization methods used by DS08 are shown in different colors/symbols. It is possible to observe a clear separation between column densities derived from Lyα-only methods and COG solutions. This difference is a direct consequence of the fact that COG solutions can only be obtained for systems showing a high order Lyman series line, whose equivalent widths are smaller than Lyα for the same $N_{\rm HI}$ and $b_{\rm HI}$ values (because they have smaller wavelengths and smaller oscillator strengths). This results in a shift of the completeness level to higher column densities for the COG solutions, as can be seen in Figure 2.5. On the other hand, no significant difference seems to be present for $b_{\rm HI}$ values between different characterization methods, which has

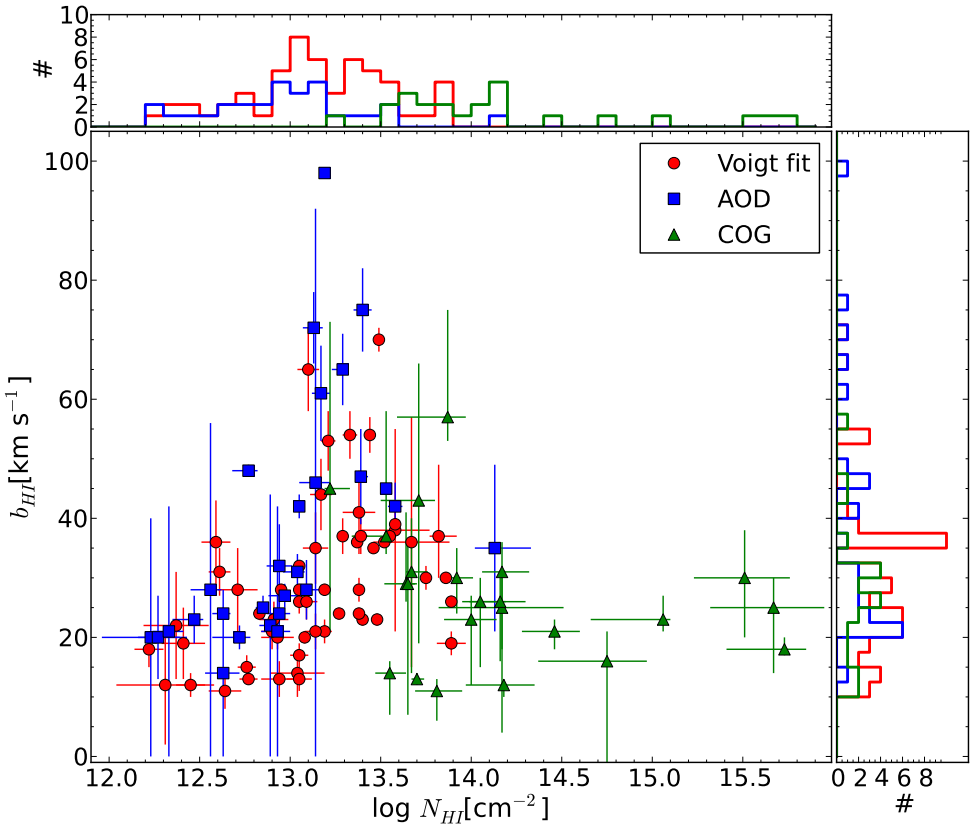

Figure 2.5: H I column density versus H I Doppler parameter for systems in our sample. Different characterization methods (see §2.4.4) are shown by different colors/symbols. Red circles correspond to systems measured by Voigt profile fits. Blue squares correspond to systems measured by the AOD method. Green triangles correspond to systems measured by a COG solution. Histograms (arbitrary binning) are also shown around the main panel using the previous color definition.

been confirmed by a KS test.

Our definitions of the different LSS samples do not depend on either $N_{\rm HI}$ nor $b_{\rm HI}$ but it rather correspond to a geometrical selection of how far away a system lies from the centre of a void. Thus, the aforementioned difference observed in $N_{\rm HI}$ should not affect the conclusions drawn from this study as long as our different LSS samples have similar contributions of systems measured with one method or the other. Assuming a Poisson distribution for the number of systems in each sample, there is a relative contribution of $82\%\pm18\%$ ($82\%\pm18\%$), $73\%\pm17\%$ ($74\%\pm17\%$) and $82\%\pm30\%$ ($79\%\pm27\%$) for systems measured by either Voigt fit or AOD methods in the void, void-edge and unclassified samples respectively. These

numbers are all consistent with each other within 1σ. Therefore we conclude that different characterization methods do not introduce an important bias in our analysis.

There is also a reported systematic effect seen when using Lyα only systems by which $N_{\rm HI}$ estimates are typically (with large scatter) low by a factor of ~ 3 while $b_{\rm HI}$ are high by a factor of ~ 2 (Shull et al., 2000; Danforth et al., 2006). This systematic effect is important for higher column densities. Our sample is dominated by systems with $N_{\rm HI} < 10^{14}$ cm^{-2}, for which the effect is smaller than the quoted numbers (see figures 2 and 3 from Danforth et al., 2006). So, despite the fact that our sample is dominated by Lyα-only measurements, we do not consider this effect to be important. Finally, as previously argued, even if any of these effects are present they will affect each of our LSS samples in roughly the same proportion.

Observables as a function of redshift

We also checked that no bias is present in our samples as a function of redshift. Figure 2.6 shows both $N_{\rm HI}$ and $b_{\rm HI}$ values as a function of redshift (middle and bottom panels). Systems belonging to different LSS samples are shown by different color/symbols. Both distributions look very uniform across the full redshift range. The top panel shows the distribution of void, void-edge and unclassified absorbers as a function of redshift. The KS test shows no significant difference between these LSS samples. We conclude that there is no evident systematic effect as a function of redshift. For simplicity we have only used LSS definitions based on X in Figure 2.6, but the previous results also hold using D instead.

The catalogue of P12 used a nearly complete, magnitude limited sample of galaxies to define the voids. Despite this, we performed an independent check by looking at the mean radius $\langle R_{\rm void} \rangle$ as a function of redshift. If the catalogue is well defined, we should expect to have this radius constant across redshift range (assuming no measurable evolution). We confirmed that this is actually the case by dividing the sample in 6 redshift bins and measuring the mean value. We

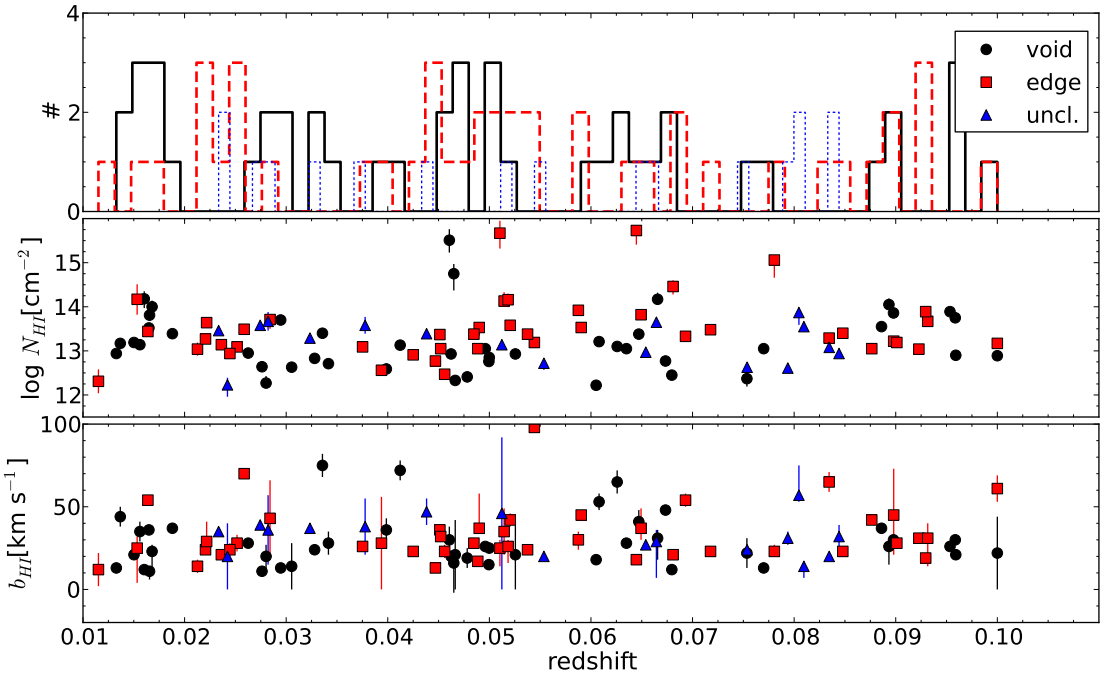

Figure 2.6: Distribution of observables as a function of redshift. Middle and bottom panels show the distribution of $N_{\rm HI}$ and $b_{\rm HI}$ respectively. Different LSS samples are shown by different color/symbols: void (black circles), void-edge (red squares) and unclassified (blue triangles). Top panel shows the distribution of different LSS as a function of redshift: void (black-solid line), void-edge (red-dashed line) and unclassified (blue-dotted line). For simplicity we only show LSS definitions based on X.

found that the mean radius is constant with $\langle R_{\rm void} \rangle \approx 13 \pm 3\ h^{-1}$Mpc in each bin.

Summary

Given that the differences between void and void-edge samples are still at $< 3\sigma$ of confidence level, we have investigated possible biases or systematic effects that could be present in our data analysis. In particular we have investigated (1) possible differences in our subsample with respect to the whole DS08 sample, (2) the effect of the different characterization methods used by DS08 to infer the gas properties, and (3) whether uniformity across our redshift range is present in our observables. From this analysis we concluded that no important biases affect our results.

2.5 Comparison with Simulations

In this section, we investigate whether current cosmological hydrodynamical simulations can reproduce our observational results presented in §2.4. For this comparison we use the *Galaxies-Intergalactic Medium Interaction Calculation* (GIMIC, Crain et al., 2009). Using initial conditions drawn from the *Millennium* simulation (Springel et al., 2005), GIMIC follows the evolution of baryonic gas within five, roughly spherical regions (radius between $18 - 25\ h^{-1}$ Mpc[5]) down to $z = 0$ at a resolution of $m_{\rm gas} \approx 10^7\ h^{-1}\ {\rm M}_\odot$. The regions were chosen to have densities deviating by $(-2, -1, 0, +1, +2)\sigma$ from the cosmic mean at $z = 1.5$, where σ is the rms mass fluctuation. The $+2\sigma$ region was additionally required to be centered on a rich cluster halo. Similarly although not imposed, the -2σ region is approximately centered on a sparse void. The rest of the Millennium simulation volume is re-simulated using only the dark matter particles at much lower resolution to account for the tidal forces. This approach gives GIMIC the advantage of probing a wide range of environments and cosmological features with a comparatively low computational expense.

GIMIC includes (i) a recipe for star formation designed to enforce a local Kennicutt-Schmidt law (Schaye and Dalla Vecchia, 2008); (ii) stellar evolution and the associated delayed release of 11 chemical elements (Wiersma et al., 2009a); (iii) the contribution of metals to the cooling of gas in the presence of an imposed UV background (Wiersma et al., 2009b); and (iv) galactic winds that pollute the IGM with metals and can quench star formation in low-mass halos (Dalla Vecchia and Schaye, 2008). Note that GIMIC does not include feedback processes associated with AGN. For further details about GIMIC we refer the reader to Crain et al. (2009).

[5]Note that GIMIC adopted a $H_0 = 100\ h$ km s^{-1} Mpc^{-1},$h = 0.73$, $\Omega_{\rm m} = 0.25$, $\Omega_\Lambda = 0.75$, $\sigma_8 = 0.9$, $k = 0$ cosmology. These parameters are slightly different than the ones used in P12.

2.5.1 Simulated H I absorbers sample

In order to obtain the properties of the simulated H I absorption systems, we placed 1000 parallel sightlines within a cube of $20\ h^{-1}$Mpc on a side centered in each individual GIMIC region at $z = 0$ (5000 sightlines in total). We have excluded the rest of the volume to avoid any possible edge effects. This roughly corresponds to 2.5 sightlines per square h^{-1}Mpc. Given this density, some sightlines could be tracing the same local LSS and therefore these are not fully independent. We consider this approach to offer a good compromise of having a large enough number of sightlines while not oversampling the limited GIMIC volumes.

We used the program SPECWIZARD[6] to generate synthetic normalized spectra associated to our sightlines using the method described by Theuns et al. (1998b). SPECWIZARD calculates the optical depth as a function of velocity along the line-of-sight, which is then converted to flux transmission as a function of wavelength for a given transition. We only used H I in this calculation. The spectra were convolved with an instrumental spread function (Gaussian) with FWHM of 6.6 km s^{-1} to match the resolution of the STIS/HST spectrograph[7]. In order to mimic the continuum fitting process in real spectra, we set the continuum level of each mock *noiseless* spectrum at the largest flux value after the convolution with the instrumental profile. Given that the lines are sparse at $z = 0$, there were almost always regions with no absorption and this last correction was almost negligible.

We used 3 different signal-to-noise ratios in order to represent our QSO sample. Out of the total of 1000 per GIMIC region, 727 sightlines were modeled with $S/N = 9$ per pixel, 182 with $S/N = 23$ and 91 $S/N = 2$. These numbers keep the proportion between the different S/N values as it is in the observed sample (see last column in Table 2.1)[8].

We fit Voigt profiles to the synthetic spectra automatically using VPFIT[9], fol-

[6]Written by Joop Schaye, Craig M. Booth and Tom Theuns.

[7]Note that the majority of the Lyα used in this work were observed with STIS/HST rather than FUSE.

[8]Note that we have divided the mean S/N per two-pixel resolution element by $\sqrt{2}$ to have an estimation per pixel.

[9]Written by R.F. Carswell and J. K. Webb (see http://www.ast.cam.ac.uk/$\sim$rfc/vpfit.html).

lowing the algorithm described by Crighton et al. (2010). First, an initial guess of several absorption lines is generated in each spectrum to minimize $\chi^2_{\rm reduced}$. If the $\chi^2_{\rm reduced}$ is greater than a given threshold of 1.1, another absorption component is added at the pixel of largest deviation and $\chi^2_{\rm reduced}$ is re-minimized. Absorption components are removed if both $N_{\rm HI} < 10^{14.3}$ cm^{-2} and $b_{\rm HI} < 0.4$ km s^{-1}. This iteration continues until $\chi^2_{\rm reduced} \leq 1.1$. Then, the Voigt fits are stored. We only kept absorption lines where the values of $\log N_{\rm HI}$ and $b_{\rm HI}$ are at least 5 times their uncertainties as quoted by VPFIT.

The fraction of hydrogen in the form of H I within GIMIC is obtained from CLOUDY (Ferland et al., 1998) after assuming an ionization background from Haardt and Madau (1996) that yields a photo-ionization rate $\Gamma = 8.59\ 10^{-14}$ s^{-1}. This ionization background is not well constrained at $z \approx 0$, so we use a post processing correction to account for this uncertainty. In the optically thin regime $\Gamma^{\rm thin} \propto \frac{1}{\tau}$, where τ is the optical depth (Gunn and Peterson, 1965). Then, scaling the optical depth values is equivalent to scaling the ionization background (e.g. Theuns et al., 1998a; Davé et al., 1999). First, we combined the five GIMIC regions using different volume weights namely: $(1/12, 1/6, 1/2, 1/6, 1/12)$ for the $(-2, -1, 0, +1, +2)\sigma$ regions respectively (see appendix 2 in Crain et al., 2009, for a justification of these weights). Then, we searched for a constant value to scale all the original optical depth values such that the mean flux of the combined sample is equal to the observed mean flux of Lyα absorption at low redshift. A second possibility is to scale the optical depth values in order to match the redshift number density of H I lines in some column density range, dN/dz, instead of the mean flux. Ideally by matching one observable the second would be also matched.

Extrapolating the double power-law fit result from Kirkman et al. (2007) to $z = 0$ (see their equation 6), the observed mean flux is $\langle F \rangle = 0.987$ with a typical statistical uncertainty of $\sigma_{\langle F \rangle} \sim 0.003$. In order to match this number in the simulation a scale of 1.16 is required in the original optical depth values (0.86 in Γ). From this correction, the redshift number density of lines in the range

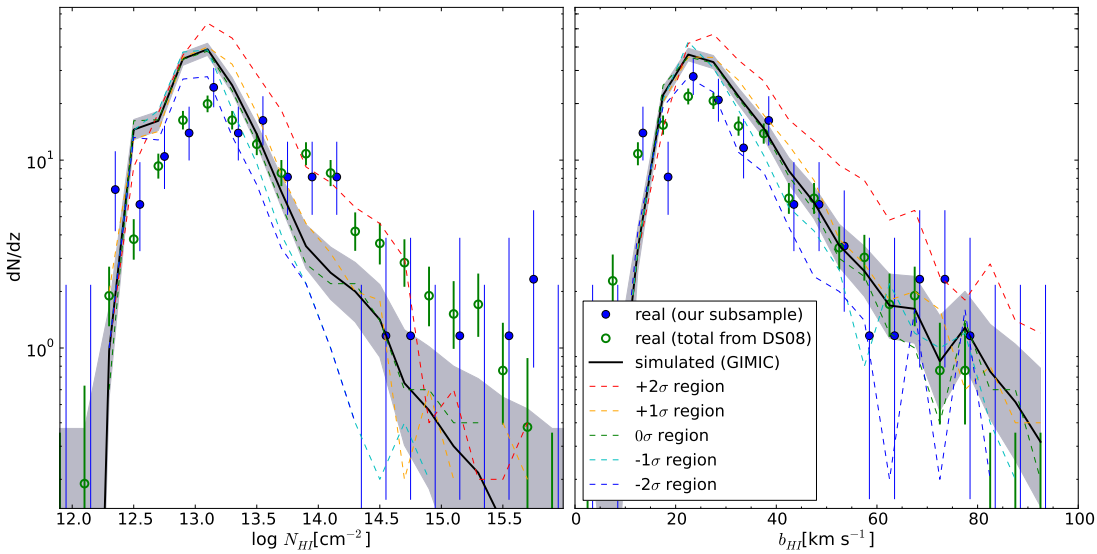

Figure 2.7: Redshift number density of H I lines as a function of both column density (left panel) and Doppler parameter (right panel) using abritrary binning of $\Delta \log N_{\rm HI} = 0.2$ dex and $\Delta b_{\rm HI} = 5$ km s^{-1} respectively. Both results have not been corrected for incompleteness. Green open-circles correspond to real data from the total sample of DS08 (657 systems) while blue filled-circles (slightly offset in the x-axes for clarity) correspond to the subsample used in this study (106 systems). The black line corresponds to the volume-weighted result from the combination of the five GIMIC regions where the shaded region correspond to the $\pm 1\sigma$ uncertainty. Dashed lines show the results from each individual GIMIC region. Error bars correspond to the Poissonian uncertainty from the analytical approximation $\sigma_n^+ \approx \sqrt{n + \frac{3}{4}} + 1$ and $\sigma_n^- \approx \sqrt{n - \frac{1}{4}}$ (Gehrels, 1986).

$10^{13.2} \leq N_{\rm HI} \leq 10^{14}$ cm^{-2} is found to be $dN/dz \approx 50$. For reference, Lehner et al. (2007) and DS08 found $dN/dz \sim 50 - 90$ over the same column density range. We have repeated the experiment for consistency, using $\langle F \rangle = \{0.984, 0.990\}$ which are within $\pm 1\sigma_{\langle F \rangle}$ of the extrapolated value. To match these, scales of 1.50 and 0.84 are required in the original optical depths values respectively (0.67 and 1.19 in Γ). From those mean fluxes we found $dN/dz \approx \{70, 35\}$ respectively along the same column density range. Therefore, a value of $\langle F \rangle = 0.990$ underpredicts the number of H I lines. On the other hand, values of $\langle F \rangle = 0.987$ and 0.984 are in good agreement with observations. In the following analysis we use $\langle F \rangle = 0.987$ unless otherwise stated.

2.5.2 Comparison between simulated and observed H I properties

Figure 2.7 shows the redshift number density of H I lines (not corrected for incompleteness) as a function of both column density (left panel) and Doppler parameter (right panel). Data from the simulation are shown by the black line (volume-weighted result) and each individual GIMIC region is shown separately by the dashed lines. For comparison, data from observations are also shown. Green open-circles correspond to the total sample from DS08 (657 systems) while blue filled-circles correspond to the subsample used in this study (106 systems that intersect the SDSS volume). There is not perfect agreement between simulated and real data. We see an excess (lack) of systems with $N_{\rm HI} \lesssim 10^{13.5}$ cm^{-2} ($N_{\rm HI} \gtrsim 10^{14}$ cm^{-2}) in the simulation compared to observations while Doppler parameters are in closer agreement, although there is still a difference at low $b_{\rm HI}$.

Assuming that the column density distribution can be modeled as a power-law, the position of the turnover at the low $N_{\rm HI}$ end give us an estimation of the completeness level of detection in the sample. As the turnover appears to be around $N_{\rm HI} \approx 10^{13}$ cm^{-2} in both simulated and real data (by design) we do not, in principle, attribute the discrepancy in the column-density distributions to a wrong choice of the simulated S/N. Raising the mean flux to a greater value than $\langle F \rangle = 0.987$ (less absorption) does not help as the dN/dz in the range $10^{13.2} \leq N_{\rm HI} \leq 10^{14}$ cm^{-2} will then be smaller than the observational result (see §2.5.1). We attempted to get a better match by using a mean flux of $\langle F \rangle = 0.984$ (more absorption), motivated to produce a better agreement at higher column densities. In order to agree at low column densities, we had to degrade the sample S/N to be composed of ~ 400, ~ 100 and ~ 500 sightlines at signal-to-noise ratios of 9, 23 and 2 respectively. It is implausible that half of the observed redshift path has such poor quality.

Another possibility to explain the discrepancy could be the fact that weak systems in observations were preferentially characterized with the AOD method,

whereas here we have only used Voigt profile fitting. In order to test this hypothesis, we have merged closely separated systems (within 150 km s^{-1}) whose summed column density is less than $10^{13.5}$ cm^{-2}. Using these constraints, 43 out of 4179 systems were merged ($\approx 1\%$). Such a small fraction does not have an appreciable effect on the discrepancy. As an extreme case, we have repeated the experiment merging *all* systems within 300 km s^{-1} independently of their column densities. From this, 555 out of 4179 systems were merged ($\approx 13\%$) but still it was not enough to fully correct the discrepancy. Given that the discrepancy is not explained by a systematic effect from different line characterization methods, we chose to keep our original simulated sample in the following analysis without merging any systems.

There is a reported systematic effect by which column densities inferred from a single Lyα line are typically (with large scatter) underestimated with respect to the curve-of-growth (COG) solution. Similarly $b_{\rm HI}$ are typically overestimated (Shull et al. 2000; Danforth et al. 2006; see also §2.4.4 for discussion on how this may affect our observational results). This effect is only appreciable for $N_{\rm HI} \gtrsim 10^{14}$ cm^{-2} and is bigger for saturated lines. Given that our simulated sample was constructed to reproduce the observed sample, this effect could be present. If so, it would in principle help to reduce the discrepancy at the high column density end. From figure 3 of Danforth et al. (2006) we have inferred a correction for systems with $N_{\rm HI} \geq 10^{13.5}$ cm^{-2} of,

$$\log N_{\rm HI}^{\rm corr} = \frac{\log N_{\rm HI}^{\rm obs} - 8.37}{1 - 0.62} \,, \tag{2.3}$$

where $N_{\rm HI}^{\rm corr}$ and $N_{\rm HI}^{\rm obs}$ are the corrected and observed $N_{\rm HI}$ values respectively. From this correction we found an increase in the number of systems at $N_{\rm HI} \gtrsim 10^{14.5}$ cm^{-2} up to values consistent with observations. This however does not help with the discrepancy at lower column densities.

At this point, it is difficult to reconcile the simulation result with the real data using only a single effect. We note that the discrepancy is a factor of ~ 2 only, so it could be in principle explained through a combination of several

observational effects. Also note that the number of observed lines at higher column densities is still small and it could be affected by low number statistics. The lack of systems with very low $N_{\rm HI}$ and $b_{\rm HI}$ values can be explained by our selection of the highest signal-to-noise being $S/N = 23$ while in real data there could be regions with higher values. It is not the aim of this section to have a perfect match between simulations and observations but rather examine the qualitative differences between simulated regions of different densities. Thus hereafter, we will use the results from the simulation in its original form (as shown in Figure 2.7), i.e., without any of the aforementioned corrections.

2.5.3 Simulated H I absorbers properties in different LSS regions

Given that GIMIC does not provide enough volume to perform a completely analogous search for voids (each region is $\sim 20\ h^{-1}$Mpc of radius), we use them only as crude guides to compare our results with. We could consider the -2σ region as representative of void regions as it is actually centered in one. Naively, we could consider the 0σ regions as representative of void-edge regions, as it is there where the mean cosmological density is reached. A direct association for the $+1\sigma$ and $+2\sigma$ is not so simple though, as they would be associated to some portions of the void-edge regions too. It seems more reasonable to use the GIMIC spheres as representative of different density environments then, where $-2\sigma/+2\sigma$ correspond to extremely under/over-dense regions and so on. For reference, the whole $(-2, -1, 0, +1, +2)\sigma$ GIMIC regions correspond to densities of $\frac{\rho}{\langle\rho\rangle} \approx (0.4, 0.6, 0.9, 1.2, 1.8)$[10] at $z = 0$ respectively, where $\langle\rho\rangle$ is the mean density of the universe (see figure A1 from Crain et al., 2009).

Figure 2.8 shows the cumulative distributions of $N_{\rm HI}$ (left panel) and $b_{\rm HI}$ (right panel). Results from each of the individual GIMIC region are shown by dashed lines. Void and void-edge absorbers are shown by solid black and red lines respectively. For simplicity we show only LSS definition based on X. Cumulative

[10]Given that we are using cubic sub-volumes centered in these spheres, these cubes should have higher density differences between them.

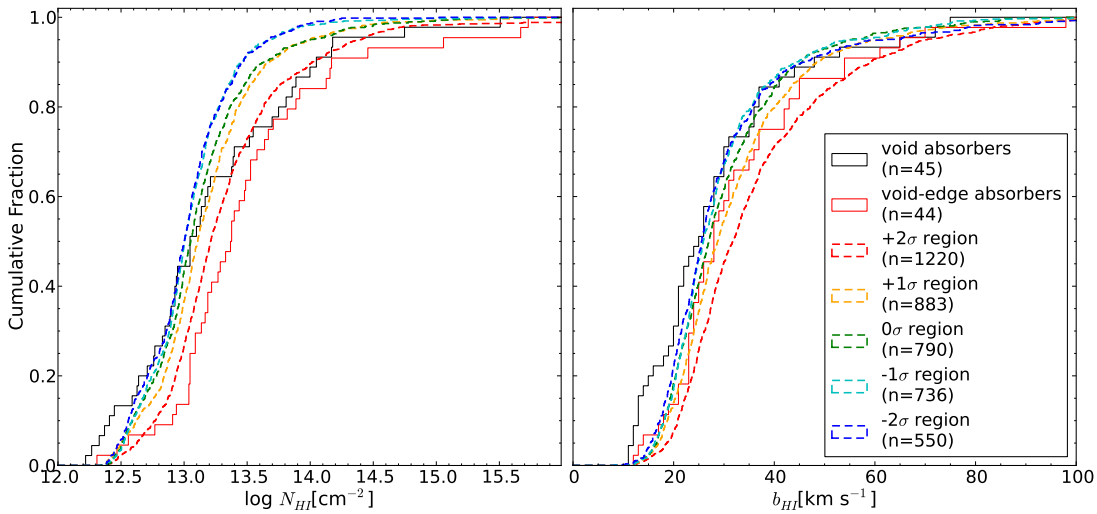

Figure 2.8: Column density (left panel) and Doppler parameter (right panel) cumulative distributions for H I. Void absorbers are shown by solid-black lines while void-edge absorbers are shown by solid-red lines. Dashed lines show the result from each individual GIMIC region. For simplicity we only show LSS definitions based on X.

distributions between real and simulated data do not agree perfectly. However, in both real and simulated data, there is an offset between column densities and Doppler parameters found in different environments. Low density environments have smaller values for both $N_{\rm HI}$ and $b_{\rm HI}$ than higher density ones (and viceversa). This trend still holds when using a $S/N = 23$ per pixel for the 5000 sightlines.

The KS test gives a significant difference between the $+2\sigma$, $+1\sigma$ and $+0\sigma$ regions, and any other GIMIC region at the $\gg 99\%$, $\gtrsim 99\%$ and $\gtrsim 95\%$ confidence level respectively in both $N_{\rm HI}$ and $b_{\rm HI}$ distributions. The KS test gives no significant difference between the -2σ and -1σ regions in both $N_{\rm HI}$ and $b_{\rm HI}$ distributions (see Table 2.4). These results do not change significantly when correcting GIMIC to match the observed $N_{\rm HI}$ distribution using a different $\langle F \rangle$ and S/N values. We do not attempt to make a more detailed comparison between distributions coming from real data (void, void-edge samples) and the different GIMIC regions as there are already known differences between them (see §2.5.2).

Table 2.4: Kolmogorov-Smirnov (KS) test probabilities between different GIMIC regions.[a]

$N_{\rm HI}$ \ $b_{\rm HI}$	-2σ	-1σ	0σ	$+1\sigma$	$+2\sigma$
-2σ	$\cdots$	30%	2%	2%	$\ll 1\%$
-1σ	77%	$\cdots$	6%	$\ll 1\%$	$\ll 1\%$
0σ	11%	0.9%	$\cdots$	0.7%	$\ll 1\%$
$+1\sigma$	$\ll 1\%$	$\ll 1\%$	0.4%	$\cdots$	$\ll 1\%$
$+2\sigma$	$\ll 1\%$	$\ll 1\%$	$\ll 1\%$	$\ll 1\%$	$\cdots$

[a]KS test probabilities for $\log N_{\rm HI}$ and $b_{\rm HI}$ distributions are shown in the left-bottom and right-upper sides respectively. $\ll 1\%$ corresponds to values $< 10^{-4}\%$.

2.6 Discussion

2.6.1 Three Lyα forest populations

Our first result is that there is a $> 99\%$ c.l. excess of Lyα systems at the edges of galaxy voids compared to a random distribution (see §2.4.1 and Figure 2.1). Our random sample was normalized to have the same density of systems in the whole volume. Then, an excess in a sub-volume means necessarily a deficit in another. Given that we found no significant difference in the number of systems in voids with respect to the random expectation, the excess is not explained by a deficit of Lyα systems inside galaxy voids. The observed excess seems more related to the lack of systems found at distances $\gtrsim 5\ h^{-1}$Mpc outside the catalogued voids (and viceversa). Thus, despite the fact that we see Lyα clustered at the edges of galaxy voids, it is not clear from this data that Lyα-voids at low-z exist at all (see Carswell and Rees 1987 for similar result at high-z; although see Williger et al. 2000). This picture is somewhat different from the case of galaxies, where galaxy voids are present even in the distribution of low mass galaxies (e.g. Peebles, 2001; Tikhonov and Klypin, 2009). There is agreement though in the sense that both Lyα system and galaxy distributions have their peaks at the edges of galaxy voids. We observe a typical scale length of the excess to be $\sim 5\ h^{-1}$Mpc, consistent with

numerical predictions for the typical radius of the filamentary structure of the 'cosmic web' ($\sim 2\ h^{-1}$Mpc; see González and Padilla, 2010; Aragón-Calvo et al., 2010; Bond et al., 2010). Note that this scale length is approximately twice the scale associated with a velocity uncertainty of $\Delta v \approx 200$ km s^{-1} at $z = 0$. Such dispersions could be present in our void-edge sample.

In our data analysis we have defined three samples of absorption systems, based on how they are located with respect to the closest galaxy void (see §2.4.2). Let us consider now a very simple model in which we have only two LSS environments: under-dense and over-dense LSS. Then we could relate all the 'random-like' Lyα forest systems found in the void sample ($X < 0.9$ and/or $D < -2$ h^{-1}Mpc) with the under-dense LSS, while all the systems associated with the excess over random ($0.9 \leq X < 1.3$ and/or $-2 \leq D < 4\ h^{-1}$Mpc) to the over-dense LSS[11]. Note that the fact that we cannot distinguish between the under-dense LSS Lyα distribution and a random distribution does not mean that the former is really random. If these Lyα forest systems follow the underlying dark matter distribution (e.g. see Croft et al., 1998), they should have a non-negligible clustering amplitude that is not observed only because of the lack of statistical power of our sample. In relative terms, considering the under-dense LSS Lyα as random seems to be a good approximation though, especially for the large scale distances involved in this work (> 1 Mpc). This is also supported by the very low auto-correlation amplitude observed in the whole population of Lyα forest systems at such scales (e.g. Croft et al., 1998; Rollinde et al., 2003; Crighton et al., 2011). This 'random' behavior of Lyα systems in the under-dense LSS can be understood as originating in structures still evolving from the primordial density perturbations in the linear regime. At $z = 0$ however, the majority of the mass resides at the edges of voids (in the 'cosmic web') whose density perturbations have reached non-linear evolution regime at higher redshifts. For reference, we expect the under-dense and over-dense LSS to have typical $\delta \gtrsim 0$ and $\delta \lesssim 0$ respectively, where δ is the density contrast defined as,

[11] We have left the unclassified systems out of this interpretation.

$$\delta \equiv \frac{\rho - \langle\rho\rangle}{\langle\rho\rangle} \ , \tag{2.4}$$

where ρ is the density and $\langle\rho\rangle$ is the mean density of the universe. Note however that these LSS environments are not defined by a particular density but rather by a topology (voids, walls, filaments).

Theoretical arguments point out that the observed column density of neutral hydrogen at a fixed z is,

$$N_{\rm HI} \propto \rho_{\rm H}^{1.5}\, T^{-0.26}\, \Gamma^{-1}\, f_g^{0.5} \ , \tag{2.5}$$

where $\rho_{\rm H}$ is the density of hydrogen, T is the temperature of the gas, Γ is the hydrogen photo-ionization rate and f_g is the fraction of mass in gas (Schaye, 2001). In the diffuse IGM it has been predicted that $T \propto \rho_{\rm H}^{\alpha}$, where $\alpha \approx 0.59$ (Hui and Gnedin, 1997). This implies that for a fixed Γ, the main dependence of $N_{\rm HI}$ is due to $\rho_{\rm H}$ as $N_{\rm HI} \propto \rho_{\rm H}^{1.4}$. Then, despite the extremely low densities inside galaxy voids we can still observe Lyα systems, although only the ones corresponding to the densest structures.

Let us consider the predicted ratio between $N_{\rm HI}$ observed inside voids and at the edges of voids as,

$$\frac{N_{\rm HI}^{\rm void}}{N_{\rm HI}^{\rm edge}} \approx \left(\frac{\rho_{\rm H}^{\rm void}}{\rho_{\rm H}^{\rm edge}}\right)^{1.4} \left(\frac{\Gamma^{\rm void}}{\Gamma^{\rm edge}}\right)^{-1} \left(\frac{f_g^{\rm void}}{f_g^{\rm edge}}\right)^{0.5} \ . \tag{2.6}$$

Given that the timescale for photons to travel along $\sim 10 - 20\ h^{-1}$Mpc is $\ll 1$ Gyr, we can consider $\Gamma^{\rm void} \approx \Gamma^{\rm edge}$. Even if we assume that the gas inside voids has not formed galaxies, $f_g^{\rm void} \gtrsim f_g^{\rm edge}$, because f_g is dominated by the dark matter. This implies that a given observed $N_{\rm HI}$ inside and at the edge of galaxy voids will correspond to similar densities of hydrogen ($\rho_{\rm H}^{\rm void} \approx \rho_{\rm H}^{\rm edge}$). This is important because it means that the Lyα forest in the under-dense LSS is not different than the over-dense LSS one, and two systems with equal $N_{\rm HI}$ are comparable, independently of its large scale environment.

If there were no galaxies, this simple model may suffice to explain the differences in the observed Lyα population. The fact that some of the Lyα systems are directly associated with galaxies cannot be neglected though. There is strong evidence from observations (e.g. Lanzetta et al., 1995; Chen et al., 1998; Stocke et al., 2006; Morris and Jannuzi, 2006; Chen and Mulchaey, 2009; Crighton et al., 2011; Prochaska et al., 2011b; Rakic et al., 2011; Rudie et al., 2012) and simulations (e.g. Fumagalli et al., 2011; Stinson et al., 2012) that $N_{\rm HI} \gtrsim 10^{15}$ cm^{-2} systems are preferentially found within a couple of hundred kpc of galaxies. Probably an appropriate interpretation of such a result is that galaxies are always found in 'local' ($\lesssim 100\ h^{-1}$kpc) high $N_{\rm HI}$ density regions. Then, a plausible scenario would require *at least* three types of Lyα forest systems: (1) containing embedded galaxies, (2) associated with over-dense LSS but with no close galaxy and (3) associated with under-dense LSS but with no close galaxy. For convenience, we will refer to the first type as 'halo-like', although with the caution that these systems may not be gravitationally bound with the galaxy.

Given that there are galaxies inside galaxy voids, the 'halo-like' Lyα systems will be present in both low and high density LSS environments (galaxies are a 'local' phenomenon). The contribution of the 'halo-like' in galaxy voids could be considered small though. Assuming this contribution to be negligible, we can estimate the fraction of Lyα systems in the under-dense LSS as $\approx 25-30\% \pm 6\%$[12]. Likewise, $\approx 70-75\% \pm 12\%$[11] of the Lyα forest population are due to a combination of systems associated with galaxies and systems associated with the over-dense LSS. We could estimate the contribution of 'halo-like' absorbers by directly looking for and counting galaxies relatively close to the absorption systems. A rough estimation can be done by assuming that galaxy halos will have only $N_{\rm HI} \geq 10^{14}$ cm^{-2} systems, leading to a contribution of $\approx 12-15\% \pm 4\%$[13] in our sample.

In summary, our results require *at least* three types of Lyα systems to explain

[12]These numbers come from $28 \pm 5 (22 \pm 5)$ and $61 \pm 8 (65 \pm 8)$ systems found at $X < 0.9$ ($D < -2$ h^{-1}Mpc) and at $0.9 \leq X < 1.3$ ($-2 \leq D < 4$ h^{-1}Mpc) respectively (see §2.4.1).

[13]From either $13/89$ (excluding the unclassified sample) or $13/106$ (including the unclassified sample). We have assumed Poisson uncertainty.

the observed Lyα forest population at low-z ($N_{\rm HI} \gtrsim 10^{12.5}$ cm^{-2}):

- *Halo-like:* Lyα with embedded nearby galaxies ($\lesssim 100\ h^{-1}$kpc) and so directly correlated with galaxies ($\approx 12 - 15\% \pm 4\%$),
- *Over-dense LSS:* Lyα associated with the over-dense LSS that are correlated with galaxies only because both populations lie in the same LSS regions ($\approx 50 - 55\% \pm 13\%$) and,
- *Under-dense LSS:* Lyα associated with the under-dense LSS with very low auto-correlation amplitude that are not correlated with galaxies ($\approx 25 - 30\% \pm 6\%$).

The relative contributions of these different Lyα populations is a function of the lower $N_{\rm HI}$ limit. Low $N_{\rm HI}$ systems dominate the Lyα column density distribution. Then, given that under-dense LSS Lyα systems tend to be of lower column density than the other two types, we expect the contribution of 'random-like' Lyα to increase (decrease) while observing at lower (higher) $N_{\rm HI}$ limits. Note that there are not sharp $N_{\rm HI}$ limits to differentiate between our three populations (see Figure 2.2). The 'halo-like' is defined by being close to galaxies while the 'LSS-like' ones are defined in terms of a LSS topology (voids, wall, filaments).

Motivated by a recently published study on the Lyα/galaxy association by Prochaska et al. (2011b), we can set a conservative upper limit to the 'halo-like' contribution. These authors have found that nearly all their observed $L \geq 0.01L^*$ galaxies ($33/37$) have $N_{\rm HI} \geq 10^{13.5}$ cm^{-2} absorption at impact parameters $< 300\ h_{72}^{-1}$kpc. If we invert the reasoning and assume an extreme (likely unrealistic) scenario where *all* the $N_{\rm HI} \geq 10^{13.5}$ cm^{-2} are directly associated with galaxies, then, the 'halo-like' contribution will have an upper limit of $< 33\% \pm 7\%$[14]. Consequently, the contribution of the over-dense LSS to the Lyα population will be $> 37 - 42\% \pm 14\%$. Still, note that we have found several systems with $10^{13.5} \lesssim N_{\rm HI} \lesssim 10^{14.5}$ cm^{-2} inside galaxy voids for which a direct association with galaxies is dubious

[14]From either $29/89$ (excluding the unclassified sample) or $35/106$ systems in our sample (including the unclassified sample). We have assumed Poisson uncertainty.

(see Figure 2.2). Also note that only $\sim 10\%$ ($\sim 0\%$) of the Lyα systems between $10^{13.5} < N_{\rm HI} < 10^{14.5}$ cm^{-2} may be associated with a galaxy at impact parameters < 300 kpc (< 100 kpc) in the Prochaska et al. (2011b) sample (see their figure 4).

Our findings are consistent with previous studies pointing out a non-negligible contribution of 'random' Lyα systems (at a similar $N_{\rm HI}$ limit) of $\approx 20 - 30\%$ (Mo and Morris, 1994; Stocke et al., 1995; Penton et al., 2002). These authors estimated that $\approx 70 - 80\%$ of the Lyα population is associated with either LSS (galaxy filaments) or galaxies. Note that Mo and Morris (1994) put a upper limit of $\approx 20\%$ being directly associated with galaxies, which is also consistent with our estimation. Our result is also in accordance with the previous estimation that $22\% \pm 8\%$ (Penton et al. 2002; based on 8 systems) and $17\% \pm 4\%$ (Wakker and Savage 2009; based on 17 systems) of the Lyα systems lie in voids (defined as locations at $> 3\ h_{70}^{-1}$ Mpc from the closest $> L^*$ galaxy). This is in contrast with early models that associated *all* Lyα systems with galaxies (e.g. Lanzetta et al., 1995; Chen et al., 1998).

Although there is general agreement with recently proposed models to explain the origin of the low-z Lyα forest (e.g. Wakker and Savage, 2009; Prochaska et al., 2011b), we emphasize that our interpretation is qualitatively different and adds an important component to the picture: the presence of the under-dense LSS ('random-like') systems. For instance, assuming infinite filaments of typical widths of $\approx 400\ h_{72}^{-1}$kpc around galaxies, Prochaska et al. (2011b) argued that *all* Lyα systems at low-z belong either to the circum-galactic medium (CGM; which includes our 'galaxy halo' definition) or the filamentary structure in which galaxies reside (equivalent to our over-dense LSS definition). Our findings are not fully consistent with this hypothesis, as neither the 'CGM model' nor the 'galaxy filament model' seem likely to explain the majority of our under-dense LSS absorbers at $N_{\rm HI} \lesssim 10^{13.5}$ cm^{-2}. To do so there would need to be a whole population of unobserved galaxies (dwarf spheroidals?) inside galaxy voids with an auto-correlation amplitude as low as the 'random-like' Lyα one. As discussed by Tikhonov and Klypin 2009, very low surface brightness dwarf

spheroidals could be a more likely explanation than dwarf irregulars because the latter should have been observed with higher incidences in recent H I emission blind surveys inside galaxy voids (e.g. HIPASS, Doyle et al. 2005). On the other hand, the formation of dwarf spheroidals inside galaxy voids is difficult to be explained from the current galaxy formation paradigm (see Tikhonov and Klypin 2009 for further discussion). As mentioned, it seems more natural to relate the majority of the under-dense LSS absorbers with the peaks of extremely low density structures inside galaxy voids, still evolving linearly from the primordial density perturbations that have not formed yet galaxies because of their low densities. Our interpretation can be tested by searching for galaxies close to our lowest $N_{\rm HI}$ void-absorbers (see Figure 2.2). Another prediction of our interpretation is that the vast majority of $N_{\rm HI} \lesssim 10^{13}$ cm^{-2} systems should reside inside galaxy voids. If the QSO sightlines used here were observed at higher sensitivities, weak Lyα systems should preferentially appear at $X < 1$ ($D < 0$ h^{-1}Mpc). Therefore, we should expect to have an *anti*-correlation between $N_{\rm HI} \lesssim 10^{13}$ cm^{-2} and galaxies.

2.6.2 $N_{\rm HI}$ and $b_{\rm HI}$ distributions

Our second result is that there is a systematic difference ($\gtrsim 98\%$ c.l.) between the column density distributions of Lyα systems found within, and those found at the edge of, galaxy voids. Void absorbers have more low column density systems than the void-edge sample (see Figure 2.3). A similar trend is found in GIMIC, where low density environments present smaller $N_{\rm HI}$ values than higher density ones (see Figure 2.8, left panel). This can be explained by the fact that baryonic matter follows the underlying dark matter distribution. Then, the highest density environments should be located at the edges of voids (in the intersection of walls and filaments), consequently producing higher column density absorption than in galaxy voids (e.g. see Schaye, 2001).

Also, by construction, there is a higher chance to find galaxies at the edges rather than inside galaxy voids. Assuming that some of the Lyα forest are associated with galaxy halos (see §2.6.1 for further discussion), then this population

should be present mainly in our void-edge sample. As galaxy halos correspond to local density peaks, we should also expect on average higher column density systems in this population. Given that galaxies may affect the properties of the surrounding gas, there could be processes that only affect Lyα systems close to galaxies. For instance, the distribution of Lyα systems around galaxy voids seems to show a two-peaked shape (see Figure 2.1). We speculate that this could be a signature of neutral hydrogen being ionized by the ultra-violet background produced by galaxies (see also Adelberger et al., 2003), mostly affecting $N_{\rm HI} \lesssim 10^{13}$ cm^{-2} inside the filamentary structure of the 'cosmic web'. Another explanation could be that in the inner parts of the filamentary structure, Lyα systems get shock heated by the large gravitational potentials, raising their temperature and ionization state (e.g. Cen and Ostriker, 1999). A third possibility is that it could be a signature of bulk outflows as the shift between peaks is consistent with a $\Delta v \approx 200 - 300$ km s^{-1}. On the other hand, the two peaks could have distinct origins as the first one may be related to an excess of $N_{\rm HI} \lesssim 10^{13}$ cm^{-2} systems, probably associated with the over-dense LSS in which galaxies reside, while the second one may be related to an excess of $N_{\rm HI} \gtrsim 10^{14}$ cm^{-2} systems, more likely associated with systems having embedded galaxies. As mentioned, we cannot prove the reality of this two-peaked signature at a high confidence level from the current sample and so we leave to future studies the confirmation or disproof of these hypotheses.

The GIMIC data analysis shows a clear differentiation of $b_{\rm HI}$ distributions in different density environments (see Figure 2.8, right panel). Low density environments have smaller $b_{\rm HI}$ values than higher density ones. We see a similar trend in the real data between our void and void-edge absorber samples, although only at a $\gtrsim 90\%$ of confidence level (i.e., not very significant; see Figure 2.4). The main mechanisms that contribute to the observed line broadening are temperature, local turbulence and bulk motions of the gas (excluding systematic effects from the line fitting process or degeneracy with $N_{\rm HI}$ for saturated lines). Naturally, in high density environments, we would expect to have greater contributions

from both local turbulence and bulk motions compared to low density ones. The gas temperature is also expected to increase from low density environments to high density ones. As previously mentioned, theoretical arguments predict the majority of the diffuse IGM will have temperatures related to the density by $T \propto \rho^{\alpha}$ with $\alpha > 0$ (Hui and Gnedin, 1997; Theuns et al., 1998b; Schaye et al., 1999). This is also seen in density-temperature diagrams drawn from current hydrodynamical cosmological simulations (e.g. Davé et al., 2010; Tepper-García et al., 2012). Therefore, our findings are consistent with current expectations.

2.6.3 Future work

The high sensitivity of the recently installed *Cosmic Origins Spectrograph* (COS/HST, Green et al., 2012) in the UV (especially the far-UV), will allow us to improve the $N_{\rm HI}$ completeness limit compared with current surveys. This will considerably increase the number of observed Lyα absorption systems at low-z. In the short term, there are several new QSO sightlines scheduled for observations (or already observed) with COS/HST that intersect the SDSS volume. Combining these with current and future galaxy void catalogues, we expect to increase the statistical significance of the results presented in this work. COS/HST will also allow observations of considerably more metal lines (especially O VI) than current IGM surveys. Again, in combination with LSS surveys, this will be very useful for studies on metal enrichment in different environments. For instance, we have identified 8 systems with observed O VI absorption from STIS/HST in our sample. Three of these lie inside voids at $X \approx \{0.6, 0.7, 0.9\}$ ($D \approx \{-5.4, -4.6, -1.6\}$ h^{-1}Mpc) respectively. The first two systems that lie inside voids correspond to the highest $N_{\rm HI}$ values ($N_{\rm HI} > 10^{14.5}$ cm^{-2}; see Figure 2.2). We have performed a search in the SDSS DR8 for galaxies in a cylinder of radius $1\ h_{71}^{-1}$ Mpc and within ± 200 km s^{-1} around these 2 absorbers (both systems belong to the same sightline and are at a similar redshift; one of them shows C IV absorption also). We found 9 galaxies with these constraints, hinting on a possible association of these systems with a void galaxy. The one at the very edge of the void limit has $N_{\rm HI} = 10^{13.14 \pm 0.07}$

cm^{-2} and $N_{\rm OVI} = 10^{13.69\pm0.18}$ cm^{-2}, and it could in principle be associated with the over-dense LSS. The other 5 O VI absorbers lie in our void-edge sample and have $N_{\rm HI} > 10^{13.5}$ cm^{-2}, so they are likely to be associated with galaxies. None of the observed O VI lie in our unclassified sample. The current sample of O VI systems is very small, and so we do not aim to draw statistical conclusions from them. However, these systems individually offer interesting cases worth further investigation. We intend to perform a carefully search for galaxies that could be associated to each of the Lyα absorbers presented in our sample in future work. In the longer term, it will be possible to extend similar analysis to well defined galaxy filaments and clusters when the new generation of galaxy surveys are released.

A scenario with three different types of Lyα forest systems, as proposed here, can help to interpret recent measurements of the cross-correlation between Lyα and galaxies (Morris et al., 1993; Chen et al., 2005; Ryan-Weber, 2006; Wilman et al., 2007; Chen and Mulchaey, 2009; Shone et al., 2010; Rudie et al., 2012). These studies come mainly from pencil beam galaxy surveys around QSO sightlines where identifying LSS such as voids or filaments is more challenging. As mentioned, different Lyα systems are not separated by well defined $N_{\rm HI}$ limits and so we suggest using our results to properly account for under-dense LSS ('random-like') absorbers in gas/galaxy cross-correlations. Truly random distributions are easy to correct for, as they lower the amplitude of the correlations at all scales. Then, acknowledging these 'random-like' absorbers, it will be possible to split the correlation power in its other two main components: gas in galaxy halos and gas in the over-dense LSS.

2.7 Summary

We have presented a statistical study of H I Lyα absorption systems found within and around galaxy voids at $z \lesssim 0.1$. We found a significant excess ($> 99\%$ c.l.) of Lyα systems at the edges of galaxy voids with respect to a random distribution,

over a $\sim 5\ h^{-1}$Mpc scale. We have interpreted this excess as being due to Lyα systems associated with both galaxies ('halo-like') and the over-dense LSS in where galaxies reside (the observed 'cosmic web'), accounting for $\approx 70 - 75\% \pm 12\%$ of the Lyα population. We found no significant difference in the number of systems inside galaxy voids compared to the random expectation. We therefore infer the presence of a third type of Lyα systems associated to the under-dense LSS with a low auto-correlation amplitude ($\approx$ random) that are not associated with luminous galaxies. These 'random-like' absorbers are mainly found in galaxy voids. We argue that these systems can be associated with structures still growing linearly from the primordial density fluctuations at $z = 0$ that have not yet formed galaxies because of their low densities. Although the presence of a 'random' population of Lyα absorbers was also inferred (or assumed) in previous studies, our work presents for the first time a simple model to explain it (see §2.6.1). Above a limit of $N_{\rm HI} \gtrsim 10^{12.5}$ cm^{-2}, we estimate that $\approx 25 - 30\% \pm 6\%$ of Lyα forest systems are 'random-like' and not correlated with luminous galaxies. Assuming that *only* $N_{\rm HI} \geq 10^{14}$ cm^{-2} systems have embedded galaxies nearby, we have estimated the contribution of the 'halo-like' Lyα population to be $\approx 12 - 15\% \pm 4\%$ and consequently $\approx 50 - 55\% \pm 13\%$ of the Lyα systems to be associated with the over-dense LSS.

We have reported differences between both the column density ($N_{\rm HI}$) and the Doppler parameter ($b_{\rm HI}$) distributions of Lyα systems found inside and at the edge of galaxy voids observed at the $> 98\%$ and $> 90\%$ of confidence level respectively. Low density environments (voids) have smaller values for both $N_{\rm HI}$ and $b_{\rm HI}$ than higher density ones (edges of voids). These trends are theoretically expected. We have performed a similar analysis using simulated data from GIMIC, a state-of-the-art hydrodynamical cosmological simulation. Although GIMIC did not give a perfect match to the observed column density distribution, the aforementioned trends were also seen. Any discrepancy between GIMIC and real data could be due to low number statistic fluctuations and/or a combination of several observational effects.

In summary, our results are consistent with the expectation that the mechanisms shaping the properties of the Lyα forest are different in different LSS environments. By focusing on a 'large scale' ($\gtrsim 10$ Mpc) point of view, our results offer a good complement to previous studies on the IGM/galaxy connection based on 'local' scales ($\lesssim 2$ Mpc).

Acknowledgments

We thank the anonymous referee for helpful comments which improved this Chapter and its published version (Tejos et al., 2012). We thank Charles Danforth for having kindly provided S/N information for the QSO spectra used in this work.

We thank contributors to SciPy[15], Matplotlib[16], and the Python programming language[17]; the free and open-source community; and the NASA Astrophysics Data System[18] for software and services.

N.T. acknowledges grant support by CONICYT, Chile (PFCHA/*Doctorado al Extranjero 1ª Convocatoria*, 72090883). This research was supported in part by the National Science Foundation under Grant NSF PHY11-25915.

[15] `http://www.scipy.org`
[16] `http://www.matplotlib.sourceforge.net`
[17] `http://www.python.org`
[18] `http://adswww.harvard.edu`

Chapter 3

The two-point correlation function

3.1 Definition

The two-point correlation function (2PCF), $\xi(r)$, is defined as the excess probability of finding two points at a given separation with respect to the expectation from a randomly distributed sample (e.g. see chapter 3 of Peebles, 1980). Under the assumption of isotropy, the correlation function is a function of the separation between points only. Let $n(\vec{r})$ be the density of points in a volume element δV at the position $\vec{r}$. Then, the density fluctuation in such a volume is defined as

$$\frac{\delta n(\vec{r})}{\langle n \rangle} \equiv \frac{n(\vec{r}) - \langle n \rangle}{\langle n \rangle} \, , \tag{3.1}$$

where $\langle n \rangle$ is the average density of the whole sample. The 2PCF would correspond then to the averaged product between the density fluctuations in two volume elements separated by a distance r:

$$\xi(r) \equiv \left\langle \frac{\delta n(\vec{r}+r)}{\langle n \rangle} \frac{\delta n(\vec{r})}{\langle n \rangle} \right\rangle = \frac{\langle n(\vec{r}+r) n(\vec{r}) \rangle}{\langle n \rangle^2} - 1 \, , \tag{3.2}$$

which corresponds to a convolution between density fluctuation functions over the whole volume. This definition allows values between $-1 \leq \xi(r) < \infty$. The points are correlated when $\xi(r) > 0$. When r is such that $-1 \leq \xi(r) < 0$, it means that there are less pairs of points than expected from a homogeneous distribution, so there is an anti-correlation at that separation. On the other hand, the sample is uncorrelated when $\xi(r) = 0$.

3.1.1 Auto- and cross-correlation

The previous definition is based on just one population of points, so it will be referred as the auto-correlation function. The cross-correlation function is equivalently defined by counting pairs composed of points from different populations. Following the previous notation, it is possible to define $\delta n_a(\vec{r})$ and $\delta n_b(\vec{r})$ as the density fluctuations for a-like and b-like points. In such a case, the cross-correlation can be written as

$$\xi_{\rm ab}(r) \equiv \left\langle \frac{\delta n_{\rm a}(\vec{r}+r)}{\langle n_{\rm a}\rangle}\frac{\delta n_{\rm b}(\vec{r})}{\langle n_{\rm b}\rangle}\right\rangle = \frac{\langle n_{\rm a}(\vec{r}+r)n_{\rm b}(\vec{r})\rangle}{\langle n_{\rm a}\rangle\langle n_{\rm b}\rangle} - 1\ . \tag{3.3}$$

Note that by the definition symmetry, $\xi_{\rm ab}(r) = \xi_{\rm ba}(r)$.

A relation between the cross-correlation and the auto-correlations can be inferred from the previous definition by considering the Cauchy-Schwarz inequality:

$$\begin{aligned}\xi^2_{\rm ab}(r) &= \left\langle \frac{\delta n_{\rm a}(\vec{r}+r)}{\langle n_{\rm a}\rangle}\frac{\delta n_{\rm b}(\vec{r})}{\langle n_{\rm b}\rangle}\right\rangle^2 \\ &\leq \left\langle \frac{\delta n_{\rm a}(\vec{r}+r)}{\langle n_{\rm a}\rangle}\frac{\delta n_{\rm a}(\vec{r})}{\langle n_{\rm a}\rangle}\right\rangle \left\langle \frac{\delta n_{\rm b}(\vec{r}+r)}{\langle n_{\rm b}\rangle}\frac{\delta n_{\rm b}(\vec{r})}{\langle n_{\rm b}\rangle}\right\rangle\ ,\end{aligned} \tag{3.4}$$

then,

$$\xi^2_{\rm ab}(r) \leq \xi_{\rm aa}\xi_{\rm bb}\ . \tag{3.5}$$

The equality only holds when both auto-correlations are linearly dependent.

3.1.2 Limited samples

So far we have implicitly defined the 2PCF over a *fair sample*, i.e., that the sample is representative of the *real* underlying distribution. However in practice, the 2PCF is measured over a finite sample which can not be always considered as *fair*.

Let us consider $\xi_{\rm ab}(r)$ as the *real* underlying cross-correlation while $\langle\xi_{\rm ab}(r)\rangle$ is the observed one over a volume V, defined by the survey. Thus,

$$\langle\xi_{\rm ab}(r)\rangle \equiv \frac{1}{V^2}\int_V \xi_{\rm ab}(r)\, d^2V \ . \tag{3.6}$$

This means that what we measure as the 2PCF will be inevitably affected by the limited nature of the sample. This issue was identified early and is usually refered to as the *Integral Constraint* bias. Still, the equation 3.5 is also valid for the measured correlations. Consider,

$$\begin{aligned}\langle\xi_{\rm ab}(r)\rangle^2 &= \left[\frac{1}{V^2}\int_V \xi_{\rm ab}(r)\, d^2V\right]^2 \\ &= \left[\frac{1}{V^2}\int_V [\xi^2_{\rm ab}(r)]^{\frac{1}{2}}\, d^2V\right]^2 \ ,\end{aligned} \tag{3.7}$$

from equation 3.5 we have,

$$\langle\xi_{\rm ab}(r)\rangle^2 \leq \left[\int_V \left(\frac{\xi_{\rm aa}(r)}{V^2}\right)^{\frac{1}{2}}\left(\frac{\xi_{\rm bb}(r)}{V^2}\right)^{\frac{1}{2}}\, d^2V\right]^2 \ , \tag{3.8}$$

and applying the Cauchy-Schwarz inequality again we have,

$$\begin{aligned}\langle\xi_{\rm ab}(r)\rangle^2 &\leq \int_V \frac{\xi_{\rm aa}(r)}{V^2}\, d^2V \int_V \frac{\xi_{\rm bb}(r)}{V^2}\, d^2V \\ &\leq \langle\xi_{\rm aa}(r)\rangle\langle\xi_{\rm bb}(r)\rangle \ .\end{aligned} \tag{3.9}$$

Two situations are worth noting from this result. First, equation 3.9 will also be valid when the survey has a more general (positive) non-constant selection function. Second, even if the equality holds for the underlying correlations (i.e. $\xi^2_{\rm ab} = \xi_{\rm aa}\xi_{\rm bb}$), it will not be necessarily the case for the observed ones.

3.2 Pairwise estimators

As can be seen from previous section, the 2PCF is a powerful tool for measuring clustering. Particularly in astronomy, important properties of the surveyed objects can be inferred from the 2PCF (e.g. halo masses, power spectrum, spatial distribution). In practice, an estimator is needed in order to measure the 2PCF. The most commonly used so far, are the so called pairwise-like estimators[1].

From equation 3.2 we have the following

$$\xi(r) = \frac{\langle n(\vec{r}+r)n(\vec{r})\rangle}{\langle n\rangle^2} - 1 \; . \tag{3.10}$$

On one hand, $\langle n(\vec{r}+r)n(\vec{r})\rangle$ is proportional to the current number of pairs at separation r over the whole survey, while on the other hand, $\langle n\rangle^2$ is proportial to the expected number of pairs from an, equaly dense, randomly distributed sample. Therefore, a way to estimate the 2PCF in a particular sample would be by counting pairs. The natural estimator is then defined as,

$$\hat{\xi}_{\rm N} \equiv \frac{DD}{RR} - 1 \; , \tag{3.11}$$

where DD, RR correspond to the number of data-data, and random-random pairs at each given separation. While the original data are directly observed, the random data must be simulated. With the advent of faster and more powerful computers this task is relatively easy to achieve. A clear advantage of this kind of estimator is that it takes into account the systematics associated with the geometry of the survey, as the random points will be confined in the same volume as the real data. Moreover, it is possible to simulate as much random data as we want, to make the statistical uncertainties in the RR value as small as it is required.

Similarly, other pairwise-like estimators have been introduced by different

[1]For a more general discussion about other kind of estimators we refer the reader to Kerscher (1999) and references therein.

authors. For instance, the Hewett estimator (Hewett, 1982), is defined as

$$\hat{\xi}_{\rm Hew} \equiv \frac{DD - DR}{RR} \,, \tag{3.12}$$

while the Davis and Peebles estimator (Davis and Peebles, 1983), is

$$\hat{\xi}_{\rm DP} \equiv \frac{DD}{DR} - 1 \,, \tag{3.13}$$

where DR data-random pairs at each separation. These estimators are equivalent than the natural, as the normalized number of data-random pairs is equal to RR. This is so because the random points are, by definition, uncorrelated, and therefore the cross-correlation between data-random pairs is equal to 0. Given that the process of counting data-random pairs is much less time-consuming than that for counting random-random pairs, the Davis and Peebles estimator is commonly preferred over either the natural or the Hewett estimator.

3.2.1 Refined pairwise estimators

Given that the 2PCF is measured over a limited sample (see also section 3.1.2), the uncertainty in the previous estimators are limited by the uncertainty in the mean density (see Hamilton, 1993; Landy and Szalay, 1993, for a demonstration). These authors have proposed refined estimators which minimize the statistical uncertainties in the 2PCF measurement. These are,

$$\hat{\xi}_{\rm Ham} \equiv \frac{DD \times RR}{(DR)^2} - 1 \,, \tag{3.14}$$

and,

$$\hat{\xi}_{\rm LS} \equiv \frac{DD - 2DR + RR}{RR} \,, \tag{3.15}$$

which are referred as the Hamilton and the Landy and Szalay estimators respectively. Both estimators have proved to be very successful (e.g. see Kerscher et al., 2000) and nowadays are widely used by the astronomy community. However,

systematic uncertainties are still important (see section 3.3) and will be present even in current samples.

3.2.2 Cross-correlation estimators

The previous definitions can be generalized to the cross-correlation between two kind of populations. Let $D_{\rm a}D_{\rm b}$ and $R_{\rm a}R_{\rm b}$ be the number of data-data and random-random cross-pairs at a given separation respectively. Analogously, let $D_{\rm a}R_{\rm b}$ and $R_{\rm a}D_{\rm b}$ be the 2 possible number of data-random cross-pairs at the same separation. By a symmetry argument, the estimators can be written as,

$$\begin{aligned}
\hat{\xi}_{\rm N} &\equiv \frac{D_{\rm a}D_{\rm b}}{R_{\rm a}R_{\rm b}} - 1 \,, \\
\hat{\xi}_{\rm Hew} &\equiv \frac{D_{\rm a}D_{\rm b} - D_{\rm a}R_{\rm b}}{R_{\rm a}R_{\rm b}} - 1 \,, \\
\hat{\xi}_{\rm DP} &\equiv \frac{D_{\rm a}D_{\rm b}}{D_{\rm a}R_{\rm b}} - 1 \,, \\
\hat{\xi}_{\rm Ham} &\equiv \frac{D_{\rm a}D_{\rm b} \times R_{\rm a}R_{\rm b}}{D_{\rm a}R_{\rm b} \times R_{\rm a}D_{\rm b}} - 1 \,, \\
\hat{\xi}_{\rm LS} &\equiv \frac{D_{\rm a}D_{\rm b} - D_{\rm a}R_{\rm b} - R_{\rm a}D_{\rm b} + R_{\rm a}R_{\rm b}}{R_{\rm a}R_{\rm b}} \,.
\end{aligned} \tag{3.16}$$

3.3 Multi-component Samples

It is straightforward to show how the observed 2PCF will be affected by the presence of different population which are differently correlated, in a given sample. Let us consider a sample of two *distinguishable* population of objects, composed of $N_{\rm a}$ and $N_{\rm b}$ points respectively which are immersed in a volume V. Let $N_{\rm a}$ and $N_{\rm b}$ be each composed by *indistinguishable* points coming from A and B different kinds of correlated populations respectively, so $N_{\rm a} = \sum_i^A N_{{\rm a}_i}$ and $N_{\rm b} = \sum_j^B N_{{\rm b}_j}$. The mean densities for each component population are then approximated by $\langle n_{{\rm a}_i} \rangle \approx N_{{\rm a}_i}/V$ for $i \in \{1, 2, .., A\}$ and $\langle n_{{\rm b}_j} \rangle \approx N_{{\rm b}_j}/V$ for $j \in \{1, 2, .., B\}$. Following the equation 3.3, the probability of finding an object of pop-a in the element $\delta V_{\rm a}$ and simultaneously an object of pop-b in the element $\delta V_{\rm b}$ at separation r will be:

$$\delta P = \sum_{i,j}^{A,B} \langle n_{\mathrm{a}_i} \rangle \langle n_{\mathrm{b}_j} \rangle [1 + \xi_{\mathrm{a}_i \mathrm{b}_j}(r)] \delta V_{\mathrm{a}} \delta V_{\mathrm{b}} \ , \tag{3.17}$$

where $\xi_{\mathrm{a}_i \mathrm{b}_j}$ is the cross-correlation function between population a_i and b_j. Likewise, the probability of finding random objects in δV_{a} and δV_{b} at separation r, for the same sample, is

$$\delta P = \sum_{i,j}^{A,B} \langle n_{\mathrm{a}_i} \rangle \langle n_{\mathrm{b}_j} \rangle \delta V_{\mathrm{a}} \delta V_{\mathrm{b}} \ . \tag{3.18}$$

The expected number of cross-correlated data-data pairs between both populations, $D_{\mathrm{a}} D_{\mathrm{b}}$, and the expected number of random-random pairs, $R_{\mathrm{a}} R_{\mathrm{b}}$, for any given separation, are then,

$$D_{\mathrm{a}} D_{\mathrm{b}} = \sum_{i,j}^{A,B} \langle n_{\mathrm{a}_i} \rangle \langle n_{\mathrm{b}_j} \rangle \int_V \int_V (1 + \xi_{\mathrm{a}_i \mathrm{b}_j}) \, dV_{\mathrm{a}} dV_{\mathrm{b}} \ , \tag{3.19}$$

and,

$$\begin{aligned} R_{\mathrm{a}} R_{\mathrm{b}} &= \sum_{i,j}^{A,B} \langle n_{\mathrm{a}_i} \rangle \langle n_{\mathrm{b}_j} \rangle \int_V \int_V \, dV_{\mathrm{a}} dV_{\mathrm{b}} \\ &= \langle n_{\mathrm{a}} \rangle \langle n_{\mathrm{b}} \rangle V^2 \ , \end{aligned} \tag{3.20}$$

respectively. From this, the *observed* cross-correlation, which is estimated from $D_{\mathrm{a}} D_{\mathrm{b}} / R_{\mathrm{a}} R_{\mathrm{b}}$, will be finally,

$$\begin{aligned} (1 + \xi_{\mathrm{ab}}^{obs}) &= \frac{D_{\mathrm{a}} D_{\mathrm{b}}}{R_{\mathrm{a}} R_{\mathrm{b}}} \\ &= \sum_{i,j}^{A,B} \left(\frac{\langle n_{\mathrm{a}_i} \rangle}{\langle n_{\mathrm{a}} \rangle} \right) \left(\frac{\langle n_{\mathrm{b}_j} \rangle}{\langle n_{\mathrm{b}} \rangle} \right) (1 + \langle \xi_{\mathrm{a}_i \mathrm{b}_j} \rangle) \ , \end{aligned} \tag{3.21}$$

then,

$$\xi_{\rm ab}^{obs} = \sum_{i,j}^{\rm A,B} \left(\frac{\langle n_{\rm a_i} \rangle}{\langle n_{\rm a} \rangle} \right) \left(\frac{\langle n_{\rm b_j} \rangle}{\langle n_{\rm b} \rangle} \right) \langle \xi_{\rm a_i b_j} \rangle \ , \tag{3.22}$$

where we have defined $\langle \xi_{\rm a_i b_j} \rangle \equiv \frac{1}{V^2} \int \int_V \xi_{\rm a_i b_j} \, dV_{\rm a} dV_{\rm b}$ (see section 3.1.2).

Given that in this section we are not interested in the statistical uncertainties associated to the measurement of the correlation function, but just the systematics when multi-component samples are present, we will assume that the sample is *fair*, so it is large enough that we can consider $\langle n_{a_i} \rangle$, $\langle n_{b_j} \rangle$ being good representations of the *real* values. Likewise, we can consider $\langle \xi_{a_i b_j} \rangle \approx \xi_{a_i b_j}$. Defining $f_{a_i} \equiv \frac{N_{a_i}}{N_{\rm a}}$ and $f_{b_j} \equiv \frac{N_{b_j}}{N_{\rm b}}$, then the observed auto-correlation will be,

$$\xi_{ab}^{obs} = \sum_{i,j}^{A,B} f_{a_i} f_{b_j} \xi_{a_i b_j} \ . \tag{3.23}$$

This result is not just valid for the natural estimator. Given that $R_{\rm a} R_{\rm b} \propto D_{\rm a} R_{\rm b} \propto D_{\rm b} R_{\rm a} \propto \langle n_{\rm a} \rangle \langle n_{\rm b} \rangle V^2$ (see section 3.2), it is easy to show that all the pairwise 2PCF estimators will be affected in the same manner.

3.3.1 Simple Examples

Since within a given observed population, there may be no way to distinguish whether a point belongs to one kind of sub-population or to another, when estimating the two-point correlation function, the observed correlation will be considerably affected by the contribution of the different sub-populations. In the following, we will consider two simple examples in order to illustrate the situation.

Example 1: contaminated auto-correlation

Let us consider a sample of $N = N_{a_1} + N_{a_2}$ points where N_{a_1} and N_{a_2} came from different but indistinguishable kind of a-like populations. According to the previous result (from equation), the observed auto-correlation will be,

$$\begin{aligned}\xi_{aa}^{obs} &= f_{a_1} f_{a_1} \xi_{a_1 a_1} + f_{a_1} f_{a_2} \xi_{a_1 a_2} + f_{a_2} f_{a_1} \xi_{a_2 a_1} + f_{a_2} f_{a_2} \xi_{a_2 a_2} \\ &= f_{a_1}^2 \xi_{a_1 a_1} + 2 f_{a_1} f_{a_2} \xi_{a_1 a_2} + f_{a_2}^2 \xi_{a_2 a_2} \ ,\end{aligned} \tag{3.24}$$

where $\xi_{a_1 a_1}$ and $\xi_{a_2 a_2}$ are the auto-correlation of each sub-population whereas $\xi_{a_1 a_2}$ ($= \xi_{a_2 a_1}$) is the cross-correlation between both of them. If we are interested in measuring the $\xi_{a_1 a_1}$, the a_2-like population would correspond to a 'contamination', and we need to correct for it. Then,

$$\xi_{a_1 a_1} = \frac{\xi_{aa}^{obs}}{f_{a_1}^2} - 2 \left(\frac{f_{a_2}}{f_{a_1}} \right) \xi_{a_1 a_2} - \left(\frac{f_{a_2}}{f_{a_1}} \right)^2 \xi_{a_2 a_2} \ . \tag{3.25}$$

If we think now of a contamination of uncorrelated points, say $\xi_{a_2 a_2} = 0$ (then also $\xi_{a_1 a_2} = 0$) and therefore the auto-correlation will be,

$$\xi_{a_1 a_1} = \frac{\xi_{aa}^{obs}}{f_{a_2}^2} \ . \tag{3.26}$$

So, the observed correlation is considerably attenuated. A contamination of $\approx 10 - 20\%$ of uncorrelated points will lead to an observed correlation $\approx 20 - 36\%$ smaller than the expected one.

Example 2: contaminated cross-correlation

Let us consider a sample of $N_{\rm a} = N_{\rm a_1} + N_{a_2}$ and $N_{\rm b}$ points belonging to different correlated populations as before. In this case, the observed cross-correlation function between both a-like and b-like objects will be,

$$\xi_{ab}^{obs} = f_{a_1} f_{\rm b} \xi_{a_1 b} + f_{a_2} f_{\rm b} \xi_{a_2 b} \ . \tag{3.27}$$

Given that in this case there is just one type of b-like points, $f_{\rm b} = 1$ and therefore,

$$\xi_{ab}^{obs} = f_{a_1} \xi_{a_1 b} + f_{a_2} \xi_{a_2 b} \ , \tag{3.28}$$

which is independent of $N_{\rm b}$. If we are interested in measuring $\xi_{a_1 b}$, the a_2-like

population would correspond to a 'contamination'. Finally, the cross-correlation must be corrected as,

$$\xi_{a_1 b} = \frac{\xi_{ab}^{obs}}{f_{a_1}} - \left(\frac{f_{a_2}}{f_{a_1}}\right) \xi_{a_2 b} \, . \tag{3.29}$$

As before, if we think of a contamination of uncorrelated points, say $\xi_{a_2 a_2} = 0$, then $\xi_{a_2 b} = 0$ and therefore the cross-correlation will be just,

$$\xi_{a_1 b} = \frac{\xi_{ab}^{obs}}{f_{a_1}} \, . \tag{3.30}$$

Although smaller than the auto-correlation case, the effect of the contamination is still significant. If we allow population b to be also contaminated, then the total attenuation will be proportional to the product between both contamination fractions.

3.3.2 Implications in astrophysical experiments

In this section we want to emphasize the importance of the contamination in some astrophysical experiments, and how it should be properly corrected according to our previous findings.

Photometric galaxy samples

Given that photometry is much cheaper than spectroscopy, photometric galaxy samples have the advantage of being very large. However, a clear disadvantage is that the photometric redshift precision is much worst than that from spectroscopy. Likewise, another disadvantage is the significant presence of contamination.

Examples of these kind of surveys are the narrow-band selected galaxies like Lyα emitters (LAEs) as well as the Lyα break galaxies (LBGs). The galaxy clustering at a particular redshift is intended to be measured from such surveys (e.g. Bielby et al., 2011). However, contamination from stars and interlopers (galaxies at different redshifts) are inevitable. Therefore, there are two sources of

contamination in this case. Following equation 3.23 and its notation, the observed auto-correlation will be,

$$\xi^{obs} = f_g^2\xi_{gg} + f_i^2\xi_{ii} + f_s^2\xi_{ss} + 2f_gf_s\xi_{gs} + 2f_gf_i\xi_{gi} + 2f_if_s\xi_{is} \ , \tag{3.31}$$

where the sub-indexes g,i and s represent galaxies, interlopers and stars respectively. A good approximation is to consider stars as an uncorrelated population so $\xi_{ss} = \xi_{gs} = \xi_{is} = 0$. The wanted galaxy-galaxy auto-correlation should be properly corrected by using

$$\xi_{gg} = \frac{\xi^{obs}}{f_g^2} - 2\left(\frac{f_i}{f_g}\right)\xi_{gi} - \left(\frac{f_i}{f_g}\right)^2\xi_{ii} \ . \tag{3.32}$$

Although it seems to be independent of the uncorrelated population, note that $f_g + f_i + f_s = 1$, so the correction does depend on it. One can assume that galaxies and interlopers come from dependent populations[2] then $\xi_{gi} = \sqrt{\xi_{gg}\xi_{ii}}$. However, one would still need an independent measure of ξ_{ii} and a f_i or f_g estimation.

'Clean' photometric and spectroscopic galaxy samples

Let us consider now 'clean' galaxy samples, which do not have contamination due to stars or interlopers. By directly looking at the spectrum of each object, spectroscopic galaxy surveys would not include this contamination. In the case of photometric galaxy surveys though, they must be properly corrected as we see in the previous sub-section. Despite these considerations, both surveys could still contain a less evident source of contamination.

Given that the selection functions are not necessarily related with the physical properties of the objects surveyed, a cut-off in observed luminosity for instance will include different types of galaxies which are not necessarily correlated in the same manner. Likewise, when correlations between the IGM and galaxies are

[2]This could be the case if both galaxy samples come from the same dark matter distribution for instance.

performed, they could be affected by less correlated gas (e.g. Hydrogen) at very low columns densities for instance. Moreover, different kind of galaxies could be differently correlated with gas depending on a particular property (their star formation activity for instance; see Chen et al., 2005) and so on. Thus, special care must be taken when constructing the sample that will lead a correlation function measurement. One needs to consider all possible sources of contamination in order to have a meaningful result.

3.3.3 Summary

In this section have investigated how the 2PCF is affected by the presence of multi-component samples. We have defined 'contamination' as an indistinguishable set of points coming from different correlated populations which are immersed in the whole sample of objects for which the real correlation is measured. We have quantified the net effect of this systematic bias in the observed correlation, which has been found to be important.

Chapter 4

The IGM–galaxy cross-correlation at $z \lesssim 1$

4.1 Overview

We present a new optical spectroscopic survey of 1777 'star-forming' ('SF') and 366 'non-star-forming' ('non-SF') galaxies (2143 in total), 22 AGN and 423 stars, observed by instruments such as DEIMOS, VIMOS and GMOS, around 3 fields containing 5 QSOs with HST UV spectroscopy. We also present a new spectroscopic survey of 173 'strong' ($10^{14} \leq N_{\rm HI} \lesssim 10^{17}$ cm^{-2}), and 496 'weak' ($10^{13} \lesssim N_{\rm HI} < 10^{14}$ cm^{-2}) well identified intervening H I absorption line systems at $z \lesssim 1$ (669 in total), observed in the spectra of 8 QSOs at $z \sim 1$ by COS and FOS on the HST. Combining these new data with previously published galaxy catalogs such as VVDS and GDDS, we have gathered a sample of 17509 galaxies and 654 H I absorption systems suitable for a two-point correlation function analysis. We present observational results on the H I–galaxy ($\xi_{\rm ag}$) and galaxy–galaxy ($\xi_{\rm gg}$) correlations at transverse scales $r_\perp \lesssim 10$ Mpc, and the H I–H I auto-correlation ($\xi_{\rm aa}$) at transverse scales $r_\perp \lesssim 2$ Mpc. The two-point correlation functions are measured both along and transverse to the line-of-sight, $\xi(r_\perp, r_\parallel)$. We also infer the shape of their corresponding 'real-space' correlation functions, $\xi(r)$, assuming power-laws of the form $\xi(r) = (r/r_0)^{-\gamma}$. Comparing the results from $\xi_{\rm ag}$, $\xi_{\rm gg}$ and $\xi_{\rm aa}$ we constrain the H I–galaxy statistical connection, as a function of both H I column density and galaxy star-formation activity. Our results are consistent with the following conclusions: (i) the bulk of H I systems around galaxies have very little velocity dispersion ($\ll 120$ km s^{-1}) with respect to the bulk of galaxies

(hence no strong galaxy outflow or inflow signal is detected); (ii) the vast majority ($\sim 100\%$) of 'strong' H I systems and 'SF' galaxies are distributed in the same locations, together with $75 \pm 15\%$ of 'non-SF' galaxies, all of which typically reside in dark matter haloes of similar masses; (iii) $25 \pm 15\%$ of 'non-SF' galaxies reside in galaxy clusters and are not correlated with 'strong' H I systems at scales $\lesssim 2$ Mpc; and (iv) $> 50\%$ of 'weak' H I systems reside within galaxy voids (hence not correlated with galaxies), and are confined in dark matter haloes of masses smaller than those hosting 'strong' systems and/or galaxies. We also found evidence that the 'absolute bias' of 'weak' H I systems might be $b < 1$.

4.2 Introduction

4.2.1 Motivation

The study of the intergalactic medium (IGM) and its connection with galaxies is key to understanding the evolution of baryonic matter in the Universe. This is because of the continuous interplay between the gas in the IGM and galaxies: (i) galaxies are formed by the condensation and accretion of primordial or enriched gas; and (ii) galaxies enrich their haloes and the IGM via galactic winds and/or merger events.

Theoretical analyses—under a Λ cold dark matter paradigm (ΛCDM)—suggest that: (i) the accretion happens in two major modes: 'hot' and 'cold' (e.g. Rees and Ostriker, 1977; White and Rees, 1978; White and Frenk, 1991; Kereš et al., 2005; van de Voort et al., 2011); and (ii) galactic winds are mostly driven by supernova (SN) and/or active galactic nuclei (AGN) feedback (e.g. Baugh et al., 2005; Bower et al., 2006; Lagos et al., 2008).

Combining 'N-body' dark matter simulations (collisionless, dissipationless) with 'semi-analytic' arguments (e.g. Baugh, 2006, and references therein) have been successful in reproducing basic statistical properties of luminous galaxies (e.g. luminosity functions, clustering, star-formation histories, among others).

However, in order to provide predictions for the signatures of 'hot'/'cold' accretion and/or AGN/SN feedback in the IGM, a *full* hydrodynamical description is required.

In practice, hydrodynamical simulations still rely on unresolved 'sub-grid physics' to lower the computational cost (e.g. Schaye et al., 2010; Scannapieco et al., 2012), whose effects are not fully understood. Therefore, observations of the IGM and galaxies in the same volume are fundamental to testing these predictions and helping to discern between different physical models (e.g. Fumagalli et al., 2011; Oppenheimer et al., 2012; Stinson et al., 2012; Hummels et al., 2013; Ford et al., 2013).

Although the IGM hosts the main reservoir of baryons at all epochs (e.g. Fukugita et al., 1998; Cen and Ostriker, 1999; Davé et al., 2010; Shull et al., 2012), its extremely low densities make its observation difficult and limited. Currently, the only feasible way to observe the IGM is through intervening absorption line systems in the spectra of bright background sources, limiting its characterization to being one-dimensional. Still, an averaged three dimensional picture can be obtained by combining multiple lines-of-sight (LOS) and galaxy surveys, which is the approach adopted in this work (see Section 4.2.2).

The advent of the Cosmic Origins Spectrograph (COS) on the Hubble Space Telescope (HST) has revolutionized the study of the IGM and its connection with galaxies at low-z. With a sensitivity ~ 10 times greater than that of its predecessors, COS has considerably increased the number of quasi-stellar objects (QSOs) for which ultra-violet (UV) spectroscopy is feasible. This capability has recently been exploited for studies of the so-called circumgalactic medium (CGM), by characterizing neutral hydrogen (H I) and metal absorption systems in the vicinity of known galaxies (e.g. Tumlinson et al., 2011; Thom et al., 2012; Werk et al., 2013; Stocke et al., 2013; Keeney et al., 2013; Lehner et al., 2013).

Studies of the CGM implicitly assume a direct one-to-one association between absorption systems and their closest observed galaxy, which might not always hold because of incompleteness in the galaxy surveys and projection effects. Given

that metals are formed and expelled by galaxies, a direct association between them seems sensible, in accordance with predictions from low-z simulations (e.g. Oppenheimer et al., 2012). However, the situation for neutral hydrogen is more complicated, as H I traces both enriched *and* primordial material.[1]

The nature of the relationship between H I and galaxies at low-z has been widely debated. Early studies have pointed out two distinct scenarios for this connection: (i) a one-to-one physical association because they both belong to the same dark matter haloes (e.g. Mo, 1994; Lanzetta et al., 1995; Chen et al., 1998); and (ii) an indirect association because they both trace the same underlying dark matter distribution but not necessarily the same haloes (e.g. Morris et al., 1991, 1993; Mo and Morris, 1994; Stocke et al., 1995; Tripp et al., 1998). More recent studies have shown the presence of H I absorption systems within galaxy voids (e.g. Grogin and Geller, 1998; Penton et al., 2002; Manning, 2002; Stocke et al., 1995; Tejos et al., 2012), hinting at a third scenario: (iii) the presence of systems that *are not* associated with galaxies (although see Wakker and Savage, 2009).[2]

If we think of galaxies as peaks in the density distribution (e.g. Press and Schechter, 1974), it is natural to expect the high column density H I systems to show a stronger correlation with galaxies than low column density ones, owing to a density-H I column density proportionality (e.g. Schaye, 2001; Davé et al., 2010). Similarly, we also expect the majority of low column density H I systems to belong to dark matter haloes that did not form galaxies. Thus, the relative importance of these three scenarios should depend, to some extent, on the H I column density. Tejos et al. (2012) estimated that these three scenarios account for $\sim 15\%$, $\sim 55\%$ and $\sim 30\%$ of the low-z H I systems at column densities $N_{\rm HI} \gtrsim 10^{12.5}$ cm^{-2}, respectively, indicating that the vast majority of H I absorption line systems are not physically associated with luminous galaxies (see also Prochaska et al., 2011b).

[1]Note that whether truly primordial H I clouds exist at low-z is still to be observationally confirmed.

[2]Note that little can be said about low surface brightness galaxies, as current spectroscopic surveys are strongly biased against these, for obvious reasons.

4.2.2 Study strategy

In this Chapter we address the statistical connection between H I and galaxies at $z \lesssim 1$ through a clustering analysis (e.g. Morris et al., 1993; Chen et al., 2005; Ryan-Weber, 2006; Wilman et al., 2007; Chen and Mulchaey, 2009; Shone et al., 2010), without considering metals. We focus only on hydrogen because it is the best IGM tracer for a statistical study. Apart from the fact that it traces both primordial and enriched material, it is also the most abundant element in the Universe. Hence, current spectral sensitivities allow us to find H I inside and outside galaxy haloes, which is not the case yet for metals at low-z (according to recent theoretical results; e.g. Oppenheimer et al., 2012).

Focusing on the second half of the history of the Universe ($z \lesssim 1$) has the advantage of allowing relatively complete galaxy surveys even at faint magnitudes ($\lesssim L^*$; elusive at higher redshifts). Faint galaxies are important for statistical analyses as they dominate the luminosity function, not just in number density, but also in total luminosity and mass. Moreover, the combined effects of structure formation, expansion of the Universe, and the reduced ionization background, allow us to observe a considerable amount of H I systems and yet resolve the so-called H I Lyα-forest into individual lines (e.g. Theuns et al., 1998a; Davé et al., 1999). This makes it possible to recover column densities and Doppler parameters through Voigt profile fitting.

One major advantage of clustering over one-to-one association analyses is that it does not impose arbitrary scales, allowing us to obtain results for both small ($\lesssim 1$ Mpc) and large scales ($\gtrsim 1 - 10$ Mpc). In this way, we can make use of *all* the H I and galaxy data available, and not only those lying close to each other. As mentioned, results from the small scale association are important to put constraints on the 'sub-grid physics' adopted in current hydrodynamical simulations. Conversely, results from the largest scales provide information unaffected by these uncertain 'sub-grid physics' assumptions (e.g. Hummels et al., 2013; Ford et al., 2013). Moreover, the physics and cosmic evolution of the diffuse IGM (traced by H I) obtained by cosmological hydrodynamical simulations (e.g. Paschos et al.,

2009; Davé et al., 2010) are in good agreement with analytic predictions (e.g. Schaye, 2001). Our results will be able to test all of these predictions.

Another advantage to using a clustering analysis is that it properly takes into account the selection functions of the surveys. Even at scales $\lesssim 300$ kpc (the typical scale adopted for the CGM), a secure or unique H I–galaxy one-to-one association is not always possible. This is because H I and galaxies are clustered at these scales and because surveys are never 100% complete. Clustering provides a proper statistical analysis, at the cost of losing details on the physics of an individual H I–galaxy pair. Thus, both one-to-one associations and clustering results are complementary, and needed, to fully understand the relationship between the IGM and galaxies.

In this Chapter we present observational results for the H I–galaxy two-point correlation function at $z \lesssim 1$. Combining data from UV HST spectroscopy of 8 QSOs in 6 different fields, with optical deep multi-object spectroscopy (MOS) surveys of galaxies around them, we have gathered a sample of 654 well identified invervening H I absorption systems and 17509 galaxies at projected separations $\lesssim 50$ Mpc from the QSO line-of-sight (LOS). This dataset is the largest sample to date for such an analysis.

Comparing the results from the H I–galaxy cross-correlation with the H I–H I and galaxy–galaxy auto-correlations, we provide constraints on their statistical connection as a function of both H I column density and galaxy star-formation activity.

This Chapter is structured as follows. Sections 4.3 and 4.4 describe the IGM and galaxy data used in this work, respectively. The IGM sample is described in Section 4.5 while the galaxy sample is described in Section 4.6. Section 4.7 describes the formalisms used to measure the H I–galaxy cross-correlation and the H I–H I and galaxy–galaxy auto-correlations. Our observational results are presented in Section 4.8 and discussed in Section 4.9. A summary is presented in Section 4.10.

All distances are in co-moving coordinates assuming $H_0 = 70$ km s^{-1} Mpc^{-1},

$\Omega_{\rm m} = 0.3$, $\Omega_\Lambda = 0.7$, $k = 0$, unless otherwise stated, where H_0, $\Omega_{\rm m}$, Ω_Λ and k are the Hubble constant, mass energy density, 'dark energy' density and spatial curvature, respectively. Our chosen cosmological parameters lie between the latest results from the Wilkinson Microwave Anisotropy Probe (Komatsu et al., 2011) and the Planck satellite (Planck Collaboration et al., 2013).

4.3 Intergalactic medium data

We used HST spectroscopy of 8 QSOs to directly characterize the diffuse IGM through the observations of intervening H I absorption line systems. We used data from COS (Green et al., 2012) taken under HST programs General Observer (GO) 12264 (PI: Morris), GO 11585 (PI: Crighton) and GO 11598 (PI: Tumlinson); and data from the Faint Object Spectrograph (FOS) (Keyes et al., 1995) taken under HST programs GO 5320 (PI: Foltz), GO 6100 (PI: Foltz) and GO 6592 (PI: Foltz).

Data from program GO 12264 were taken to study the statistical relationship between H I absorption line systems and galaxies at redshift $z \lesssim 1$. We selected four QSOs at $z_{\rm QSO} \sim 1$ (namely J020930.7-043826, J100535.24+013445.7, J135726.27+043541.4 and J221806.67+005223.6) lying in fields of view that were already surveyed for their galaxy content by the Very Large Telescope (VLT) Visible Multi-Object Spectrograph (VIMOS) Deep Survey (VVDS) (Le Fèvre et al., 2005) and the Gemini Deep Deep Survey (GDDS) (Abraham et al., 2004). Data from programs GO 5320, GO 6100, GO 6592 and GO 11585 contain spectroscopy of three QSOs (namely Q0107-025A, Q0107-025B and Q0107-0232) whose LOS are separated by $\sim 0.4 - 1$ Mpc. This triple QSO field is ideal for measuring the characteristic sizes of the H I absorption systems but it can also be used to address the connection between H I systems and galaxies (e.g. Crighton et al., 2010). Data from program GO 11598 were originally taken to investigate the properties of the CGM by targeting QSOs whose LOS lie within $\lesssim 150$ kpc of a known galaxy. In this work we used one QSO observed under program GO 11598 (namely J102218.99+013218.8), for which we have conducted our own galaxy

Table 4.1: Properties of the observed QSOs.

QSO Name	Field Name	R.A. (hr min sec)	Dec. (deg min sec)	$z_{\rm QSO}$	Magnitude Visual (Band)	NUV (AB)	FUV (AB)
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Q0107-025A	Q0107	01 10 13.10	−02 19 52.0	0.96000	18.1 (B)	18.1	19.3
Q0107-025B	Q0107	01 10 16.20	−02 18 50.0	0.95600	17.4 (V)	17.5	18.6
Q0107-0232	Q0107	01 10 14.51	−02 16 57.5	0.72600	18.4 (B)	18.9	20.1
J020930.7-043826	J0209	02 09 30.74	−04 38 26.3	1.12800	17.2 (g)	17.5	18.5
J100535.24+013445.7	J1005	10 05 35.26	+01 34 45.6	1.08090	16.8 (g)	17.4	18.6
J102218.99+013218.8	J1022	10 22 18.99	+01 32 18.8	0.78900	16.8 (V)	17.2	18.1
J135726.27+043541.4	J1357	13 57 26.27	+04 35 41.4	1.23176	17.2 (g)	17.8	19.2
J221806.67+005223.6	J2218	22 18 06.69	+00 52 23.7	1.27327	17.8 (V)	18.6	24.0[a]

(1) Name of the QSO. (2) Name of the field. (3) Right ascension (J2000). (4) Declination (J2000). (5) Redshift of the QSO. (6) Apparent visual magnitude; the band is given in parentheris. (7) Apparent near-UV magnitude from GALEX. (8) Apparent far-UV magnitude from GALEX.

[a] The sudden decrease in flux is due to the presence of a Lyman Limit System.

survey around its LOS (see Section 4.4). Given that this LOS contains only one pre-selected galaxy, this selection will not affect our results on the IGM–galaxy statistical connection.

Table 4.1 summarizes our QSO sample while Table 4.2 gives details on their HST observations.

4.3.1 Data reduction

COS data

Individual exposures from COS were downloaded from the Space Telescope Science Institute (STScI) archive and reduced using CALCOS v2.18.5 in combination with Python routines developed by the authors.[3] A full description of the reduction process will be presented in Finn et al. (2013, in prep.) and so here we present a summary.

Individual files corresponding to single central wavelength setting, stripe and FP-POS (i.e. `x1d` files) were obtained directly from CALCOS. The source extraction was performed using a box of 25 pixels wide along the spatial direction for all G130M exposures, and 20 pixels for all G160M and G230L exposures. The background extraction was performed using boxes encompassing as much of the background signal as possible, whilst avoiding regions close to the detector edges. We set the background smoothing length in CALCOS to 1 pixel and performed our own background smoothing procedure masking out portions of the spectra affected by strong geocoronal emission lines (namely the H I Lyα and O I $\lambda\lambda 1302, 1306$) and pixels with bad data quality flags[4]. We interpolated across the gaps to get the background level in these excluded regions. The background smoothing lengths were set to 1000 pixels for the far ultra-violet (FUV)A stripes, 500 pixels for the FUVB stripes and 100 pixels for all near ultra-violet (NUV) stripes, along the dispersion direction.

The error array was calculated in the same way as in CALCOS, but using

[3] Available at `https://github.com/cwfinn/COS/`

[4] `http://www.stsci.edu/hst/cos/pipeline/cos_dq_flags`

Table 4.2: Summary of the QSO observations (HST spectroscopy).

QSO Name	Instrument	Grating	Wavelength range (Å)	FWHM (Å)	Dispersion (Å/pixel)	$\langle S/N \rangle$ (per pixel)	Exposure time (h)	Program ID
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Q0107-025A	COS	G130M	1135–1460	0.07	0.01	9	7.8	11585
	COS	G160M	1460–1795	0.09	0.01	8	12.3	11585
	FOS	G190H	1795–2310	1.39	0.36	28	7.5	5320, 6592
	FOS	G270H	2310–3277	1.97	0.51	32	2.4	6100
Q0107-025B	COS	G130M	1135–1460	0.07	0.01	9	5.9	11585
	COS	G160M	1460–1795	0.09	0.01	7	5.9	11585
	FOS	G190H	1795–2310	1.39	0.36	28	1.8	5320, 6592
	FOS	G270H	2310–3277	1.97	0.51	32	1.8	6100
Q0107-0232	COS	G160M	1434[a]1795	0.09	0.01	7	23.2	11585
	FOS	G190H	1795–2310	1.39	0.36	18	9.1	11585
J020930.7-043826	COS	G130M	1277[a]1460	0.07	0.01	12	3.9	12264
	COS	G160M	1460–1795	0.09	0.01	10	7.8	12264
	COS	G230L	1795–3084	0.79	0.39	12	4.0	12264
J100535.24+013445.7	COS	G130M	1135–1460	0.07	0.01	9	3.9	12264
	COS	G160M	1460–1795	0.09	0.01	9	6.2	12264
J102218.99+013218.8	COS	G130M	1135–1460	0.07	0.01	6	0.6	11598
	COS	G160M	1460–1795	0.09	0.01	5	0.8	11598
J135726.27+043541.4	COS	G130M	1135–1460	0.07	0.01	9	3.9	12264
	COS	G160M	1460–1795	0.09	0.01	7	7.8	12264
	COS	G230L	1795–3145	0.79	0.39	11	4.0	12264
J221806.67+005223.6	COS	G230L	2097[b]3084	0.79	0.39	10	5.6	12264

(1) Name of the QSO. (2) Instrument. (3) Grating. (4) Wavelength range used for a given setting. (5) Full-width at half maximum of the line spread function of the spectrograph. (6) Dispersion. (7) Average signal-to-noise ratio per pixel over the given wavelenght range. (8) Exposure time of the observations. (9) HST program ID of the observations.

[a] Due to the presence of a Lyman Limit System blocking shorter wavelengths.

[b] Due to poor signal-to-noise data at shorter wavelengths.

our new background estimation. Each spectrum was then flux calibrated using sensitivity curves provided by STScI.

Co-alignment was performed by cross-correlating regions centred on strong Galactic absorption features (namely, C II $\lambda 1334$, Al II $\lambda 1670$, Si II $\lambda 1260$, Si II $\lambda 1526$ and Mg II $\lambda\lambda 2796, 2803$ Å). For each grating we pick the central wavelength setting and FP-POS position with the most accurately determined wavelength solutions from STScI as a reference. These are FP-POS $= 3$ for all gratings, central wavelengths of 1309 and 1600 Å for the G130M and G160M gratings respectively, and 2950 Å (using only the 'B' stripe) for the G230L grating. All other settings for each grating are then cross-correlated on these ones, assuming the reference and comparison settings both contain one of the absorption features specified. Wavelengths offsets are then applied to the comparison settings to match the reference ones. These offsets typically amount to a resolution element or less. For those settings that could not be aligned on any of the Galactic features specified, we manually searched for other strong absorption lines on which to perform the cross-correlation. Strong absorption lines were always found. We then scaled the fluxes of the comparison setting such that its median flux value matches that of either the reference or the already calibrated setting in the overlapped region.

At this point we changed some pixel values according to their quality flags: flux and error values assigned to pixels with bad data quality flags were set to zero, while pixels with warnings had their exposure times reduced by a factor of two. We then re-scaled the wavelength scale of each exposure to have a constant spacing equal to the dispersion for the grating, using nearest-neighbour interpolation. The combined wavelength scale is therefore three wavelength scales, one for the G130M grating ($\lambda < 1460$ Å), one for the G160M grating ($1460 \leq \lambda < 1795$ Å) and one for the G230L grating ($\lambda \geq 1795$Å).

The co-addition was then performed via modified exposure time weighting. Finally, the combined FUV and NUV spectra were re-binned to ensure Nyquist sampling (two pixels per resolution element). Both are binned onto a linear wavelength scale with spacing equal to 0.0395 Å for the FUV, and a spacing equal

to 0.436 Å for the NUV.

FOS data

Individual exposures from FOS were downloaded from the STScI archive and reduced using the standard CALFOS pipeline. Wavelength corrections given by Petry et al. (2006) were applied to each individual exposure. As described by Petry et al., these corrections were determined using a wavelength calibration exposure taken contemporaneously with the G190H grating science exposures, and were verified using Galactic Al II and Al III absorption features. The shortest wavelength region of the FOS G190H settings overlap with the longest wavelength COS settings, and we confirmed that the wavelength scales in these overlapping regions were consistent between the two instruments. Then we combined all individual exposures together, resampling to a common wavelength scale of 0.51 Å per pixel.

4.3.2 Continuum fitting

We fit the continuum of each QSO in a semi-automatized and iterative manner: (i) we first divide each spectrum in multiple chunks, typically of 12 Å at wavelengths shorter than that of the H I Lyα emission from the QSOs (at larger wavelengths we used much longer intervals but these are not relevant for the present work); (ii) we then fit straight line segments through the set of points given by the central wavelength and the median flux values for each chunk; (iii) we then removed pixels with flux values falling $3\times$ their uncertainty below the fit value; (iv) we repeat steps (ii) and (iii) until a convergent solution is reached; (v) we fit a cubic spline through the final set of median points to get a smooth continuum. The success of this method strongly depends on the presence of emission lines, and on number and positions of the chosen wavelength chunks. Therefore, we visually inspect the solution and improve it by adding and/or removing points accordingly, making sure that the distribution of flux values above the continuum

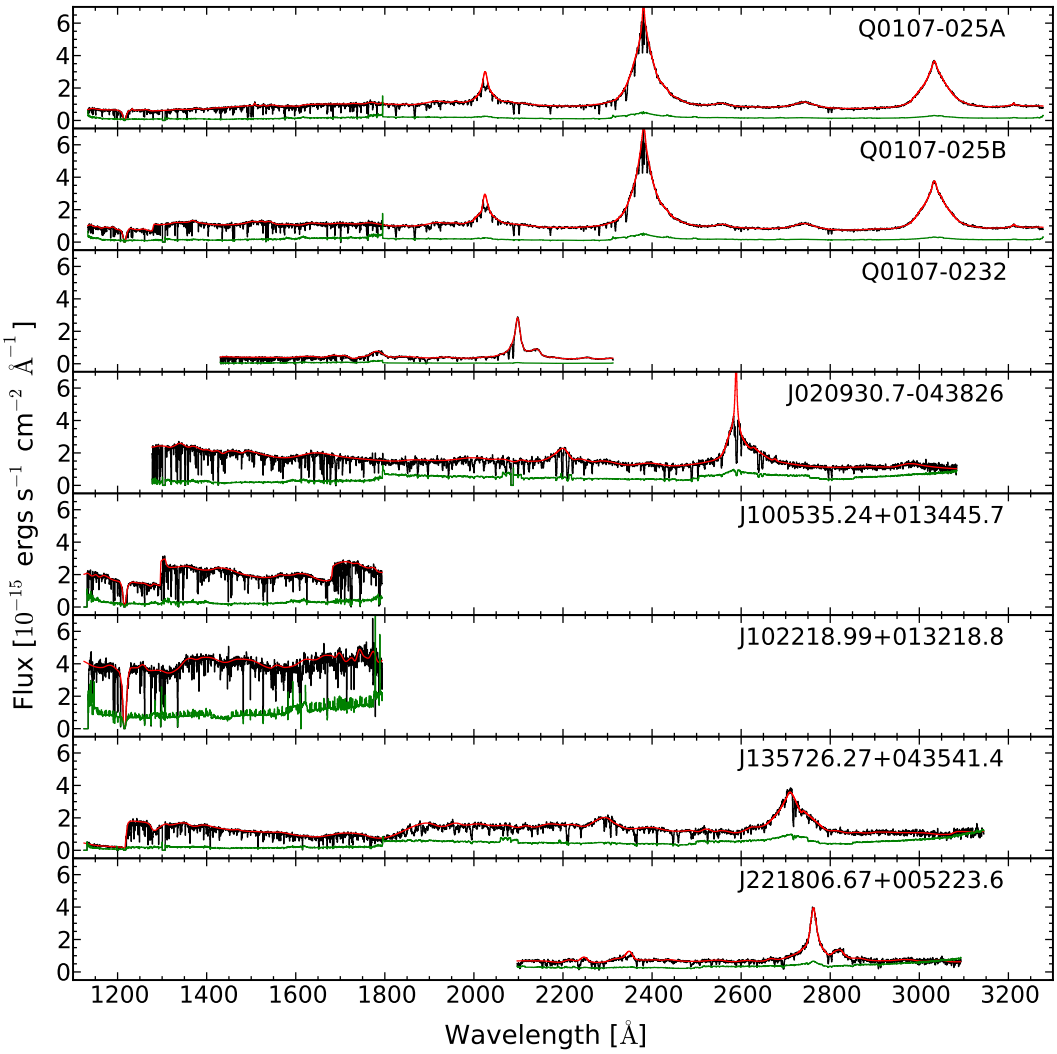

Figure 4.1: Observed spectra of our sample of QSOs: flux (black lines), uncertainty (multiplied by a factor of 5 for clarity; green lines) and continuum fit (red lines). Wavelengths $\lambda < 1795$ Å and $\lambda \geq 1795$ Å correspond to data from the FUV and NUV channels respectively (see Table 4.2). The FUV spectra have been re-binned to match the resolution of the NUV spectra for clarity.

fit is consistent with a Gaussian tail. We checked that the use of these subjective steps does not affect the final results significantly (see Section 4.5.4).

Figure 4.1 shows our QSO spectra (black lines) with their corresponding uncertainties (green lines) and continuum fit (red lines). Note the similarities between the Q0107-025A and Q0107-025B spectrum, and in some parts with that of Q0107-0232. This is due to the fact that these three QSO LOS intersect the same coherent large scale structure (LSS) (see above). This will allow us to constrain the H I–H I auto-correlation on transverse scales given by the separation between the LOS, although with the caveat that our results will come from a single field that might be affected by cosmic variance (see Section 4.8).

4.4 Galaxy data

Our chosen QSOs are at $z_{\rm QSO} \sim 0.7 - 1.3$, so we aim to target galaxies at $z \lesssim 1$, corresponding to the last ~ 7 Gyr of cosmic evolution. The majority of these QSOs lie in fields already surveyed by their galaxy content. We used archival galaxy data from: the VVDS (Le Fèvre et al., 2005), GDDS (Abraham et al., 2004) and the Canada France Hawaii Telescope (CFHT) MOS survey published by Morris and Jannuzi (2006). Despite the existence of some galaxy data around our QSO fields we have also performed our own galaxy surveys using MOS to increase the survey completeness[5]. We acquired new galaxy data from different ground-based MOS, namely: the Visible Multi-Object Spectrograph (VIMOS) (Le Fèvre et al., 2003) on the VLT under programs 086.A-0970 (PI: Crighton) and 087.A-0857 (PI: Tejos); the Deep Imaging Multi-Object Spectrograph (DEIMOS) (Faber et al., 2003) on Keck under program A290D (PIs: Bechtold and Jannuzi); and the Gemini Multi-Object Spectrograph (GMOS) (Davies et al., 1997) on Gemini under program GS-2008B-Q-50 (PI: Crighton). Table 4.3 summarizes the observations taken to construct our galaxy samples.

The following sections provide detailed descriptions of the observations, data

[5]Note that the largest of these surveys, the VVDS, has a completeness of only about $20 - 25\%$.

Table 4.3: Summary of galaxy observations (spectroscopy).

Field Name	Instrument	Grating	Wavelength range (Å)	Dispersion (Å/pixel)	Exposure time (h)	Reference
(1)	(2)	(3)	(4)	(5)	(6)	(7)
Q0107	DEIMOS	1200l/mm	6400-9100	0.3	0.99	This work
	GMOS	R400	5000-9000	0.7	0.90	This work
	VIMOS	LR_red	5500-9500	7.3	0.96	This work
	CFHT-MOS	O300	5000-9000	3.5	0.83	Morris and Jannuzi (2006)
J0209	GMOS	R150_G5306	5500-9200	3.4	21	GDDS
J1005	VIMOS	LR_red	5500-9500	7.3	0.96	This work
	VIMOS	LR_red	5500-9500	7.3	0.83	VVDS
J1022	VIMOS	LR_red	5500-9500	7.3	0.96	This work
J1357	VIMOS	LR_red	5500-9500	7.3	0.83	VVDS
J2218	VIMOS	LR_red	5500-9500	7.3	0.96	This work
	VIMOS	LR_red	5500-9500	7.3	0.83	VVDS

(1) Name of the field. (2) Instrument. (3) Grating. (4) Wavelength range. (5) Dispersion. (6) Exposure time of the observations. (7) Reference of the observations.

reduction, selection functions and construction of our new galaxy samples. We also give information on the subsamples of the previously published galaxy surveys used in this work.

4.4.1 VIMOS data

Instrument setting

We used the low-resolution (LR) grism with 1.0 arcsecond slits ($R \equiv \frac{\lambda}{\Delta\lambda} \approx 200$) due to its high multiplex factor (up to 4). As we needed to target galaxies up to the QSOs redshifts ($z_{\rm QSO} \sim 0.7 - 1.3$), we used that grism in combination with the OS-red filter giving coverage between $5500 - 9500$ Å.

Target selection, mask design and pointings

We used R-band pre-imaging to observe objects around our QSO fields and SEXTRACTOR v2.5 (Bertin and Arnouts, 1996) to identify them and assign R-band magnitudes, using zero points given by ESO. For fields J1005, J1022 and J2218 we added a constant shift of ~ 0.38 magnitudes to match those reported by the VVDS survey in objects observed by both surveys (see Section 4.4.5 and Figure 4.4). No correction was added to the Q0107 field. For objects in fields J1005, J1022 and J2218 we targeted objects at $R < 23.5$, giving priority to those with $R < 22.5$. For objects in field Q0107 we targeted objects at $R < 23$, giving priority to those with $R < 22$. We did not impose any morphological star/galaxy separation criteria, given that unresolved galaxies will look like point sources (see Section 4.6.3). The masks were designed using the VMMPS (Bottini et al., 2005) using the 'Normal Optimization' method (random) to provide a simple selection function. We targeted typically $\sim 70 - 80$ objects per mask per quadrant, equivalent to $\sim 210 - 320$ objects per pointing. We used three pointings, of one mask each, shifted by ~ 2.5 arcminutes centred around the QSO.

Data reduction for field Q0107

The spectroscopic data were taken in 2011 and the reduction was performed using VIPGI (Scodeggio et al., 2005) using standard parameters. We took three exposures per pointing of 1155 s, followed by lamps. The images were bias corrected and combined using a median filter. Wavelength calibration was performed using the lamp exposures, and further corrected using five skylines at 5892, 6300, 7859, 8347 and 8771 Å (Osterbrock et al., 1996, 1997). Finally, the slits were spectrophotometrically calibrated using standard star spectra (Oke, 1990; Hamuy et al., 1992, 1994) taken at similar dates to our observations. The extraction of the one-dimensional (1D) spectra was performed by collapsing objects along the spatial axis, following the optimal weighting algorithm presented in Horne (1986). Our wavelength solutions per slit show a quadratic mean $rms \lesssim 1$ Å in more than 75% of the slits and a $rms \lesssim 2$ Å in all the cases. We consider these as good solutions, given that the pixel size for the low resolution mode is ~ 7 Å. These data were taken before the recent update of the VIMOS charge-coupled devices (CCDs) on August 2010, and so fringing effects considerably affected the quality of the data at $\gtrsim 7500$ Å. We attempted to correct for this with no success.

Data reduction for fields J1005, J1022 and J2218

The spectroscopic data were taken on 2011 and the reduction was performed using ESOREX v.3.9.6. All three pointings of fields J1005 and J1022 were observed, while only 'pointing 3' of J2218 was observed. Due to a problem with focus, data from 'quadrant 3' of 'pointing 1' and 'pointing 3' of field J1022 were not usable. 'Pointing 2' (middle one) of fields J1005 and J1022 were observed twice to empirically asses the redshift uncertainty (see Section 4.4.1). We took three exposures per pointing of 1155 s followed by lamps. The reduction was performed using a `peakdetection` parameter (threshold for preliminary peak detection in counts) of 500 when possible, and decreasing it when needed to minimize the number of slits lost (we typically lost ~ 1 slit per quadrant). We also set the `cosmics`

parameter to 'True' (cleaning cosmic ray events) and stacked our 3 images using the median. Wavelength calibration was further corrected using four skylines at 5577.34, 6300.30, 8827.10 and 9375.36 Å (Osterbrock et al., 1996, 1997) with the `skyalign` parameter set to 1 (1st order polynomial fit to the expected positions). The slits were spectrophotometrically calibrated using standard star spectra (Oke, 1990; Hamuy et al., 1992, 1994) taken at similar dates to our observations. The extraction of the one-dimensional (1D) spectra was performed by collapsing the objects along the spatial axis, following the optimal weighting algorithm presented in Horne (1986). Our wavelength solutions per slit show a quadratic mean $rms \lesssim 1$ Å in more than 90% of the cases, which we considered as satisfactory for a pixel size of ~ 7 Å. These data were taken after a recent update to the VIMOS CCDs on August 2010, and so no important fringing effect was present.

Redshift determination

Redshifts for our new galaxy survey were measured by cross-correlating galaxy, star, and QSO templates with each observed spectrum. We used templates from the Sloan Digital Sky Survey (SDSS)[6] degraded to the lower resolution of our VIMOS observations. Galaxy templates were redshifted from $z = 0$ to $z = 2$ using intervals of $\Delta z = 0.001$. The QSO template was redshifted between $z = 0$ to $z = 4$ using larger intervals of $\Delta z = 0.01$. Star templates were shifted ± 0.005 around $z = 0$ using intervals of $\Delta z = 0.0001$ to help improve the redshift measurements and quantify the redshift uncertainty (see below). We improved the redshift solution by fitting a parabola to the 3 redshift points with the largest cross-correlation values around each local maximum. This technique gives comparable redshift solutions (within the expected errors) to that obtained by decreasing the redshift intervals by a factor ~ 10, but at a much lower computational cost. Before computing the cross-correlations, we masked out regions at the very edges of the wavelength coverage (< 5710 and > 9265 Å) and those associated with strong sky emission/absorption features (between $5870-5910$, $6275-6325$ and $7550-7720$ Å).

[6] `http://www.sdss.org/dr7/algorithms/spectemplates/`

For the Q0107 field we additionally masked out the red part at > 7550 Å because of fringing problems. We visually inspected each 1-dimensional and 2-dimensional spectrum and looked for the 'best' redshift solution (see below).

Redshift reliability

For each targeted object we manually assigned a redshift reliability flag. We used a very simple scheme based on three labels: 'a' ('secure'), 'b' ('possible') and 'c' ('no idea'). As a general rule, spectra assigned with 'a' flags have at least 3 well identified spectral features (either in emission or absorption) or 2 well identified emission lines; spectra assigned with 'c' flag are those which do not show clear spectral features either due to a low signal-to-noise ratio or because of an intrinsic lack of such lines observed at the VIMOS resolution (e.g. some possible A, F and G type stars have been included in this category); spectra assigned with 'b' flags are those that lie in between the two aforementioned categories.

Uncertainty of the semi-automatized process

The process includes subjective steps (determining the 'best' template and redshift, and assigning a redshift reliability). This uncertainty was estimated by comparing two sets of redshifts obtained independently by three of the authors (N.T. versus S.L.M. and N.T. versus N.H.M.C.) in two subsamples of the data. We found discrepancies in $\lesssim 5\%$ of the cases, the vast majority of which were for redshifts labelled as 'b'.

Further redshift calibration for fields J1005, J1022 and J2218

Even though the wavelength calibration from the ESOREX reduction was generally satisfactory, we found a ~ 1 pixel systematic discrepancy between the obtained and expected wavelength for some skylines in localized areas of the spectrum (particularly towards the red end). This effect was most noticeable in quadrant 3, where the redshift difference between objects observed twice showed a distribution displaced from zero by ~ 0.001 (~ 1 pixel). A careful inspection revealed

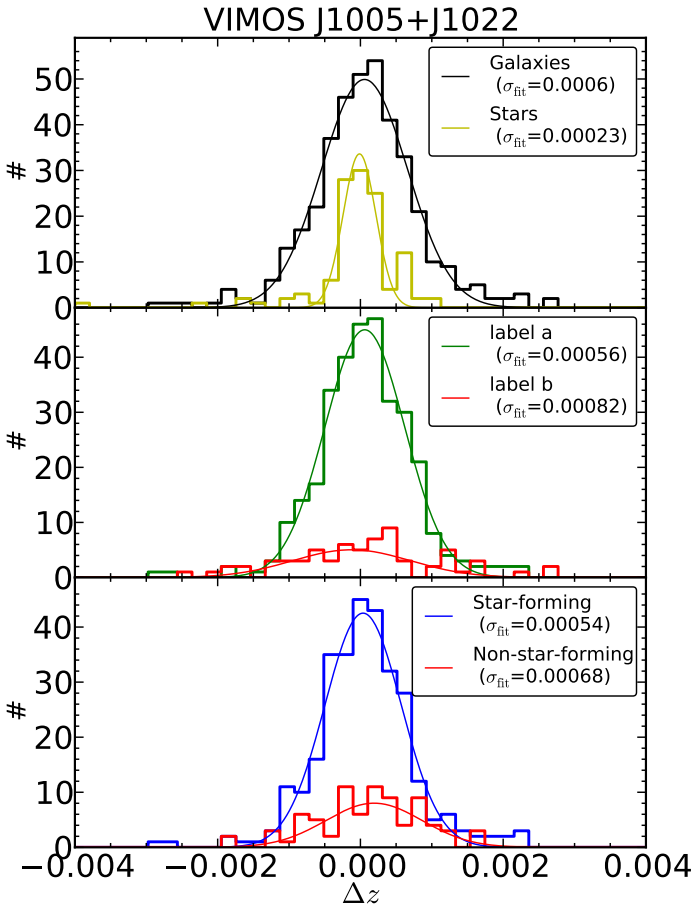

Figure 4.2: Histograms of the measured redshift difference between two independent observations of a same object in fields J1005 and J1022. Top panel shows it for all identified galaxies (blue lines) and stars (yellow lines). Middle panel shows it for galaxies with 'secure' redshifts (label 'a'; green lines) and with 'possible' redshifts (label 'b'; red lines). Bottom panel shows it for galaxies classified as 'star-forming' (blue lines) and/or 'non-star-forming' (red lines; see Section 4.6.1). Best Gaussian fits to the histograms are also shown.

that the other quadrants also showed a similar but less strong effect ($\lesssim 0.5$ pixel). We corrected for this effect using the redshift solution of the stars. For a given quadrant we looked at the mean redshift of the stars and applied a systematic shift of that amount to all the objects in that quadrant. This correction placed the mean redshift of stars at zero, and therefore corrected the redshift of all objects accordingly.

Redshift statistical uncertainty for fields J1005, J1022 and J2218

In order to assess the redshift uncertainty for these fields, we measured a redshift difference between two independent observations of the same object. These objects were observed twice, and come mainly from our 'pointing 2' in fields J1005 and J1022, but there is also a minor contribution ($\lesssim 10\%$) of objects that were observed twice using different pointings. Figure 4.2 shows the observed redshift differences for all galaxies and stars (top panel); galaxies with 'secure' and 'possible' redshifts (middle panel); and galaxies classified as 'star-forming' ('SF') or 'non-star-forming' ('non-SF') based on the presence of current, or recent, star formation (see Section 4.6.1; bottom panel). All histograms are centred around zero and do not show evident systematic biases. The redshift difference of all galaxies show a standard deviation of ≈ 0.0006. A somewhat smaller standard deviation is observed for galaxies with 'secure' redshifts and/or those classified as 'SF' (note that there is a large overlap between these two samples), and consequently a somewhat larger standard deviation is observed for galaxies with 'possible' redshift and/or classified as 'non-SF'. This behaviour is of course expected, as it is simpler to measure redshifts for galaxies with strong emission lines (for which the peak in the cross-correlation analysis is also better constrained) than for galaxies with only absorption features (at a similar signal-to-noise ratio). From this analysis we take $\approx 0.0006/\sqrt{2} = 0.0004$ as the representative redshift uncertainty of our VIMOS galaxy survey in these fields. This uncertainty corresponds to $\approx 120 - 60$ km s^{-1} at redshift $z = 0 - 1$. This uncertainty is $\gtrsim 2$ times smaller than that claimed for the VVDS survey (Le Fèvre et al., 2005).

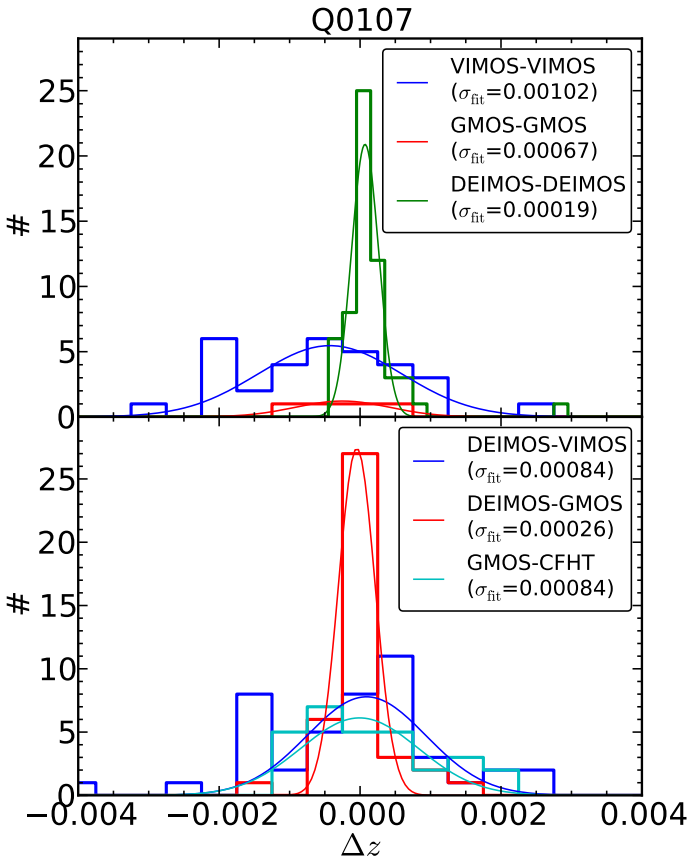

Figure 4.3: Histograms of the measured redshift difference between two independent observations of a same object in field Q0107. Top panel shows it for galaxies observed twice by the same instrument: VIMOS-VIMOS (blue lines), GMOS-GMOS (red lines) and DEIMOS-DEIMOS (green lines). Bottom panel shows it for objects observed twice by different instruments, after shifting to match the DEIMOS mean: DEIMOS-VIMOS (blue lines), DEIMOS-GMOS (red lines) and GMOS-CFHT (cyan lines). Best Gaussian fits to the histograms are also shown.

Further redshift calibration for field Q0107

We did not see systematic differences between quadrants, as was seen for fields J1005, J1022 and J2218. VIMOS observations of the Q0107 field were reduced differently, and the data come mainly from the blue part of the spectrum. Therefore, such an effect might not be present or, if present, might be more difficult to detect. However, we did find a systematic shift between the redshifts measured from VIMOS compared to those measured from DEIMOS. Given the much higher resolution of DEIMOS, we used its frame as reference for all our Q0107 observations.

Thus, we corrected the Q0107 VIMOS redshifts to match the DEIMOS frame. This correction was ~ 0.0008 ($\lesssim 1$ VIMOS pixel) and the result is shown in the bottom panel of Figure 4.3 (blue lines).

Redshift statistical uncertainty for field Q0107

In order to assess the redshift uncertainty, we used objects that were observed twice in the Q0107 field. We found a distribution of redshift differences centred at ~ 0 with a standard deviation of ≈ 0.001 (see top panel of Figure 4.3), corresponding to a single VIMOS uncertainty of $\approx 0.001/\sqrt{2} \approx 0.0007$. Another way to estimate the VIMOS uncertainty in the Q0107 field is by looking at the redshift difference for objects that were observed twice, once by VIMOS and another time by DEIMOS (44 in total; see bottom panel of Figure 4.3). In this case, the distribution shows a standard deviation of $\approx \sqrt{0.00084}$, corresponding to a single VIMOS uncertainty of $\sqrt{0.00084^2 - 0.00013^2} \approx 0.0008$, given that the uncertainty of a DEIMOS single measurement is ≈ 0.00013 (see below). So, we take a value of ≈ 0.00075 as the representative redshift uncertainty of a single VIMOS observation in the Q0107 field. This uncertainty is somewhat higher than that of fields J1005, J1022 and J2218, consistent with the poorer quality detector being used.

4.4.2 DEIMOS data

Instrument setting

We patterned our DEIMOS observations to resemble the Deep Extragalactic Evolutionary Probe 2 (DEEP2) '1 hour' survey (Coil et al., 2004). We used the 1200 line mm^{-1} grating with a 1.0 arcsecond slit giving a resolution of $R \sim 5000$ over the wavelength range $6400 - 9100$ Å.

Target selection

We used B, R and I bands pre-imaging to select objects around our Q0107 field. We used SEXTRACTOR v2.5 (Bertin and Arnouts, 1996) to identify them and assign B, R and I magnitudes to them. We used color cuts as in Coil et al. (2004, see also Newman et al. 2013) to target galaxies[7]:

$$\begin{aligned} B - R &\leq 2.35(R - I) - 0.45 \text{ or} \\ R - I &\geq 1.15 \text{ or} \\ BR &\geq 0.5 \, . \end{aligned} \tag{4.1}$$

We also gave priority to objects within 1 arcminute of the Q0107-025A LOS. We targeted objects up to $R = 24.5$ magnitudes, but we assigned higher priorities to the brightest ones. In an attempt to be efficient, we also imposed a star/galaxy morphological criteria of $\texttt{CLASS\_STAR} < 0.97$ (although see Section 4.6.3).[8]

Data reduction

The observations were taken in 2007 and 2008. The reduction was performed using the DEEP2 DEIMOS Data Pipeline[9] (Newman et al., 2013), from which galaxy redshifts were also obtained.

Redshift reliability

The redshift reliability for DEIMOS data was originally based on four subjective categories: (0) 'still needs work', (1) 'not good enough', (2) 'possible', (3) 'good' and (4) 'excellent'. In order to have a unified scheme we matched those DEIMOS labels with our previously defined VIMOS ones (see Section 4.4.1) as follows:

[7] Note that Coil et al. (2004) presented $B - R \leq 0.5$ but should have been $B - R \geq 0.5$, which is what we used.

[8] The parameter CLASS_STAR assigns a value of 1 to objects that morphologically look like stars, and a value of 0 to objects that look like galaxies. Values in between 1 and 0 are assigned for less certain objects (Bertin and Arnouts, 1996).

[9] `http://astro.berkeley.edu/~cooper/deep/spec2d/`

DEIMOS label 4 is matched to label 'a' ({4}→ {'a'}); DEIMOS labels 3 and 2 are matched to label 'b' ({3,2}→ {'b'}); and DEIMOS labels 1 and 0 are matched to label 'c' ({1,0}→ {'c'}).

Redshift statistical uncertainty for field Q0107

In order to assess the redshift uncertainty, we used objects that were observed twice in the Q0107 field. We found a distribution of redshift differences centred at ~ 0 with a standard deviation of ≈ 0.00019 (see top panel of Figure 4.3), corresponding to a single DEIMOS uncertainty of $\approx 0.00019/\sqrt{2} \approx 0.00013$. So, we take a value of ≈ 0.00013 as the representative redshift uncertainty of a single DEIMOS observation in the Q0107 field.

4.4.3 GMOS data

Instrument setting

We used the R400 grating centred on a wavelength of 7000 Å with a 1.5 arcseconds slit giving a resolution of $R = 639$.

Target selection, mask design and pointings

We used R-band pre-imaging to select objects around our Q0107 field. We used SEXTRACTOR v2.5 (Bertin and Arnouts, 1996) to identify objects and assign them R-band magnitudes. The masks were designed using GMMPS[10]. Top priority was given to objects with $R < 22$, followed by those with $22 \leq R < 23$ and last priority to those with $23 \leq R < 24$. We typically targeted ~ 40 objects per mask. Six masks were taken, three around QSO C, two around QSO B, and one around QSO A, where many objects had already been targeted in previous observations.

[10] http://www.gemini.edu/?q=node/10458

Data reduction

The observations were taken in 2008 . Three 1080 s offset science exposures were taken for each mask, dithered along the slit to cover the gaps in the CCD detectors. Arcs were taken contemporaneously to the science exposures. We used the Gemini Image Reduction and Analysis Facility (IRAF) package to reduce the spectra. A flat-field lamp exposure was divided into each bias-subtracted science exposure to remove small-scale variations across the CCDs, and the fringing pattern seen at red wavelengths. The dithered images (both arcs and science) were then combined into a single exposure. The spectrum for each mask was wavelength-calibrated by identifying known arc lines and fitting a polynomial to match pixel positions to wavelengths. Finally the wavelength-calibrated 2-d spectra were extracted to produced 1-d spectra. The typical rms scatter of the known arc line positions around the polynomial fit ranged from 0.5 to 1.0 Å, depending on how many arc lines were available to fit (bluer wavelength ranges tended to have fewer arc lines). A 0.75 Å rms scatter corresponds to a velocity error of 38 km s^{-1} at 6000 Å.

Redshift determination and reliability

We determined redshifts by using the same method to that of the VIMOS spectra: plausible redshifts were identified as peaks in the cross-correlation measured between the GMOS spectra and spectral templates (see Section 4.4.1 for further details). Redshifts reliabilities were also assigned following the definitions in our VIMOS sample.

Further redshift calibration

We found a systematic shift of the redshifts measured from GMOS with respect to those measured from DEIMOS, for the 40 objects observed by these two instruments. Given the much higher resolution of DEIMOS we used its frame as reference for our Q0107 observations. Thus, we corrected all GMOS redshifts to match the DEIMOS frame. This correction was ~ 0.0004 or ~ 80 km s^{-1} ($\lesssim 1$

GMOS pixel) and the result is shown in the bottom panel of Figure 4.3 (red lines).

Redshift statistical uncertainty for field Q0107

There were only 3 objects that were observed twice using GMOS (see top panel of Figure 4.3), and so we did not take the uncertainty from such an small sample. Instead, we use objects observed by both GMOS and DEIMOS to estimate the GMOS redshift uncertainty. The distribution of redshift differences for objects with both GMOS and DEIMOS spectra (see bottom panel of Figure 4.3) shows a standard deviation of ≈ 0.00027. Given that the uncertainty of DEIMOS alone is ≈ 0.00013 we estimate the GMOS uncertainty to be $\approx \sqrt{0.00027^2 - 0.00013^2} \approx 0.00024$.

4.4.4 CFHT MOS data

We used the CFHT galaxy survey of the Q0107 field presented by Morris and Jannuzi (2006). There are 61 galaxies in this sample, 29 of which were also observed by our GMOS survey. We use only redshift information from this sample without assigning a particular template or redshift label. We refer the reader to Morris and Jannuzi (2006) for details on the data reduction and construction of the galaxy sample.

4.4.5 VVDS

As mentioned, three of the QSOs presented in this work (namely: J100535.24+013445.7, J135726.27+043541.4 and J221806.67+005223.6) were particularly chosen because they lie in fields already surveyed for galaxies by the VVDS survey (Le Fèvre et al., 2005). For our purposes, we use a subsample of the whole VVDS survey, selecting only galaxies in those fields. We refer the reader to Le Fèvre et al. (2005) for details on the data reduction and construction of these galaxy catalogs.

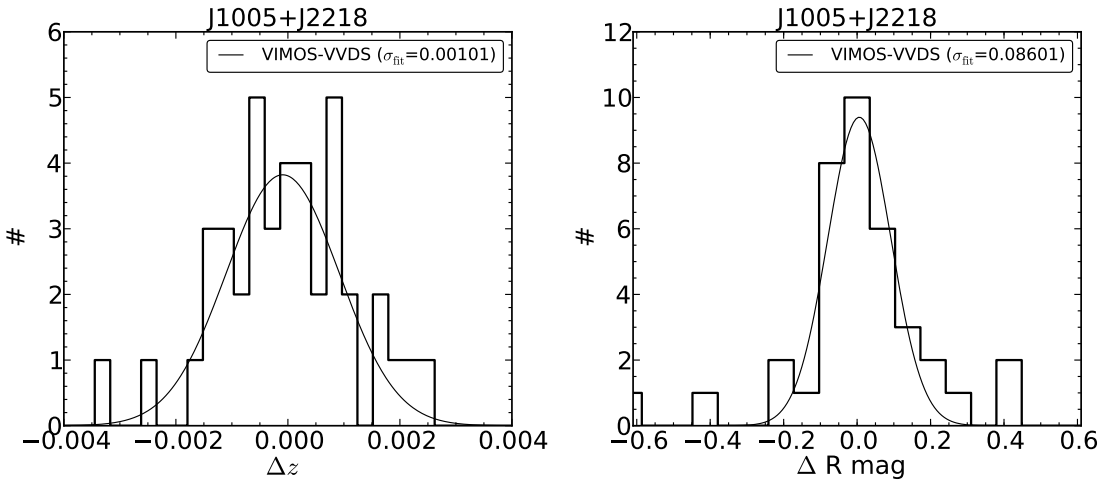

Figure 4.4: Difference in redshift (left panel) and R-band magnitude (right panel) measurements for galaxies in common between our VIMOS sample and the VVDS survey in fields J1005 and J2218. We see a good agreement in both redshift and magnitude measurements between the two surveys. The redshift difference distribution has a mean of $\lesssim 0.0001$ and a standard deviation of $\sigma_{\Delta z} \approx 0.001$, while the magnitude difference distribution has a mean of ≈ 0.006 with a standard deviation of $\sigma_{\Delta \rm R} \approx 0.09$ magnitudes. See Section 4.4.5 for further details.

Redshift reliability

The redshift reliability for VVDS data was originally based on six categories: (0) 'no redshift', (1) '50% confidence'; (2) '75% confidence'; (3) '95% confidence'; (4) '100% confidence'; (8) 'single emission line' (assumed to be [O II]); and (9) 'single isolated emission line' (Le Fèvre et al., 2005). They expanded this classification system for secondary targets (objects which are present by chance in the slits) by the use of the prefix '2'. Similarly the prefix '1' means 'primary QSO target', while the prefix '21' means 'secondary QSO target'. In order to have a unified scheme we matched those VVDS labels with our previously defined VIMOS ones (see Section 4.4.1) as follows: VVDS label 4, 3 and their corresponding extensions are matched to label 'a' ({4,14,24,214,3,13,23,213}$\rightarrow$ {'a'}); VVDS labels 2, 9 and their corresponding extensions are matched to label 'b' ({2,12,22,212,9,19,29,219}$\rightarrow$ {'b'}); and VVDS labels 1, 0 and their corresponding extensions are matched to label 'c' ({1,11,21,211,0,10,20,210}$\rightarrow$ {'c'}).

Consistency check between our VIMOS and VVDS sample

We performed a consistency check by comparing the redshifts and R-band magnitudes obtained for galaxies in common between our VIMOS sample and the VVDS survey in fields J1005 and J2218 (the only ones with such overlap). We found a good agreement in redshift measurements between the two surveys, with a mean of the distribution being ≈ 0.0003 and a standard deviation of $\sigma_{\Delta z} \approx 0.001$. This standard deviation is consistent with the quadratic sum of the typical VVDS uncertainty ($\sim 0.0013/\sqrt{2}$) and our VIMOS one ($\sim 0.0006/\sqrt{2}$), as $\sim \sqrt{0.0006^2 + 0.0013^2}/\sqrt{2} \approx 0.001$. In order to place all galaxies in a single consistent frame we shifted the VVDS redshifts by 0.0003. The left panel of Figure 4.4 shows the distribution of these redshift differences after applying the correction.

The right panel of Figure 4.4 shows the distribution of R-band magnitude differences. We also see a good agreement in the magnitude difference distribution (by construction, see Section 4.4.1), with a mean of 0.006 and a standard deviation of $\sigma_{\Delta z} \approx 0.09$. We note that this standard deviation is greater than $\sqrt{2}\times$ the typical magnitude uncertainty as given by SEXTRACTOR of ~ 0.02. Thus, we caution the reader that our reported R-band magnitude uncertainties might be underestimated by a factor of ~ 3.

4.4.6 GDDS

One of the QSOs presented in this work (namely: J020930.7-043826) was particularly chosen because it lies in a field already surveyed for galaxies by the GDDS survey. For our purposes we use a subsample of the whole GDDS survey selecting only galaxies in this field. We refer the reader to Abraham et al. (2004) for details on data reduction and construction of this galaxy catalog.

Redshift reliability

The redshift reliability for GDDS data was originally based on five subjective categories: (0) 'educated guess', (1) 'very insecure'; (2) 'reasonable secure' (two or

more spectral features); (3) 'secure' (two or more spectral features and continuum); (4) 'unquestionably correct'; (8) 'single emission line' (assumed to be [O II]); and (9) 'single emission line' (Abraham et al., 2004). In order to have a unified scheme we matched those GDDS labels with our previously defined VIMOS ones (see Section 4.4.1) as follows: GDDS label 4 and 3 are matched to label 'a' ({4,3}→ {'a'}); GDDS labels 2, 8 and 9 are matched to label 'b' ({2,8,9}→ {'b'}); and GDDS labels 1 and 0 are matched to label 'c' ({1,0}→ {'c'}).

4.5 IGM samples

4.5.1 Absorption line search

The search of absorption line systems in the continuum normalized QSO spectra was performed manually (eyeballing), based on a iterative process described as follows: (i) we first searched for all possible features (H I and metal lines) at redshift $z = 0$ and $z = z_{\rm QSO}$, and labelled them accordingly. (ii) We then searched for strong H I absorption systems, from $z = z_{\rm QSO}$ until $z = 0$, showing at least 2 transitions (e.g. Lyα and Lyβ or Lyβ and Lyγ, and so on). This last condition allowed us to identify (strong) H I systems at redshifts greater than $z > 0.477$ even for spectra without NUV coverage ($\lambda > 1795$ Å). (iii) When a H I system is found, we labelled all the Lyman series transitions accordingly and looked for possible metal transitions at the same redshift. (iv) We then performed a search for 'high-ionization' doublets (Ne VIII, O VI, N V, C IV and Si IV), from $z = z_{\rm QSO}$ until $z = 0$, independently of the presence of H I. (v) For the remaining unidentified features we assumed them to be H I Lyα and repeated step (iii). For all of the identified transitions we set initial guesses in number of velocity components, column densities and Doppler parameters, for a subsequent Voigt profile fitting.

This algorithm allowed us to identify the majority but not all the absorption line systems observed in our QSO spectral sample. The remaining unidentified features are typically very narrow and inconsistent with being H I (assuming a

minimum temperature of the diffuse IGM of $T \sim 10^4$ K, implies a $b_{\rm HI} \sim 10\,{\rm km\,s^{-1}}$; e.g. Davé et al. 2010), so we are confident that our H I sample is fairly complete.

4.5.2 Voigt profile fitting

We fit Voigt profiles to the identified absorption line systems using VPFIT[11]. We accounted for the non-Gaussian COS line spread function (LSF), by interpolating between the closest COS LSF tables provided by STScI[12] at a given wavelength. We used the guesses provided by the absorption line search (see Section 4.5.1) as the initial input of VPFIT, and modified them when needed to reach satisfactory solutions. For intervening absorption systems we kept solutions having the least number of velocity components needed to minimize the reduced χ^2.[13] For fitting H I systems, we used at least two spectral regions associated to their Lyman series transitions when the spectral coverage allowed it. This means that for H I systems showing only Lyα transition, we also included their associated Lyβ regions (even though they do not show evident absorption) when available. This last step provides confident upper limits to the column density of these systems. For strong H I systems we used regions associated to as many Lyman series transitions, but excluding regions of poor signal-to-noise ($S/N \lesssim 1$) spectra.

In the following we will present only results for H I systems; a catalog of metal systems will be published elsewhere.

4.5.3 Absorption line reliability

For each H I absorption system we assigned a reliability flag. We used a scheme based on three labels:

- *Secure ('a'):* systems at redshifts that allow the detection of either Lyα and Lyβ or Lyβ and Lyγ transitions in a given spectrum, whose $\log N_{\rm HI}$ values are greater than $30\times$ their uncertainties as quoted by VPFIT.

[11] http://www.ast.cam.ac.uk/~rfc/vpfit.html
[12] http://www.stsci.edu/hst/cos/performance/spectral_resolution
[13] Our typical reduced χ^2 values are of the order $\lesssim 1.2$.

- *Probable ('b'):* systems at redshifts that only allow the detection of the Lyα transition in a given spectrum, whose $\log N_{\rm HI}$ values are greater than $30\times$ their uncertainties as quoted by VPFIT.

- *Uncertain ('c'):* systems at any redshift, whose $\log N_{\rm HI}$ values are smaller than $30\times$ their uncertainties as quoted by VPFIT. Systems in this category will be excluded from the correlation analyses presented in this work.

4.5.4 Consistency check of subjective steps

The whole process of finding and characterizing IGM absorption lines involves subjective steps. We checked that this fact does not affect our final results by comparing redshift, column density and Doppler parameter values for H I systems obtained independently—including the continuum fitting—by two of the authors (N.T. versus C.W.F.) in the J020930.7-043826 QSO spectrum. We found values consistent with one another at the 1σ level in $\sim 90\%$ of cases for $\log N_{\rm HI}$ and $b_{\rm HI}$, and in 100% of cases for redshifts. The vast majority of discrepancies were driven by weak absorption systems close to the level of detectability, for which the differences in the continuum fitting are more important.

4.5.5 $N_{\rm HI}$ and $b_{\rm HI}$ distributions and completeness

Figure 4.5 shows the observed H I column density ($N_{\rm HI}$; left panel) and Doppler parameter ($b_{\rm HI}$; middle panel) distributions for 'secure' systems ('a' label; black solid lines), 'secure' plus 'probable' systems ('a+b' labels; dashed black lines), and 'uncertain' systems ('c' label; dotted red lines; see Section 4.5.3). We see sudden decreases in the number of systems at $N_{\rm HI} \lesssim 10^{13}$ cm^{-2} and $b_{\rm HI} \lesssim 10$ km s^{-1}, which indicate the observational completeness limits of our sample and/or our selection (shown as grey shaded areas in Figure 4.5).

Theoretical results point out that the H I column density distribution is well described by a power law of the form $f(N_{\rm HI}) \propto N_{\rm HI}^{-\beta}$ with $\beta \sim -1.7 - 1.8$, extending significantly below $\sim 10^{13}$ cm^{-2} (e.g. Paschos et al., 2009; Davé et al.,

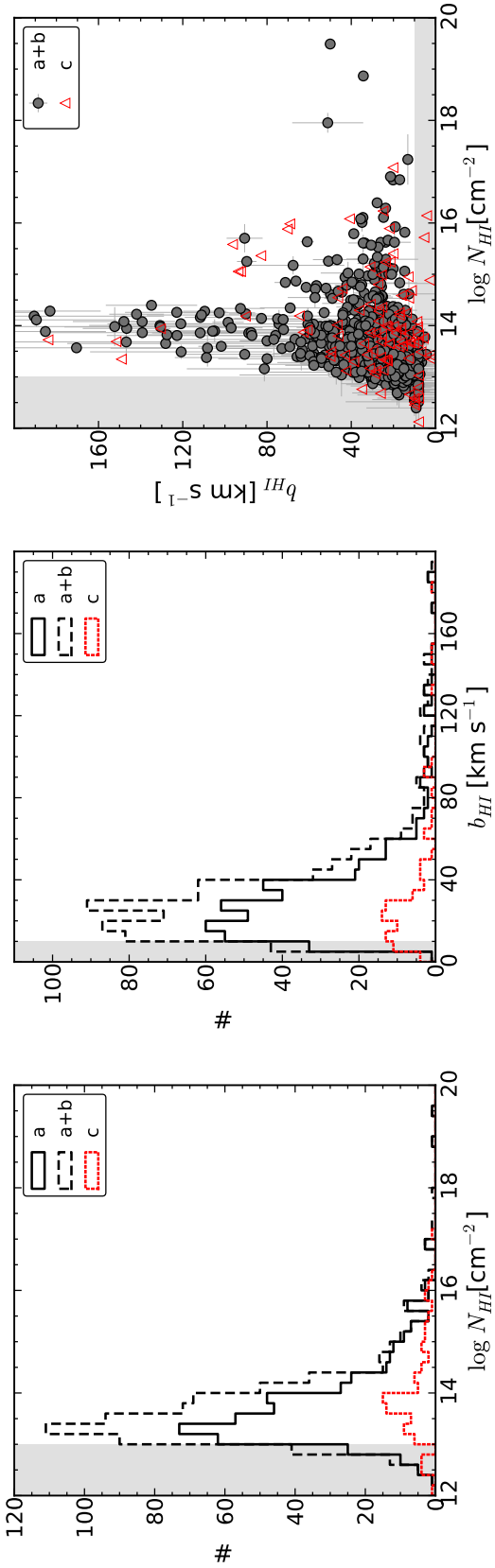

Figure 4.5: The first two panels show the observed H I column density ($N_{\rm HI}$; left panel) and Doppler parameter ($b_{\rm HI}$; middle panel) distributions for 'secure' systems ('a' label; black solid lines), 'secure' plus 'probable' systems ('a+b' labels; dashed black lines), and 'uncertain' systems ('c' label; dotted red lines; see Section 4.5.3 for definitions of these labels). The right panel shows the distribution of Doppler parameters as a function of column density for 'secure' plus 'probable' systems ('a+b' labels; grey circles), and 'uncertain' systems ('c' label; red open triangles; uncertainties not shown). Grey shaded areas show regions with low completeness levels. For further details see Section 4.5.5.

2010; Tepper-García et al., 2012). This has been observationally confirmed from higher signal-to-noise ratio data ($S/N \sim 20 - 40$) at least down to $N_{\rm HI} \sim 10^{12.3}$ cm^{-2} (Williger et al., 2010). Our current $N_{\rm HI}$ completeness limit is therefore not physical, and driven by the signal-to-noise ratio of our sample. Indeed, using the results from Keeney et al. (2012), the expected minimum rest-frame equivalent width for H I lines detected in the FUV-COS—in which the majority of weak lines are detected—at the 3σ confidence level, for our typical signal-to-noise ratio ($S/N \sim 10$; see Table 4.2), is ~ 40 mÅ. This limit corresponds to $N_{\rm HI} \sim 10^{13}$ cm^{-2} for a typical Doppler parameter of $b_{\rm HI} \sim 30$ km s^{-1}, which is consistent with what we observe.

The same theoretical results point out that the H I Doppler parameter distribution for the diffuse IGM peaks at $\sim 20 - 40$ km s^{-1}, with almost negligible contribution of lines with $b_{\rm HI} < 10$ km s^{-1} (Paschos et al., 2009; Davé et al., 2010; Tepper-García et al., 2012). Given that the FUV-COS and FOS data have spectral resolutions of about ~ 16 km s^{-1}, these samples should include the vast majority of real H I systems at $N_{\rm HI} \gtrsim 10^{13}$ cm^{-2}. On the other hand, the NUV-COS data (3 QSOs; see Table 4.2) have spectral resolutions of about ~ 100 km s^{-1}, which introduces some unresolved unphysically broad lines. Unresolved blended systems also add some unphysically broad lines in all our data. This observational effect explains, in part, the tail at large $b_{\rm HI}$ (see middle panel of Figure 4.5). We note that very broad lines can also be explained by physical mechanisms, such as temperature, turbulence, Jeans smoothing and Hubble flow broadenings (e.g. Rutledge, 1998; Hui and Rutledge, 1999; Davé et al., 2010; Tepper-García et al., 2012). There are a total of $58/766 \sim 8\%$ of systems with $b_{\rm HI} \geq 80$ km s^{-1}. Such a small fraction should not affect our results significantly.

We also note that the typical $b_{\rm HI}$ uncertainties are of the order of ~ 10 km s^{-1}, and so scatter of similar amount should be expected in the $b_{\rm HI}$ distributions. This explains the presence of lines with $b_{\rm HI} \lesssim 10$ km s^{-1}, all of which are consistent with 10 km s^{-1} within the errors. However, as we do not use the actual $b_{\rm HI}$ values in any further analysis, this uncertainty does not affect our results.

The right panel of Figure 4.5 shows the distribution of $b_{\rm HI}$ as a function of $\log N_{\rm HI}$ for 'secure' plus 'probable' systems ('a+b' labels; grey circles), and 'uncertain' systems ('c' label; red open triangles; uncertainties not shown). We see that there are not strong correlations between these values, apart from the upper and lower $b_{\rm HI}$ envelopes. The upper envelope is consistent with an observational effect, as higher $N_{\rm HI}$ values are required to observe lines with larger $b_{\rm HI}$, for a fixed signal-to-noise ratio (e.g. Paschos et al., 2009; Williger et al., 2010). The lower envelope is consistent with a physical effect, driven by the temperature-density relation of the diffuse IGM: H I systems with larger $N_{\rm HI}$ probe, on average, denser regions for which the temperature—a component of the $b_{\rm HI}$—is also, on average, larger (e.g. Hui and Gnedin, 1997; Paschos et al., 2009; Davé et al., 2010; Tepper-García et al., 2012). A proper analysis of these two effects is beyond the scope of the present work.

4.5.6 Column density classification

One of our goals is to test whether the cross-correlation between H I absorption systems and galaxies depends on the H I column density. To do so, we split our H I sample into subcategories based on a column density limit. We define 'strong' systems as those with column densities $N_{\rm HI} \geq 10^{14}$ cm^{-2}, and 'weak' systems as those with $N_{\rm HI} < 10^{14}$ cm^{-2}. The transition column density of 10^{14} cm^{-2} is somewhat arbitrary but was chosen such that: (i) the H I–galaxy cross-correlation for 'strong' systems and the galaxy–galaxy auto-correlation have similar amplitudes; and (ii) the 'strong' systems sample is large enough to measure the cross-correlation at relatively high significance. A larger column density limit (e.g. $\sim 10^{15-16}$ cm^{-2}) does indeed give a stronger H I–galaxy clustering amplitude, but it also increases the noise of the measurement.

We note that there might not necessarily be a physical mechanism providing a sharp $N_{\rm HI}$ limit. However, recent theoretical results (e.g. Davé et al., 2010) suggest that there might still be a physical meaning for such a column density limit. We will discuss more on this issue in Section 4.9.2.

Table 4.4: Summary of the H I survey used in this work.[a]

	Secure ('a')	Probable ('b')	Uncertain ('c')	Total
Q0107-025A				
H I	76	29	15	120
Strong	26	1	10	37
Weak	50	28	5	83
Q0107-025B				
H I	45	6	16	67
Strong	22	1	2	25
Weak	23	5	14	42
Q0107-0232				
H I	26	20	4	50
Strong	19	6	0	25
Weak	7	14	4	25
J020930.7-043826				
H I	74	60	22	156
Strong	17	10	6	33
Weak	57	50	16	123
J100535.24+013445.7				
H I	70	61	8	139
Strong	9	8	5	22
Weak	61	53	3	117
J102218.99+013218.8				
H I	50	10	6	66
Strong	5	5	0	10
Weak	45	5	6	56
J135726.27+043541.4				
H I	86	46	10	142
Strong	23	9	4	36
Weak	63	37	6	106
J221806.67+005223.6				
H I	5	12	9	26
Strong	5	8	9	22
Weak	0	4	0	4
Total				
H I	453	216	97	766
Strong	126	47	37	210
Weak	327	169	60	556

[a] See Section 4.5.3 and Section 4.5.6 for definitions.

4.5.7 Summary

Our IGM data is composed of HST data from the COS and FOS instruments taken on 8 different QSOs (see Tables 4.1 and 4.2). We have split our H I absorption line system sample into 'strong' and 'weak' based on a column density limit of 10^{14} cm^{-2}. Our survey is composed by a total of 669 well identified H I systems with $N \gtrsim 10^{13}$ cm^{-2}.[14] Table 4.4 shows a summary of our H I survey. Tables A.1 to A.8 present the survey in detail.

4.6 Galaxy samples

In this section we describe our galaxy samples. In the following, we will refer to our new galaxy surveys in terms of the instrument used (either VIMOS, DEIMOS and GMOS), to distinguish them from the VVDS or GDDS surveys.

4.6.1 Spectral type classification

One of our goals is to test whether the cross-correlation between H I absorption systems and galaxies depends on the galaxy spectral type (either absorption or emission line dominated; e.g. Chen et al., 2005; Chen and Mulchaey, 2009). To do so we need to classify our galaxy sample accordingly.

We took a conservative approach by considering only two galaxy subsamples: those which have not undergone important star formation activity over their past ~ 1 Gyr and those which have. In terms of their spectral properties the former type has to show strong a D4000 break and no significant emission lines (including Hα and [O II]). The later type are the complementary galaxies, i.e. those with measurable emission lines. We therefore name these subsamples as 'non-star-forming' ('non-SF') and 'star-forming' ('SF') galaxies respectively, deliberately avoiding the misleading terminology of 'early' and 'late' types. Summarizing,

[14] Note that only 654 of those are used in our correlation analysis (see Section 4.8).

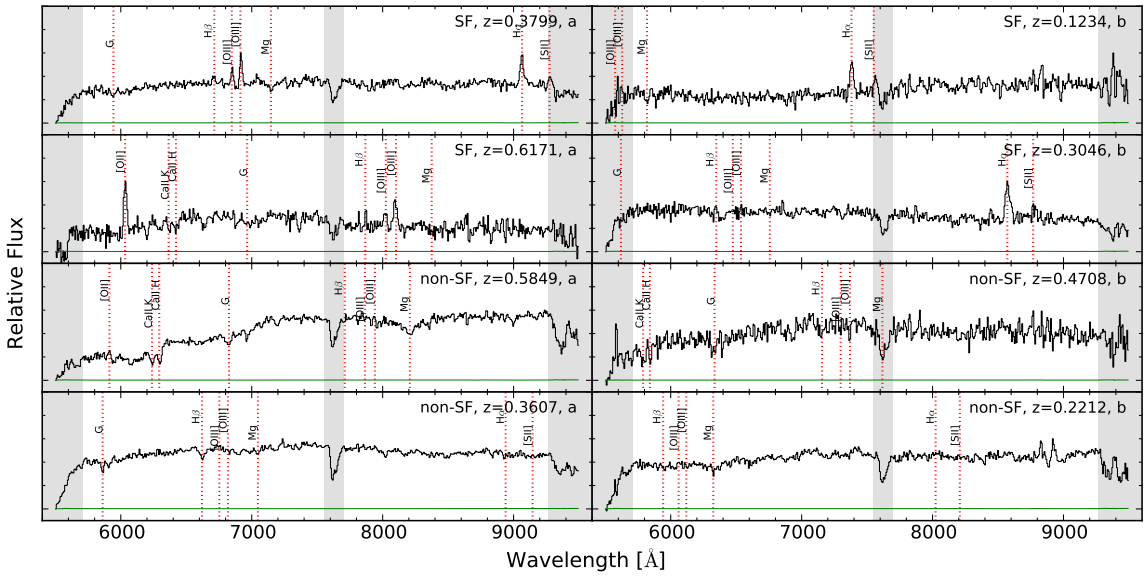

Figure 4.6: Examples of galaxy spectra taken with VIMOS (black lines) and its uncertainties (green lines). The left panels show spectra with 'secure' redshifts ('a' labels) while the right panels show spectra with 'possible' redshifts ('b' labels). The top four panels show examples of 'SF' galaxies while the bottom four panels show examples of 'non-SF' galaxies. Grey shaded areas show regions affected by poor sensitivity (edges) and by telluric absorption (middle) excluded from the redshift determination process. Red dotted lines show the position of some spectral features for each galaxy spectrum.

- *Non-star-forming galaxies ('non-SF'):* those galaxies which show no measurable star formation activity over the past $\gtrsim 1$ Gyr (e.g. Early, Bulge, Elliptical, Red Luminous Galaxy and S0 templates).

- *Star-forming galaxies ('SF'):* those galaxies which show evidence of current or recent ($\lesssim 1$ Gyr) star formation activity (e.g. Late, Sa, Sb, Sc, SBa, SBb, SBc and Starburst templates).

We note that we are not classifying galaxies on morphology, even though the template names might suggest that. Our classification is based solely on the presence or absence of spectral features associated with star-formation activity. As an example, Figure 4.6 shows 8 galaxies with a variety of signal-to-noise ratios, redshifts, redshift reliabilities, and spectral classifications.

This template matching scheme was used only for our VIMOS and GMOS galaxies because in both the redshifts were determined using template matches.

For the rest of our data we used different approaches, described in the following sections.

DEIMOS data

The DEIMOS reduction pipeline provides three weights from a principal component analysis: w_1 ('absorption-like'), w_2 ('emission-like') and w_3 ('star-like'). Thus, for DEIMOS data we use these weights to define star-forming and non-star-forming galaxies as follows: if $\max(fw_1, w_2) = fw_1$ we assigned that object to be a 'non-SF' galaxy; if $\max(fw_1, w_2) = w_2$ we assigned that object to be a 'SF' galaxy; and if $z < 0.005$ we assigned that object to be a 'star' (this last condition takes precedence over the previous ones). We used $f = 0.2$ to be conservative in the definition of 'non-SF' galaxies. This value also minimizes the 'uncertain-identification rate' in field Q0107 (see below). We did not use the information provided by w_3 because we found 7 objects with $z > 0.005$ (galaxies) showing $\max(w_1, w_2, w_3) = w_3$, probably because of their low signal-to-noise spectra.

CFHT data

In the case of the CFHT survey, we did not performe a spectral type split, and so we will only use these galaxies for results involving the whole galaxy population. We note that there is a large overlap between our GMOS and the CFHT samples and that the CFHT sample is comparatively small (61 galaxies). Thus, this choice does not compromise our analysis.

VVDS data

In the case of the VVDS survey we used a colour cut to split the sample into red and blue galaxies. We chose this approach because the current VVDS survey does not provide spectral classification for galaxies in the fields used in this work. We used a single color limit of $B - R = 2.15$ (no 'k-correction' applied[15]) to split

[15]If we knew the spectral type of the galaxies we would have not required the color split in the first place.

our sample. Thus, galaxies with $B - R < 2.15$ were assigned to our 'SF' sample, whereas those with $B - R \geq 2.15$ were assigned to our 'non-SF' sample. We chose this limit as it gives the same proportion of 'non-SF'/'SF' galaxies as in the rest of our sample. Objects with no $B - R$ color measurement were left out of this classification, and so these will only contribute to the results involving the whole galaxy population.

GDDS data

The GDDS survey provides spectral classification based on three binary digits, each one referring to 'young' ('100'), 'intermediate-age' ('010') and 'old' ('001') stellar populations (Abraham et al., 2004). The GDDS spectral classification also allowed for objects dominated by one or more types, so '101' could mean that the object has strong D4000 break and yet some strong emission lines. In order to match GDDS galaxies to our spectral classification we proceeded in the following way. Galaxies classified as 'old' were matched to our 'non-SF' sample ({'001'}$\rightarrow$ {'non-SF' }); and galaxies classified as not being 'old' where matched to our 'SF' sample ({$\neq$'001'}$\rightarrow$ {'SF' }).

Uncertainty in the spectral classification scheme

We quantified the uncertainty in this spectral classification by looking at the 'uncertain-classification rate', i.e. the fraction of (duplicate) galaxies that were not consistently classified as either 'SF' or 'non-SF' over the total number of (duplicate) galaxies. For fields J1005, J1022 and J2218 this uncertainly-classification rate corresponds to $11/667 \sim 2\%$. None of these uncertainly-classified galaxies show redshift differences $\gtrsim 0.005$ (catastrophic). For the Q0107 field this uncertain-classification rate corresponds to $25/280 \sim 9\%$. From these, $4/25$ show redshift differences $\gtrsim 0.005$, all of which are galaxies labelled as 'b' ('possible'); and $19/25$ were driven by observations using different instruments. The higher uncertain-identification rate for Q0107 is therefore mostly driven by the inhomogeneity of our samples.

For fields J1005 and J2218 we also checked whether the color cut limit used to split the VVDS sample (see Section 4.6.1) gives consistency with the actual spectral classification of our VIMOS sample, for common objects observed by these two surveys. In this case, the uncertain-classification rate corresponds to $2/40 \sim 5\%$, all of which were conservative in the sense that the VVDS classification (uncertain) was 'SF' whereas the VIMOS one (reliable) was 'non-SF'.

4.6.2 Treatment of duplicates

For objects observed with different instruments and/or showing different redshift confidences, we combined their redshift information considering the following priorities:

- *Redshift label priority:* we gave primary priority to redshifts labelled as 'a', 'b' and 'c', in that order.

- *Instrument priority:* we gave secondary priority to redshifts measured with DEIMOS, GMOS, VIMOS and CFHT, in that order. We based this choice on spectral resolution.

We therefore chose the redshift given by the highest priority and took the average when 2 or more observations had equivalent priorities. The spectral classification of uncertainly-classified objects (i.e., being classified as both 'SF' *and* 'non-SF') was set to be 'SF', ensuring a conservative 'non-SF' classification.

4.6.3 Star/galaxy morphological separation

Our DEIMOS observations deliberately avoided star-like (unresolved) objects, based on the `CLASS_STAR` parameter provided by SEXTRACTOR (Section 4.4.2). We found that this selection misses a number of faint, unresolved galaxies and so it might introduce an undesirable bias selection (see also Prochaska et al., 2011a). This motivated our subsequent VIMOS and GMOS selection, for which

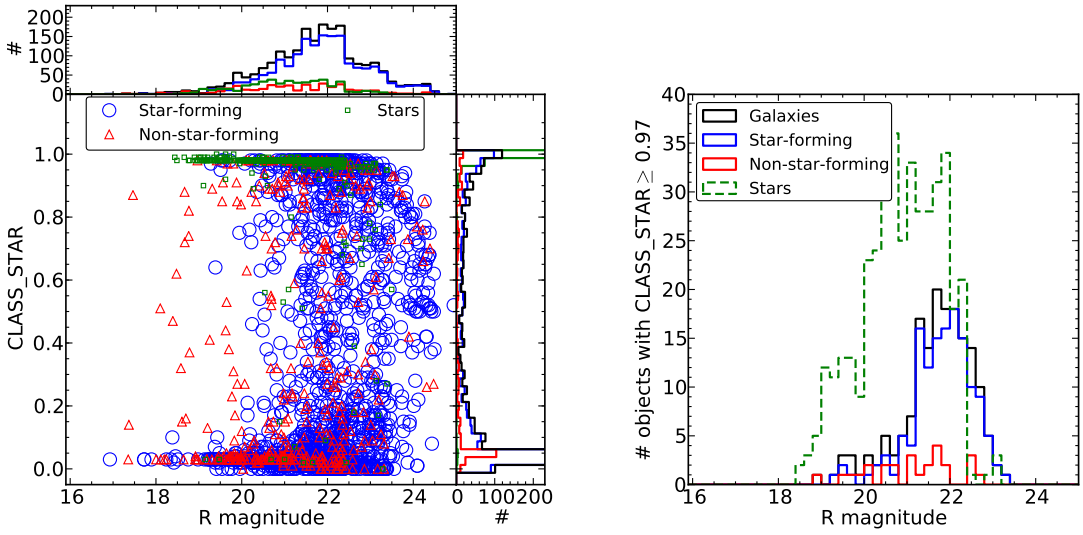

Figure 4.7: Left: SEXTRACTOR CLASS_STAR as a function of R-band magnitude for objects with spectroscopic redshifts: 'SF' galaxies (big blue open circles), 'non-SF' galaxies (small red open triangles) and stars (small green squares). Histograms are shown around the main panel truncated at 230 counts. The sudden decreases of objects at $R \sim 22$ and $R \sim 23$ are due to our target selection (see Section 4.4). Right: Histogram of objects with CLASS_STAR ≥ 0.97: all galaxies (solid black), 'SF' galaxies (solid blue), 'non-SF' galaxies (solid red) and stars (dashed green). We see a significant number of unresolved galaxies at $R \gtrsim 21$ mag (see Section 4.6.3 for further discussion).

no morphological criteria was imposed (see Section 4.4.1 and Section 4.4.3). Here we summarize our findings regarding this issue.

The left panel of Figure 4.7 shows CLASS_STAR values as a function of R-band magnitude for objects with spectroscopic redshifts: 'SF' galaxies (big blue open circles), 'non-SF' galaxies (small red open triangles) and stars (small green squares). The sudden decrease of objects at $R \sim 22.5$, $R \sim 23.5$ and $R \sim 24.5$ magnitudes are due to our target selection (see Section 4.4). The fraction of 'non-SF' with respect to 'SF' galaxies is higher at brighter magnitudes (see Section 4.6.4). We see a bimodal distribution of objects having CLASS_STAR ~ 0 (resolved) and CLASS_STAR ~ 1 (unresolved). The vast majority of resolved objects are galaxies but some stars also fall in this category due to the non-uniform point spread function (PSF) that varies across the imaging field of view. On the other hand, the vast majority of bright unresolved objects are stars, but a significant fraction

of faint ones are galaxies. The right panel of Figure 4.7 shows a histogram of objects with CLASS_STAR ≥ 0.97 as a function of R-band magnitude. Such objects are typically excluded from galaxy spectroscopic surveys. We find unresolved galaxies over a wide range of magnitudes, but more importantly at $R \gtrsim 21$. At $R \gtrsim 22$ unresolved galaxies dominate over stars, and so a CLASS_STAR < 0.97 criteria indeed introduces an undesirable selection bias. Even at magnitudes brighter than $R \sim 21$, where the fraction of unresolved galaxies is small, this morphological bias is still undesirable for galaxy-absorber direct association studies. In our survey, 2(7) out of 33(82) $R \leq 21$ ($R \leq 24$) unresolved galaxies lie at ≤ 300 kpc (physical) from a QSO LOS which might have been left out based on a morphological selection. As mentioned, our DEIMOS survey is indeed affected by this selection effect, but our VIMOS and GMOS surveys are not, which allowed us to overcome this potential problem in all our fields, including Q0107.

Neither the VVDS nor the GDDS data are affected in this way. The VVDS survey targeted objects based only on magnitude limits, while the GDDS survey used photometric redshifts to avoid low-z galaxies, with no morphological criteria imposed.

4.6.4 Completeness

The completeness of a survey is defined as the fraction of detected objects with respect to the total number of objects that could be observed. In the case of our galaxy survey the completeness can be decomposed in: (i) the fraction of objects with successful redshift determination with respect to the total number of targeted objects; (ii) the fraction of targeted objects with respect to the total number of objects detected by SEXTRACTOR; and (iii) the fraction of objects detected by SEXTRACTOR with respect to the total number of objects that could be observed. In the following we will focus only on the first of these terms for our new galaxy data. For the completeness of VVDS, GDDS and CFHT surveys we refer the reader to Le Fèvre et al. (2005), Abraham et al. (2004) and Morris and Jannuzi (2006) respectively.

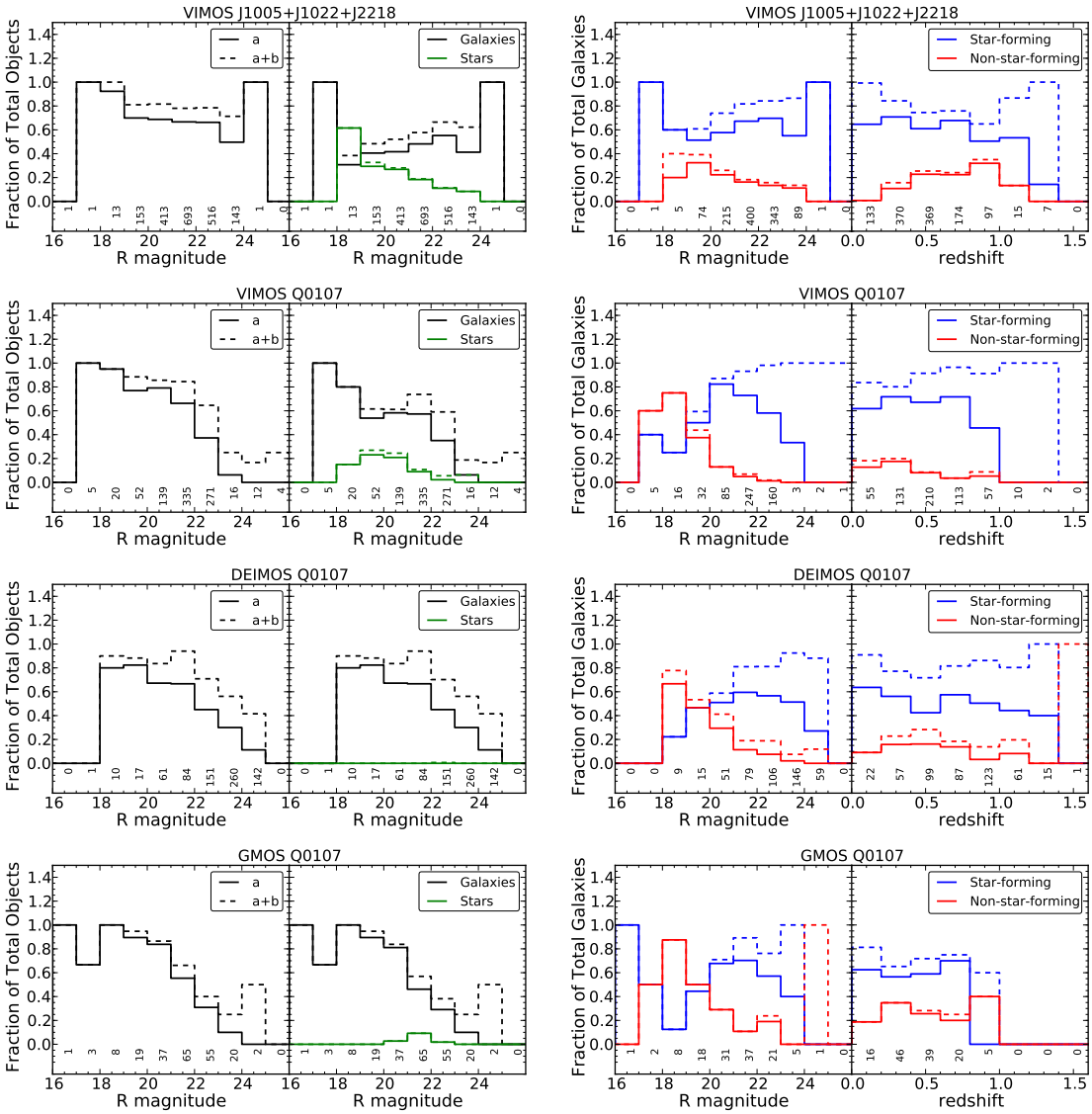

Figure 4.8: Success rate of assigning redshift for our new galaxy surveys. From top to bottom: VIMOS (J1005, J1022 and J2218); VIMOS (Q0107); DEIMOS (Q0107); and GMOS (Q0107). The left and second-left panels show the fraction of targeted objects with assigned redshift and the fraction of those that were identified as galaxies (black lines) and/or stars (green lines), as function of apparent R-band magnitude, respectively. The second-right and right panels show the fraction of galaxies that were classified as 'star-forming' (blue lines) and/or 'non-star-forming' (red lines), as a function of R-band magnitude and redshift, respectively. All these fractions are shown for both objects with high ('a' label; solid lines) and/or any ('a+b' label; dashed lines) redshift confidence. The number of objects corresponding to a fraction of 1 (total) are labeled at the bottom of each bin. See Section 4.6.4 for further discussion.

Figure 4.8 shows the success rate of assigning redshifts as a function of R-band apparent magnitude, for all objects (left panel) and for galaxies and/or stars (second-left panel). We present them separately for each of our new galaxy surveys because of their different selection functions. From top to bottom: VIMOS (J1005, J1022 and J2218), VIMOS (Q0107), DEIMOS (Q0107), and GMOS (Q0107). All of these fractions are computed for objects whose redshifts have been measured at high (label 'a', solid lines) and/or any confidence (label 'a+b', dashed lines). We see that our surveys have a $\sim 70-90\%$ success rate for objects with $R \lesssim 22$ magnitudes, and a $\lesssim 40\%$ success rate for objects with $22 \lesssim R \lesssim 24$, except for our VIMOS survey of fields J1005, J1022 and J2218, which shows a $\sim 70-90\%$ success rate even for faint objects. As mentioned in Section 4.4 our VIMOS, GMOS and DEIMOS surveys were limited at $R = 23-23.5$, $R = 24$, $R = 24.5$ respectively, and so the small contribution of objects fainter than those limits correspond to untargeted objects that happened to lie within the slits. These objects correspond to a very small fraction of the total, and so we left them in. The higher success rate for brighter objects is expected given the higher signal-to-noise ratio of those spectra. For objects brighter than $R \sim 22$ magnitudes, the fraction of identified galaxies is $\gtrsim 50\%$, and the fraction of identified stars varies: from $\sim 0\%$ in our DEIMOS survey (by construction; see Section 4.4.2), $\lesssim 10\%$ in our GMOS survey, to $\sim 20-10\%$ in our VIMOS surveys. The fraction of identified galaxies and stars at fainter magnitudes is $\lesssim 50\%$ and $\lesssim 10\%$ respectively.

Figure 4.8 also shows how the galaxy completeness depends on our galaxy spectral type classification (see Section 4.6.1). The second-right and right panels show the fraction of galaxies classified as 'SF' (blue lines) and/or 'non-SF' (red lines) over the total number of galaxies as a function of R-band magnitude and redshift respectively. Excluding magnitude bins with < 10 galaxies, we see that the fraction of 'non-SF' galaxies decreases with R-band apparent luminosity, consistent with the higher signal-to-noise ratio spectra for the brighter objects. The fraction of 'SF' galaxies shows a flatter behavior because the redshift determination depends more on the signal-to-noise of the emission lines than the

signal-to-noise of the continuum. The fraction of 'non-SF' galaxies dominates over 'SF' ones at $R \lesssim 19$ (see also left panel Figure 4.7), with a contribution of $\sim 50 - 70\%$, although these bins have typically < 20 objects. At fainter magnitudes ($R \gtrsim 20$), 'SF' galaxies dominate over 'non-SF' ones with a contribution of $\sim 60 - 90\%$. Despite these magnitude trends, we see that our galaxy sample is dominated by the 'SF' type over the whole redshift range (except for the one galaxy observed at $z > 1.4$ in the DEIMOS survey), as might have been expected from our conservative spectral classification (Section 4.6.1). 'SF' ('non-SF') galaxies account for $\sim 60 - 80\%$ ($\sim 20 - 30\%$) of the total galaxy fraction at $z \lesssim 1$, with a mild decrease (increase) with redshift. This redshift trend is most apparent in our VIMOS survey of fields J1005, J1022 and J2218, which we explain as follows. The D4000 Å break becomes visible at 5500 Å for redshifts ~ 0.4 and moves towards wavelength ranges of higher spectral quality ($\sim 6000 - 7500$ Å) at $z \sim 0.7 - 0.9$. Simultaneously, Hα and [O III] emission lines are shifted towards poor quality spectral ranges ($\gtrsim 8000$ Å; due to the presence of sky emission lines) at $z \sim 0.2$ and $z \sim 0.6$, and are out of range at $z \gtrsim 0.4$ and $z \gtrsim 0.8$ respectively. At $z \gtrsim 1$ the only emission line available is [O II] which explains the rise in the fraction of low redshift confidence ('b' labels) 'SF' galaxies.

4.6.5 Summary

Our galaxy data is composed of a heterogeneous sample obtained by 4 different instruments (see Table 4.3), taken around 8 different QSO LOS in 6 different fields (see Figure 4.9 and Table 4.1). For fields with observations from more than one instrument, we have made sure that the redshift frames are all consistent. We have also split the galaxies into 'star-forming' ('SF') and 'non-star-forming' ('non-SF'), based on either spectral type (for those lying close to the QSO LOS, i.e., VIMOS, DEIMOS, GMOS and GDDS samples) or color (VVDS sample). Table 4.5 shows a summary of our galaxy survey. Tables A.9 to A.12 present our new galaxy survey in detail. We refer the reader to Le Fèvre et al. (2005), Abraham et al. (2004) and Morris and Jannuzi (2006) for retrieving the VVDS, GDDS and CFHT data

Table 4.5: Summary of the galaxy surveys used in this work.

	Secure ('a')	Possible ('b')	No idea ('c')	Undefined ('n')	Total
Our new survey					
Galaxies	1634	509	0	0	2143
'SF'	1336	441	0	0	1777
'non-SF'	298	68	0	0	366
Stars	451	42	0	0	493
AGN	2	20	0	0	22
Unknown	0	0	893	0	893
GGDS survey[a]					
Galaxies	41	12	0	0	53
'SF'	32	11	0	0	43
'non-SF'	9	1	0	0	10
Stars	1	0	0	0	1
AGN	1	0	0	0	1
Unknown	0	0	5	0	5
VVDS survey[b]					
Galaxies	9458	7903	0	0	17361
'SF'	3766	3179	0	0	6945
'non-SF'	789	639	0	0	1428
Stars	1	2	0	0	3
AGN	138	131	0	0	269
Unknown	0	0	8394	0	8394
CFHT survey					
Galaxies	0	0	0	31	31
Total					
Galaxies	11133	8424	0	31	19588
'SF'	5134	3631	0	0	8765
'non-SF'	1096	708	0	0	1804
Stars	453	44	0	0	497
AGN	141	151	0	0	292
Unknown	0	0	9292	0	9292

[a] Only objects in field J0209.
[b] Only objects in fields J1005, J1357 and J2218.

respectively.

Our final dataset comprises 19588 (11133) galaxies with good (excellent) spectroscopic redshifts at $z \lesssim 1$ around QSO LOS with 669 (453) good (excellent) H I absorption line systems. This is by far the largest sample suitable for a statistical analysis on the IGM–galaxy connection to date.

4.7 Correlation analysis

The main goal of this work is to address the connection between the IGM traced by H I absorption systems and galaxies in a statistical manner. To do so, we focus on a two-point correlation analysis rather than attempting to associate individual H I systems with individual galaxies.

The two-point correlation function, $\xi(r)$, is defined as the probability excess of finding a pair of objects at a distance r with respect to the expectation from a randomly distributed sample.[16] Combining the results from the H I–galaxy cross-correlation with those from the H I–H I and galaxy–galaxy auto-correlations for different subsamples of H I systems and galaxies, we aim to get further insights on the relationship between the IGM and galaxies.

4.7.1 Two-dimensional correlation measurements

In order to measure these spatial correlation functions we converted all H I system and galaxy positions given in (RA, DEC, z) coordinates into a Cartesian co-moving system (X, Y, Z). We first calculated the radial co-moving distance to an object at redshift z as,

$$R(z) = \frac{c}{H_0} \int_0^z \frac{1}{\sqrt{\Omega_{\rm m}(1+z') + \Omega_\Lambda}} \, {\rm d}z' \, . \tag{4.2}$$

where c is the speed of light, H_0 is the Hubble constant and, $\Omega_{\rm m}$ and Ω_Λ are the mass and 'dark energy' density parameters, respectively. Let (RA_0, DEC_0)

[16] Assuming isotropy, ξ is a function of distance only.

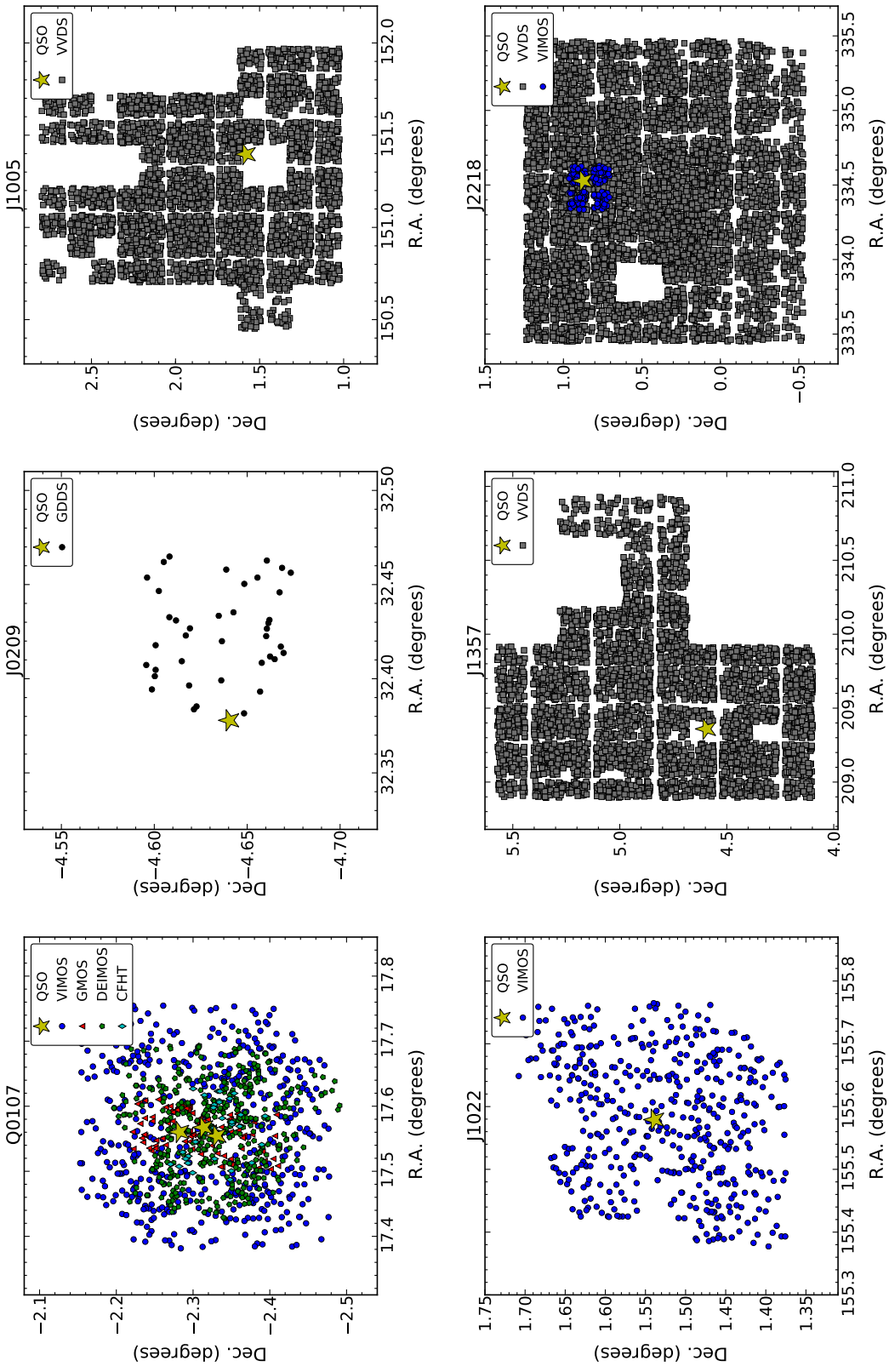

Figure 4.9: Distribution on the sky of galaxies and background QSOs (yellow stars) for each field. Blue circles, red triangles and green pentagons correspond to our new VIMOS, GMOS and DEIMOS galaxies respectively; while black circles, grey squares and cyan diamonds correspond to GDDS, VVDS and CFHT MOS galaxies respectively.

be the central coordinates of a given independent field. Then, we transformed (RA,DEC,z) to (X, Y, Z) as follows:

$$\begin{aligned} X &\equiv R(z)\cos(\Delta\delta)\cos(\Delta\alpha) \\ Y &\equiv R(z)\cos(\Delta\delta)\sin(\Delta\alpha) \\ Z &\equiv R(z)\sin(\Delta\delta)\ , \end{aligned} \tag{4.3}$$

where $\Delta\delta \equiv (\mathrm{DEC} - \mathrm{DEC}_0)$ and $\Delta\alpha \equiv (\mathrm{RA} - \mathrm{RA}_0)\cos(\mathrm{DEC}_0)$, both in radians. Note that all our fields are far away from the poles and each of them have small angular coverage ('pencil beam' surveys), making this transformation accurate. For fields with only one QSO we chose $(\mathrm{RA}_0, \mathrm{DEC}_0) = (\mathrm{RA}_{\mathrm{QSO}}, \mathrm{DEC}_{\mathrm{QSO}})$, while for the triple QSO field we took the average position as the central one.

Given that peculiar velocities add an extra component to redshifts (in addition to cosmological expansion), our (X, Y, Z) will be affected differently, producing distortions even from actually isotropic signals. This is because the X coordinate is parallel to the LOS, while the Y and Z coordinates are perpendicular to it. Let $R(z)$ be the radial co-moving distance at redshift z (Equation 4.2) and $\Delta\theta$ a small ($\ll 1$) angular separation in radians. The transverse co-moving separation can be then approximated by $\approx R(z)\Delta\theta$, implying that our X coordinate will be affected a factor of $\approx 1/\Delta\theta$ times that of the Y and Z coordinates for a fixed redshift difference. As an example, a redshift difference of $\Delta z = 0.0007$ at $z = 0.5$ ($\approx 140\ \mathrm{km\,s^{-1}}$) will roughly correspond to a radial co-moving difference of ≈ 2 Mpc, while only to a $\lesssim 0.02$ Mpc difference in the transverse direction for co-moving separations $\lesssim 20$ Mpc. We therefore measured the auto- and cross-correlations both along and transverse to the LOS, $\xi(r_\perp, r_\parallel)$, independently. In terms of our Cartesian coordinates we have that $r_{\parallel,ij} \equiv |X_i - X_j|$ and $r_{\perp,ij} \equiv \sqrt{|Y_i - Y_j|^2 + |Z_i - Z_j|^2}$, are the along the LOS and transverse to the LOS distances between two objects at positions (X_i, Y_i, Z_i) and (X_j, Y_j, Z_j) respectively. Deviations from an isotropic signal in our $(r_\perp, r_\parallel)$ coordinates can then be attributed to redshift uncertainties, peculiar velocities and/or LSS bulk motions between the objects in the sample.

We used the Landy and Szalay (1993) estimator to calculate the galaxy–galaxy auto-correlation as,

$$\xi^{\rm LS}_{\rm gg}(r_\perp, r_\parallel) = \frac{D_{\rm g}D_{\rm g}/n^{\rm DD}_{\rm gg} - 2D_{\rm g}R_{\rm g}/n^{\rm DR}_{\rm gg}}{R_{\rm g}R_{\rm g}/n^{\rm RR}_{\rm gg}} + 1 \, , \tag{4.4}$$

where $D_{\rm g}D_{\rm g}$ is the number of observed 'data-data' galaxy–galaxy pairs, $R_{\rm g}R_{\rm g}$ is the number of 'random-random' galaxy–galaxy pairs and $D_{\rm g}R_{\rm g}$ is the number of 'data-random' galaxy–galaxy pairs, all of which are measured at the given $(r_\perp, r_\parallel)$ scales; and $n^{\rm DD}_{\rm gg}$, $n^{\rm DR}_{\rm gg}$ and $n^{\rm RR}_{\rm gg}$ are the normalization factors for each respective pair count. Let $N^{\rm real}_{\rm gal}$ and $N^{\rm rand}_{\rm gal} \equiv \alpha_{\rm gal} N^{\rm real}_{\rm gal}$ be the total number of real and random galaxies respectively, then

$$\begin{aligned} n^{\rm DD}_{\rm gg} &= N^{\rm real}_{\rm gal}(N^{\rm real}_{\rm gal} - 1)/2 \\ n^{\rm DR}_{\rm gg} &= \alpha_{\rm gal}(N^{\rm real}_{\rm gal})^2 \\ n^{\rm RR}_{\rm gg} &= \alpha_{\rm gal} N^{\rm real}_{\rm gal}(\alpha_{\rm gal} N^{\rm real}_{\rm gal} - 1)/2. \end{aligned} \tag{4.5}$$

The H I–H I auto-correlation, $\xi^{\rm LS}_{\rm aa}$, was calculated in the same fashion as $\xi^{\rm LS}_{\rm gg}$, so

$$\xi^{\rm LS}_{\rm aa}(r_\perp, r_\parallel) = \frac{D_{\rm a}D_{\rm a}/n^{\rm DD}_{\rm aa} - 2D_{\rm a}R_{\rm a}/n^{\rm DR}_{\rm aa}}{R_{\rm a}R_{\rm a}/n^{\rm RR}_{\rm aa}} + 1 \, , \tag{4.6}$$

where $D_{\rm a}D_{\rm a}$ is the number of observed 'data-data' absorber-absorber pairs, $R_{\rm a}R_{\rm a}$ is the number of 'random-random' absorber-absorber pairs and $D_{\rm a}R_{\rm a}$ is the number of 'data-random' absorber-absorber pairs, all of which measured at the given $(r_\perp, r_\parallel)$ scales; and $n^{\rm DD}_{\rm aa}$, $n^{\rm DR}_{\rm aa}$ and $n^{\rm RR}_{\rm aa}$ are the normalization factors for each respective pair count. Let $N^{\rm real}_{\rm abs}$ and $N^{\rm rand}_{\rm abs} \equiv \alpha_{\rm abs} N^{\rm real}_{\rm abs}$ be the total number of real and random H I systems respectively, then

$$
\begin{aligned}
n_{\rm aa}^{\rm DD} &= N_{\rm abs}^{\rm real}(N_{\rm abs}^{\rm real}-1)/2 \\
n_{\rm aa}^{\rm DR} &= \alpha_{\rm abs}(N_{\rm abs}^{\rm real})^2 \\
n_{\rm aa}^{\rm RR} &= \alpha_{\rm abs}N_{\rm abs}^{\rm real}(\alpha_{\rm abs}N_{\rm abs}^{\rm real}-1)/2.
\end{aligned}
\tag{4.7}
$$

The H I–galaxy cross-correlation, $\xi_{\rm ag}^{\rm LS}$, was calculated using a generalized version of the Landy and Szalay (1993) estimator,

$$
\xi_{\rm ag}^{\rm LS}(r_\perp, r_\parallel) = \frac{D_{\rm a}D_{\rm g}/n_{\rm ag}^{\rm DD} - D_{\rm a}R_{\rm g}/n_{\rm ag}^{\rm DR} - R_{\rm a}D_{\rm g}/n_{\rm ag}^{\rm RD}}{R_{\rm a}R_{\rm g}/n_{\rm ag}^{\rm RR}} + 1 \tag{4.8}
$$

(e.g. Adelberger et al., 2003), where $D_{\rm a}D_{\rm g}$ is the number of observed 'data-data' absorber–galaxy pairs, $R_{\rm a}R_{\rm g}$ is the number of 'random-random' absorber–galaxy pairs, $D_{\rm a}R_{\rm a}$ and $R_{\rm a}D_{\rm g}$ are the number of 'data-random' and 'random-data' absorber–galaxy pairs respectively, all of which are measured at the given $(r_\perp, r_\parallel)$ scales. Following previous conventions the normalization factors in this case are,

$$
\begin{aligned}
n_{\rm ag}^{\rm DD} &= N_{\rm abs}^{\rm real}N_{\rm gal}^{\rm real} \\
n_{\rm ag}^{\rm DR} &= \alpha_{\rm gal}N_{\rm abs}^{\rm real}N_{\rm gal}^{\rm real} \\
n_{\rm ag}^{\rm RD} &= \alpha_{\rm abs}N_{\rm abs}^{\rm real}N_{\rm gal}^{\rm real} \\
n_{\rm ag}^{\rm RR} &= \alpha_{\rm gal}\alpha_{\rm abs}N_{\rm abs}^{\rm real}N_{\rm gal}^{\rm real}\ .
\end{aligned}
\tag{4.9}
$$

This approach makes the random samples a crucial component of the analysis. A detailed description of the random generator algorithms is presented in Section 4.7.2.

Landy and Szalay (1993) showed that $\xi^{\rm LS}$ minimizes the observed variance and so is preferable over other proposed estimators (e.g. Sharp, 1979; Hewett, 1982; Davis and Peebles, 1983; Hamilton, 1993). Given the limited nature of any survey, all estimators are biased towards lower correlation amplitudes. This is because the mean densities of our two populations are estimated from the survey itself. In

order for us to measure a positive correlation in a certain scale, the measured ξ needs to be negative at another. This leads to an observed correlation amplitude which is lower than the underlying real one, $\xi^{\rm real}$, assumed to be positive. This is a well known bias commonly referred to as the 'integral constraint'. Landy and Szalay (1993) showed that $\xi^{\rm LS}$ and $\xi^{\rm real}$ are related as

$$1 + \xi^{\rm LS} = \frac{1 + \xi^{\rm real}}{1 + \xi_V} \, , \tag{4.10}$$

where ξ_V is the 'integral constraint' (scalar) defined as

$$\xi_V \equiv \int_V G(r) \xi^{\rm real}(r) \, {\rm d}^2 V \, . \tag{4.11}$$

Here $G(r)$ is a normalized geometric window function (positive) which gives the probability of having two volume elements separated by a distance r in the survey. In the case of our auto- and cross-correlations G is given by $G_{\rm gg} \approx R_{\rm g} R_{\rm g} / n^{\rm gg}_{\rm RR}$, $G_{\rm aa} \approx R_{\rm a} R_{\rm a} / n^{\rm aa}_{\rm RR}$ and $G_{\rm ag} \approx R_{\rm a} R_{\rm g} / n^{\rm ag}_{\rm RR}$. Although we cannot know a priori the amplitude of $\xi^{\rm real}$ we made a small correction using

$$\xi = (1 + \tilde{\xi}_V)(1 + \xi^{\rm LS}) - 1 \, , \tag{4.12}$$

where $\tilde{\xi}_V \equiv \int_V G(r) \xi^{\rm LS}(r) \, {\rm d}^2 V$, which still helps because of the discrete nature of all our cross-pair counts (including the randoms).

The computation of $\xi^{\rm LS}_{\rm gg}$, $\xi^{\rm LS}_{\rm aa}$ and $\xi^{\rm LS}_{\rm ag}$ was performed *after* summing all the cross-pairs from our $N_{\rm f} = 6$ independent fields,

$$D_{\rm g} D_{\rm g}(r_\perp, r_\parallel) = \sum_i^{N_{\rm f}} D_{\rm g} D_{\rm g}(r_\perp, r_\parallel)_i \, , \tag{4.13}$$

where $(D_{\rm g} D_{\rm g})_i$ is the number of 'data-data' galaxy–galaxy pairs in the i-th field; and so on for the rest of the cross-pair counts. In contrast to measuring $\xi^{\rm LS}$ for each independent field and then taking a weighted average, our adopted approach reduces the 'shot noise'.

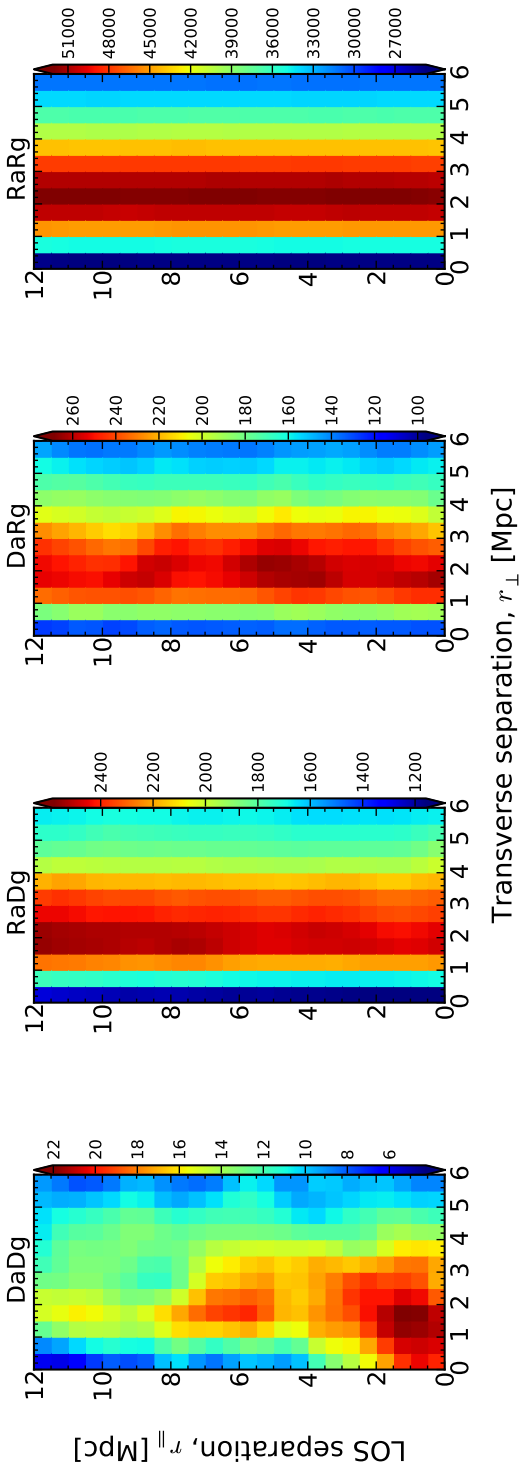

Figure 4.10: Total number of cross-pairs between H I absorption systems and galaxies as a function of separations along ($r_\parallel$; y-axes) and transverse to the line-of-sight ($r_\perp$; x-axes). We used an arbitrary binning of 0.5 Mpc with a Gaussian smoothing of standard deviation of 0.5 Mpc along both directions.

Another way to reduce the 'shot noise' is by using large bin sizes for counting the cross-pairs, but this will limit the spatial resolution of our ξ measurements. Therefore, we have chosen to compute the cross-pairs at scales $r_\perp < 10$ Mpc using a linear grid of 0.5 Mpc in both $(r_\perp, r_\parallel)$ coordinates and apply a Gaussian filter of 0.5 Mpc standard deviation (in both directions) to smooth the final counts distribution obtained from Equation 4.13 *before* applying Equations 4.4, 4.6, and 4.8. We treated the edges of the grid as if they were mirrors for the smoothing. As an example, Figure 4.10 shows the number of cross-pairs between H I absorption systems and galaxies for our 'Full Sample' (defined in Section 4.8) using our adopted binning and smoothing.

An isotropic smoothing is desirable to avoid introducing artificial distortions, especially at the smallest scales. The use of a smoothing filter is justified by assuming that the underlying matter distribution that gives rise to H I absorption systems and galaxies (and hence to the data-data cross-pairs) is also smooth. Our approach offers a compromise between reducing the 'shot noise' while keeping a relatively small bin size. We caution though that if the geometry of H I clouds does contain sharp edges at scales smaller than our adopted binning or smoothing length we would not be able to detect such a feature.

4.7.2 Random samples

One of the crucial steps for a correlation analysis is the construction of the random samples. In order to cancel out any possible bias we preserved the sensitivity function of the real survey in our random samples. A detailed description of the random generator algorithms for H I absorption systems and galaxies is presented in the following sections.

Random absorption lines

We created random samples for individual observations made with a given instrument and/or instrument setting (i.e. resolution, wavelength coverage, etc).

This means that we treat the two channels of COS (FUV and NUV) independently for the creation of the random samples, and also for FOS. For a given absorption system with (RA, DEC, $z_{\rm abs}$, $N_{\rm HI}$, $b_{\rm HI}$) we create $\alpha_{\rm abs}$ random ones, varying the redshift but preserving the rest of its parameters.

The random redshifts were chosen based on the properties of the spectrum in which the original absorption system was observed. We first estimated the minimum rest-frame equivalent width of a transition that could have been observed in the spectrum at a redshift z. For unresolved features, the minimum equivalent width for a line to be observed at wavelength λ is

$$W_{\rm min}(\lambda) \approx sl\ \frac{\rm FWHM}{\langle S/N \rangle_\lambda}\ , \tag{4.14}$$

where sl is the significance level of the detection in standard deviation units, FWHM is the 'full-width at half maximum' of the line spread function (LSF) of the spectrograph in Å, and $\langle S/N \rangle_\lambda$ is the average signal-to-noise per resolution element. Transforming λ coordinates to redshift coordinates for a given rest-frame transition at λ_0 (i.e. $\lambda\ \rightarrow\ z = \frac{\lambda}{\lambda_0} - 1$), and assuming a constant spectral resolution $R \equiv \frac{\lambda}{\rm FWHM}$, the rest-frame minimum equivalent width is then given by

$$W_{\rm r,min}(z) \approx \frac{sl\lambda_0}{R\langle S/N \rangle_z}\ . \tag{4.15}$$

Finally, for a given absorber with equivalent width, $W_{\rm HI}^{\rm obs}$, we compare it with $W_{\rm r,min}(z)$ and place $\alpha_{\rm abs}$ random absorbers uniformly at redshifts where the condition $W_{\rm HI}^{\rm obs} \geq W_{\rm r,min}(z)$ is satisfied. We masked out spectral regions over a velocity window of ± 200 km s^{-1} around the position where strong Galactic absorption could have been detected (namely: C II, N V, O I, Si II, P III, S II and Fe II) *before* the random redshifts are assigned.

Even though we have direct measurements of the equivalent widths for the real absorption systems, we do not use them directly in order to avoid confusion from blended systems. We use instead the approximation given by Draine (2011, see his equation 9.27) to convert the inferred $N_{\rm HI}$ and $b_{\rm HI}$ to a $W_{\rm HI}^{\rm obs}$. Note that

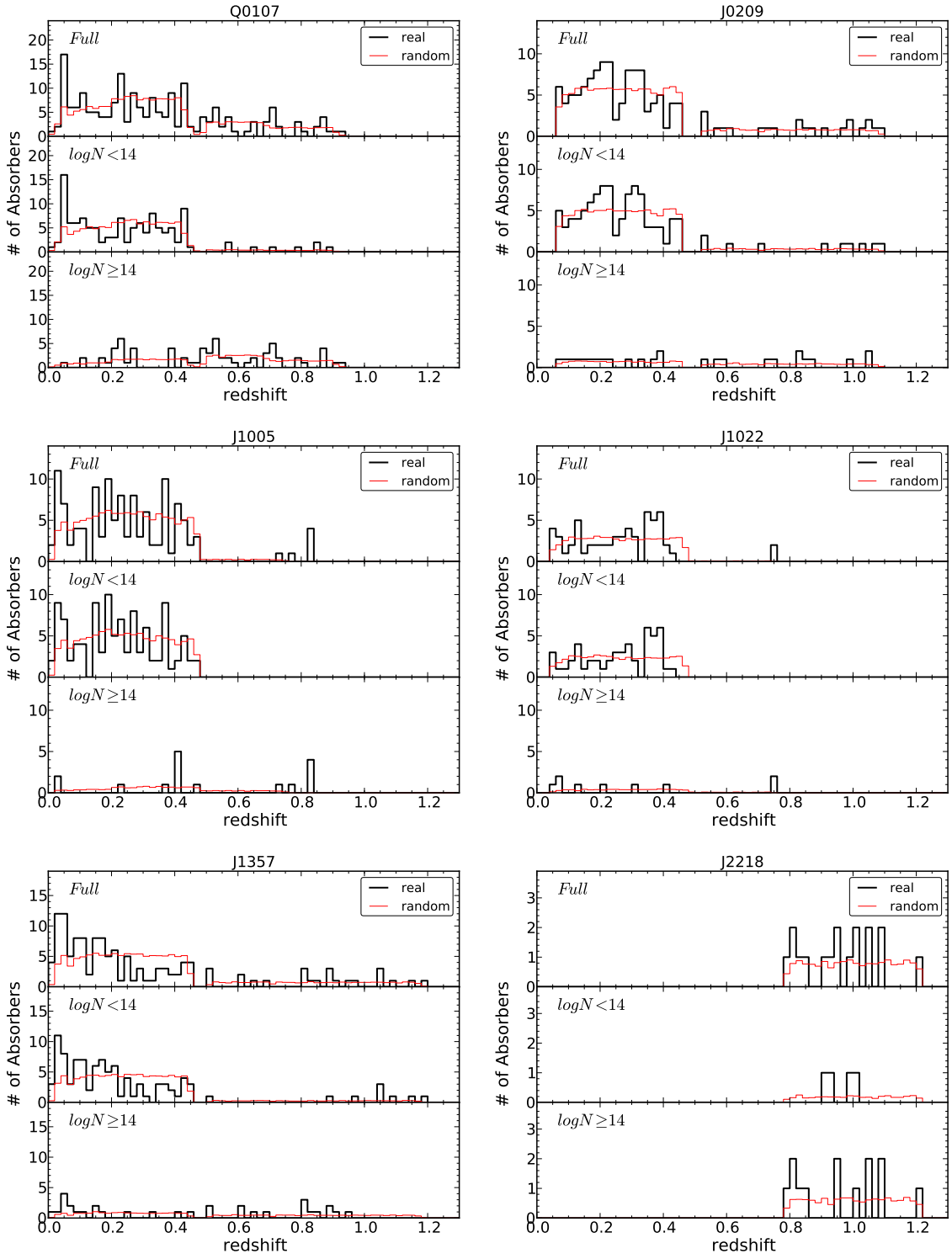

Figure 4.11: Histograms of the H I absorption systems redshift distribution for our different fields (0.02 binning). The black thick solid lines correspond to the real distributions whereas the red thin solid lines correspond to the normalized random expectation drawn from samples of 200× the real sample sizes. A full description of the random generator algorithm can be found in Section 4.7.2. Top panels show the full H I samples while the middle and bottom panels show subsamples based on $N_{\rm HI}$ cuts.

passing from $W_{\rm HI} \rightarrow (N_{\rm HI}, b_{\rm HI})$ is not always robust because of the flat part of the curve-of-growth, but passing from $(N_{\rm HI}, b_{\rm HI}) \rightarrow W_{\rm HI}$ is.

We mainly based our search of H I absorption systems on the Lyα transition (for which $\lambda_0 = 1215.67$ Å), but in some cases we extended it to Lyβ in spectral regions with no Lyα coverage. For the Lyβ detected systems, we applied the same method described above but changing the transition parameters accordingly.

Figure 4.11 presents the redshift distribution of real (black lines) and random (red lines) absorbers in each of our independent fields using $\alpha_{\rm abs} = 200$.

Random galaxies

The random galaxies were created for each field and instrument independently. This means that we treat different galaxy surveys independently for the creation of the random samples, even when the galaxy surveys come from the same field. For a given observed galaxy with (RA, DEC, $z_{\rm gal}$, magnitude, spectral type, etc.) we create $\alpha_{\rm gal}$ random ones, varying the redshift, but preserving the rest of its parameters. This approach ensures the selection function is well matched by the random galaxies.

The random redshifts ($z_{\rm gal}^{\rm rand}$) were chosen based on the observed redshift distribution. We made sure that our randoms resembled the observed galaxy distribution independently of the observed magnitude of the galaxies. To do so, we selected multiple subsamples of galaxies at different magnitude bins, whose empirical redshift distributions are used as proxies for the redshift selection function. We used magnitude bins of size 1, shifted by 0.5 magnitudes, ranging from 15 to 25. For the brighter and fainter ends of the subsamples we increased the magnitude bin sizes to ensure a minimum of 20 galaxies. For each magnitude subsample, we computed histograms using redshift bins of $\Delta z = 0.01$ (arbitrary), which were then smoothed with a Gaussian filter of standard deviation $\sigma = 0.1$ (roughly corresponding to a co-moving scale of ≈ 300 Mpc at redshift $z = 0.5$). This large smoothing length is important to get rid of the LSS spikes and valleys present in the real redshift distributions. The final redshift probability distribution

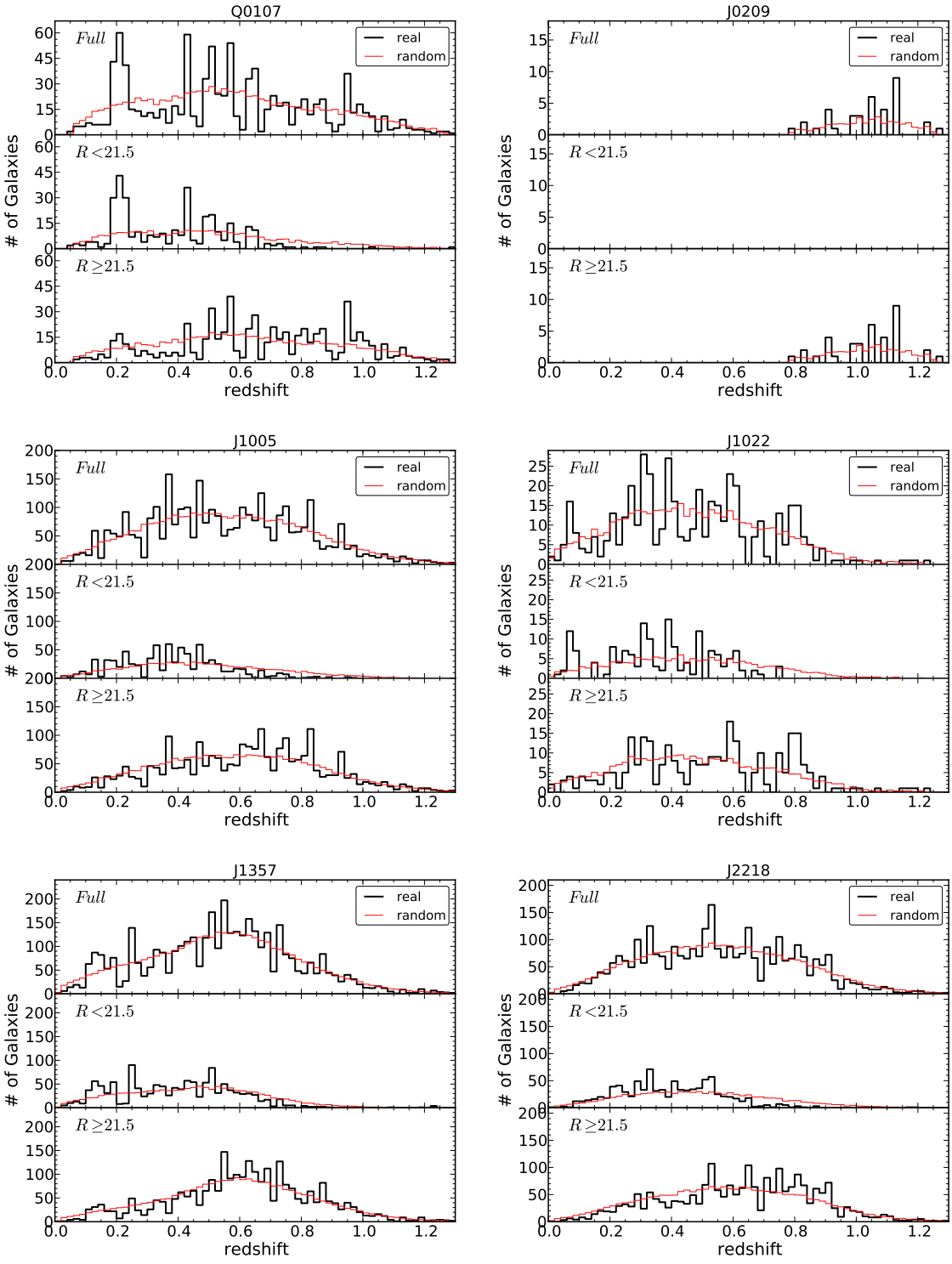

Figure 4.12: Histograms of the galaxy redshift distribution for our different fields (0.02 binning). The black thick solid lines correspond to the real distributions whereas the red thin solid lines correspond to the normalized random expectation drawn from samples of $20\times$ the real sample sizes. A full description of the random generator algorithm can be found in Section 4.7.2. Top panels show the full galaxy samples while the middle and bottom panels show subsamples based on R-band magnitude cuts.

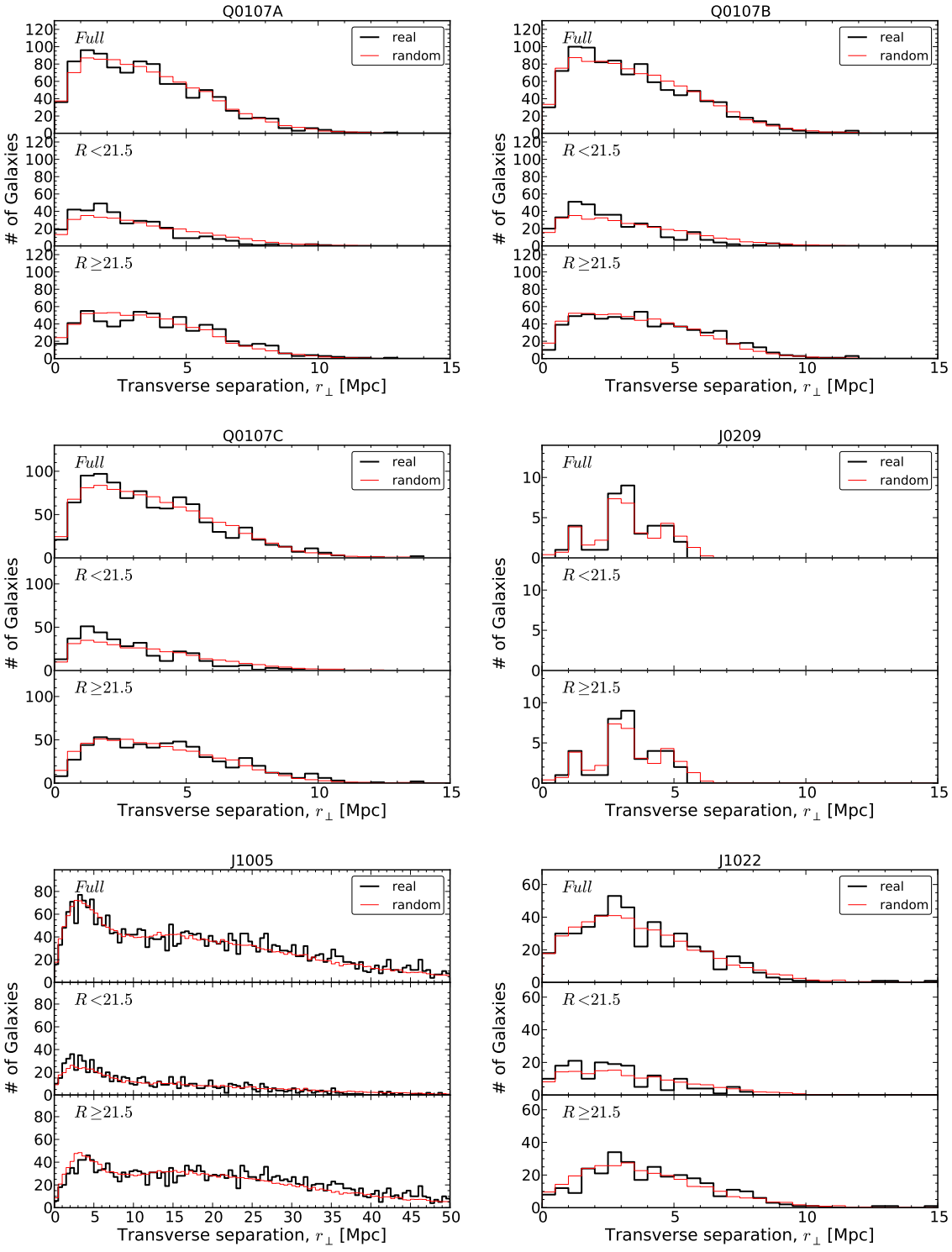

Figure 4.13: Histograms of the galaxy transverse separation distribution for our different fields (0.5 Mpc bining). The black thick solid lines correspond to the real distributions whereas the red thin solid lines correspond to the normalized random expectation drawn from samples of $20\times$ the real sample sizes. A full description of the random generator algorithm can be found in Section 4.7.2. Top panels show the full galaxy samples while the middle and bottom panels show subsamples based on R-band magnitude cuts.

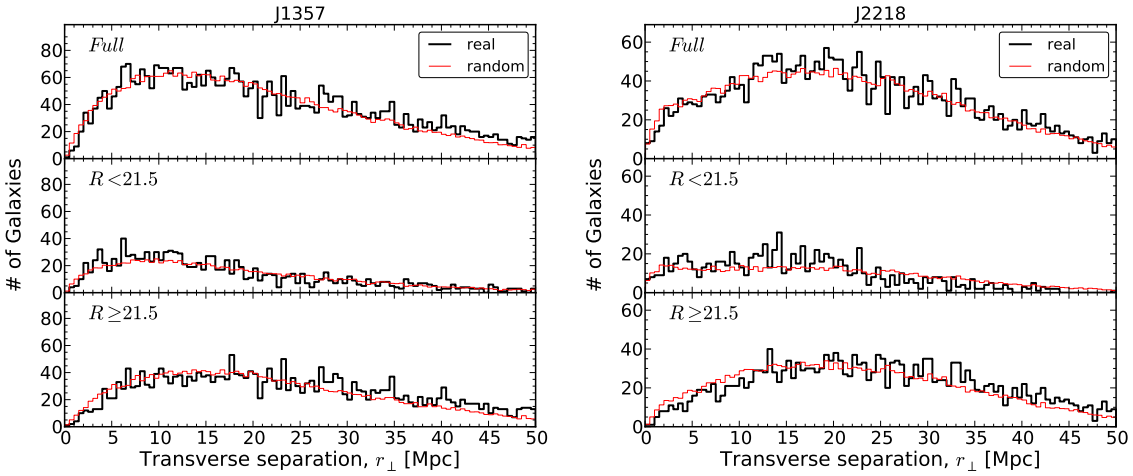

Figure 4.13: *Continued.*

of a given magnitude bin is provided by a cubic spline interpolation over the smoothed histograms. Thus, for a given galaxy with observed magnitude m, we placed $\alpha_{\rm gal}$ randoms according to the spline fit associated with the subsample of galaxies centred on the closest magnitude bin to m. We also imposed the redshifts of the random galaxies to lie between $z_{\rm min} < z_{\rm gal}^{\rm random} < z_{\rm max}$, where $z_{\rm min}$ and $z_{\rm max}$ are the minimum and maximum galaxy redshifts of the real sample.

Figure 4.12 presents the redshift distribution of real (black lines) and random (red lines) galaxies in each of our independent fields using $\alpha_{\rm gal} = 20$. Similarly, Figure 4.13 presents the distribution in transverse separations of real (black lines) and random (red lines) galaxies with respect to their respective QSO LOS. The fact that we have not randomized the (RA, DEC) of our galaxies should not have an important effect in the quality of the random samples. This is so because the geometry of the survey in each field (i.e. a large but narrow cone) allows to erase any coherent structure observed at a fixed transverse distance by only randomizing the redshifts. The goodness of this method can be appreciated in Figure 4.13, excepting perharps in field J0209. The contribution of field J0209 to the final galaxy sample is minimal however, and so it should not compromise our correlation analysis.

4.7.3 Projected correlations along the line-of-sight

A useful quantity to compute from the two-dimensional correlation functions is the projected correlation function along the LOS,

$$\Xi(r_\perp) = 2\int_0^\infty \xi(r_\perp, r_\parallel)\,\mathrm{d}r_\parallel \,, \tag{4.16}$$

as it will be insensitive to redshift distortions, at least for the transverse separations involved in this work (Davis and Peebles, 1983). Therefore, one can find a relation between the 'real-space' correlation function (distortion free), $\xi(r = \sqrt{r_\parallel^2 + r_\perp^2})$, and $\Xi(r_\perp)$, as

$$\begin{aligned}\Xi(r_\perp) &= 2\int_0^\infty \xi(r)\,\mathrm{d}r_\parallel \\ &= 2\int_{r_\perp}^\infty \xi(r)\frac{r\,\mathrm{d}r}{\sqrt{r^2 - r_\perp^2}} \,,\end{aligned} \tag{4.17}$$

which leads to $\xi(r)$ being given by the inverse Abel transform,

$$\xi(r) = -\frac{1}{\pi}\int_r^\infty \frac{\mathrm{d}\Xi(r_\perp)}{\mathrm{d}r_\perp}\frac{\mathrm{d}r_\perp}{\sqrt{r_\perp^2 - r^2}} \,. \tag{4.18}$$

Davis and Peebles (1983) showed that when $\xi(r)$ is described by a power-law of the form,

$$\xi(r) = \left(\frac{r}{r_0}\right)^{-\gamma} \,, \tag{4.19}$$

then Equation 4.17 leads to

$$\Xi(r_\perp) = A(r_0, \gamma) r_\perp^{1-\gamma} \,, \tag{4.20}$$

where $A(r_0, \gamma) = r_0^\gamma \Gamma(1/2)\Gamma[(\gamma-1)/2]/\Gamma(\gamma/2)$ and Γ is the Gamma function. Therefore, r_0 and γ of $\xi(r)$ can be directly obtained from a power-law fit to $\Xi(r_\perp)$ using Equation 4.20. Note that this method is only valid for $\gamma > 1$.

In practice, we will use $r_\parallel^{\max} = 20$ Mpc as the integration limit in Equation 4.16.

A larger integration limit will increase the 'shot noise' while not adding much correlation power. As long as the vast majority of correlated pairs are included in the integration limit (which is the case), this approach will suffice (e.g. Davis and Peebles, 1983; Ryan-Weber, 2006). In order to further reduce the 'shot noise', we summed all the cross-pairs along the LOS, e.g. $D_{\rm a}D_{\rm g}(r_\perp) = \sum_i D_{\rm a}D_{\rm g}(r_\perp, r_{\parallel,i})$ (and so on for the others), and then computed the Landy and Szalay estimators, $\xi^{\rm LS}(r_\perp)$, using these collapsed cross-pairs, so

$$\Xi(r_\perp) = 2r_\parallel^{\rm max}\xi^{\rm LS}(r_\perp)\ . \tag{4.21}$$

This approach is justified given the cylindrical geometry of our survey, for which the 'random-random' pairs (denominator of the LS estimator) is almost constant along the $r_\parallel$-axis for the scales involved in this study (e.g. see right panel of Figure 4.10). We compared the absolute values of Ξ from our adopted approach with that of a direct integration (as in Equation 4.16 using $r_\parallel^{\rm max} = 20$ Mpc as the integration limit). We obtained differences of $\lesssim 5\%$ in the correlation amplitudes, indicating that our approach is appropiate.

4.7.4 Relations between auto- and cross-correlations

We use the Cauchy–Schwarz inequality,

$$\xi_{\rm ag}^2 \leq \xi_{\rm gg}\xi_{\rm aa}, \tag{4.22}$$

as the main tool to address the connection between H I and galaxies. The equality only holds when the density fluctuations that give rise to H I absorption systems and galaxies are linearly dependent. However, in the most general case, the product of the auto-correlation functions does not necessarily equals $\xi_{\rm ag}^2$. If we do assume that both H I absorption systems and galaxies trace the same underlying dark matter density distribution (e.g. Ryan-Weber, 2006), we have

$$\begin{aligned} \xi_{\rm gg} &= b_{\rm g}^2 \xi_{\rm DM} \\ \xi_{\rm aa} &= b_{\rm a}^2 \xi_{\rm DM} \\ \xi_{\rm ag} &= b_{\rm a} b_{\rm b} \xi_{\rm DM}, \end{aligned} \tag{4.23}$$

where $\xi_{\rm DM}$ is the dark matter auto-correlation function (assumed positive) and $b_{\rm g}$ and $b_{\rm a}$ are the galaxy and H I 'absolute biases' (also positives), respectively. If these biases are independent of the scale (i.e. linear biases), then the equality of Equation 4.22 holds. If that is the case, one can use the ratio between the correlation functions to infer the dark matter halo masses of one population relative to the other (e.g. Mo et al., 1993; Ryan-Weber, 2006). On the other hand, if $\xi_{\rm ag}^2 < \xi_{\rm gg}\xi_{\rm aa}$ we can no longer assume such a simplistic model. In such a case, the observed difference with respect to $\xi_{\rm ag}^2 = \xi_{\rm gg}\xi_{\rm aa}$ can be used to: (i) get insights on the baryonic physics affecting H I absorption systems and/or galaxies, assuming that the standard cosmological paradigm is correct; or (ii) put constraints on the current cosmological paradigm, assuming that the baryonic physics is fully understood. In this work we will focus on the former.

Adelberger et al. (2003) showed a third possibility, having $\xi_{\rm ag}^2$ exceeding $\xi_{\rm gg}\xi_{\rm aa}$ for correlation functions measured from discrete and volume limited samples. In the hypothetical case of an H I–galaxy one-to-one correspondence, then $\xi_{\rm gg} = \xi_{\rm aa}$, but $\xi_{\rm ag}$ will appear higher at the very small scales because in the case of auto-correlations we exclude the correlation of an object with itself, whereas in $\xi_{\rm ag}$ that correlation is present (Adelberger et al., 2003, see their appendix A). Such a behaviour between auto- and cross-correlations will indicate that the two populations of objects are indeed the same physical entities. The geometry of our survey might not be suitable for testing this idea, as we are only mapping H I absorption systems along single LOS for which the completeness of galaxies close to these absorbers is far from being complete. Still, we will bear this result in mind for the interpretation of our results.

4.7.5 Uncertainty estimation

When dealing with cross-correlations it is important to realize that the uncertainties will be dominated by the *smallest* sample. If we consider a sample composed of a single object and another sample composed of 100 objects, the number of cross-pairs is 100 but none of these pairs are truly independent as they all share a common object. Therefore, assuming Poisson uncertainty for the number of pairs (as commonly seen in the literature; e.g. Chen et al., 2005; Chen and Mulchaey, 2009) is not optimal, as it will underestimate the true uncertainty. For *correlated* distributions, none of the pairs are independent because the number of systems at a given scale will depend on the number of systems at all other scales, and deviations from the Poissonian expectation will be more important at the scales where the correlation signal is large. Indeed, Landy and Szalay (1993) showed that the variance of $\xi^{\rm LS}$ can be approximated by (using our notation),

$$\Delta^2_{\rm LS}(\xi^{\rm LS}) \approx \frac{(1+\xi^{\rm LS})^2}{n_{\rm DD}(RR/n_{\rm RR})} \approx \frac{(1+\xi^{\rm LS})^3}{DD} \,. \tag{4.24}$$

This variance is greater than the commonly used

$$\Delta^2_{\rm DD}(\xi^{\rm LS}) = \frac{1+\xi^{\rm LS}}{DD} \,, \tag{4.25}$$

by a factor of $\sim (1+\xi)^2$ and so, we caution the use of the latter as it might still underpredict the real uncertainty.

In order to test whether the uncertainty given by Equation 4.24 is reasonable for our survey, we also computed the 'jackknife' and 'bootstrap' variances. The 'jackknife' variance was computed as

$$\Delta^2_{\rm JK}(\xi) = \frac{1}{N_{\rm f}(N_{\rm f}-1)} \sum_i^{N_{\rm f}} (\xi^*_i - \bar{\xi}^*)^2 \,, \tag{4.26}$$

where ξ^*_i is the i-th 'pseudo-value' of the correlation function, $\xi^*_i \equiv N_{\rm f}\xi - (N_{\rm f}-1)\xi_{-i}$, with ξ_{-i} being the value of the correlation function measured when the

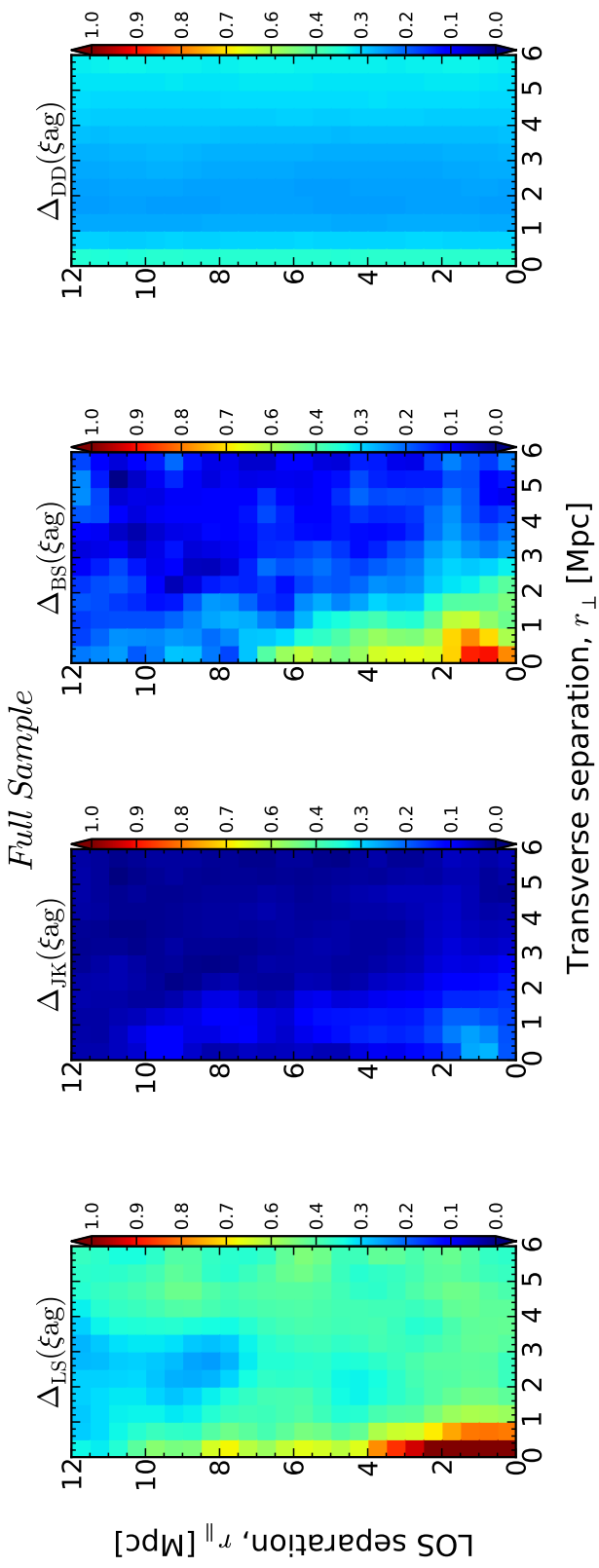

Figure 4.14: Uncertainty estimations (squared root of variances) of the H I-galaxy cross-correlation, $\Delta(\xi^{\rm LS}_{\rm ag})$, measured from our 'Full Sample' as a function of separations both along the line-of-sight ($r_\parallel$; y-axes) and transverse to the line-of-sight ($r_\perp$; x-axes). From left to right: uncertainty from the Landy and Szalay (1993) analytical approximation, $\Delta_{\rm LS}$ (equation 4.24); uncertainty from a 'jackknife' resampling, $\Delta_{\rm JK}$ (equation 4.26); uncertainty from a 'bootstrap' resampling, $\Delta_{\rm BS}$ (equation 4.27); and the commonly used Poissonian uncertainty, $\Delta_{\rm DD}$ (equation 4.25).

i-th field is removed from the sample, and $\bar{\xi}^*$ is the mean of the 'pseudo-values'. The 'bootstrap' variance was computed by creating $N_{\rm bs} = 500$ sets of $N_{\rm f}$ fields randomly chosen (with repetition) from the set of real fields[17], so

$$\Delta^2_{\rm BS}(\xi) = \frac{1}{N_{\rm bs}} \sum_{i}^{N_{\rm bs}} (\xi_i - \bar{\xi})^2 \,, \tag{4.27}$$

where ξ_i is the correlation measured from the i-th random set and $\bar{\xi}$ is the mean of these 'bootstrap' measurements. Uncertainties for the projected correlations, Ξ, and the ratio $(\xi_{\rm ag})^2/(\xi_{\rm gg}\xi_{\rm aa})$, were calculated analogously.

As an example, Figure 4.14 shows these 4 uncertainty estimations (squared root of the variances) for our measurements of $\xi^{\rm LS}_{\rm ag}(r_\perp, r_\parallel)$. From left to right: $\Delta_{\rm LS}$, $\Delta_{\rm JK}$, $\Delta_{\rm BS}$ and $\Delta_{\rm DD}$. We see that all these uncertainty estimations are within ~ 1 order of magnitude consistent with each other, but systematic trends are present. We see that $\Delta_{\rm LS}$ and $\Delta_{\rm BS}$ give the largest uncertainties while $\Delta_{\rm JK}$ and $\Delta_{\rm DD}$ give the smallest. We also observe that $\Delta_{\rm LS}$, $\Delta_{\rm JK}$, $\Delta_{\rm BS}$ peak at the smallest scales (where the correlation amplitudes are greater) while $\Delta_{\rm DD}$ does not. Similar behaviors are observed for the uncertainties associated to our $\xi^{\rm LS}_{\rm gg}(r_\perp, r_\parallel)$ measurements (not plotted).

Figure 4.15 shows these 4 uncertainty estimations for our measurements of the projected correlations $\Xi(r_\perp)$, for both H I–galaxy (squares) and galaxy–galaxy (circles): $\Delta_{\rm LS}$ (green lines), $\Delta_{\rm JK}$ (red lines), $\Delta_{\rm BS}$ (blue lines) and $\Delta_{\rm DD}$ (yellow lines). The top panel shows the absolute values for these different uncertainties, while the bottom panel shows the ratio of a given uncertainty estimation and $\Delta_{\rm BS}$. As before, we observe systematic trends, but all uncertainties are consistent within ~ 1 order of magnitude of each other. In contrast to the two-dimensional uncertainties, $\Delta_{\rm BS}$ is the largest in this case. Focusing at the smallest scales (where the correlation amplitudes are greater) we see that $\Delta_{\rm JK}$ and $\Delta_{\rm LS}$ are in closer agreement to $\Delta_{\rm BS}$ than $\Delta_{\rm DD}$.

These results suggest that $\Delta_{\rm LS}$ should be preferable over $\Delta_{\rm DD}$ and even over

[17]Note that for 6 fields, the total number of possible combinations is $\frac{(6+6-1)!}{6!(6-1)!} = 462$.

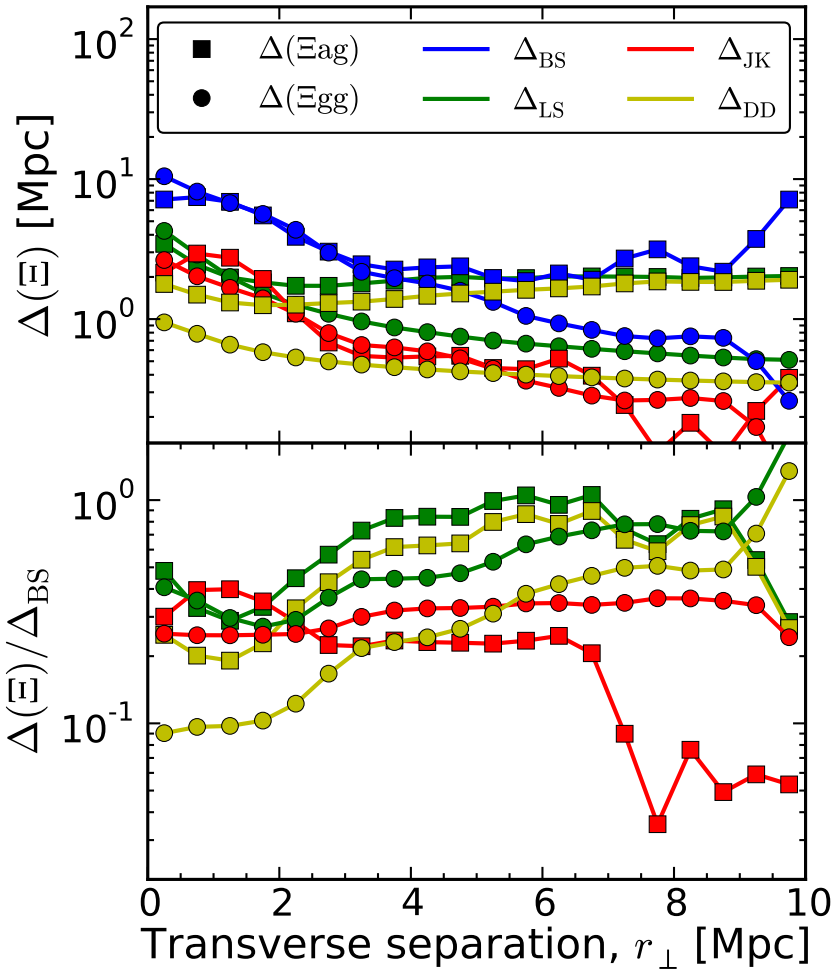

Figure 4.15: Uncertainty estimations (squared root of variances) of the projected H I-galaxy cross-correlation ($\Xi_{\rm ag}$; squares) and galaxy auto-correlation ($\Xi_{\rm gg}$; circles) measured from our full sample as a function of separations transverse to the line-of-sight ($r_\perp$). The top panel shows the uncertainty from the Landy and Szalay (1993) analytical approximation, $\Delta_{\rm LS}$ (Equation 4.24; green lines); uncertainty from a 'jackknife' resampling, $\Delta_{\rm JK}$ (Equation 4.26; red lines); uncertainty from a 'bootstrap' resampling, $\Delta_{\rm BS}$ (Equation 4.27; blue lines); and the commonly used Poissonian uncertainty, $\Delta_{\rm DD}$ (Equation 4.25; yellow lines). The bottom panel shows the ratio between these uncertainties and $\Delta_{\rm BS}$.

$\Delta_{\rm JK}$ (at least when the number of independent fields is small, like in our case). A proper study of the error estimation for auto- (e.g. Norberg et al., 2009) and cross-correlations is beyond the scope of the present work and so we did not explore this further.

For the results of this work, we will adopt uncertainties given by $\Delta_{\rm BS}$. As has been shown, $\Delta_{\rm BS}$ gives, in general, the most conservative uncertainty estimation at all scales. An exception to this rule was found for $\xi_{\rm aa}(r_\perp, r_\parallel)$ and $\Xi_{\rm aa}(r_\perp)$, in which $\Delta_{\rm LS} > \Delta_{\rm BS}$. This is due to the combination of the special survey geometry in which $\xi_{\rm aa}$ is measured. Thus, for such a sample we adopted $\Delta_{\rm LS}$ as the uncertainty.

4.7.6 Calibration between galaxy and H I redshift frames

Before computing the final two-point correlation functions, we calibrated the redshift frames between our H I absorption systems and galaxies, using the idea presented by Rakic et al. (2011): that in an isotropic Universe, the mean H I absorption profile around galaxies should be symmetric. Thus, we measured the H I–galaxy cross-correlation using $r_{\parallel,ij} \equiv X_i - X_j$ instead of $r_{\parallel,ij} \equiv |X_i - X_j|$, and applied a constant redshift shift to all our galaxies such that the cross-correlation appears symmetric with respect to the $r_\parallel = 0$ axis at the scales involved in this analysis. This redshift shift corresponded to $+0.0002$ (smaller than the galaxy redshift uncertainty). Note that this shift has not been added to the redshifts reported in Tables A.9 to A.12. The final two-point correlation functions were still calculated using $r_{\parallel,ij} \equiv |X_i - X_j|$ in order to reduce the 'shot noise'.

4.8 Results

In this section we present the results of the two point correlation analysis, following the formalisms described in Section 4.7. We used H I and galaxy samples described in Sections 4.3, 4.4, 4.5 and 4.6, but excluding: (i) H I and galaxies falling in their respective 'c' categories (see Sections 4.4.1 and 4.5.3); (ii) H I and galaxies at $z < 0.01$ and at $z > 1.3$; (iii) H I systems at redshifts within 5000 km s^{-1} the

Table 4.6: Summary of the 'Full Sample' used for the cross-correlation analysis, as a function of $r_\perp$.

	< 2 Mpc (1)	< 10 Mpc (2)	< 50 Mpc (3)	Total (4)
Galaxies	1354	6871	19509	17509
'SF'	997	4756	9963	8293
'non-SF'	193	779	2011	1743
H I	...	...	...	654
'strong'	...	...	...	165
'weak'	...	...	...	489

(1): Number of galaxies at transverse distances $r_\perp < 2$ Mpc from a QSO LOS. (2) Number of galaxies at transverse distances $r_\perp < 10$ Mpc from a QSO LOS. (3) Number of galaxies at distances $r_\perp < 50$ Mpc from a QSO LOS. (4) Total number of galaxies and H I absorption systems in the 'Full Sample'. Note that the vast majority of galaxies in the triple QSO field Q0107 have been counted three times in columns (1), (2) and (3).

redshift of the QSO in which the absorption line was observed; and (iv) galaxies at projected distances greater than 50 Mpc from the centre of their closest field. We will refer to this sample as the 'Full Sample', which comprises: 654 H I absorption systems, of which, 165 are classified as 'strong' and 489 as 'weak' (see Section 4.5.6 for definitions); and 17509 galaxies, of which, 8293 are classified as 'SF' and 1743 as 'non-SF' (see Section 4.6.1 for definitions).

Table 4.6 summarizes relevant information regarding our 'Full Sample'. The following results were computed with random samples $200\times$ and $20\times$ larger than the real H I and galaxy samples, respectively. Even though we have galaxies up to 50 Mpc from the QSO LOS, we will focus only on clustering at scales $r_\perp < 10$ Mpc, as at larger scales our results get considerably noisier. Galaxies at $r_\perp > 10$ Mpc are still used for the galaxy–galaxy auto-correlation though. In the case of the H I–H I auto-correlation, we only focus on scales $r_\perp < 2$ Mpc as we have no data sampling larger scales.

4.8.1 Two-dimensional correlations

Full Sample

Figure 4.16 shows the two-dimensional correlation functions (top panels) and their respective uncertainties (bottom panels) for our 'Full Sample' of H I-absorption systems and galaxies. The first three panels, from left to right, show the H I–galaxy cross-correlation ($\xi^{\rm LS}_{\rm ag}$; Equation 4.8), the galaxy–galaxy auto-correlation ($\xi^{\rm LS}_{\rm gg}$; Equation 4.4) and the H I–H I auto-correlation ($\xi^{\rm LS}_{\rm aa}$; Equation 4.6), respectively. We see that the amplitudes of $\xi_{\rm ag}$ and $\xi_{\rm aa}$ are comparable (within uncertainties), whereas the amplitude of $\xi_{\rm gg}$ is greater than these two (see also Table 4.7). Also, the fact that both $\xi_{\rm gg}$ and $\xi_{\rm ag}$ peak at the smallest separations confirms that the redshift frames for H I absorption systems and galaxies are self consistent (by construction; see Section 4.7.6). The decrease of the $\xi_{\rm aa}$ signal at the smallest $r_{\parallel}$ separations is because we cannot always resolve two real absorption systems separated by less than the typical width of an absorption feature.

Our sample of H I absorption systems is not large enough to measure $\xi_{\rm ag}$ or $\xi_{\rm aa}$ anisotropies at a high confidence level. Still, we can obtain qualitative features by looking at the corresponding 'iso-correlation' contours. We observe deviations from an isotropic signal in both $\xi_{\rm ag}$ and $\xi_{\rm gg}$. Apart from a decrease of the $\xi_{\rm aa}$ signal at the smallest $r_{\parallel}$ separations, we do not see significant anisotropies in $\xi_{\rm aa}$. The typical uncertainty for our single galaxy redshift determination, $\Delta z_{\rm gal} \approx 0.0006/\sqrt{2}$, is equivalent to $\sim 1.7 - 1.4$ Mpc at $z = 0.1 - 0.5$, which corresponds to an 'anisotropy ratio' of $\sim 3:1$ for pixels of 0.5 Mpc each. If the observed anisotropies are dominated by redshift uncertainties, we should expect the $\xi_{\rm ag}$ contours to be consistent with this ratio (neglecting the much smaller contribution from the H I redshift uncertainty) and the $\xi_{\rm gg}$ one to be $\sim 4:1$ (greater by a factor of $\sqrt{2}$). These expectations are consistent with what we see in our 'Full Sample' for the smallest scales whereas for scales $\gtrsim 4$ Mpc the anisotropy looks somewhat reduced. We do not detect compression along the LOS at larger scales either (e.g. Kaiser, 1987). The only anisotropy observed can be fully explained by galaxy

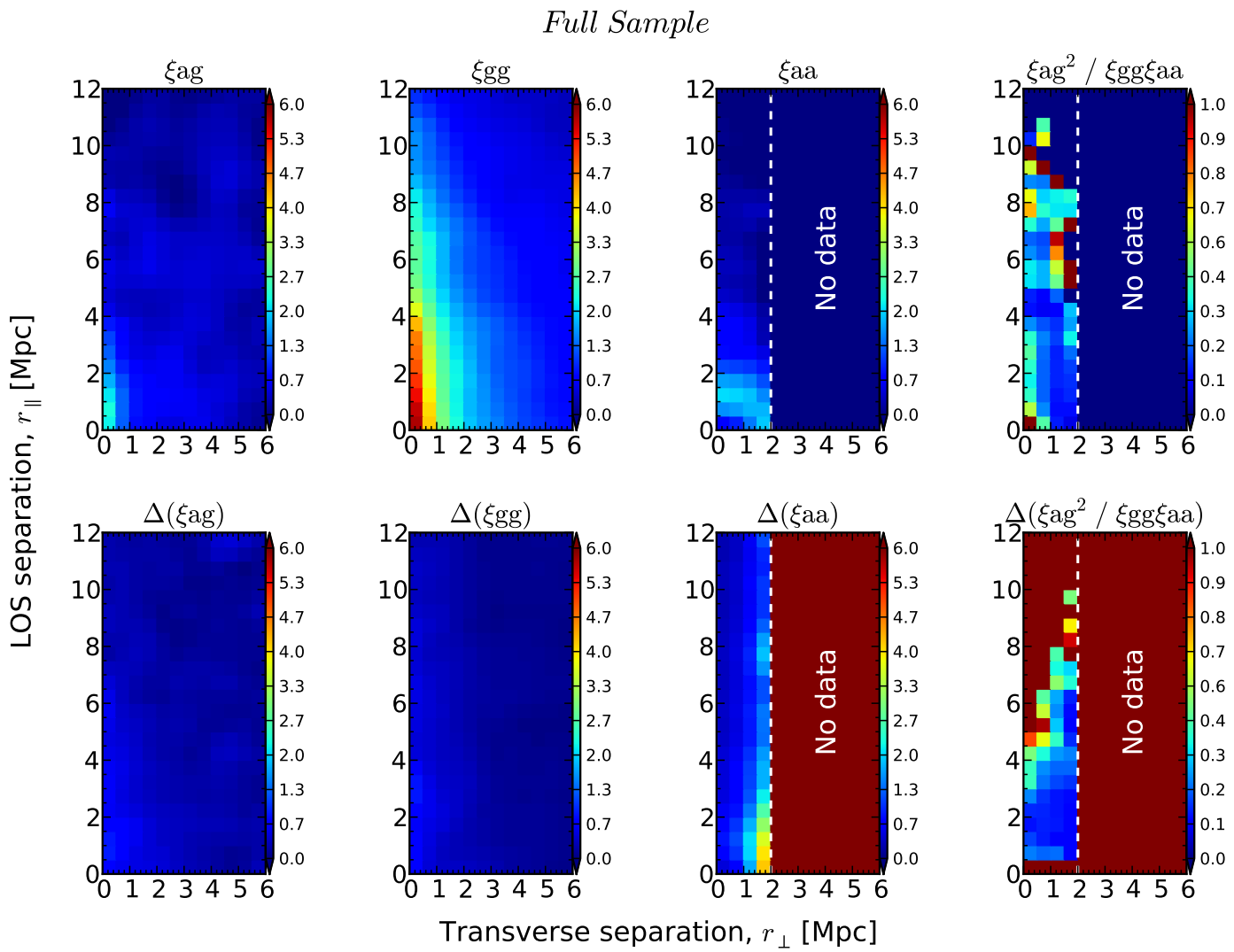

Figure 4.16: Two-dimensional correlation functions for galaxies and H I absorption system (top panels) and their respective uncertainties (bottom panels), as a function of separations along ($r_\parallel$; y-axes) and transverse to the line-of-sight ($r_\perp$; x-axes). From left to right: the galaxy-H I cross-correlation ($\xi^{\rm LS}_{\rm ag}$; equation 4.8), the galaxy-galaxy auto-correlation ($\xi^{\rm LS}_{\rm gg}$; equation 4.4), the H I-H I auto-correlation ($\xi^{\rm LS}_{\rm aa}$; equation 4.6) and the ratio, $(\xi^{\rm LS}_{\rm ag})^2/(\xi^{\rm LS}_{\rm gg}\xi^{\rm LS}_{\rm aa})$. Note that our data are not suitable for measuring the $\xi^{\rm LS}_{\rm aa}$ and $(\xi^{\rm LS}_{\rm ag})^2/(\xi^{\rm LS}_{\rm gg}\xi^{\rm LS}_{\rm aa})$ at scales $r_\perp > 2$ Mpc. The correlation functions in this figure were calculated using an arbitrary binning of 0.5 Mpc with cross-pairs counts smoothed with a Gaussian filter of standard deviation of 0.5 Mpc along both directions. See Sections 4.7.1 and 4.8.1 for further details.

redshift uncertainties.

The fourth panel of Figure 4.16 shows the ratio, $(\xi_{\rm ag}^{\rm LS})^2/(\xi_{\rm gg}^{\rm LS}\xi_{\rm aa}^{\rm LS})$. We see that the majority of the bins at smallest scales have values $(\xi_{\rm ag})^2/(\xi_{\rm gg}\xi_{\rm aa}) < 1$. Contrary to what is usually assumed, this result suggests that the population of H I absorption systems (as a whole) and galaxies do not linearly trace the same underlying dark matter distributions (see Section 4.7.4). Note however that the different redshift uncertainties and peculiar velocities of H I absorption systems and galaxies will smooth the correlation signals in different ways, which makes the interpretation of $(\xi_{\rm ag}^{\rm LS})^2/(\xi_{\rm gg}^{\rm LS}\xi_{\rm aa}^{\rm LS})$ not straightforward.

In the following, we will split the H I absorber sample into 'strong' ($N_{\rm HI} \geq 10^{14}$ cm^{-2}) and 'weak' ($N_{\rm HI} < 10^{14}$ cm^{-2}), and the galaxy sample into 'SF' and 'non-SF'. In this way we can isolate the contribution of each sub-population of H I and galaxies to the correlation functions, and to the $(\xi_{\rm ag})^2/(\xi_{\rm gg}\xi_{\rm aa})$ ratio.

'Strong' H I systems and 'SF' galaxies

Figure 4.17 is analogous to Figure 4.16 but for 'strong' H I systems ($N_{\rm HI} \geq 10^{14}$ cm^{-2}) and 'SF' galaxies. We see that in this case the $\xi_{\rm ag}$, $\xi_{\rm gg}$ and $\xi_{\rm aa}$ are all comparable within the errors (see also Table 4.7). Anisotropy signals behave in the same way as for our 'Full Sample', i.e. dominated by our galaxy redshift uncertainty and with no detected compression along the LOS at large scales.

In this case, the ratio $(\xi_{\rm ag})^2/(\xi_{\rm gg}\xi_{\rm aa})$ seems consistent with 1, suggesting that 'SF' galaxies and $N_{\rm HI} \geq 10^{14}$ cm^{-2} systems do trace the same underlying dark matter distribution. The comparable clustering amplitudes may also indicate that they typically belong to dark matter halo of similar masses. We will address these points more quantitatively in Section 4.8.2

'Strong' H I systems and 'non-SF' galaxies

Figure 4.18 shows the correlation functions for 'strong' H I systems ($N_{\rm HI} \geq 10^{14}$ cm^{-2}) and 'non-SF' galaxies. In this case, $\xi_{\rm ag}$, $\xi_{\rm gg}$ and $\xi_{\rm aa}$ are all comparable within the errors (see also Table 4.7), but $\xi_{\rm gg}$ appears systematically larger. As before, the

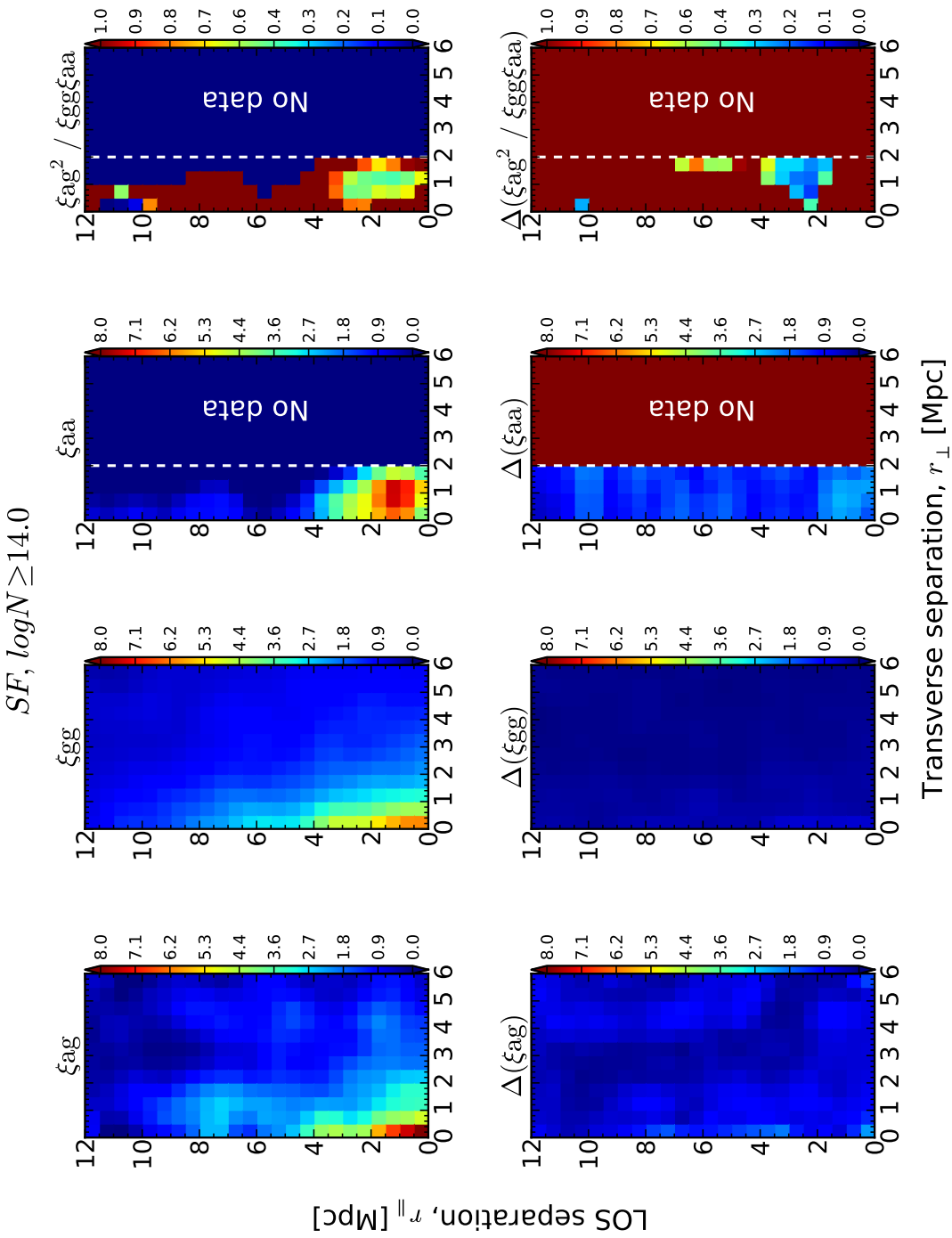

Figure 4.17: Same as Figure 4.16 but for 'SF' galaxies and H I absorption systems with $N_{\rm HI} \geq 10^{14}$ cm^{-2}.

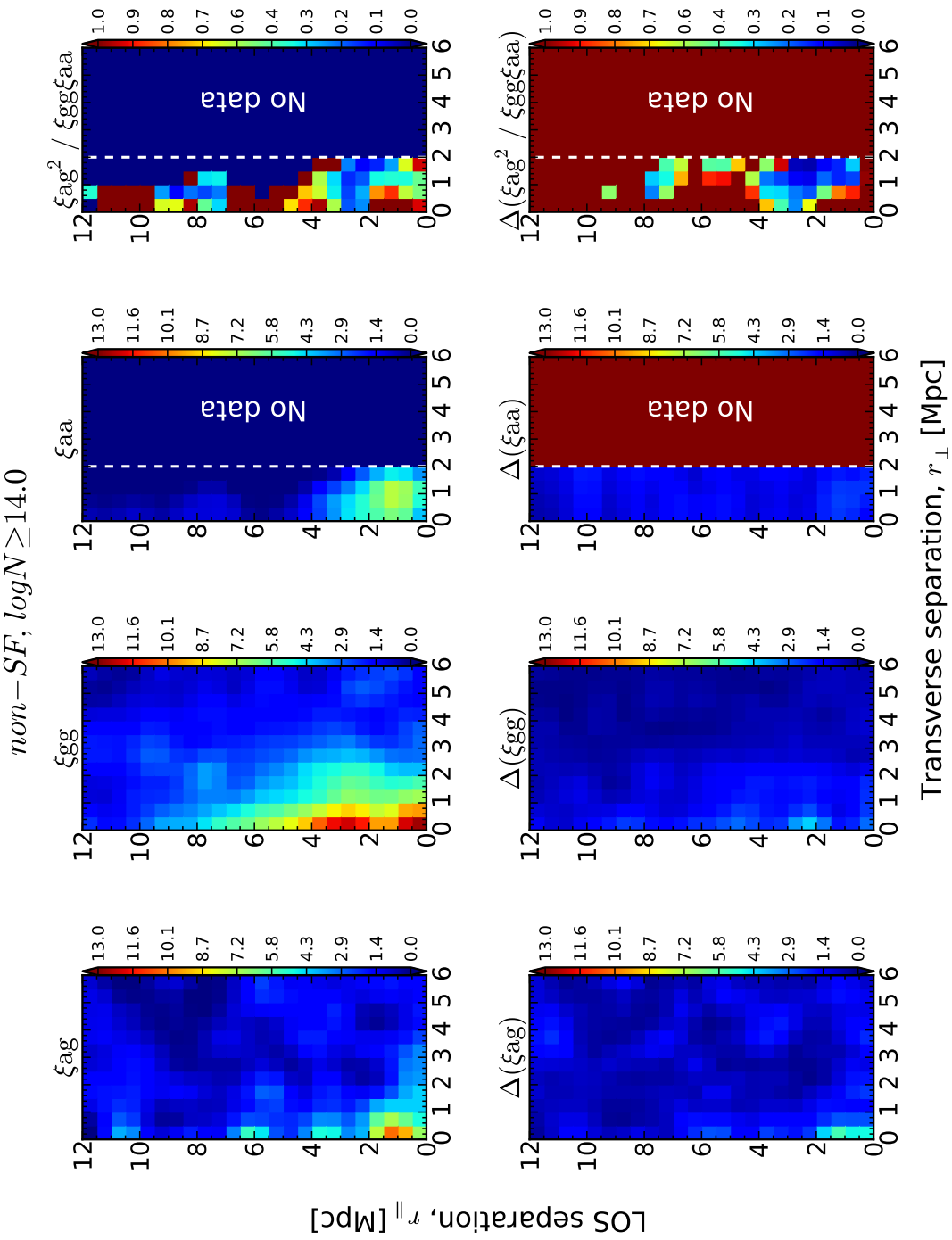

Figure 4.18: Same as Figure 4.16 but for 'non-SF' galaxies and H I absorption systems with $N_{\rm HI} \geq 10^{14}$ cm^{-2}.

anisotropy is dominated by the galaxy redshift uncertainty and no (significant) compression along the LOS at large scales is detected.

Interestingly, there is a displacement of the $\xi_{\rm ag}$ peak relative to the smallest bin. This signal also appears symmetric with respect to the $r_\parallel = 0$ axis, after computing $\xi_{\rm ag}$ using $r_{\parallel,ij} \equiv X_i - X_j$ (not plotted) instead of $r_{\parallel,ij} \equiv |X_i - X_j|$. We also checked that the signal remained using only 'a' labelled H I systems and galaxies. This suggest that this feature might be real. We note that a similar (although more uncertain than ours) feature was observed by Wilman et al. (2007), from their observation of the H I-'absorption-line-dominated galaxy' cross-correlation (see their figure 4). Pierleoni et al. (2008) also reported a similar signal from hydrodynamical simulations (see their figure 7), although their samples of H I and galaxies are not directly comparable to our 'strong' and 'non-SF' ones. A more detailed comparison between our results and those from previous studies will be presented in Section 4.9.1.

The ratio $(\xi_{\rm ag})^2/(\xi_{\rm gg}\xi_{\rm aa})$ seems also consistent with 1, which suggests that 'non-SF' galaxies and $N_{\rm HI} \geq 10^{14}$ cm^{-2} systems trace the same underlying dark matter distribution linearly.

Comparing Figures 4.17 and 4.18 we see that $\xi_{\rm gg}$ for 'non-SF' galaxies is larger than that of 'SF' galaxies (as has been shown by many authors). Given that $\xi_{\rm aa}$ is the same in both cases, one would expect $\xi_{\rm ag}$ to be also larger for 'non-SF' than that of 'SF' galaxies. Although within uncertainties our results indicate that the $\xi_{\rm ag}$ amplitude is independent of galaxy type, we do see a somewhat larger cross-correlation signal for 'non-SF' galaxies (see Table 4.7). We will address these points more quantitatively in Section 4.8.2.

'Weak' H I systems and galaxies

Figures 4.19 and 4.20 show the two-dimensional correlation functions for 'weak' H I absorption systems ($N_{\rm HI} < 10^{14}$ cm^{-2}) and 'SF' and 'non-SF' galaxies, respectively. These results are dramatically different than those for 'strong' H I systems and galaxies. In particular, $\xi_{\rm ag}$ is significantly weaker than $\xi_{\rm gg}$ but also weaker

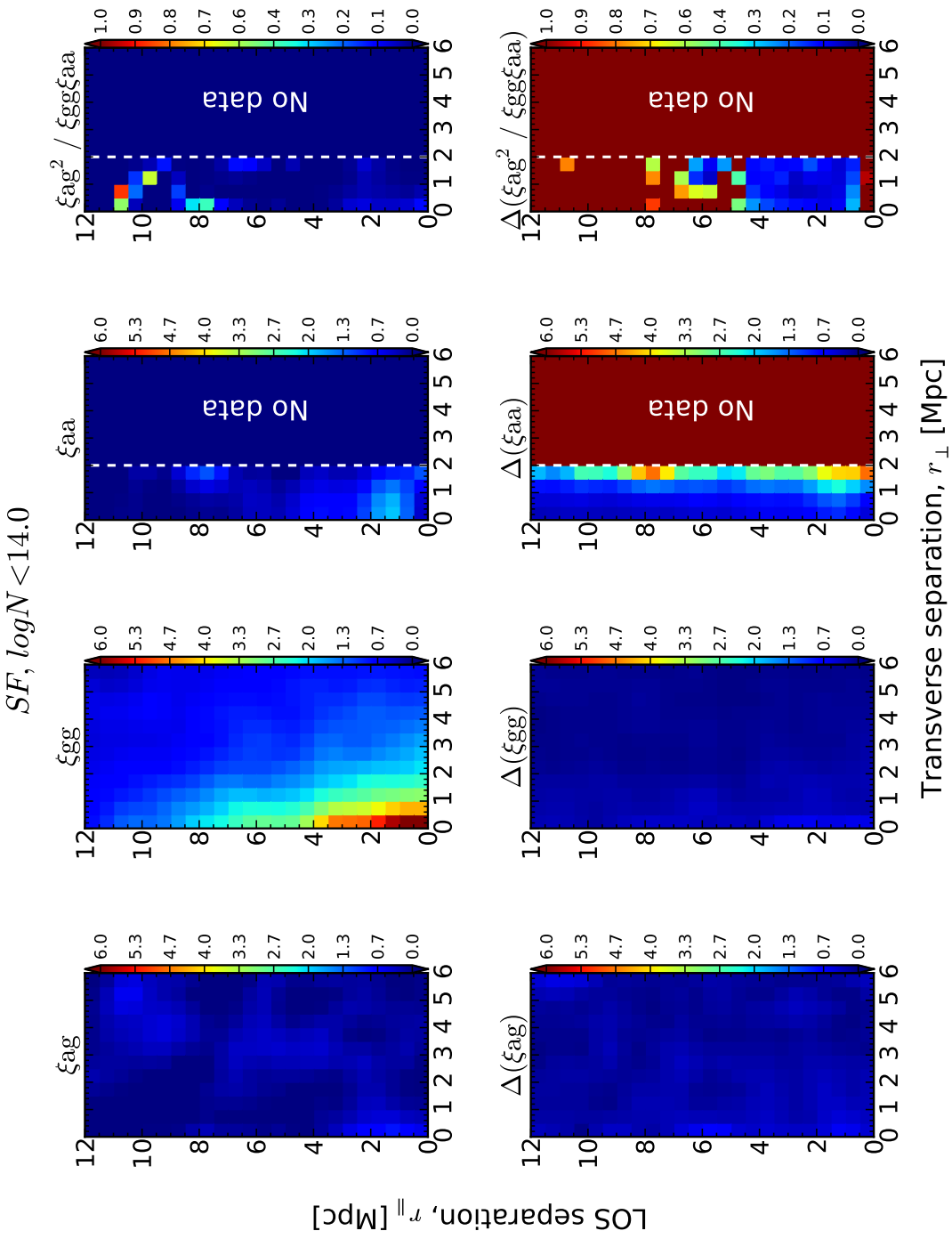

Figure 4.19: Same as Figure 4.16 but for 'SF' galaxies and H I absorption systems with $N_{\rm HI} < 10^{14}$ cm^{-2}.

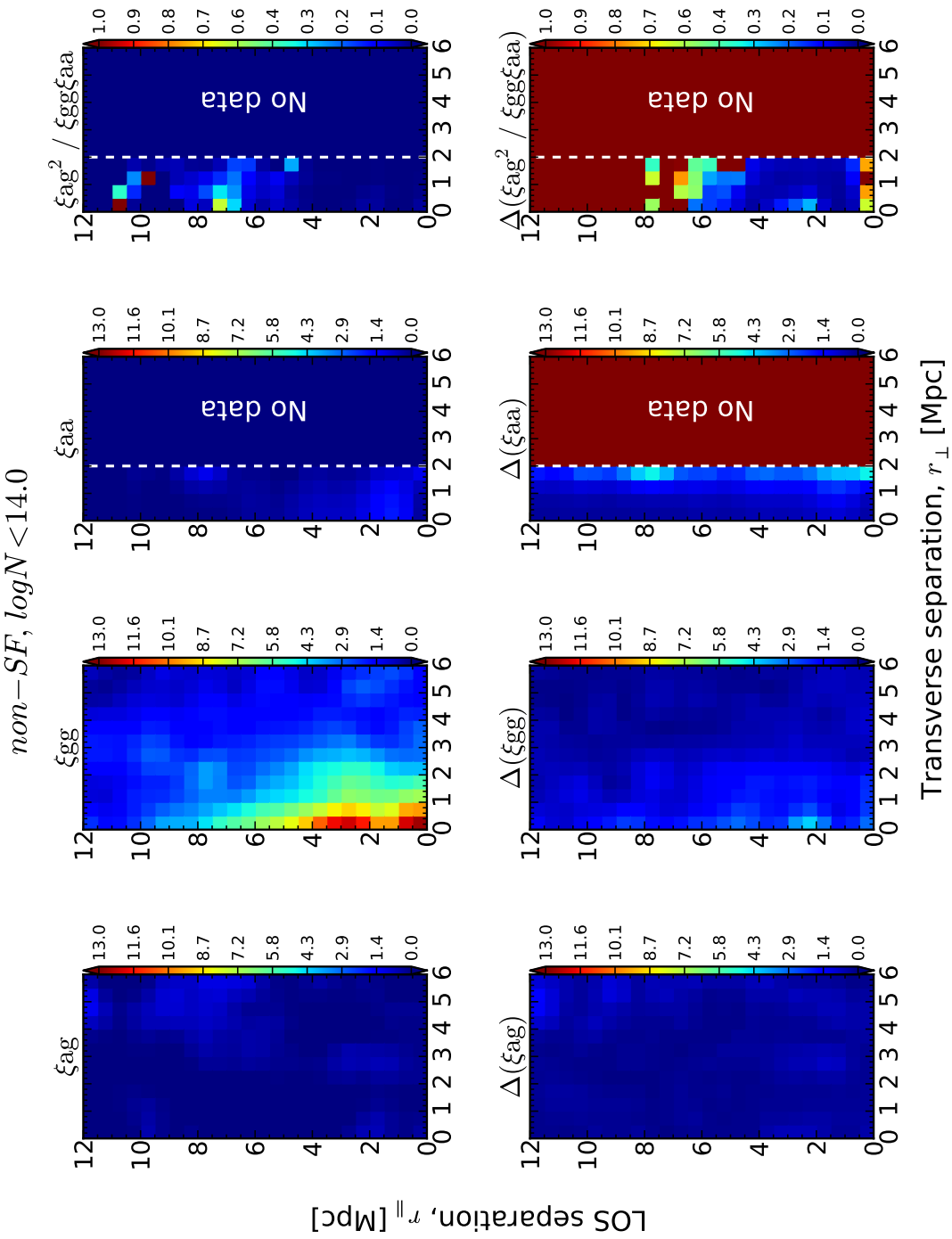

Figure 4.20: Same as Figure 4.16 but for 'non-SF' galaxies and H I absorption systems with $N_{\rm HI} < 10^{14}$ cm^{-2}.

Table 4.7: Strength of the two-dimensional correlations, $\xi(r_\perp, r_\parallel)$, at their peaks.[a]

	$\xi_{\rm ag}^{\rm peak}$	$\xi_{\rm gg}^{\rm peak}$	$\xi_{\rm aa}^{\rm peak}$
Full Sample	2.3 ± 0.9	5.7 ± 0.7	2.1 ± 0.9
'Strong'–'SF'	8.3 ± 2.2	6.1 ± 0.6	7.5 ± 2.3
'Strong'–'non-SF'	10.3 ± 5.6	12.6 ± 3.0	7.5 ± 2.3
'Weak'–'SF'	0.9 ± 0.6	6.1 ± 0.6	1.9 ± 0.9
'Weak–'non-SF'	0.6 ± 0.5	12.6 ± 3.0	1.9 ± 0.9

[a] Note that peaks are not necessarily at the smallest scale bins (see Figures 4.16 to 4.20). Also note that the different redshift uncertainties and peculiar velocities of H I absorption systems and galaxies will smooth the correlation signals in different ways, which makes a direct comparison between them not straightforward.

than $\xi_{\rm aa}$. Consequently, the ratios $(\xi_{\rm ag})^2/(\xi_{\rm gg}\xi_{\rm aa})$ are both smaller than one. This is a very strong indication (given the comparatively smaller uncertainties) that the underlying baryonic matter distributions giving rise to 'weak' H I absorption systems and galaxies are not linearly dependent. Given that the signal in the $\xi_{\rm ag}$ is marginally consistent with zero, we do not observe anisotropies either.

To summarize, 'strong' systems and galaxies are consistent with tracing the same underlying dark matter distribution linearly, whereas 'weak' systems are not. Therefore, the fact that $(\xi_{\rm ag})^2/(\xi_{\rm gg}\xi_{\rm aa}) < 1$ in the 'Full Sample' should be primarily driven by the presence of H I systems with $N_{\rm HI} < 10^{14}$ cm^{-2}. We also note that the amplitude of $\xi_{\rm aa}$ is weaker for 'weak' systems than that of 'strong' systems.

As mentioned, because redshift uncertainties and peculiar velocities affect $\xi_{\rm ag}$, $\xi_{\rm gg}$ and $\xi_{\rm aa}$ in different ways, the interpretation of the two-dimensional $(\xi_{\rm ag})^2/(\xi_{\rm gg}\xi_{\rm aa})$ is not straightforward. In the following we present the results for the projected correlation functions, which are not affected by velocity distortions along the LOS.

4.8.2 Correlations projected along the line-of-sight

Full Sample

Figure 4.21 shows the projected (along the LOS; see Equation 4.16) correlation functions divided by the transverse separation, $\Xi(r_\perp)/r_\perp$, for our 'Full Sample' of H I absorption systems and galaxies. Different symbols/colors show our different measurements: the blue squares correspond to the H I–galaxy cross-correlation ($\Xi_{\rm ag}$), the black circles to the galaxy–galaxy auto-correlation ($\Xi_{\rm gg}$), and the red triangles to the H I–H I auto-correlation ($\Xi_{\rm aa}$; slightly shifted along x-axis for the sake of clarity). The lines correspond to the best power-law fits (Equation 4.20) to the data, from a non-linear least squares analysis. The parameters r_0 and γ correspond to those of the real-space correlation function, $\xi(r)$, when described as a power-law of the form presented in Equation 4.19. Uncertainties in these fits include the variances and covariances of both parameters. From this figure we see that a power-law fit is a good description of the data, hence justifying the use of Equations 4.19 and 4.20.[18] Table 4.8 summarizes the best power-law fit parameters for our different samples.

We find that $\xi_{\rm ag}(r)$ has a correlation length of $r_0^{\rm ag} = 1.6 \pm 0.2$ Mpc and slope $\gamma^{\rm ag} = 1.4 \pm 0.1$, whereas $\xi_{\rm gg}(r)$ and $\xi_{\rm aa}(r)$ have correlation lengths of $r_0^{\rm gg} = 3.9 \pm 0.1$ Mpc and $r_0^{\rm aa} = 0.3 \pm 0.3$ Mpc, and slopes $\gamma^{\rm gg} = 1.7 \pm 0.1$, $\gamma^{\rm aa} = 1.1 \pm 0.1$, respectively. Thus, the clustering of H I absorption systems and galaxies is weaker than the clustering of galaxies with themselves, and that the clustering of H I systems with themselves is weaker still. We also see that the slopes are inconsistent with each other at the 1σ confidence level (c.l.), which is in tension with the assumption that these objects trace the same underlying dark matter distribution linearly (see Section 4.7.4). Moreover, the slope of the $\xi_{\rm aa}(r)$ being consistent with $\gamma = 1$ indicates that this distribution is at the limit in which the methodology adopted here is valid (see Section 4.7.3).

[18] We note that there might be some tension in fitting $\Xi_{\rm gg}$ with a single power-law function. We did not explore more complicated fits in order to keep the analysis and further comparisons as simple as possible.

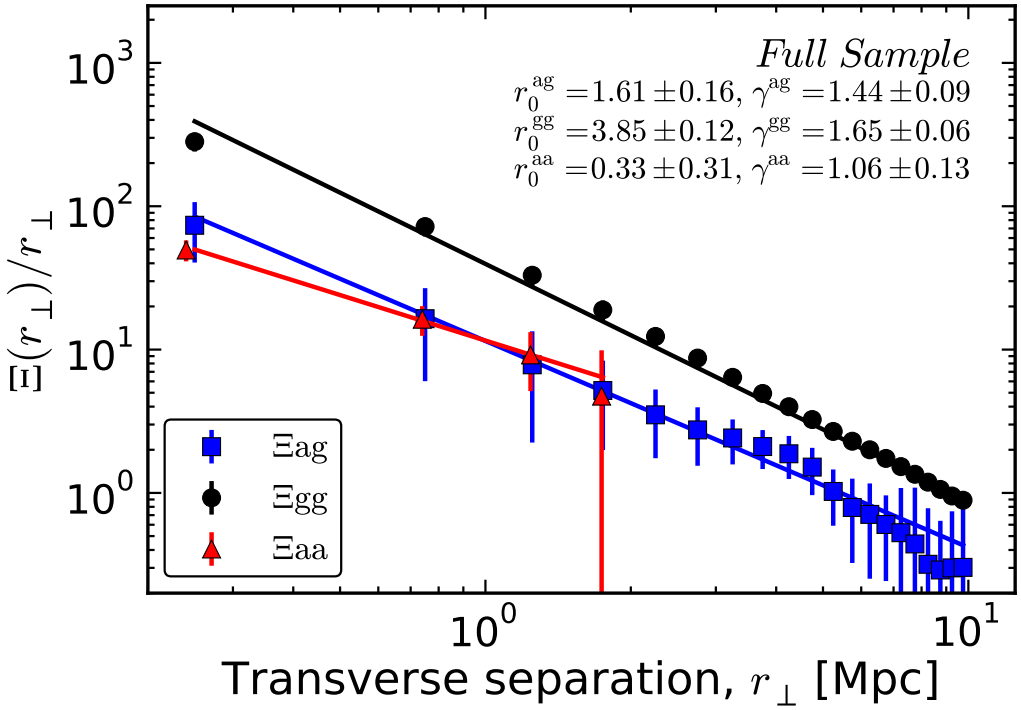

Figure 4.21: Projected (along the line-of-sight; see equation 4.16) correlation functions divided by the transverse separation, $\Xi(r_\perp)/r_\perp$, for our 'Full Sample' of galaxies and H I absorption system. Different symbols/colors show to our different measurements: the blue squares correspond to the galaxy-H I cross-correlation (Ξ_{ag}); the black circles to the galaxy-galaxy auto-correlation (Ξ_{gg}); and the red triangles to the H I-H I auto-correlation (Ξ_{aa}; slightly shifted along x-axis for the sake of clarity). The lines correspond to the best power-law fits (Equation 4.20) to the data, from a non-linear least squares analysis. The parameters r_0 and γ correspond to those of the real-space correlation function, $\xi(r)$, when described as a power-law of the form presented in Equation 4.19. Note that points and uncertainties are both correlated, and that uncertainties smaller than the symbols are not shown.

As was the case for the two-dimensional results, in the following we will split the H I and galaxy samples into 'weak' and 'strong', and 'SF' and 'non-SF', respectively, in order to isolate different contributions from these sub-populations into the observed correlations.

'Strong' H I systems and 'SF' galaxies

The top left panel of Figure 4.22 shows the projected correlation functions for our 'strong' H I systems ($N_{\rm HI} \geq 10^{14}$ cm^{-2}) and 'SF' galaxies. In this case we find that the $\xi_{\rm ag}(r)$ has a correlation length of $r_0^{\rm ag} = 3.8 \pm 0.2$ Mpc and slope $\gamma^{\rm ag} = 1.7 \pm 0.1$, whereas $\xi_{\rm gg}(r)$ and $\xi_{\rm aa}(r)$ have correlation lengths of $r_0^{\rm gg} = 3.9 \pm 0.1$ Mpc and $r_0^{\rm aa} = 3.1 \pm 0.7$ Mpc, and slopes $\gamma^{\rm gg} = 1.6 \pm 0.1$, $\gamma^{\rm aa} = 1.3 \pm 0.4$, respectively (see also Table 4.8). Thus, all have correlation lengths and slopes agreeing with each other at the 1σ c.l.. The fact that all have comparable correlation lengths and slopes supports the hypothesis that 'strong' H I systems and 'SF' galaxies do trace the same underlying dark matter distribution.

'Strong' H I systems and 'non-SF' galaxies

The bottom left panel of Figure 4.22 shows the projected correlation functions for our 'strong' H I systems ($N_{\rm HI} \geq 10^{14}$ cm^{-2}) and 'non-SF' galaxies. In this case we find that $\xi_{\rm ag}(r)$ has a correlation length of $r_0^{\rm ag} = 4.0 \pm 0.3$ Mpc and slope $\gamma^{\rm ag} = 1.7 \pm 0.1$, whereas $\xi_{\rm gg}(r)$ has a correlation length of $r_0^{\rm gg} = 6.2 \pm 0.2$ Mpc and slope $\gamma^{\rm gg} = 1.6 \pm 0.1$ (see also Table 4.8). The parameters for $\xi_{\rm aa}(r)$ are the same than in the previous case (see Section 4.8.2). The fact that the slopes are all consistent supports the idea that 'strong' H I systems and 'non-SF' galaxies also trace the same underlying dark matter distribution. This is an expected result in view of what was observed for the case of 'strong' H I systems and 'SF' galaxies, and because it is well known that 'SF' and 'non-SF' do trace the same underlying dark matter distribution as well. We also see that the galaxy–galaxy auto-correlation is significantly larger than the H I–galaxy cross-correlation and the H I–H I auto-correlation. The most simple explanation for such a difference

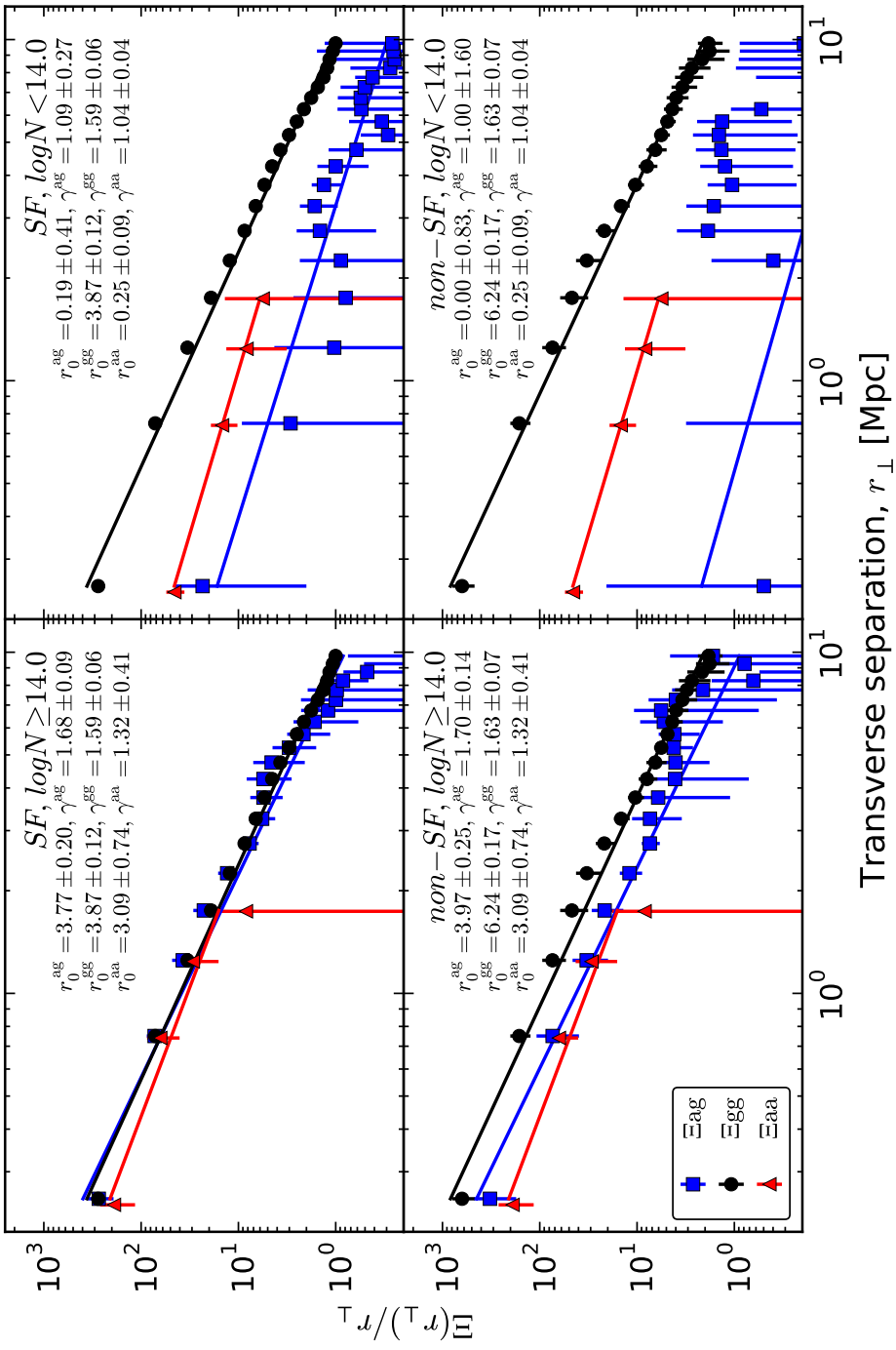

Figure 4.22: Same as Figure 4.21 but for our different subsamples: 'SF' galaxies and 'strong' ($N_{\rm HI} \geq 10^{14}$ cm^{-2}) H I absorption systems (top left); 'SF' galaxies and 'weak' ($N_{\rm HI} < 10^{14}$ cm^{-2}) H I absorption systems (top right); 'non-SF' galaxies and 'strong' H I absorption systems (bottom left); and 'non-SF' galaxies and 'weak' H I absorption systems (bottom right). Note that points and uncertainties are both correlated, and that uncertainties smaller than the symbols are not shown.

is that the linear bias (see Section 4.7.4) of 'non-SF' is greater than that of 'SF' galaxies. This has been commonly interpreted as 'non-SF' galaxies belonging, on average, to more massive dark matter haloes than 'SF' galaxies. The fact that the correlation length for the 'strong' H I–galaxy cross-correlation is (marginally) larger for 'non-SF' than 'SF' galaxies is also expected because the H I population is the same in both cases. However, we will see in Section 4.8.2 that this length is smaller than what is expected from the linear dependence hypothesis.

'Weak' H I systems and galaxies

The top right panel of Figure 4.22 shows the projected correlation functions for our 'weak' H I systems ($N_{\rm HI} < 10^{14}$ cm^{-2}) and 'SF' galaxies. In this case we find that $\xi_{\rm ag}(r)$ has a correlation length of $r_0^{\rm ag} = 0.2 \pm 0.4$ Mpc and slope $\gamma^{\rm ag} = 1.1 \pm 0.3$, whereas $\xi_{\rm aa}(r)$ has a correlation length of $r_0^{\rm aa} = 0.3 \pm 0.1$ Mpc and slope $\gamma^{\rm aa} = 1.0 \pm 0.1$ (see also Table 4.8). The parameters for $\xi_{\rm gg}(r)$ are the same than in Section 4.8.2. These results are dramatically different from those involving 'strong' H I systems. In particular for the H I–galaxy cross-correlation, not only is the power-law fit questionable, but also the correlation length is smaller than both galaxy–galaxy and H I–H I auto-correlations. Moreover, the correlation length of the cross-correlation is consistent with $r_0 = 0$, i.e., no correlation.

The result for 'weak' H I systems and 'non-SF' galaxies are even more dramatic. The bottom right panel of Figure 4.22 shows the projected correlation functions for these samples. In this case we find that $\xi_{\rm ag}(r)$ has a correlation length of $r_0^{\rm ag} = 0.0 \pm 0.8$ Mpc and slope $\gamma^{\rm ag} = 1.0 \pm 1.6$. Although consistent within errors with the 'weak' H I-'SF' galaxy cross-correlation, this correlation length is even smaller. This result goes in the opposite direction to what would be expected in the case of linear dependency, because the clustering of 'non-SF' galaxies with themselves is stronger than that of 'SF'. Therefore, these results are a strong indication that 'weak' H I systems and galaxies do not trace the same underlying dark matter distribution linearly.

Table 4.8: Best fit parameters to the real-space correlation function assuming power-law of the form presented in Equation 4.19.

.

	$\xi_{\rm ag}(r)$		$\xi_{\rm gg}(r)$		$\xi_{\rm aa}(r)$	
	$r_0\,(\mathrm{Mpc})$	γ	$r_0\,(\mathrm{Mpc})$	γ	$r_0\,(\mathrm{Mpc})$	γ
Full Sample	1.6 ± 0.2	1.4 ± 0.1	3.9 ± 0.1	1.7 ± 0.1	0.3 ± 0.3	1.1 ± 0.1
'Strong'–'SF'	3.8 ± 0.2	1.7 ± 0.1	3.9 ± 0.1	1.6 ± 0.1	3.1 ± 0.7	1.3 ± 0.4
'Strong'–'non-SF'	4.0 ± 0.3	1.7 ± 0.1	6.2 ± 0.2	1.6 ± 0.1	3.1 ± 0.7	1.3 ± 0.4
'Weak'–'SF'	0.2 ± 0.4	1.1 ± 0.3	3.9 ± 0.1	1.6 ± 0.1	0.3 ± 0.1	1.0 ± 0.1
'Weak–'non-SF'	0.0 ± 0.8	1.0 ± 1.6	6.2 ± 0.2	1.6 ± 0.1	0.3 ± 0.1	1.0 ± 0.1

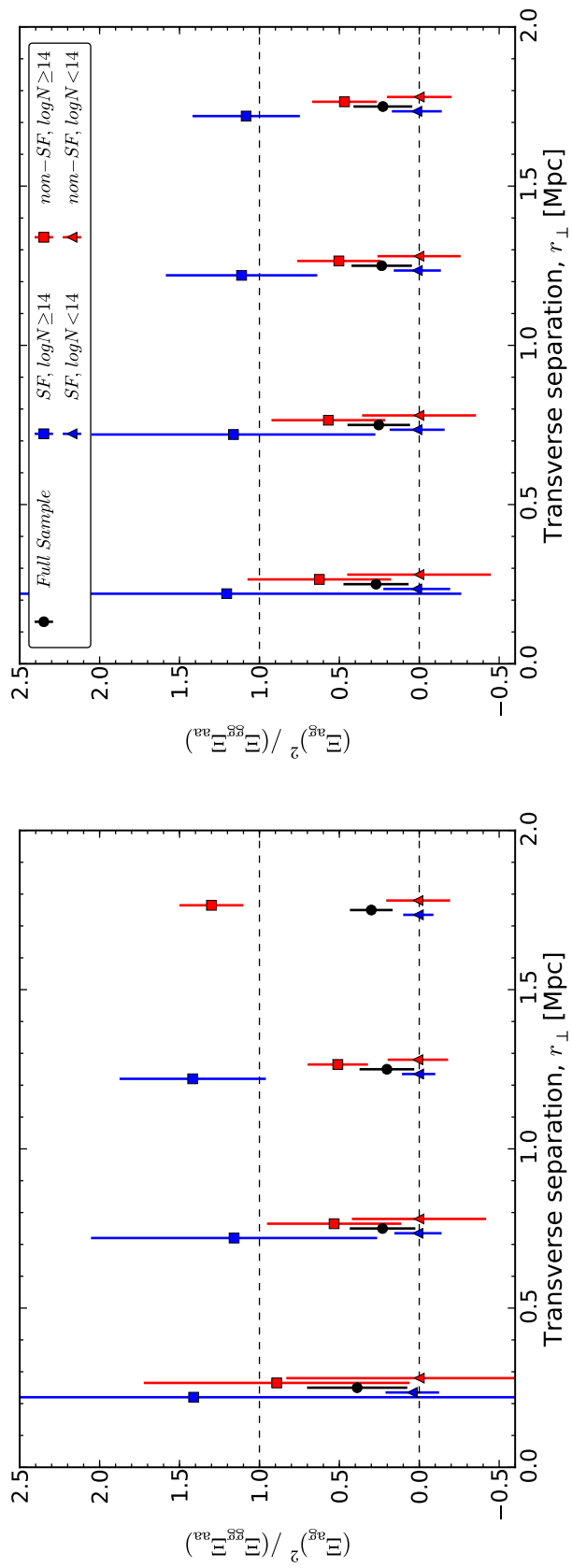

Figure 4.23: The ratio $(\Xi_{ag})^2/(\Xi_{gg}\Xi_{aa})$ as a function of transverse separation, $r_\perp$. Results from different samples of galaxies and H I absorption systems are shown by different colors/symbols. The black circles correspond to our 'Full Sample'; blue and red symbols correspond to 'SF' and 'non-SF' galaxies respectively; and squares and triangles correspond to 'strong' ($N_{\rm HI} \geq 10^{14}$ cm^{-2}) and 'weak' ($N_{\rm HI} < 10^{14}$ cm^{-2}) H I absorption systems respectively. The left panel shows the results from our adopted Gaussian smoothing of 0.5 Mpc standard deviation while the right panel shows it applying a Gaussian smoothing of 1 Mpc standard deviation. The smoothings were applied to the cross-pairs only, *before* calculating the different Ξ and the corresponding ratios (see Section 4.7). Uncertainties were obtained directly from the 'bootstrap' resampling technique of our independent fields (see Section 4.8.2.

Ratio $(\Xi_{\rm ag})^2/(\Xi_{\rm gg}\Xi_{\rm aa})$

Figure 4.23 shows the ratio $(\Xi_{\rm ag})^2/(\Xi_{\rm gg}\Xi_{\rm aa})$ for our different samples. The black circles correspond to our 'Full Sample'; blue and red symbols correspond to 'SF' and 'non-SF' galaxies respectively; and squares and triangles correspond to 'strong' ($N_{\rm HI} \geq 10^{14}$ cm^{-2}) and 'weak' ($N_{\rm HI} < 10^{14}$ cm^{-2}) H I absorption systems respectively. The left panel shows the results from our adopted Gaussian smoothing of 0.5 Mpc standard deviation. Given that the points are all correlated, we expected this ratio to be roughly independent of the scale, at least below $\lesssim 2$ Mpc. Thus, we attribute the large variation seen in the left panel of Figure 4.23 to 'shot noise' and repeated the calculation using a Gaussian smoothing of 1 Mpc standard deviation. The right panel of Figure 4.23 show the results from this last calculation. We note that the smoothings were applied to the cross-pairs only, *before* calculating the different Ξ and the corresponding ratios (see Section 4.7 for details), and that the uncertainties were obtained directly from the 'bootstrap' resampling of our independent fields (see Section 4.7.5).

These results are consistent with what we found for the two-dimensional correlations. We see that the 'Full Sample' have ratios inconsistent with 1. Taking the bin at 1.25 Mpc as representative, we find that $(\Xi_{\rm ag})^2/(\Xi_{\rm gg}\Xi_{\rm aa}) \approx 0.2 \pm 0.2$, which gives a high significance confidence level (c.l; $> 3\sigma$) for ruling out the hypothesis of linear dependency between the underlying matter distribution giving rise to H I and galaxies. The same is true for our samples of 'weak' H I systems and 'SF' galaxies, for which $(\Xi_{\rm ag})^2/(\Xi_{\rm gg}\Xi_{\rm aa}) \approx 0.02 \pm 0.2$. In the case of 'weak' H I systems and 'non-SF' galaxies, we find $(\Xi_{\rm ag})^2/(\Xi_{\rm gg}\Xi_{\rm aa}) \approx 0.001 \pm 0.3$ which is also inconsistent with 1, but the significance is somewhat reduced. Apart from the fact that 'weak' systems and galaxies have this ratio inconsistent with 1, it is also interesting to note that all are consistent with 0. This result supports the conclusion that many 'weak' H I systems *are not* correlated with galaxies at scales $\lesssim 2$ Mpc.

On the other hand, in the case of 'strong' H I systems and 'SF' galaxies this ratio appears $(\Xi_{\rm ag})^2/(\Xi_{\rm gg}\Xi_{\rm aa}) \approx 1.1 \pm 0.6$. Thus, we find consistency with the linear

dependency hypothesis although with large uncertainty. The ratio for 'strong' H I systems and 'non-SF' is $(\Xi_{\rm ag})^2/(\Xi_{\rm gg}\Xi_{\rm aa}) \approx 0.5 \pm 0.3$, which is neither consistent with 1 nor 0 (at least at the 1σ c.l.). Given the large uncertainty in this case, no strong conclusion can be drawn. Still, if we believe this ratio to be < 1, it would mean that a fraction of 'non-SF' would not be correlated with 'strong' H I systems either. This fraction can be estimated from the actual value of $(\Xi_{\rm ag})^2/(\Xi_{\rm gg}\Xi_{\rm aa})$ (e.g. see Section 4.9.2).

Results assuming a fixed slope $\gamma = 1.6$

As mentioned in Section 4.7.4, if we assume that H I and galaxies do trace the same underlying dark matter distribution linearly one can use the different correlation lengths to obtain the relative linear biases between populations (e.g. Mo et al., 1993; Ryan-Weber, 2006). For this method to work, we require the slopes of the correlation functions to be the same. Even though we have shown that this assumption is not always valid (at a $> 2\sigma$ c.l.), in this section we fix the slope of the real-space correlations and repeat the analysis. We do this for illustrative purposes, so these results should not be taken as conclusive.

Figure 4.24 is the same as Figure 4.22 but using a fixed slope of $\gamma = 1.6$. Judging from the plots, the fits work reasonably well for the galaxy–galaxy auto-correlations and the H I–galaxy cross-correlations for the 'strong' H I systems, but they fail to represent the H I–H I auto-correlations and the 'weak' H I–galaxy cross-correlations. These are expected results given what we observed in the previous analysis.

The top left panel of Figure 4.24 shows the results for our samples of 'strong' H I systems and 'SF' galaxies. We see that the $\xi_{\rm ag}(r)$ has a correlation length of $r_0^{\rm ag} = 3.6 \pm 0.2$ Mpc, whereas $\xi_{\rm gg}(r)$ and $\xi_{\rm aa}(r)$ have correlation lengths of $r_0^{\rm gg} = 3.9 \pm 0.1$ Mpc and $r_0^{\rm aa} = 3.0 \pm 0.4$ Mpc respectively. All these correlation lengths are consistent with each other within the uncertainties, indicating that 'strong' H I systems and 'SF' galaxies trace the same underlying dark matter distribution linearly. In fact, the ratio $(\xi_{\rm ag})^2/(\xi_{\rm gg}\xi_{\rm aa}) \approx 1.1 \pm 0.2$. From Equation 4.23 we have

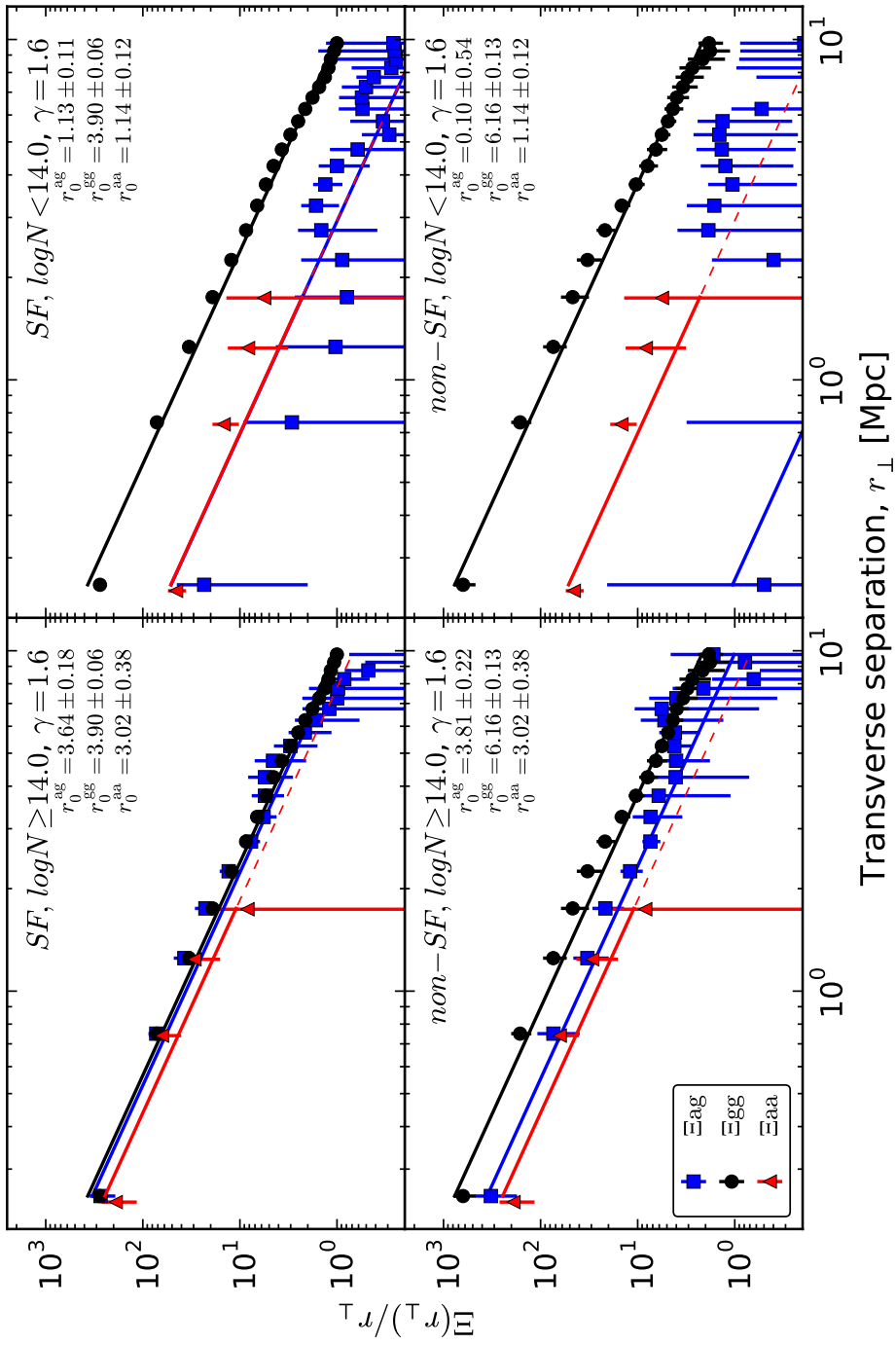

Figure 4.24: Same as Figure 4.22 but using a fixed slope, $\gamma = 1.6$, for the power-law fits. These fits are for illustrative purposes only (see Section 4.8.2 for further details). Note that points and uncertainties are both correlated, and that uncertainties smaller than the symbols are not shown.

that the relative linear biases should be,

$$\left(\frac{b_{\rm g}}{b_{\rm a}}\right) = \left(\frac{r_0^{\rm gg}}{r_0^{\rm aa}}\right)^{\frac{\gamma}{2}} = \left(\frac{r_0^{\rm gg}}{r_0^{\rm ag}}\right)^{\gamma}, \tag{4.28}$$

where $b_{\rm g}$ and $b_{\rm a}$ are the 'SF' and 'strong' H I biases respectively. Replacing the correlation lengths we get that $b_{\rm g} : b_{\rm a} \sim 1.1 - 1.2$, which implies these objects belonging to dark matter haloes of similar masses.

The bottom left panel of Figure 4.24 shows the results for our samples of 'strong' H I systems and 'non-SF' galaxies. We see that $\xi_{\rm ag}(r)$ has a correlation length of $r_0^{\rm ag} = 3.8 \pm 0.2$ Mpc, whereas $\xi_{\rm gg}(r)$ has a correlation length of $r_0^{\rm gg} = 6.2 \pm 0.1$ Mpc. The correlation length for $\xi_{\rm aa}(r)$ is the same as before. In contrast to the 'SF' case, the correlation length of 'non-SF' galaxies with themselves is significantly larger ($> 3\sigma$ c.l.). In this case, the ratio $(\xi_{\rm ag})^2/(\xi_{\rm gg}\xi_{\rm aa}) \approx 0.8 \pm 0.1$. Consequently, Equation 4.28 is at the limit of its validity. Applying this equation we find that $b_{\rm g} : b_{\rm a} \sim 1.8 - 2.2$.

The top right panel of Figure 4.24 shows the results for our samples of 'weak' H I systems and 'SF' galaxies. We see that $\xi_{\rm ag}(r)$ has a correlation length of $r_0^{\rm ag} = 1.1 \pm 0.1$ Mpc, whereas the $\xi_{\rm aa}(r)$ has $r_0^{\rm aa} = 1.1 \pm 0.1$ Mpc. The correlation length for $\xi_{\rm gg}(r)$ is the same as previously mentioned (two paragraphs above). In this case, the ratio $(\xi_{\rm ag})^2/(\xi_{\rm gg}\xi_{\rm aa}) \approx 0.3 \pm 0.1$. Consequently, Equation 4.28 should not hold. Still, if we apply this equation anyways we find that $b_{\rm g} : b_{\rm a} \sim 2.6 - 7.6$.

The bottom right panel of Figure 4.24 shows the results for our samples of 'weak' H I systems and 'non-SF' galaxies. We see that the $\xi_{\rm ag}(r)$ has a correlation length of $r_0^{\rm ag} = 0.1 \pm 0.5$ Mpc. The parameters for $\xi_{\rm aa}(r)$ and $\xi_{\rm gg}(r)$ are the same as previously mentioned (one and two paragraphs above respectively). In this case, the ratio $(\xi_{\rm ag})^2/(\xi_{\rm gg}\xi_{\rm aa}) \approx 0.001 \pm 0.1$. Consequently, Equation 4.28 should not hold either. Still, if we apply this equation we find that $b_{\rm g} : b_{\rm a} \sim 4 - 700$.

4.8.3 Consistency checks

In order to check whether our results are robust, we have repeated the analysis using only H I systems and galaxies in their respective 'a' categories (i.e., best quality; see Sections 4.4.1 and 4.5.3). We found qualitative agreement with all our previous results, but a systematic increase in the correlation amplitudes by $\lesssim 10\%$ (with larger statistical uncertainties) was observed. Such a difference is expected due to the presence of random contamination in our 'Full Sample' (e.g. catastrophic failures, missidentification of H I systems, etc.). Still, within uncertainties, the results from both analyses are fully consistent.

We also checked the effect of the Gaussian smoothing by repeating the analysis without smoothing at all (but still using the same linear grid). As expected, the new results for r_0 and γ had increased their statistical uncertainties but were all consistent with our previously reported values. We note that the slopes obtained from this new analysis were systematically larger by $\sim 10\%$ in most of the cases, but a $\sim 30\%$ increase was found for $\gamma^{\rm aa}$ and $\gamma^{\rm ag}$ in samples involving 'weak' H I systems.

4.9 Discussion

4.9.1 Comparison with previous results

In this section we compare our results with those published in recent other studies considering the H I–galaxy two-point correlation function at $z \lesssim 1$.[19]

Comparison with Ryan-Weber (2006) results ($z \sim 0$)

Ryan-Weber (2006) measured the H I–galaxy cross-correlation at $z < 0.04$ using H I data from literature (Impey et al., 1999; Penton et al., 2000; Bowen et al., 2002; Penton et al., 2004; Williger et al., 2006) and galaxy data from the H I Parkes All

[19] Note that we do not directly compare our results with those presented in Chen et al. (2005) because their data and analyses are included in Chen and Mulchaey (2009).

Sky Survey (HIPASS) (Doyle et al., 2005). Their total sample comprised 129 H I absorption systems with $10^{12.5} \lesssim N_{\rm HI} \lesssim 10^{15}$ cm^{-2}, from 27 QSO LOS; and 5317 gas-rich galaxies.

Our results are in contrast with theirs. First, they found a strong 'finger-of-god' signal in the two-dimensional H I–galaxy cross-correlation, extending up to ~ 10 h_{100}^{-1}Mpc (see their figure 3), corresponding to an 'anisotropy ratio' of $\sim 10:1$. This anisotropy signal is also larger than what they observed for the galaxy–galaxy auto-correlation, see their figure 2), meaning that it can not be explained by the galaxy redshift uncertainties. This result is in contrast to ours in that we do not see such a significant 'finger-of-god' signal, and the only anisotropy that we observe is consistent with being drawn by the galaxy redshift uncertainty.

Another difference between our results and theirs is the correlation length of the real-space correlation. They found $r_0^{\rm ag} = 7.2 \pm 1.4$ h_{100}^{-1}Mpc (which in our adopted cosmology corresponds to $r_0^{\rm ag} \approx 10.3 \pm 2.0$ h_{70}^{-1}Mpc) imposing $\gamma^{\rm ag}$ to be equal than that of the $\xi_{\rm gg}(r)$, $\gamma^{\rm ag} \equiv \gamma^{\rm gg} = 1.9 \pm 0.3$. Although the slope is marginally consistent with what we find (see Section 4.8.2), the correlation length is more than 3σ c.l. away from any of our values. If we set the slope of our correlations to be $\gamma = 1.9$, we do not find consistency either (also note that a power-law fit for such a slope is not a good representation of our data). Ryan-Weber (2006) used this result to rule out 'mini-haloes' for the confinement of H I absorption systems. In view of our new results we consider that this conclusion must be revisited (see Section 4.9.2).

Another intriguing result from Ryan-Weber (2006) is the fact the amplitude of $\xi_{\rm ag}(r)$ is greater than that of $\xi_{\rm gg}(r)$. They found a $\xi_{\rm gg}(r)$ correlation length of $r_0^{\rm gg} = 3.5 \pm 0.7$ h_{100}^{-1}Mpc ($\approx 5.0 \pm 1.0$ h_{70}^{-1}Mpc), which is somewhat larger but marginally consistent with our findings. In order to explain the larger $r_0^{\rm ag}$ value with respect to $r_0^{\rm gg}$, the $\xi_{\rm aa}(r)$ should be greater than both $\xi_{\rm gg}(r)$ and $\xi_{\rm ag}(r)$. This hypothesis is difficult to understand within the current cosmological paradigm, and in fact, it is not supported by our results on the H I–H I auto-correlation either.

We note that our surveys have important differences, in particular regarding

the galaxy samples. Given that HIPASS selected galaxies based on H I emission, it includes low-surface brightness galaxies that might be lacking in ours. The clustering of these galaxies should be lower than that of brighter galaxies in our sample though, which goes in the opposite trend than what we observe. The much lower redshift range in their sample might also have an impact in the clustering, as structures are more collapsed. This might help to increase the correlation lengths, but it should not make the $\xi_{\rm ag}(r)$ amplitude greater than $\xi_{\rm gg}(r)$ by itself. Another possibility is that this might be a case in which the ratio $(\xi_{\rm ag})^2/(\xi_{\rm gg}\xi_{\rm aa}) > 1$, meaning that the H I and galaxies observed actually correspond to the same physical objects (see Section 4.7.4). Such an effect should be most noticeable at the smallest scales, but this is in contrast to their results. Indeed, there is a flattening in their reported $\Xi_{\rm ag}(r_\perp)/r_\perp$ at $\lesssim 1\ h_{100}^{-1}$Mpc (see their figure 5) which makes the H I–galaxy cross-correlation consistent with the galaxy–galaxy auto-correlation at these scales.[20]

Comparison with Wilman et al. (2007) results ($z \lesssim 1$)

Wilman et al. (2007) measured the H I–galaxy cross-correlation at $z \lesssim 1$ using data from Morris and Jannuzi (2006). Their total sample comprised 381 H I absorption systems with $10^{13} \lesssim N_{\rm HI} \lesssim 10^{19}$ cm^{-2}, from 16 QSO LOS; and 685 galaxies at $\lesssim 2\ h_{100}^{-1}$Mpc from the QSO LOS, of which, 225 were classified as 'absorption-line-dominated' and 406 as 'emission-line-dominated'.

We find qualitative agreements with their observational results in the following senses: (i) no strong 'finger-of-god' effect is seen in the observed H I–galaxy cross-correlation; (ii) the larger the $N_{\rm HI}$, the stronger the clustering with galaxies; and (iii) no evidence for 'emission-line-dominated' galaxies cluster more strongly with H I systems than 'absorption-line-dominated' ones (in contrast to what was reported by Chen and Mulchaey, 2009, see below).

Wilman et al. (2007) also performed a comparison with a cosmological hydrodynamical simulation. In contrast to their observational results, they did

[20]Note that a flattening in $\Xi(r_\perp)/r_\perp$ means an that $\Xi(r_\perp) \propto r_\perp$.

find a strong 'finger-of-god' effect (similar to that found by Ryan-Weber, 2006) in their simulated data (see their figure 6). This prediction is in contrast to our observations, as we do not find such a strong anisotropy feature.

Comparison with Pierleoni et al. (2008) results (simulations)

Pierleoni et al. (2008) investigated the observational results from Ryan-Weber (2006) and Wilman et al. (2007) from the point of view of a cosmological hydrodynamical simulation. They selected samples of simulated H I absorption systems and galaxies trying to match those from Ryan-Weber (2006).

Contrary to the Ryan-Weber (2006) observational results (and the prediction from Wilman et al., 2007), they did not find a strong 'finger-of-god' signal in the mock H I–galaxy cross-correlation (see their figure 7), which agrees with our observational result. In contrast, they find a compression along the LOS at scales $\gtrsim 4\ h_{100}^{-1}$Mpc, similar to the expectation from the 'Kaiser effect' (Kaiser, 1987). We did not detect such a feature however.

They also found that the peak in the two-dimensional H I–galaxy cross-correlation was offset along the LOS by about $\sim 1\ h_{100}^{-1}$Mpc. A similar signal was observed in our sample of 'strong' H I systems and 'non-SF' galaxies (see Figure 4.18), but these two results are not directly comparable. Indeed, we do not observe such a feature in our 'Full Sample'. Still, we caution the reader that a ~ 1 h_{100}^{-1}Mpc displacement in the LOS direction is comparable to our galaxy redshift uncertainty ($\sim 1.4 - 1.7\ h_{70}^{-1}$Mpc), and so such a signal might get easily diluted.

Another qualitative agreement between our results and those from Pierleoni et al. (2008) is that the amplitude of the H I–galaxy cross-correlation is smaller than that of the galaxy–galaxy auto-correlation, and that the H I–H I auto-correlation one is smaller still (see their figures 3 and 9). Quantitatively, they found that $\xi_{\rm ag}(r)$ and $\xi_{\rm gg}(r)$ have correlation lengths of $r_0^{\rm ag} = 1.4 \pm 0.1\ h_{100}^{-1}$Mpc ($\approx 2.0 \pm 0.1\ h_{70}^{-1}$Mpc) and $r_0^{\rm gg} = 3.1 \pm 0.2\ h_{100}^{-1}$Mpc ($\approx 4.4 \pm 0.2\ h_{70}^{-1}$Mpc), and slopes $\gamma^{\rm ag} = 1.29 \pm 0.03$ and $\gamma^{\rm gg} = 1.46 \pm 0.03$ respectively. These values are marginally consistent with our findings (see Table 4.8). Moreover, they also predict a flattening of the $\xi_{\rm aa}(r)$

at scales $\lesssim 1\ h_{100}^{-1}$Mpc which is consistent with our observations.

Finally, we also find agreement in the sense that the amplitude of the H I–galaxy cross-correlation significantly increases for high column density absorbers, but little variation is observed for different galaxy samples selected by mass (see their figure 4). Even though we do not have direct measurements of galaxy masses in our galaxy samples, the significantly larger auto-correlation amplitude of 'non-SF' galaxies with respect to 'SF' suggests that, on average, 'non-SF' galaxies typically belong to more massive dark matter haloes than 'SF' galaxies (see also Section 4.9.2).

Comparison with Chen and Mulchaey (2009) results ($z \lesssim 0.5$)

Chen and Mulchaey (2009) measured the H I–galaxy cross-correlation at $z \lesssim 0.5$ from their own H I and galaxy survey (including data from Chen et al., 2005). Their total sample comprised 195 H I absorption systems with $10^{12.5} \lesssim N_{\rm HI} \lesssim 10^{16}$ cm^{-2}, from 3 QSO LOS; and 670 galaxies at $\lesssim 4\ h_{100}^{-1}$Mpc from the QSO LOS, of which, 222 are classified as 'absorption-line-dominated' and 448 as 'emission-line-dominated'.

In this case, we find both agreements and disagreements. Our results agree with theirs in the sense that the clustering of 'strong' H I systems ($N_{\rm HI} \geq 10^{14}$ cm^{-2}) with galaxies is stronger than that of 'weak' H I systems and galaxies (see their figure 13), and that 'strong' H I systems and 'emission-line-dominated' galaxies have comparable clustering amplitudes. However, our results disagree with their claim that 'strong' H I system cluster more strongly with 'emission-line-dominated' than with 'absorption-line-dominated' (see their figure 13). In fact, our findings are consistent with the amplitude of the H I–galaxy cross-correlation being independent of spectral type (within the statistical uncertainties). Moreover, we find that the H I–galaxy cross-correlation for 'non-SF' galaxies is systematically stronger than that of 'SF' galaxies, which is the opposite to what Chen and Mulchaey (2009) found.

Quantitatively, they reported a $\sim 6\times$ smaller clustering amplitude between

'strong' H I absorption systems and 'SF' galaxies than that of 'non-SF' galaxies with themselves, whereas we find this difference to be a factor of ~ 2 only. We note that their quoted statistical errors are Poissonian, which underestimate the true uncertainties. The Poissonian uncertainty is typically ~ 1 order of magnitude smaller than our adopted 'bootstrap' one (see Section 4.7.5). Thus, there is still room for their results to agree with ours after taking this fact into account. There is also the possibility that cosmic or sample variance is significantly affecting their results. We note that one of the three QSO LOS used by them passes at ~ 2 Mpc from the Virgo Cluster. Even though this single cluster is not likely to explain our discrepancy, any sightline passing through it is also probing an unusually high overdensity in the local Universe (which extends beyond the Virgo Cluster itself).

Comparison with Shone et al. (2010) results ($z \sim 1$)

Shone et al. (2010) measured the H I–galaxy cross-correlation at $0.7 \lesssim z \lesssim 1.5$ from their own H I and galaxy survey. Their total sample comprised 586 H I absorption systems with $10^{13.2} \lesssim N_{\rm HI} \lesssim 10^{17}$ cm^{-2}, from 2 QSO LOS; and 193 galaxies at $\lesssim 4$ h_{70}^{-1}Mpc from the QSO LOS (196 absorber–galaxy pairs used).

They found the peak in the two-dimensional H I–galaxy cross-correlation to be $\xi_{\rm ag}^{\rm peak} = 1.9 \pm 0.6$ (although displaced from the smallest separation bin by ~ 5 h_{70}^{-1}Mpc along the LOS; see their figure 12), whereas $\xi_{\rm gg}^{\rm peak} = 10.7 \pm 1.4$ for the galaxy–galaxy auto-correlation (see their figure 13). Our results agree with theirs qualitatively in the sense that the clustering of H I and galaxies is weaker than that of the galaxies with themselves.

4.9.2 Interpretation of the results

In this section we will provide our preferred interpretation for our observational results, under the assumption that our measured correlation functions are representative of the real ones, i.e. they come from fair samples of H I systems and galaxies (we do not have evidence suggesting the contrary).

Probabilistic interpretation (model independent)

The clustering analysis provides an essentially model independent statistic. As mentioned, the amplitude of the two-point correlation function corresponds to the probability excess of finding a pair with respect to the Poissonian expectation. Thus, our results point towards the following conclusions:

- The probability of finding a 'SF' galaxy at a distance $\lesssim 10$ Mpc from another 'SF' galaxy, is $\sim 2\times$ smaller than that of finding a 'non-SF' galaxy at that same distance from another 'non-SF' galaxy.

- The probability of finding a H I absorption system with $N_{\rm HI} \geq 10^{14}$ cm^{-2} at a distance $\lesssim 10$ Mpc from a 'SF' galaxy, is approximately the same than that of finding a 'SF' galaxy at that same distance from another 'SF' galaxy.

- The probability of finding a H I absorption system with $N_{\rm HI} < 10^{14}$ cm^{-2} at a distance $\lesssim 10$ Mpc from a 'SF' galaxy, is $\sim 10\times$ smaller than that of finding a 'SF' galaxy at that same distance from another 'SF' galaxy.

- The probability of finding a H I absorption system with $N_{\rm HI} \geq 10^{14}$ cm^{-2} at a distance $\lesssim 10$ Mpc from a 'non-SF' galaxy, is $\sim 2\times$ smaller than that of finding a 'non-SF' galaxy at that same distance from another 'non-SF' galaxy.

- The probability of finding a H I absorption system with $N_{\rm HI} < 10^{14}$ cm^{-2} at a distance $\lesssim 10$ Mpc from a 'non-SF' galaxy, is $\gtrsim 100\times$ smaller than that of finding a 'non-SF' galaxy at that same distance from another 'non-SF' galaxy.

- The probability of finding a H I absorption system with $N_{\rm HI} < 10^{14}$ cm^{-2} at a distance $\lesssim 2$ Mpc from another $N_{\rm HI} < 10^{14}$ cm^{-2} system is $\sim 4\times$ smaller than that of finding a $N_{\rm HI} \geq 10^{14}$ cm^{-2} system at that same distance from another $N_{\rm HI} \geq 10^{14}$ cm^{-2} system.

Any physical model aiming to explain the connection between H I absorption systems and galaxies at $z \lesssim 1$ will need to take these relations into account.

Velocity dispersion between H I and galaxies

We find that the two-dimensional H I–galaxy cross-correlations do not show detectable velocity distortions along the LOS larger than those expected from the galaxy redshift uncertainties. As mentioned, the typical uncertainty for our single galaxy redshift determination is $\Delta z_{\rm gal} \approx 0.0006/\sqrt{2}$, which is equivalent to rest-frame velocity differences of $\Delta v \sim 120-60$ km s^{-1} at $z = 0.1-1$, respectively. Any velocity dispersion between H I systems and galaxies greater than, or of the order of, this value, would have been noticeable in the two-dimensional H I–galaxy cross-correlation signals. Therefore, we conclude that the bulk of H I systems on $\sim$ Mpc scales, have little velocity dispersion ($\lesssim 120$ km s^{-1}) with respect to the bulk of galaxies. Hence, no strong galaxy outflow or inflow signal is detected in our data.

We emphasize that our results are based on H I only. Given that H I does not exclusively trace gas originating in galaxy outflows or inflows, we do not necessarily expect to find the same signatures as those traced by metals. Moreover, our results are dominated by scales somewhat larger than those typically associated to the CGM, in which the outflow or inflow signal is expected to be maximized. In view of these considerations, it is not surprising that no strong outflow or inflow signal is detected in our data.

We also emphasize that the cross-correlation analysis provides an averaged statistical result; individual galaxies having strong H I inflows/outflows might still be present, but our results indicate that these do not dominate the cross-correlation signal at $z \lesssim 1$.

Spatial distribution of H I and galaxies

The absolute and relative clustering amplitudes of our different populations of H I and galaxies can be used to give us an idea of their spatial distribution. Our

conclusions on this are as follows:

- The fact that 'strong' H I systems and 'SF' galaxies have similar amplitudes and slopes for the auto- and cross-correlation, indicates that these are distributed roughly in the same locations.

- The fact that the auto-correlation of 'non-SF' has also the same slope but a larger amplitude, indicates that there are sub-locations—within those where galaxies and 'strong' H I systems reside—with a higher density of 'non-SF' galaxies than 'SF' galaxies and/or 'strong' H I systems. This interpretation also explains the fact that the ratio $(\xi_{\rm ag})^2/\xi_{\rm gg}\xi_{\rm aa}$ for 'strong' H I systems and 'non-SF' galaxies is neither consistent with 1 nor 0 (see Section 4.9.2).

- The facts that the self-clustering of 'weak' systems is not zero, and that 'weak' H I systems and galaxies have a ratio $(\xi_{\rm ag})^2/\xi_{\rm gg}\xi_{\rm aa} \approx 0$, indicate that these are not distributed in the same locations. Therefore, there are locations containing 'weak' H I systems but roughly devoid of 'strong' H I systems and galaxies of any kind.

This picture fits well with the recent results presented in (Tejos et al., 2012) (see Chapter 2), from their study of the distribution of H I absorption systems within and around galaxy voids at $z \lesssim 0.1$. They showed that galaxy voids are not empty, and in fact contain about $\sim 20-40\%$ of H I absorption line systems with $N_{\rm HI} \gtrsim 10^{12.5}$ cm^{-2}. The remaining $\sim 60-80\%$ were found at the edges of voids, hence sharing locations with galaxies.

Even though it seems natural to identify our 'weak' systems with those systems found in galaxy voids, not all 'weak' systems need to be unassociated with galaxies. Despite the fact that Tejos et al. (2012) reported a (tentative) difference in the column density distributions between H I absorbers within and around galaxy voids (at the $\sim 2\sigma$ c.l.), they did not find sharp $N_{\rm HI}$ transitions between their samples. The most important difference came from the presence of 'extremely weak' H I systems, $N_{\rm HI} \lesssim 10^{13}$ cm^{-2}, that were present within galaxy

voids but not without (see their figures 2 and 3). Such a low column density is at the limit of our current completeness (see Section 4.5.5) and so we are not able to give confident results on the clustering of these 'extremely weak' H I systems either with themselves or with galaxies. Restricting the column density range to $10^{13} \leq N_{\rm HI} < 10^{14}$ cm^{-2}, there are $19/50 \sim 40\%$ systems within galaxy voids in the Tejos et al. (2012) sample. In the following we will estimate the fraction of 'weak' systems that could be still associated with galaxies in our current sample.

It is straightforward to show that the two-point correlation function between two populations, a and b, each one composed by sub-populations a_i where $i \in \{0, 1, ..., N_a\}$, and b_j where $j \in \{0, 1, ..., N_b\}$, respectively, is

$$\xi_{ab} = \sum_{i}^{N_a} \sum_{j}^{N_b} f_i f_j \xi_{a_i b_j} \ , \tag{4.29}$$

where $\xi_{a_i b_j}$ is the cross-correlation between the a_i and b_j sub-populations (assumed positive), and f_i and f_j are the fractions of a_i and b_j objects over the total samples a and b, respectively. Thus, if we think of 'weak' absorbers being composed by two kind of populations, using Equation 4.29 we have,

$$\xi_{aa}^{\rm weak} = f_{a_1}^2 \xi_{a_1 a_1} + f_{a_2}^2 \xi_{a_2 a_2} + 2 f_{a_1} f_{a_2} \xi_{a_1 a_2} \ . \tag{4.30}$$

If we consider a scenario in which one of these populations cluster in the same way as 'strong' H I systems ($\xi_{a_1 a_1} \equiv \xi_{aa}^{\rm strong}$) and the other is completely random ($\xi_{a_2 a_2} = \xi_{a_1 a_2} \equiv 0$), then,

$$\xi_{aa}^{\rm weak} = f_{a_1}^2 \xi_{aa}^{\rm strong} \ , \tag{4.31}$$

From this, we can estimate the fraction of 'weak' systems that could be clustered like 'strong' ones as $f_{a_1} = \sqrt{\xi_{aa}^{\rm weak} / \xi_{aa}^{\rm strong}} \sim 50\%$. We note that the assumption that one of the sub-populations has $\xi_{a_2 a_2} \equiv 0$ might be unrealistic, because $\xi_{aa}^{\rm weak}$ and $\xi_{aa}^{\rm strong}$ have marginally different slopes, and a random component does not

change the slope but only the amplitude of the correlation function. Also, if both populations lie exclusively in different locations, the cross-correlation should be $\xi_{a_1 a_2} < 0$, which makes $\xi_{a_1 a_2} \equiv 0$ unrealistic too. These two effects go in opposite directions for the final fraction estimation however, which might in the end compensate each other. With this caveat in mind, this rough estimation seems consistent with what Tejos et al. (2012) found for systems in the range $10^{13} \lesssim N_{\rm HI} \lesssim 10^{14}$ cm^{-2} ($\sim 60\%$; see above).

H I and non-star-forming galaxies

Our results point towards 'strong' H I systems and 'non-SF' galaxies having a ratio $(\Xi_{\rm ag})^2/(\Xi_{\rm gg}\Xi_{\rm aa}) \approx 0.5 \pm 0.3$, which is neither consistent with 1 nor 0 at the $\sim 1\sigma$ c.l.. In order to explain this result we will consider the presence of two types of 'non-SF' galaxies: one type (g_1) that correlates linearly with 'strong' H I absorbers, and another type (g_2) that does not. Thus,

$$\begin{aligned} \frac{(\xi_{ag_1})^2}{(\xi_{aa})(\xi_{g_1 g_1})} &\equiv 1 \ , \\ \frac{(\xi_{ag_2})^2}{(\xi_{aa})(\xi_{g_2 g_2})} &\equiv 0 \ . \end{aligned} \tag{4.32}$$

Let f_{g_1} and f_{g_2} be the fraction of 'non-SF' galaxies of type g_1 and g_2, respectively, such that $f_{g_1} + f_{g_2} = 1$. Then, from Equation 4.29 we have,

$$\begin{aligned} \xi_{ag} &= f_{g_1}\xi_{ag_1} + f_{g_2}\xi_{ag_2} \\ &= f_{g_1}\xi_{ag_1} \ , \end{aligned} \tag{4.33}$$

because $\xi_{ag_2} = 0$. Similarly,

$$\begin{aligned} \xi_{gg} &= f_{g_1}^2\xi_{g_1 g_1} + f_{g_2}^2\xi_{g_2 g_2} + 2 f_{g_1} f_{g_2}\xi_{g_1 g_2} \\ &= f_{g_1}^2\xi_{g_1 g_1} + f_{g_2}^2\xi_{g_2 g_2} \ , \end{aligned} \tag{4.34}$$

because $\xi_{g_1 g_2} \approx 0$ also. Our observational results indicate that,

$$\frac{(\xi_{ag})^2}{(\xi_{aa})(\xi_{gg})} = \alpha \, , \tag{4.35}$$

with $0 < \alpha < 1$. Combining these relations we find the following quadratic equation for f_{g_1},

$$(1 - \alpha - \alpha\beta) f_{g_1}^2 + 2\alpha\beta f_{g_1} - \alpha\beta = 0 \, . \tag{4.36}$$

where $\beta \equiv \xi_{g_2 g_2}/\xi_{g_1 g_1}$. Solving Equation 4.36 for a positive solution smaller than 1, gives us our estimation of the required fraction of 'non-SF' galaxies that are correlated with 'strong' H I systems linearly, for the given $(\xi_{ag})^2/(\xi_{aa})(\xi_{gg})$ and $\xi_{g_2 g_2}/\xi_{g_1 g_1}$ ratios.

Our proposed scenario aims to approximate what might be the case for galaxy clusters, which contain an important fraction of 'non-SF' galaxies but whose diffuse IGM or CGM can get destroyed by baryonic physics (e.g. Morris et al., 1993; Lopez et al., 2008; Padilla et al., 2009; Yoon et al., 2012). In such a case, $\xi_{g_2 g_2}/\xi_{g_1 g_1} \gg 1$ because galaxy clusters represent the most massive dark matter haloes. Measurements and predictions for the auto-correlation of galaxy clusters point towards correlation length of $r_0^{\rm cc} \sim 20 - 30$ Mpc (e.g. Colberg et al., 2000; Estrada et al., 2009; Hong et al., 2012), which would imply a $\xi_{g_2 g_2}/\xi_{g_1 g_1} \sim 10 \pm 5$ (assuming a slope of $\gamma = 1.6$). Using this value together with $(\xi_{ag})^2/(\xi_{aa})(\xi_{gg}) = 0.5 \pm 0.3$, we find the fraction $f_{g_1} \approx 0.75 \pm 0.15\%$ and consequently $f_{g_2} \approx 0.25 \pm 0.15\%$.[21]

Therefore, our results suggest that an important fraction of 'non-SF' galaxies ($\sim 60 - 90\%$) trace the same underlying dark matter distribution as 'strong' H I systems and 'SF' galaxies at scales $\lesssim 2$ Mpc. This is in contrast with what can be inferred from the results reported by Chen and Mulchaey (2009), in which 'strong' H I systems cluster more weakly with 'non-SF' than 'SF' galaxies (see

[21]Note that the functional form of the solution of Equation 4.36 gives relatively well constrained results, even for $(\xi_{ag})^2/(\xi_{aa})(\xi_{gg})$ and $\xi_{g_2 g_2}/\xi_{g_1 g_1}$ ratios with large uncertainties (as in our case).

Section 4.9.1). In such a case, $(\xi_{ag})^2/(\xi_{aa})(\xi_{gg}) \approx 0$[22], implying a fraction close to $f_{g_1} \sim 0\%$.

Our simple interpretation agrees quite well with the recent observational results presented by Thom et al. (2012). These authors found that $11/15 \sim 70\%$ of their sample of 'non-star-forming-galaxies' at low-z, have H I absorption with rest-frame equivalent widths, $W_r > 0.3$ Å (equivalent to $\gtrsim 10^{14}$ cm^{-2}), within 300 km s^{-1} from their systemic redshifts, and at impact parameters $\lesssim 200$ kpc (see their figures 2 and 3). By definition, these H I systems should be associated with the CGM of these galaxies. However, because of incompleteness in the galaxy surveys, it is not certain that this gas is purely associated to these 'non-star-forming-galaxies' (less luminous 'star-forming-galaxies' could have been missed by their target selection; e.g. Stocke et al., 2013). Still, both Thom et al. (2012) and our results point towards the conclusion that a significant fraction of 'non-star-forming-galaxies' share locations with 'strong' H I systems at scales $\lesssim 2$ Mpc. Thus, our results indicate that the 'cold gas' (traced by 'strong' H I) around 'non-SF' galaxies could be the rule rather than the exception.

Column density limit

Our results show that there is not a sharp $N_{\rm HI}$ limit determining the fate of the H I–galaxy relationship, at least in the range $10^{13} \lesssim N_{\rm HI} \lesssim 10^{16}$ cm^{-2}. Our choice of a 10^{14} cm^{-2} limit was arbitrary (see Section 4.5.6). As mentioned, when we increase the limit for defining 'strong' versus 'weak' systems from 10^{14} to $\sim 10^{15-16}$ cm^{-2} we get larger cross-correlation amplitudes and slopes (although with larger uncertainties due to the reduced number of systems above such limits in our sample) for 'strong' compared to those from 'weak' systems. Similarly, when we decrease the limit from 10^{14} to $\sim 10^{13}$ cm^{-2} we observe a decrease in the cross-correlation amplitudes and slopes of 'strong' systems. Such a behaviour can be explained by assuming that the fraction of systems that are not correlated with galaxies decrease with an increase in the minimum column density limit.

[22]Otherwise $\xi_{\rm gg}^{\rm non-SF} < \xi_{\rm gg}^{\rm SF}$, which is in contradiction with their observations.

Any change in the amplitude of the correlation functions can be understood as a change in the 'linear bias' and/or the fraction of 'random contamination' present. Changes in the slope of the correlations (like the one we have observed in this work) would necessarily require the addition of baryonic physics under our current cosmological paradigm.

Despite the fact that a 10^{14} cm^{-2} limit does not provide a sharp transition in either the correlation signals or in the spatial distribution of H I absorption systems, it might still have a physical meaning. Davé et al. (2010) used a cosmological hydrodynamical simulation to study the properties of H I absorption systems from $z = 2$ to $z = 0$. They found an interesting bimodality in the distribution of $\log N_{\rm HI}$ per unit path length at $\langle z \rangle \approx 0.25$, where $N_{\rm HI} < 10^{14}$ cm^{-2} systems are dominated by the diffuse IGM and $N_{\rm HI} > 10^{14}$ cm^{-2} are dominated by the condensed IGM associated with galaxy halos (see their figure 10).

According to the results from Davé et al. (2010), the diffuse IGM roughly follows,

$$\frac{\rho}{\bar{\rho}} \approx 50 \left(\frac{N_{\rm HI}}{10^{14}\,{\rm cm}^{-2}} \right)^{0.74} 10^{-0.37z}, \tag{4.37}$$

where $\rho/\bar{\rho}$ is the local baryonic density in units of the cosmic mean (see their equation 3 and figure 9), which gives us an idea of the overdensities involved (see also Schaye, 2001, for a similar relationship from analytical arguments). A change of one order of magnitude in column density, corresponds to a factor of ~ 5 (directly proportional) in $\rho/\bar{\rho}$, whereas a change of one unit redshift to a factor of ~ 2 (inversely proportional). Thus, a limit of 10^{14} cm^{-2} would correspond to overdensities of $\sim 50\times$ and $\sim 25\times$ the cosmic mean at $z = 0$ and $z = 1$, respectively. Similarly, limits of 10^{13} and 10^{15} cm^{-2} would correspond to $\sim 5\times$ more and less those values, respectively. We emphasize that there are large scatters involved in this relation: roughly one order of magnitude in overdensity for a fixed H I column density, and roughly half an order of magnitude in H I column density for a fixed overdensity. Therefore, from these theoretical results a sharp $N_{\rm HI}$ limit is not expected either.

Dark matter halo masses hosting H I systems and galaxies

It is common practice to compare the observed clustering amplitudes of extra-galactic objects (e.g. galaxies, galaxy clusters, IGM absorbers, etc.) with that of the expected theoretical (cold) dark matter in a given cosmological framework, in order to infer a typical dark matter halo mass for the confinement of such objects (e.g. Mo et al., 1993; Ryan-Weber, 2006). This method is model dependent and it is only applicable to narrow cosmological epochs (narrow redshift ranges).

Our sample is composed of objects at $0 \lesssim z \lesssim 1$, which corresponds to about half of the history of the Universe. Thus, a direct link between the clustering amplitudes reported in this work with a single dark matter halo mass is not meaningful. Still, simple reasoning leads to the conclusion that the typical dark matter haloes for the confinement of H I systems and galaxies, should follow the same trends as the amplitudes of their correlation functions. Therefore, the most massive ones should correspond to 'non-SF' galaxies, followed by 'SF' galaxies, 'strong' H I systems (both comparable) and 'weak' H I systems, in that same order.

Figure 4.25 is the same as Figure 4.22 with the prediction for the dark matter clustering at $z \lesssim 1$ (thick dashed line). The shaded regions enclose the expected dark matter clustering between redshift $z = 1$ (lower envelope) and $z = 0$ (upper envelope) while the dashed lines themselves correspond to the expectation at $z = 0.5$. These predictions were obtained from the dark matter power spectrum[23] provided by CAMB[24] (Lewis et al., 2000), with (thick dashed lines and dark shaded regions) and without (thin dashed lines and light shaded regions) using the non-linear corrections of Smith et al. (2003), for our adopted cosmological parameters and $\sigma_8 = 0.8$. We see that the shape of the correlations for 'strong' H I systems and galaxies are somewhat consistent with that of the predicted dark matter for the non-linear regime. Their somewhat larger amplitudes hint towards 'absolute biases' $b \gtrsim 1$. On the other hand, the shape of the 'weak' H I is marginally in disagreement with that of the dark matter expectation in the non-linear regime. In

[23]Note that the power spectrum is the Fourier transform of the correlation function (and viceversa).

[24]www.camb.info

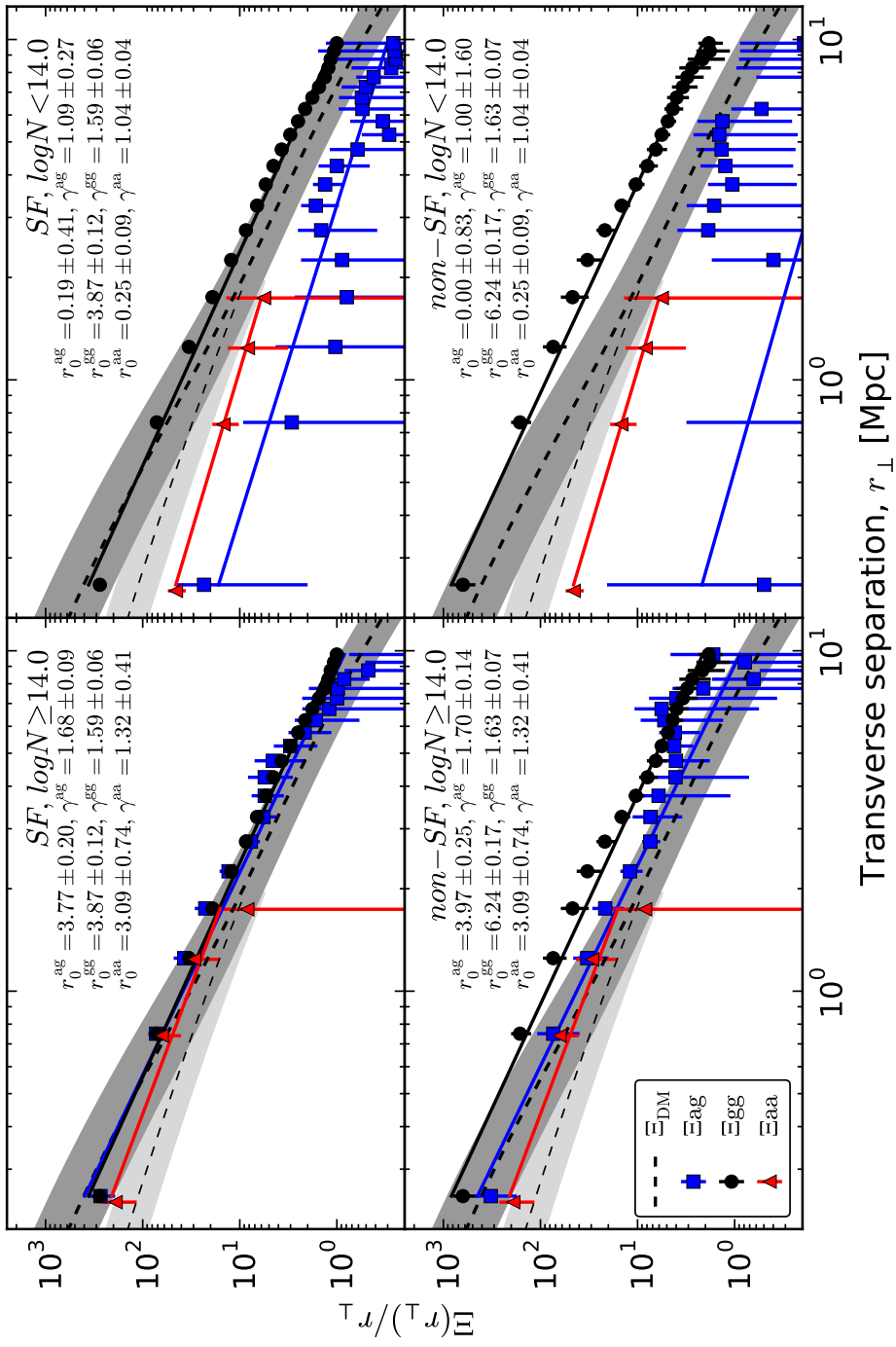

Figure 4.25: Same as Figure 4.22 with the prediction for the dark matter clustering at $z \lesssim 1$ (dashed line). The shaded regions enclose the expected dark matter clustering between redshift $z = 1$ (lower envelope) and $z = 0$ (upper envelope) while the dashed lines themselves correspond to the expectation at $z = 0.5$. These predictions were obtained from the dark matter power spectrum provided by CAMB (Lewis et al., 2000), with (thick dashed lines and dark shaded regions) and without (thin dashed lines and light shaded regions) using the non-linear corrections of Smith et al. (2003), for our adopted cosmological parameters and $\sigma_8 = 0.8$. See Section 4.9.2 for further details.

this case, the lower amplitude compared to that of the dark matter hints towards an 'absolute bias' $b < 1$. We note that for the case of 'weak' systems, a linear approximation for the dark matter clustering (i.e. neglecting the correction of Smith et al., 2003), gives a somewhat better match in terms of slopes, although still with amplitudes marginally above our observed ones. If a significant fraction of 'weak' H I systems reside within galaxy voids, a linear evolution should be expected even at $z \approx 0$. We speculate that H I systems within galaxy voids might have clustering amplitudes consistent with that of the dark matter when the Universe was much younger than $z \sim 1$, still evolving in the linear regime even at scales $\lesssim 2$ Mpc.

In view of these results, we revisit the claim by Ryan-Weber (2006) that H I absorption at $\lesssim 10^{15}$ cm^{-2} reside preferentially in dark matter halo of masses $M \sim 10^{13.6} - 10^{14.5}\ h_{100}^{-1}\mathrm{M}_{\odot}$, analogous to those of massive galaxy groups. Given the significantly lower clustering amplitude of our full sample of H I systems compared to that of galaxies, we conclude that H I absorption systems are preferentially found in dark matter haloes of masses smaller than those populated by galaxies. At most, 'SF' galaxies and 'strong' H I systems are typically found in dark matter haloes of similar masses. Moreover, a significant fraction of 'weak' H I systems might reside in underdense regions with 'absolute biases' $b < 1$.

Three types of relationships between H I and galaxies

We have reported a significant ($> 3\sigma$ c.l.) rejection of the hypothesis that H I absorption systems and galaxies (as a whole) trace the same underlying dark matter distribution linearly (see Section 4.8). We have found that this is mostly driven by H I absorption systems with column densities $N_{\mathrm{HI}} < 10^{14}$ cm^{-2} ('weak' systems), which show little (consistent with 0) correlation with galaxies. On the other hand, H I systems with $N_{\mathrm{HI}} \geq 10^{14}$ cm^{-2} ('strong' systems) are consistent with such an hypothesis. Thus, this indicates the presence of at least two types of relationships between H I and galaxies: (i) linear correlation, and (ii) no correlation.

A third type of relationship comes from the fact that at small enough scales,

H I systems and galaxies are different manifestation of the same physical object; a galaxy is also a very strong H I absorption system and, depending on the galaxy definition, the other way around also applies. Our survey was not designed for studying scales $\lesssim 0.5$ Mpc, and so it is not surprising that we do not observe a characteristic signal of a one-to-one association (see Section 4.7.4). Thus, we can not neglect the fact that this relationship exists and should be included in our interpretation. Still, the contribution of this one-to-one correlation to the total fraction of H I systems at $10^{13} \lesssim N_{\rm HI} \lesssim 10^{17}$ cm^{-2} is quite low.

This picture fits well with what was early presented by Mo (1994), and is in contrast to the commonly adopted interpretation presented by Lanzetta et al. (1995) which claims that the majority of low-z H I systems belong to the extended haloes of luminous galaxies.

4.9.3 Prospects and future work

In this section we will enumerate some of the projects that are directly linked to our current study but that we have not performed here either because of lack of observational data or limited time. We aim to address them in the near future.

Comparison with simulations

Even though many of our results are in good agreements with those presented by Pierleoni et al. (2008, see Section 4.9.1), others have not been properly compared with the predictions from simulations yet. For instance, one of our key results is the fact that 'weak' H I systems and galaxies cluster more weakly than 'weak' H I systems with themselves or than galaxies with themselves. As discussed in Section 4.9.2, this would imply that 'weak' H I systems and galaxies do not trace the same underlying matter distribution linearly. It is still to be seen if current cosmological hydrodynamical simulations can reproduce this and *all* our observational results.

Redshift evolution

A complete picture of the relationship between the IGM and galaxies requires understanding not only their statistical connection at a given epoch but also their cosmological evolution. Combining our results with those from higher redshifts ($z \sim 2-3$; e.g. Adelberger et al., 2003, 2005; Crighton et al., 2011; Rudie et al., 2012; Rakic et al., 2012; Tummuangpak et al., 2013), such an evolution can be studied. It is important to keep in mind that: (i) galaxy samples in these high-z studies are strongly biased against 'non-star-forming-galaxies', and (ii) the lower the redshift, the higher the (average) overdensity traced by a fixed $N_{\rm HI}$ limit (e.g. Schaye, 2001; Davé et al., 2010, see Equation 4.37). Thus, any evolutionary analysis has to properly take into account such differences.

We also note that because of observational limitations, the redshift range between $z \sim 1-2$ is currently unexplored for studies of the IGM–galaxy connection. This is a very important cosmological time—as it is when the star formation density starts to decline (e.g. Hopkins and Beacom, 2006)—which we hope will be covered in the near future.

Splitting by H I Doppler parameter

In our current analysis we have completely ignored the information provided by the Doppler parameters of the H I systems. Current hydrodynamical simulations suggest that above a limit of $b_{\rm HI} \sim 50$ km s^{-1}, an important fraction of H I lines trace the warm-hot intergalactic medium (WHIM) (e.g. Davé et al., 2010, see their figure 11). The WHIM is currently the best candidate to host the majority of the 'missing baryons' at low-z (e.g. Cen and Ostriker, 1999; Davé et al., 2010; Shull et al., 2012). However, because of their expected large $b_{\rm HI}$ and low $N_{\rm HI}$ ($\lesssim 10^{13}$ cm^{-2}), its direct observation through H I has been extremely difficult. In fact, H I can appear undetectable in such conditions (Savage et al., 2010). Still, the H I–galaxy cross-correlation could provide an indirect way to observe the WHIM by splitting the samples by $b_{\rm HI}$, and applying a similar reasoning than that presented

in Section 4.9.2.

Cross-correlations for the CGM

Our current statistical results seem adequate for constraining the H I–galaxy connection on scales $\sim 0.5 - 10$ Mpc. An obvious improvement would be to increase the galaxy completeness level at scales $\lesssim 0.5$ Mpc. In this way the two-point correlation function results can be directly linked to the studies of the CGM based on one-to-one absorber–galaxy associations (e.g. Prochaska et al., 2011b; Tumlinson et al., 2011; Thom et al., 2012; Werk et al., 2013; Stocke et al., 2013). Correlations between metals and galaxies will also provide a useful complement for such studies. Similarly, a better characterization of the galaxies (e.g. stellar masses, specific star-formation rates, morphology, etc.) in these samples will allow us to isolate their relative contributions (hence importance) to the observed correlation amplitudes.

'Extremely weak' H I systems

Our current data quality is not enough to observe 'extremely weak' H I systems ($N_{\rm HI} \lesssim 10^{13}$ cm^{-2}), but studying the H I–galaxy cross-correlation at such low column densities is definitively worth exploring. There is strong observational evidence that the vast majority of these absorbers reside within galaxy voids (e.g. Manning, 2002; Tejos et al., 2012). In such a case an *anti*-correlation between 'extremely weak' H I absorption systems and galaxies should be expected, but this has not been observationally confirmed (or refuted) yet. There is also the interesting possibility that these absorbers may represent a completely different type of H I absorption systems than those found co-existing with galaxies. If true, such systems are good candidates for testing our current galaxy formation paradigm (e.g. Manning, 2002, 2003).

4.10 Summary

We have presented a new optical spectroscopic galaxy survey of 2143 galaxies at $z \lesssim 1$, around 3 fields containing 5 QSO with HST UV spectroscopy.[25] These galaxies were observed by optical multi-object spectroscopy instruments such as DEIMOS, VIMOS and GMOS, and were mostly selected based on magnitude limits ($R \sim 23 - 24$ mag; no morphological criteria imposed). This selection also lead to the detection of 423 stars and 22 AGN within those fields. Out of our new 2143 galaxies, 1777 have detectable star formation activity within their past ~ 1 Gyr (referred as 'SF'), while the remaining 366 have not (referred as 'non-SF').

We have also presented a new spectroscopic survey of 669 well identified intervening H I absorption line systems at $z \lesssim 1$, observed in the spectra of 8 QSO at $z \sim 1$. These systems were detected in high-resolution UV HST spectroscopy from instruments such as COS and FOS. Out of these 669 H I systems, 173 have column densities $10^{14} \leq N_{\rm HI} \lesssim 10^{17}$ cm^{-2} (referred as 'strong'), while the remaining 496 have $10^{13} \lesssim N_{\rm HI} < 10^{14}$ cm^{-2} (referred as 'weak').

Combining these new data with previously published galaxy catalogs from the VVDS (Le Fèvre et al., 2005), GDDS (Abraham et al., 2004) and Morris and Jannuzi (2006) surveys, we have gathered a sample of 17509 galaxies with redshifts between $0.01 < z < 1.3$, and at transverse separations < 50 Mpc from their respective fields centres; and 654 H I absorption systems at redshifts between $0.01 < z < z_{\rm max}$, where $z_{\rm max}$ is the redshift corresponding to 5000 km s^{-1} bluewards the redshift of their respective QSO. Out of those 17509 galaxies, 8293 were classified as 'SF' and 1743 as 'non-SF'; while out of those 654 H I systems, 165 were classified as 'strong' and 489 as 'weak'.

Using these data, we have investigated the statistical connection between the intergalactic medium (IGM) and galaxies through a clustering analysis. This dataset is the largest sample to date for such an analysis. We presented observational results for the H I–galaxy cross-correlation and both the galaxy–galaxy

[25] Note that one of the fields has 3 QSOs.

and H I–H I auto-correlations at $z \lesssim 1$. The two-point correlation functions have been measured both along and transverse to the LOS, $\xi(r_\perp, r_\parallel)$, in a linear grid of 0.5 Mpc in both directions. We have constrained the H I–galaxy ($\xi_{\rm ag}$) and galaxy–galaxy ($\xi_{\rm gg}$) correlations at transverse scales $r_\perp \lesssim 10$ Mpc, and the H I–H I auto-correlation ($\xi_{\rm aa}$) at transverse scales $r_\perp \lesssim 2$ Mpc. We have integrated these correlations along the LOS up to 20 Mpc, and used the projected results to infer the shape of their corresponding 'real-space' correlation functions, $\xi(r)$, assuming power-laws of the form $\xi(r) = (r/r_0)^{-\gamma}$. Comparing the results from the H I–galaxy cross-correlation with the H I–H I and galaxy–galaxy auto-correlations, we provided constraints on their statistical connection, as a function of both H I column density and galaxy star-formation activity.

We summarize our observational results as follows:

- Two-dimensional correlations, $\xi(r_\perp, r_\parallel)$:
  - *Full Sample:* the H I–galaxy two-dimensional cross-correlation has comparable clustering amplitudes than those of the H I–H I auto-correlation, which are lower than those of the galaxy–galaxy auto-correlation. The peaks of these correlation functions were found to be $\xi_{\rm ag} = 2.3 \pm 0.9$, $\xi_{\rm aa} = 2.1 \pm 0.9$ and $\xi_{\rm gg} = 5.7 \pm 0.7$, respectively.
  - *'Strong' H I systems and 'SF' galaxies:* the H I–galaxy, H I–H I and galaxy–galaxy two-dimensional correlations have all comparable amplitudes. The peaks of these correlation functions were found to be $\xi_{\rm ag} = 8.3 \pm 2.2$, $\xi_{\rm aa} = 7.5 \pm 2.3$ and $\xi_{\rm gg} = 6.1 \pm 0.6$, respectively.
  - *'Strong' H I systems and 'non-SF' galaxies:* the H I–galaxy two-dimensional cross-correlation has comparable clustering amplitudes than those of the galaxy–galaxy auto-correlation, which are marginally higher than those of the H I–H I auto-correlation. The peaks of the correlation functions were found to be $\xi_{\rm ag} = 10.3 \pm 5.6$, $\xi_{\rm gg} = 12.6 \pm 3.0$ and $\xi_{\rm aa} = 7.5 \pm 2.3$, respectively.

  - *'Weak' H* I *systems and 'SF' galaxies:* the two-dimensional H I–galaxy clustering has much lower amplitudes than those of the galaxy–galaxy and H I–H I auto-correlations. The H I-H I auto-correlation has also lower amplitudes than those of the galaxy–galaxy auto-correlation. The peaks of the correlation functions were found to be $\xi_{\rm ag} = 0.9 \pm 0.6$, $\xi_{\rm gg} = 6.1 \pm 0.6$ and $\xi_{\rm aa} = 1.9 \pm 0.9$, respectively.

  - *'Weak' H* I *systems and 'non-SF' galaxies:* the two-dimensional H I–galaxy clustering has much lower amplitudes than those of the galaxy–galaxy and H I–H I auto-correlations. The H I-H I auto-correlation has also lower amplitudes than those of the galaxy–galaxy auto-correlation. The peaks of the correlation functions were found to be $\xi_{\rm ag} = 0.6 \pm 0.5$, $\xi_{\rm gg} = 12.6 \pm 3.0$ and $\xi_{\rm aa} = 1.9 \pm 0.9$, respectively.

- Real space correlations, $\xi(r) \equiv (r/r_0)^{-\gamma}$:

  - *Full Sample:* the H I–galaxy cross-correlation has comparable clustering amplitudes than those of the H I–H I auto-correlation, which are lower than those of the galaxy–galaxy auto-correlation. The correlation lengths and slopes are found to be $r_0^{\rm ag} = 1.6 \pm 0.2$ Mpc and $\gamma^{\rm ag} = 1.4 \pm 0.1$, $r_0^{\rm aa} = 0.3 \pm 0.3$ Mpc and $\gamma^{\rm aa} = 1.1 \pm 0.1$, and $r_0^{\rm gg} = 3.9 \pm 0.1$ Mpc and $\gamma^{\rm gg} = 1.7 \pm 0.1$, respectively.

  - *'Strong' H* I *systems and 'SF' galaxies:* the H I–galaxy, H I–H I and galaxy–galaxy correlations have all comparable amplitudes. The correlation lengths and slopes are found to be $r_0^{\rm ag} = 3.8 \pm 0.2$ Mpc and $\gamma^{\rm ag} = 1.7 \pm 0.1$, $r_0^{\rm aa} = 3.1 \pm 0.7$ Mpc and $\gamma^{\rm aa} = 1.3 \pm 0.4$, and $r_0^{\rm gg} = 3.9 \pm 0.1$ Mpc and $\gamma^{\rm gg} = 1.6 \pm 0.1$, respectively.

  - *'Strong' H* I *systems and 'non-SF' galaxies:* the H I–galaxy cross-correlation has comparable clustering amplitudes than those of the galaxy–galaxy auto-correlation, which are higher than those of the H I–H I auto-correlation. The correlation lengths and slopes found to be $r_0^{\rm ag} =$

4.0 ± 0.3 Mpc and $\gamma^{\rm ag} = 1.7 \pm 0.1$, $r_0^{\rm gg} = 6.2 \pm 0.2$ Mpc and $\gamma^{\rm gg} = 1.6 \pm 0.1$, and $r_0^{\rm aa} = 3.1 \pm 0.7$ Mpc and $\gamma^{\rm aa} = 1.3 \pm 0.4$, respectively.

- *'Weak' H* I *systems and 'SF' galaxies:* the H I–galaxy cross-correlation has much lower amplitudes than those of the galaxy–galaxy and H I–H I auto-correlations. The H I-H I auto-correlation has also lower amplitudes than those of the galaxy–galaxy auto-correlation. The correlation lengths and slopes are found to be $r_0^{\rm ag} = 0.2 \pm 0.4$ Mpc and $\gamma^{\rm ag} = 1.1 \pm 0.3$, $r_0^{\rm gg} = 3.9 \pm 0.1$ Mpc and $\gamma^{\rm gg} = 1.6 \pm 0.1$, and $r_0^{\rm aa} = 0.3 \pm 0.1$ Mpc and $\gamma^{\rm aa} = 1.0 \pm 0.1$, respectively. We note however that a power-law fit for H I–galaxy cross-correlation might not be a good description of the observations.

- *'Weak' H* I *systems and 'SF' galaxies:* the H I–galaxy cross-correlaiton has much lower amplitudes than those of the galaxy–galaxy and H I–H I auto-correlations. The H I-H I auto-correlation has also lower amplitudes than those of the galaxy–galaxy auto-correlation. The correlation lengths and slopes are found to be $r_0^{\rm ag} = 0.0 \pm 0.8$ Mpc and $\gamma^{\rm ag} = 1.0 \pm 1.6$, $r_0^{\rm gg} = 6.2 \pm 0.2$ Mpc and $\gamma^{\rm gg} = 1.6 \pm 0.1$, and $r_0^{\rm aa} = 0.3 \pm 0.1$ Mpc and $\gamma^{\rm aa} = 1.0 \pm 0.1$, respectively. We note however that a power-law fit for the real-space H I–galaxy cross-correlation might not be a good description of the observations.

- Amplitudes:

  - *H* I*–galaxy cross-correlations:* The H I–galaxy cross-correlation amplitudes are systematically higher for 'strong' systems than 'weak' systems, and for 'non-SF' galaxies than 'SF' galaxies, with a much stronger dependence on H I column density than galaxy star-formation activity. This is true either for the two-dimensional or the real-space correlations (see numbers above).

  - *Galaxy auto-correlations:* The galaxy–galaxy auto-correlation amplitudes

are systematically higher for 'non-SF' galaxies than 'SF' galaxies. This is true either for the two-dimensional or the real-space correlations (see numbers above).

– *H* I *auto-correlations:* The H I–H I auto-correlation amplitudes are systematically higher for 'strong' systems than 'weak' systems. This is true either for the two-dimensional and real-space correlations (see numbers above).

- Velocity distortions:

  – The two-dimensional H I–galaxy cross-correlations do not show significant velocity distortions along the LOS, apart from those expected by the galaxy redshift uncertainties.

  – The peak in the two-dimensional H I–galaxy cross-correlation for 'strong' systems and 'non-SF' galaxies appears ~ 1 Mpc shifted along the LOS from 0, and there is marginal evidence (not significant) that this might be a real feature.

- Two-dimensional ratios, $(\xi_{\rm ag})^2/(\xi_{\rm gg}\xi_{\rm aa})$:

  – *Full Sample:* the ratio $(\xi_{\rm ag})^2/(\xi_{\rm gg}\xi_{\rm aa})$ appears marginally inconsistent with 1 on scales $\lesssim 2$ Mpc.

  – *'Strong' H* I *systems and galaxies:* the ratio $(\xi_{\rm ag})^2/(\xi_{\rm gg}\xi_{\rm aa})$ appears roughly consistent (large uncertainties) with 1 on scales $\lesssim 2$ Mpc, independently of the galaxy star-formation activity.

  – *'Weak' H* I *systems and galaxies:* the ratio $(\xi_{\rm ag})^2/(\xi_{\rm gg}\xi_{\rm aa})$ appears inconsistent with 1 on scales $\lesssim 2$ Mpc, independently of the galaxy star-formation activity.

- Projected along the LOS ratios, $(\Xi_{\rm ag})^2/(\Xi_{\rm gg}\Xi_{\rm aa})$:

- *Full Sample:* we find $(\Xi_{\rm ag})^2/(\Xi_{\rm gg}\Xi_{\rm aa}) \approx 0.2 \pm 0.2$ on scales $\lesssim 2$ Mpc. Therefore, ruling out the hypothesis that H I systems and galaxies (as a whole) trace the same underlying dark matter distribution linearly on those scales, at a high statistical significance ($> 3\sigma$ c.l.).
- *'Strong' H I systems and 'SF' galaxies:* we find $(\Xi_{\rm ag})^2/(\Xi_{\rm gg}\Xi_{\rm aa}) \approx 1.1 \pm 0.6$ on scales $\lesssim 2$ Mpc. Therefore, roughly consistent (large uncertainties) with the hypothesis that 'strong' H I systems and 'SF' galaxies trace the same underlying dark matter distribution linearly on those scales.
- *'Strong' H I systems and 'non-SF' galaxies:* we find $(\Xi_{\rm ag})^2/(\Xi_{\rm gg}\Xi_{\rm aa}) \approx 0.5 \pm 0.3$ on scales $\lesssim 2$ Mpc. Therefore, marginally ruling out the hypothesis that 'strong' H I systems and 'non-SF' galaxies trace the same underlying dark matter distribution linearly on those scales (only at the $\sim 2\sigma$ c.l.).
- *'Weak' H I systems and 'SF' galaxies:* we find $(\Xi_{\rm ag})^2/(\Xi_{\rm gg}\Xi_{\rm aa}) \approx 0.02 \pm 0.2$ on scales $\lesssim 2$ Mpc. Therefore, ruling out the hypothesis that 'weak' H I systems and 'SF' galaxies trace the same underlying dark matter distribution linearly on those scales, at a high statistical significance ($> 3\sigma$ c.l.).
- *'Weak' H I systems and 'non-SF' galaxies:* we find $(\Xi_{\rm ag})^2/(\Xi_{\rm gg}\Xi_{\rm aa}) \approx 0.001 \pm 0.4$ on scales $\lesssim 2$ Mpc. Therefore, marginally ruling out the hypothesis that 'weak' H I systems and 'non-SF' galaxies trace the same underlying dark matter distribution linearly on those scales (only at the $\sim 2\sigma$ c.l.).

- 'Absolute biases':
  - *'Strong' H I systems and galaxies:* their 'absolute biases' are consistent with $b \gtrsim 1$.
  - *'Weak' H I systems:* their 'absolute biases' are consistent with $b < 1$.

Our interpretation of these results has lead us to the following conclusions:

- The bulk of H I systems around galaxies have very little velocity dispersion ($\ll 120$ km s^{-1}) with respect to the bulk of galaxies. Hence, no strong galaxy outflow or inflow signal is detected.

- The vast majority of 'strong' H I systems and 'SF' galaxies are distributed in the same locations. We have identified these locations with the 'overdense large-scale structure'.

- A fraction of 'non-SF' galaxies are distributed in roughly the same way as 'strong' H I systems and 'SF' galaxies but there are sub-locations—within those where galaxies and 'strong' H I systems reside—with a much higher density of 'non-SF' galaxies than 'strong' H I systems and/or 'SF' galaxies. We have identified such locations as galaxy clusters. We estimated that only a $25 \pm 15\%$ of 'non-SF' galaxies reside in galaxy clusters and that the remaining $75 \pm 15\%$ co-exist with 'strong' H I and 'SF' at scales $\lesssim 2$ Mpc, following the same underlying dark matter distribution, i.e. the 'overdense large-scale structure'.

- An important fraction of 'weak' systems could reside in locations devoid of galaxies of any kind. We have identified such locations as galaxy voids, i.e. the 'underdense large-scale structure'. At a limit of $N_{\rm HI} \geq 10^{13}$ cm^{-2}, we have estimated that roughly $\sim 50\%$ of 'weak' systems reside within galaxy voids. At lower $N_{\rm HI}$ limits this fraction is likely to increase.

- The vast majority of 'strong' H I absorption systems at low-z reside in dark matter haloes of masses comparable to those hosting galaxies.

- At least $\sim 50\%$ of 'weak' H I absorption systems with $N_{\rm HI} \geq 10^{13}$ cm^{-2}, reside in dark matter haloes less massive than those hosting 'strong' H I systems and/or galaxies. At lower $N_{\rm HI}$ limits this fraction is likely to increase.

- We speculate that H I systems within galaxy voids might be still evolving in the linear regime even at scales $\lesssim 2$ Mpc and at $z \lesssim 1$.

- We conclude that there are *at least* three types of relationship between H I absorption systems and galaxies at low-z: (i) one-to-one physical association; (ii) association because they both follow the same underlying dark matter distribution; and (iii) no association at all.

Acknowledgments

We thank the anonymous referee for helpful comments which improved this Chapter and its accepted for publication version (Tejos et al., 2013). We thank Pablo Arnalte-Mur, Rich Bielby, John Lucey, Peder Norberg and Tom Shanks for helpful discussions. We thank Pablo Arnalte-Mur for his help in obtaining the prediction for the dark matter clustering at $z \lesssim 1$. We thank Mark Swinbank for his help in the reduction of GMOS data.

We thank contributors to SciPy[26], Matplotlib[27], and the Python programming language[28]; the free and open-source community; and the NASA Astrophysics Data System[29] for software and services.

N.T. acknowledges grant support by CONICYT, Chile (PFCHA/*Doctorado al Extranjero 1ª Convocatoria*, 72090883).

This work was mainly based on observations made with the NASA/ESA Hubble Space Telescope, obtained at the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5-26555; and on observations collected at the European Southern Observatory, Chile. Some of the data presented herein were obtained at the W.M. Keck Observatory, which is operated as a scientific partnership among the California Institute of Technology, the University of California and the NASA.

[26] `http://www.scipy.org`
[27] `http://www.matplotlib.sourceforge.net`
[28] `http://www.python.org`
[29] `http://adswww.harvard.edu`

The Observatory was made possible by the generous financial support of the W.M. Keck Foundation. The authors wish also to recognize and acknowledge the very significant cultural role and reverence that the summit of Mauna Kea has always had within the indigenous Hawaiian community. We are most fortunate to have the opportunity to conduct observations from this mountain. This work was partially based on observations obtained at the Gemini Observatory, which is operated by the Association of Universities for Research in Astronomy, Inc., under a cooperative agreement with the NSF on behalf of the Gemini partnership: the NSF (United States), the National Research Council (Canada), CONICYT (Chile), the Australian Research Council (Australia), Ministério da Ciência, Tecnologia e Inovação (Brazil) and Ministerio de Ciencia, Tecnología e Innovación Productiva (Argentina).

Chapter 5

Summary & Conclusion

5.1 Summary

In this thesis we have explored the relationship between the intergalactic medium (IGM) and galaxies at $z \lesssim 1$, in a statistical manner. Galaxies have been mostly surveyed in emission using optical spectroscopy, while the IGM has been mostly surveyed in absorption in the ultra-violet (UV) spectra of background QSO. We have presented observational results for such a connection using two complementary methods:

- In Chapter 2 we used galaxy voids as tracers of both underdense and overdense regions. We used archival data to study the properties of H I absorption line systems within and around galaxy voids at $z < 0.1$. Typical galaxy voids have sizes $\gtrsim 10$ Mpc and so our results constrained the very large-scale association. This sample was composed by 106 H I absorption systems and 1054 galaxy voids.

- In Chapter 4 we used a sample H I systems and galaxies from pencil beam surveys to measure the H I–galaxy cross-correlation at $z \lesssim 1$. Our sample is composed by a combination of archival and new data taken by the author and collaborators. This survey properly covers transverse separations between H I and galaxies from ~ 100 kpc (proper) up to ~ 10 Mpc, filling the gap between the very large scales and those associated to the so-called circumgalactic medium (CGM). This sample was composed by 654 H I absorption systems and 17509 galaxies.

In this way, we have studied the IGM–galaxy statistical relationship over a wide range of scales and environments. Our main results are summarized as follows:

- We found a significant excess of H I absorption line systems at the edges of galaxy voids, consistent with the excess of galaxies that define such voids. We found no significant deficit of H I systems inside galaxy voids compared to the random expectation. This is inconsistent with the expectation from the galaxy distribution, as the density of galaxies inside galaxy voids is significantly lower than that the random expectation.

- We found differences in the H I column density ($N_{\rm HI}$) and Doppler parameter ($b_{\rm HI}$) distribution between H I systems inside and outside galaxy voids. Low density environments (voids) have smaller values for both $N_{\rm HI}$ and $b_{\rm HI}$ than higher density ones (edges of voids). These trends are theoretically expected and found in GIMIC, a state-of-the-art hydrodynamical simulation. We found no sharp $N_{\rm HI}$ or $b_{\rm HI}$ transitions with environment however, except perhaps for $N_{\rm HI} \lesssim 10^{13}$ cm^{-2}. We estimate that the vast majority of H I systems with $N_{\rm HI} \lesssim 10^{13}$ cm^{-2} reside in galaxy voids.

- We found that H I systems with $10^{14} \leq N_{\rm HI} \lesssim 10^{17}$ cm^{-2} ('strong') cluster with 'star-forming' ('SF') galaxies as strongly as 'SF' galaxies with themselves, and that H I systems with $10^{13} \lesssim N_{\rm HI} < 10^{14}$ cm^{-2} ('weak') cluster much more weakly with galaxies than would be expected if these H I systems and galaxies trace the same underlying dark matter distribution linearly. Our interpretation of this result is that a significant fraction (at least $\sim 50\%$) of 'weak' systems are not correlated at all with galaxies. This picture is consistent with our previous findings in Chapter 2.

- We found that the amplitude of the auto- and cross-correlations of 'strong' H I systems and 'non-SF' galaxies is not consistent with the hypothesis that both samples trace the same underlying dark matter distribution linearly. This result is consistent with an scenario in which $\sim 25\%$ of 'non-SF' galaxies

reside in galaxy clusters, with a significant lack of 'strong' H I systems within ~ 2 Mpc, and the remaining $\sim 75\%$ reside in locations where 'strong' H I is abundant. In this picture, cold gas (as traced by 'strong' H I systems) around 'non-SF' galaxies is the rule rather than the exception.

- We found that the amplitude of the H I–galaxy cross-correlation scales with both H I column density and the presence of star-formation activity in the galaxy, but that the dependence on H I column density is much stronger than that star-formation. Still, we do not find sharp transitions in $N_{\rm HI}$ for the observed H I–galaxy cross-correlation amplitude for column densities in the range $10^{13} \lesssim N_{\rm HI} \lesssim 10^{15}$ cm^{-2}. This result is also consistent with our findings of Chapter 2. Because of a lack of observational data we could not test the prediction that H I systems with $N_{\rm HI} \lesssim 10^{13}$ cm^{-2} are preferentially found in galaxy voids. (In such a scenario an *anti*-correlation between H I and galaxies is expected.)

- We found that the amplitude of the H I auto-correlation for 'strong' systems is larger than that of 'weak' systems. Similarly, the amplitude of the galaxy auto-correlation for 'non-SF' galaxies is larger than that for 'SF' galaxies.

- We found that the two-dimensional H I–galaxy cross-correlations do not show significant velocity distortions along the line-of-sight, apart from those expected by the galaxy redshift uncertainties. In consequence, we concluded that the bulk of H I systems around galaxies has a low velocity dispersion ($\ll 120$ km s^{-1}) with respect to the bulk of galaxies (from the point of view of H I). This result places an important constraint on the extent of galaxy feedback present at $z \lesssim 1$, as no strong galaxy outflow is detected. We did not detect a signal associated with inflowing material either, at the same velocity limit.

- When comparing our clustering results with those of the cold dark matter theoretical expectation, we found that the auto-correlation amplitudes of

the 'strong' H I systems and galaxies are consistent with an 'absolute bias' $b \geq 1$. On the other hand, the amplitudes of the 'weak' H I auto-correlation are lower than that of the dark matter, hence $b < 1$. This might be evidence that H I systems inside galaxy voids are evolving more slower than their counterparts located in higher density environments.

For a more quantitative description of our observational results we refer the reader to Sections 2.4 and 4.8. Similarly, for further discussion on the interpretation of these results we refer the reader to Sections 2.6 and 4.9.

At this point we emphasize that the observational results from the analysis presented in Chapter 2 and Chapter 4, can both be interpreted under the same simplified model. This is remarkable because these two experiments are quite different in design, not only in the way that the data was analysed, but also concerning their targeted mean scales and cosmic epochs.

5.2 Conclusion

In conclusion, our results hint towards a picture in which there are *at least* three types of association between the the diffuse gas in the Universe and galaxies at $z \lesssim 1$:

- *One-to-one direct association* because galaxies do contain diffuse gas. Even though in this thesis we did not directly constrain this association, this is an observational fact that can not be neglected.

- *Indirect association* because both the IGM and galaxies trace the same underlying dark matter distribution. Our results have provided quantitative evidence for this association to exist. Moreover, we have shown that not all galaxies are related to the diffuse gas in the same way. In particular, a non negligible fraction of 'non-star-forming' galaxies might reside in environments devoid of diffuse H I.

- *No association* because there are regions in the Universe that contain a significant amount of diffuse gas but that are devoid of galaxies. Our results have provided quantitative evidence that this scenario exists.

We have also shown that the last two scenarios are not defined in terms of sharp H I column density limits in the range $10^{13} \lesssim N_{\rm HI} \lesssim 10^{14}$ cm^{-2}, but rather in a probabilistic manner. For instance, we found that there is a $\sim 50\%$ chance that a H I system with column density within the above range belongs to a galaxy void. At much higher column densities a direct association with galaxy haloes is expected. In this way, H I systems with $N_{\rm HI} \gtrsim 10^{15}$ cm^{-2} would likely be found close to galaxies rather than in galaxy voids. On the other hand, the opposite would be true for H I systems with $N_{\rm HI} \lesssim 10^{13}$ cm^{-2}. Based on the results from Chapter 2, we predict that there should be an *anti*-correlation between H I systems with $N_{\rm HI} \lesssim 10^{13}$ cm^{-2} and galaxies.

Despite the comparatively simple physics of the diffuse IGM compared to that of galaxies, our observational results suggest that there are still aspects that are not yet accounted for in our current models of the diffuse IGM. For instance, it is usually assumed that, at large scales, *all* baryonic matter follows the same underlying distribution in a fairly linear way. It is only at the very small scales ($\lesssim 100$ kpc) that baryonic physics is invoked to explain the non-linear dependence between the observed correlations (for both IGM and galaxies) and that of the theoretical cold dark matter. Our results indicate that such non-linear dependence is also present at scales of at least $\lesssim 2$ Mpc. Moreover, according to our prediction that H I systems with $N_{\rm HI} \lesssim 10^{13}$ cm^{-2} are preferentially found in galaxy voids, this non-linear dependence can be present even at scales as large as ~ 10 Mpc.

This large-scale non-linear dependence primarily arises from a geometrical effect: there are well defined regions containing diffuse gas, but, for some reason, are comparatively devoid of galaxies. Did galaxies never form in such regions? Did they form there, but moved toward a higher density environment in less than a Hubble time? Did they form there, but somehow were destroyed or faded

(e.g. by the UVB and/or the expansion of the Universe)? It seems more natural and simple to assume that galaxies never formed in these low density environments, but all the other possibilities should be properly ruled out. Understanding why galaxies are not present in such environments will provide extra clues for understanding the evolution of baryonic matter in the Universe.

5.3 Future prospects and final word

In Sections 2.6.3 and 4.9.3, we have provided specific future prospects for the work presented in their respective Chapters. In a more broad perspective, we consider that the study of baryonic matter in the Universe will require a comprehensive understanding of how the large-scale environment shapes the IGM–galaxy relationship. Observational studies of the IGM in well defined large-scale structures—like galaxy clusters, galaxy filaments and galaxy voids—are natural continuations of the present work. Such an ambitious project is currently feasible by combining the capabilities of: (i) the Cosmic Origins Spectrograph (COS) on Hubble Space Telescope (HST); (ii) the wide and extremely large spectroscopic galaxy surveys; and (iii) deep multi-object pencil-beam like spectroscopic surveys.

Pursing that goal, we have been recently awarded HST time to observe a carefully chosen QSO with COS, whose sightline passes through 6 potential cosmological filaments—defined by galaxy clusters pairs separated by < 20 Mpc and at similar redshifts—and 3 independent galaxy clusters at impact parameters $\lesssim 1.5$ Mpc at $z \lesssim 0.4$ (GO 12958, PI: Tejos). This data will significantly increase the number of well identified 'filaments' observed in absorption and therefore will help to generate *statistical* conclusions about the nature of the H I absorption systems found in overdense environments. Similarly, we are re-submitting another complementary COS HST proposal to observe the most denser regions in the Universe, by targeting QSO at impact parameters $\lesssim 1$ Mpc from foreground galaxy clusters. In this way, we can compare the statistical properties of H I absorption systems found in quite different cosmological environments.

We also consider that the aforementioned projects can shed light on the long-lasting problem of the 'missing baryons' at low-z. As mentioned, the best candidate for hosting those baryons is the warm-hot intergalactic medium (WHIM), and our project on cosmological filaments and clusters can naturally lead to such detections if this environment is present.

A better theoretical description of the physics of the IGM in the underdense regions could also help to solve the missing baryon problem. Current estimations of the baryon content in the H I Lyα forest at low-z model all these absorption lines as coming from the same physical entities (e.g. isothermal spheres). If this assumption is not valid, and the underdense regions host a different type of H I absorption systems compared to those observed in the overdense regions, then such estimates will also change. Similarly, current hydrodynamical cosmological simulations might not be accounting for all the physics in the underdense regions because of resolution limitations.

Despite the important information on the IGM–galaxy statistical connection provided by this thesis, new questions have arisen. Our present work represents a modest but valuable contribution towards a better understanding of the evolution of baryonic matter in the Universe.

Nicolas Tejos

Durham, 2013

Appendix A | *Data Tables*

The following tables correspond to the new galaxy and H I absorption system catalogs used in Chapter 4. To avoid overloading this printable version of the thesis, we only show the first entries of each table unless otherwise stated. The full tables will be available online once Chapter 4 is published in the Monthly Notices of the Royal Astronomical Society. In the meantime, these tables can be requested from the author.

Table A.1: H I absorption systems in QSO Q0107-025A.

z (1)	$\log(N_{\rm HI}/{\rm cm}^{-2})$ (2)	$b_{\rm HI}$ (km s^{-1}) (3)	label (4)
0.003736 ± 0.000023	13.46 ± 0.09	30 ± 11	b
0.053847 ± 0.000019	13.58 ± 0.06	44 ± 8	b
0.005063 ± 0.000030	13.34 ± 0.10	37 ± 13	b
0.054636 ± 0.000030	13.08 ± 0.13	25 ± 14	b
0.006143 ± 0.002014	14.01 ± 9.45	47 ± 219	c
0.056460 ± 0.000020	13.24 ± 0.08	30 ± 9	b
0.006291 ± 0.000409	15.14 ± 6.02	31 ± 205	c
0.060568 ± 0.000006	13.91 ± 0.05	26 ± 3	b
0.006427 ± 0.007203	13.65 ± 55.41	34 ± 650	c
0.062980 ± 0.000008	13.98 ± 0.02	55 ± 3	b
0.028710 ± 0.000009	13.49 ± 0.06	19 ± 4	b
0.063400 ± 0.000018	12.95 ± 0.12	14 ± 10	b
0.031353 ± 0.000005	13.92 ± 0.23	14 ± 4	b
0.076709 ± 0.000081	13.86 ± 0.10	139 ± 43	b
0.036048 ± 0.000004	14.24 ± 0.58	14 ± 5	c
0.077981 ± 0.000008	17.08 ± 0.78	20 ± 3	c
0.040448 ± 0.000010	13.58 ± 0.05	21 ± 5	b
0.081904 ± 0.000011	13.94 ± 0.07	30 ± 5	b
0.040644 ± 0.000045	12.66 ± 0.51	12 ± 30	c
0.082419 ± 0.000028	13.44 ± 0.09	36 ± 12	b
0.040971 ± 0.000048	13.55 ± 0.10	73 ± 26	b
0.085376 ± 0.000016	13.52 ± 0.05	37 ± 7	b
0.041486 ± 0.000046	13.45 ± 0.26	24 ± 15	b
0.094779 ± 0.000008	13.53 ± 0.04	21 ± 3	b
0.041669 ± 0.000025	13.81 ± 0.11	25 ± 7	b
0.097074 ± 0.000014	12.93 ± 0.11	10 ± 8	b
0.043067 ± 0.000014	13.43 ± 0.05	37 ± 6	b
0.099364 ± 0.000021	13.11 ± 0.08	28 ± 9	b
0.047181 ± 0.000257	13.56 ± 0.36	108 ± 46	b
0.109475 ± 0.000004	14.25 ± 0.24	19 ± 3	a
0.047422 ± 0.000032	13.65 ± 0.29	51 ± 18	b
0.112880 ± 0.000012	13.73 ± 0.04	35 ± 5	a
0.050436 ± 0.000021	12.93 ± 0.17	9 ± 14	b
0.113778 ± 0.000005	13.88 ± 0.25	12 ± 3	a
0.050640 ± 0.000010	13.62 ± 0.05	22 ± 5	b
0.114403 ± 0.000012	13.40 ± 0.06	24 ± 5	a
0.050869 ± 0.000020	13.10 ± 0.12	17 ± 12	b
0.114844 ± 0.000040	13.28 ± 0.11	54 ± 21	a
0.052836 ± 0.000087	14.03 ± 0.27	74 ± 15	b
0.115532 ± 0.000008	16.23 ± 0.90	25 ± 5	c
0.052887 ± 0.000013	15.24 ± 2.95	22 ± 21	c
0.115884 ± 0.000036	13.36 ± 0.15	27 ± 14	a
0.053279 ± 0.000064	13.29 ± 0.37	33 ± 24	b
0.120679 ± 0.000021	13.44 ± 0.06	45 ± 8	a

Note. Only a portion of this table is shown. The full table will be available in the online version of the paper. (1) and (5): H I redshift. (2) and (6): H I column density from Voigt profile fitting. (3) and (7): H I Doppler parameter from Voigt profile fitting. (4) and (8): Confidence label: (a) 'secure'; (b) 'probable'; and (c) 'uncertain' (see Section 4.5.3 for definitions). See Section 4.5 for further details.

Table A.2: H I absorption systems in QSO Q0107-025B.

z (1)	$\log(N_{\rm HI}/{\rm cm}^{-2})$ (2)	$b_{\rm HI}$ (km s^{-1}) (3)	label (4)
0.006340 ± 0.000004	14.30 ± 0.08	31 ± 3	b
0.261066 ± 0.000004	14.75 ± 0.01	61 ± 1	a
0.013367 ± 0.000010	13.35 ± 0.05	22 ± 5	b
0.294558 ± 0.000012	13.81 ± 0.03	38 ± 4	a
0.041753 ± 0.000012	13.51 ± 0.04	42 ± 6	b
0.314209 ± 0.000046	13.74 ± 0.09	56 ± 12	a
0.053188 ± 0.000011	13.64 ± 0.04	33 ± 5	b
0.314527 ± 0.000014	13.65 ± 0.10	24 ± 5	a
0.060677 ± 0.000013	13.32 ± 0.06	28 ± 6	b
0.321764 ± 0.000008	13.75 ± 1.83	6 ± 7	c
0.063119 ± 0.000008	13.47 ± 0.04	24 ± 3	b
0.333328 ± 0.000013	13.75 ± 0.04	33 ± 4	a
0.109629 ± 0.000014	13.26 ± 0.06	26 ± 6	a
0.383118 ± 0.000009	13.32 ± 0.08	12 ± 3	a
0.115300 ± 0.000001	13.51 ± 0.04	32 ± 4	a
0.399112 ± 0.000003	16.84 ± 0.02	20 ± 1	a
0.115714 ± 0.000003	14.53 ± 0.06	32 ± 1	a
0.399165 ± 0.000011	15.26 ± 0.12	51 ± 3	a
0.120307 ± 0.000007	13.51 ± 0.04	22 ± 3	a
0.412355 ± 0.000007	13.91 ± 0.03	26 ± 3	a
0.120734 ± 0.000008	13.82 ± 0.02	48 ± 3	a
0.427684 ± 0.000029	13.40 ± 0.07	38 ± 9	a
0.136385 ± 0.000004	13.85 ± 0.03	26 ± 2	a
0.434283 ± 0.000012	13.75 ± 0.03	35 ± 4	a
0.154647 ± 0.000021	15.89 ± 0.59	22 ± 2	c
0.436121 ± 0.000024	13.74 ± 0.04	62 ± 7	a
0.155130 ± 0.000232	13.39 ± 0.27	64 ± 16	a
0.467456 ± 0.000013	14.04 ± 0.03	44 ± 3	a
0.200199 ± 0.000003	15.62 ± 0.01	26 ± 1	a
0.499367 ± 0.000009	15.00 ± 0.08	30 ± 3	a
0.202519 ± 0.000004	14.94 ± 0.02	43 ± 1	a
0.499540 ± 0.000069	14.72 ± 0.17	57 ± 7	a
0.203027 ± 0.000003	14.76 ± 0.03	26 ± 1	a
0.512259 ± 0.000097	13.64 ± 0.54	22 ± 58	c
0.211922 ± 0.000010	13.28 ± 0.04	26 ± 4	a
0.517071 ± 0.000011	14.65 ± 0.02	44 ± 3	a
0.226692 ± 0.000003	14.97 ± 0.04	28 ± 1	a
0.523835 ± 0.000051	14.08 ± 0.29	44 ± 33	a
0.227140 ± 0.000011	13.92 ± 0.02	49 ± 4	a
0.535355 ± 0.000008	15.18 ± 0.04	29 ± 2	a
0.254161 ± 0.000026	13.36 ± 0.06	56 ± 11	a
0.555576 ± 0.000124	13.68 ± 1.64	26 ± 171	c
0.258088 ± 0.000014	13.28 ± 0.04	31 ± 4	a
0.578527 ± 0.000077	13.76 ± 0.11	33 ± 23	a

Note. Only a portion of this table is shown. The full table will be available in the online version of the paper. (1) and (5): H I redshift. (2) and (6): H I column density from Voigt profile fitting. (3) and (7): H I Doppler parameter from Voigt profile fitting. (4) and (8): Confidence label: (a) 'secure'; (b) 'probable'; and (c) 'uncertain' (see Section 4.5.3 for definitions). See Section 4.5 for further details.

Table A.3: H I absorption systems in QSO Q0107-0232 .

z (1)	$\log(N_{\rm HI}/{\rm cm}^{-2})$ (2)	$b_{\rm HI}$ (km s^{-1}) (3)	label (4)
0.198946 ± 0.000014	14.73 ± 0.11	52 ± 5	b
0.417517 ± 0.000379	13.94 ± 0.54	131 ± 45	c
0.203349 ± 0.000015	14.79 ± 0.15	42 ± 6	b
0.423807 ± 0.000030	13.55 ± 0.06	46 ± 9	a
0.227096 ± 0.000001	14.00 ± 0.01	50 ± 1	b
0.424577 ± 0.000012	14.03 ± 0.04	36 ± 4	a
0.244724 ± 0.000001	14.00 ± 0.01	50 ± 1	b
0.425130 ± 0.000014	13.93 ± 0.04	37 ± 5	a
0.261403 ± 0.000017	14.02 ± 0.05	43 ± 6	b
0.428308 ± 0.000005	16.01 ± 0.18	35 ± 2	a
0.266532 ± 0.000003	14.53 ± 0.07	49 ± 4	b
0.436369 ± 0.000014	13.76 ± 0.06	23 ± 10	a
0.268371 ± 0.000011	13.28 ± 0.14	9 ± 5	b
0.441843 ± 0.000056	14.26 ± 0.03	208 ± 17	a
0.268636 ± 0.000006	13.80 ± 0.07	16 ± 2	b
0.444586 ± 0.000040	13.70 ± 0.05	80 ± 13	a
0.281634 ± 0.000029	13.23 ± 0.11	40 ± 14	b
0.487116 ± 0.000219	14.08 ± 0.06	384 ± 76	a
0.294486 ± 0.000025	13.86 ± 0.08	28 ± 7	b
0.499733 ± 0.000136	14.03 ± 0.32	61 ± 95	a
0.308847 ± 0.000017	13.46 ± 0.05	34 ± 8	b
0.534585 ± 0.000006	16.11 ± 0.10	25 ± 2	a
0.331159 ± 0.000022	13.74 ± 0.06	40 ± 7	b
0.557390 ± 0.000003	19.49 ± 0.04	50 ± 1	a
0.350499 ± 0.000189	13.12 ± 0.45	43 ± 49	c
0.578538 ± 0.000022	14.83 ± 0.32	17 ± 8	a
0.355650 ± 0.000039	13.32 ± 0.08	47 ± 13	b
0.578750 ± 0.000096	15.02 ± 0.22	42 ± 13	a
0.357978 ± 0.000015	13.53 ± 0.05	27 ± 5	b
0.621996 ± 0.000017	14.49 ± 0.04	40 ± 4	a
0.365404 ± 0.000014	13.58 ± 0.05	30 ± 5	b
0.648707 ± 0.000005	15.53 ± 0.09	27 ± 2	a
0.375868 ± 0.000027	13.50 ± 0.07	41 ± 9	b
0.649744 ± 0.000322	13.81 ± 0.10	123 ± 33	a
0.380223 ± 0.000011	13.15 ± 0.07	13 ± 4	b
0.683657 ± 0.000082	14.22 ± 0.02	327 ± 25	a
0.380989 ± 0.000007	13.49 ± 0.09	11 ± 3	b
0.689792 ± 0.000024	14.87 ± 0.04	66 ± 3	a
0.381378 ± 0.000005	13.95 ± 0.30	11 ± 3	b
0.690043 ± 0.000018	14.30 ± 0.12	14 ± 6	a
0.401665 ± 0.000028	13.69 ± 0.76	8 ± 9	c
0.699929 ± 0.000380	13.21 ± 0.20	51 ± 1	a
0.416660 ± 0.001123	13.69 ± 0.98	152 ± 173	c
0.701110 ± 0.000190	13.43 ± 0.13	50 ± 1	a

Note. Only a portion of this table is shown. The full table will be available in the online version of the paper. (1) and (5): H I redshift. (2) and (6): H I column density from Voigt profile fitting. (3) and (7): H I Doppler parameter from Voigt profile fitting. (4) and (8): Confidence label: (a) 'secure'; (b) 'probable'; and (c) 'uncertain' (see Section 4.5.3 for definitions). See Section 4.5 for further details.

Table A.4: H I absorption systems in QSO J020930.7-043826.

z (1)	$\log(N_{\rm HI}/{\rm cm}^{-2})$ (2)	$b_{\rm HI}$ (km s^{-1}) (3)	label (4)
0.060451 ± 0.000004	16.08 ± 0.79	41 ± 8	c
0.142066 ± 0.000005	14.30 ± 0.03	38 ± 2	b
0.061037 ± 0.000006	13.96 ± 0.05	26 ± 3	b
0.153342 ± 0.000018	13.31 ± 0.07	29 ± 7	b
0.062140 ± 0.000008	13.45 ± 0.06	16 ± 4	b
0.153903 ± 0.000008	13.61 ± 0.05	18 ± 3	b
0.066439 ± 0.000014	13.96 ± 0.14	21 ± 8	b
0.154236 ± 0.000010	13.61 ± 0.04	28 ± 4	b
0.066676 ± 0.000016	14.07 ± 0.09	27 ± 6	b
0.158239 ± 0.000055	13.44 ± 0.09	91 ± 28	b
0.068032 ± 0.000026	13.39 ± 0.17	16 ± 9	b
0.159197 ± 0.000017	13.60 ± 0.04	51 ± 6	b
0.068200 ± 0.000043	13.32 ± 0.26	24 ± 18	b
0.160967 ± 0.000005	14.17 ± 0.03	32 ± 2	b
0.068780 ± 0.000018	14.88 ± 19.64	3 ± 15	c
0.161315 ± 0.000046	13.05 ± 0.32	30 ± 21	b
0.090942 ± 0.000021	12.80 ± 0.11	17 ± 10	b
0.161665 ± 0.000020	13.96 ± 0.04	60 ± 6	b
0.095413 ± 0.000008	14.05 ± 0.07	28 ± 4	b
0.166490 ± 0.000040	13.38 ± 0.05	109 ± 15	b
0.099244 ± 0.000011	13.66 ± 0.03	42 ± 5	b
0.176407 ± 0.000013	12.97 ± 0.13	10 ± 8	b
0.099635 ± 0.000032	12.72 ± 0.17	18 ± 15	b
0.176854 ± 0.000114	13.49 ± 0.14	121 ± 43	b
0.106732 ± 0.000009	13.45 ± 0.05	22 ± 4	b
0.177641 ± 0.000124	13.16 ± 0.25	81 ± 37	b
0.107026 ± 0.000030	13.19 ± 0.09	41 ± 13	b
0.181080 ± 0.000036	12.87 ± 0.13	30 ± 15	b
0.111795 ± 0.000185	17.95 ± 0.19	51 ± 17	b
0.181325 ± 0.000016	12.98 ± 0.12	10 ± 8	b
0.113543 ± 0.000022	13.04 ± 0.08	31 ± 10	b
0.181485 ± 0.000044	12.65 ± 0.26	17 ± 20	b
0.113866 ± 0.000017	12.85 ± 0.10	14 ± 8	b
0.182500 ± 0.000006	13.93 ± 0.03	26 ± 2	b
0.121290 ± 0.000006	13.90 ± 0.02	39 ± 2	b
0.182847 ± 0.000019	13.79 ± 0.04	56 ± 6	b
0.129062 ± 0.000006	13.54 ± 0.03	24 ± 2	b
0.184403 ± 0.000003	14.04 ± 0.05	21 ± 2	b
0.135900 ± 0.000064	13.65 ± 0.30	34 ± 12	b
0.188538 ± 0.000014	13.21 ± 0.05	29 ± 6	b
0.136159 ± 0.000025	14.36 ± 0.07	42 ± 5	b
0.198091 ± 0.000016	13.21 ± 0.06	26 ± 6	b
0.136700 ± 0.000018	13.24 ± 0.07	31 ± 8	b
0.201414 ± 0.000014	13.06 ± 0.06	23 ± 6	b

Note. Only a portion of this table is shown. The full table will be available in the online version of the paper. (1) and (5): H I redshift. (2) and (6): H I column density from Voigt profile fitting. (3) and (7): H I Doppler parameter from Voigt profile fitting. (4) and (8): Confidence label: (a) 'secure'; (b) 'probable'; and (c) 'uncertain' (see Section 4.5.3 for definitions). See Section 4.5 for further details.

Table A.5: H I absorption systems in QSO J100535.24+013445.7.

z (1)	$\log(N_{\rm HI}/{\rm cm}^{-2})$ (2)	$b_{\rm HI}$ (km s^{-1}) (3)	label (4)
0.004126 ± 0.001790	13.87 ± 6.19	63 ± 209	c
0.056062 ± 0.000022	13.10 ± 0.17	10 ± 14	b
0.004283 ± 0.000106	14.55 ± 1.30	46 ± 30	c
0.062474 ± 0.000023	13.39 ± 0.11	26 ± 9	b
0.005998 ± 0.000037	13.01 ± 0.14	31 ± 17	b
0.062733 ± 0.000019	13.65 ± 0.06	35 ± 8	b
0.017673 ± 0.000020	13.21 ± 0.09	25 ± 10	b
0.083817 ± 0.000023	13.04 ± 0.12	17 ± 11	b
0.018720 ± 0.000008	13.53 ± 0.45	8 ± 6	b
0.091821 ± 0.000018	12.76 ± 0.13	9 ± 11	b
0.020953 ± 0.000021	13.23 ± 0.08	29 ± 10	b
0.093141 ± 0.000015	12.84 ± 0.12	9 ± 9	b
0.023897 ± 0.000052	13.19 ± 0.24	28 ± 22	b
0.093358 ± 0.000017	12.86 ± 0.11	13 ± 9	b
0.024157 ± 0.000012	14.30 ± 0.15	27 ± 9	b
0.099344 ± 0.004656	13.91 ± 62.47	17 ± 787	c
0.024347 ± 0.000026	13.23 ± 0.29	8 ± 14	b
0.099825 ± 0.001999	15.98 ± 80.21	69 ± 1018	c
0.025422 ± 0.000017	13.10 ± 0.09	19 ± 9	b
0.100238 ± 0.047492	14.80 ± 412.41	27 ± 3682	c
0.030210 ± 0.000011	13.17 ± 0.22	8 ± 9	b
0.103008 ± 0.000123	13.25 ± 0.37	40 ± 50	b
0.030732 ± 0.000015	13.75 ± 0.05	35 ± 6	b
0.109920 ± 0.000022	13.39 ± 0.05	54 ± 9	a
0.031179 ± 0.000059	13.09 ± 0.23	32 ± 31	b
0.115162 ± 0.000008	13.20 ± 0.08	10 ± 5	a
0.032172 ± 0.000013	13.77 ± 0.07	18 ± 5	b
0.115374 ± 0.000007	13.48 ± 0.04	21 ± 3	a
0.032359 ± 0.000010	14.22 ± 0.21	20 ± 5	b
0.140988 ± 0.000009	13.09 ± 0.11	8 ± 5	a
0.036397 ± 0.000032	13.43 ± 0.12	35 ± 14	b
0.142199 ± 0.000012	13.07 ± 0.07	15 ± 5	a
0.041237 ± 0.000010	13.88 ± 0.03	48 ± 4	b
0.145845 ± 0.000011	13.42 ± 0.04	35 ± 4	a
0.042913 ± 0.000032	13.42 ± 0.07	60 ± 14	b
0.149751 ± 0.000014	13.02 ± 0.16	8 ± 9	a
0.044827 ± 0.000012	13.70 ± 0.04	38 ± 5	b
0.153657 ± 0.000020	12.93 ± 0.12	14 ± 9	a
0.045967 ± 0.000029	13.53 ± 0.06	63 ± 12	b
0.155088 ± 0.000022	13.03 ± 0.09	24 ± 9	a
0.050678 ± 0.000012	13.37 ± 0.37	8 ± 8	b
0.155443 ± 0.000009	13.30 ± 0.06	15 ± 4	a
0.055899 ± 0.000037	13.02 ± 0.20	19 ± 19	b
0.155690 ± 0.000010	13.35 ± 0.05	20 ± 5	a

Note. Only a portion of this table is shown. The full table will be available in the online version of the paper. (1) and (5): H I redshift. (2) and (6): H I column density from Voigt profile fitting. (3) and (7): H I Doppler parameter from Voigt profile fitting. (4) and (8): Confidence label: (a) 'secure'; (b) 'probable'; and (c) 'uncertain' (see Section 4.5.3 for definitions). See Section 4.5 for further details.

Table A.6: H I absorption systems in QSO J102218.99+013218.8.

z (1)	$\log(N_{\rm HI}/{\rm cm}^{-2})$ (2)	$b_{\rm HI}$ (km s^{-1}) (3)	label (4)
0.050692 ± 0.000028	13.54 ± 0.08	46 ± 12	b
0.197224 ± 0.000019	13.10 ± 0.09	20 ± 8	a
0.056892 ± 0.000017	17.24 ± 0.49	13 ± 2	b
0.209025 ± 0.000017	13.34 ± 0.21	10 ± 8	a
0.057127 ± 0.000205	13.49 ± 0.61	50 ± 55	c
0.219244 ± 0.000005	14.80 ± 0.03	35 ± 1	a
0.058275 ± 0.000052	13.43 ± 0.14	49 ± 24	b
0.232322 ± 0.000012	13.66 ± 0.06	20 ± 5	a
0.058626 ± 0.000014	13.82 ± 0.11	20 ± 6	b
0.232584 ± 0.000009	13.81 ± 0.14	15 ± 4	a
0.072400 ± 0.000009	14.09 ± 0.06	33 ± 4	b
0.240138 ± 0.000018	13.39 ± 0.38	7 ± 6	a
0.074368 ± 0.000022	14.21 ± 0.07	34 ± 7	b
0.241623 ± 0.000016	13.61 ± 0.05	32 ± 6	a
0.074596 ± 0.000035	13.48 ± 0.24	18 ± 12	b
0.245815 ± 0.000025	13.13 ± 0.10	24 ± 10	a
0.093134 ± 0.000018	13.18 ± 0.10	19 ± 9	b
0.270943 ± 0.000017	13.73 ± 0.04	42 ± 6	a
0.116075 ± 0.000017	13.19 ± 0.10	16 ± 9	a
0.278673 ± 0.000045	13.33 ± 0.13	36 ± 17	a
0.119165 ± 0.000014	13.61 ± 0.10	18 ± 6	a
0.279029 ± 0.000018	13.71 ± 0.07	27 ± 7	a
0.124133 ± 0.000009	13.81 ± 0.05	25 ± 4	a
0.279424 ± 0.000045	13.99 ± 0.80	23 ± 17	c
0.127075 ± 0.000020	13.34 ± 0.10	22 ± 8	a
0.279616 ± 0.000930	13.43 ± 3.51	43 ± 278	c
0.133211 ± 0.000011	13.37 ± 0.09	14 ± 5	a
0.279935 ± 0.000149	13.13 ± 1.03	25 ± 32	c
0.134934 ± 0.000030	13.26 ± 0.12	27 ± 13	a
0.290005 ± 0.000018	13.48 ± 0.07	24 ± 6	a
0.137339 ± 0.000005	14.29 ± 0.09	20 ± 2	a
0.293121 ± 0.000016	13.68 ± 0.05	32 ± 6	a
0.149280 ± 0.000033	13.41 ± 0.09	46 ± 13	a
0.293685 ± 0.000038	13.45 ± 0.15	24 ± 13	a
0.160057 ± 0.000016	13.02 ± 0.10	14 ± 8	a
0.293969 ± 0.000044	13.43 ± 0.15	29 ± 14	a
0.166105 ± 0.000012	13.35 ± 0.06	20 ± 5	a
0.303449 ± 0.000016	13.73 ± 0.06	25 ± 6	a
0.180369 ± 0.000025	13.16 ± 0.10	28 ± 11	a
0.303737 ± 0.000020	13.42 ± 0.10	16 ± 8	a
0.196702 ± 0.000195	13.17 ± 0.82	31 ± 38	c
0.306712 ± 0.000025	14.14 ± 0.10	23 ± 9	a
0.196872 ± 0.000074	13.33 ± 0.57	23 ± 15	c
0.340110 ± 0.000018	13.28 ± 0.18	10 ± 8	a

Note. Only a portion of this table is shown. The full table will be available in the online version of the paper. (1) and (5): H I redshift. (2) and (6): H I column density from Voigt profile fitting. (3) and (7): H I Doppler parameter from Voigt profile fitting. (4) and (8): Confidence label: (a) 'secure'; (b) 'probable'; and (c) 'uncertain' (see Section 4.5.3 for definitions). See Section 4.5 for further details.

Table A.7: H I absorption systems in QSO J135726.27+043541.4.

z (1)	$\log(N_{\rm HI}/{\rm cm}^{-2})$ (2)	$b_{\rm HI}$ (km s^{-1}) (3)	label (4)
0.017131 ± 0.000257	13.91 ± 0.73	61 ± 42	c
0.042974 ± 0.000009	14.01 ± 0.23	14 ± 8	b
0.017248 ± 0.000014	14.08 ± 0.34	24 ± 15	b
0.043146 ± 0.000011	14.08 ± 0.07	21 ± 5	b
0.017598 ± 0.000126	13.40 ± 0.81	50 ± 103	c
0.043397 ± 0.000005	14.95 ± 2.39	13 ± 11	c
0.018668 ± 0.000011	13.37 ± 0.07	16 ± 6	b
0.043605 ± 0.000009	13.90 ± 0.04	25 ± 3	b
0.018958 ± 0.000060	13.24 ± 0.26	33 ± 27	b
0.045631 ± 0.000007	14.08 ± 0.02	44 ± 3	b
0.019134 ± 0.000011	13.77 ± 0.15	14 ± 5	b
0.048113 ± 0.000006	14.38 ± 0.22	27 ± 4	b
0.020295 ± 0.000029	13.46 ± 0.06	64 ± 13	b
0.051101 ± 0.000011	13.45 ± 0.44	8 ± 8	b
0.021041 ± 0.000016	13.14 ± 0.09	17 ± 8	b
0.051261 ± 0.000008	13.97 ± 1.89	9 ± 11	c
0.027518 ± 0.000011	14.22 ± 0.13	22 ± 4	b
0.051461 ± 0.000032	13.56 ± 0.21	19 ± 12	b
0.027706 ± 0.000022	13.73 ± 0.11	25 ± 7	b
0.051839 ± 0.000013	13.22 ± 0.13	11 ± 8	b
0.028369 ± 0.000027	12.96 ± 0.11	29 ± 14	b
0.052407 ± 0.000015	13.81 ± 0.04	54 ± 6	b
0.028676 ± 0.000008	13.33 ± 0.06	14 ± 4	b
0.059531 ± 0.000018	13.66 ± 0.04	54 ± 7	b
0.029380 ± 0.000012	13.40 ± 0.04	33 ± 5	b
0.060329 ± 0.000013	13.30 ± 0.07	19 ± 6	b
0.031349 ± 0.000005	13.82 ± 0.84	8 ± 5	c
0.076092 ± 0.000009	13.76 ± 0.12	15 ± 5	b
0.032295 ± 0.000031	13.23 ± 0.07	58 ± 14	b
0.076345 ± 0.000011	14.00 ± 0.04	36 ± 5	b
0.034097 ± 0.000008	13.66 ± 0.03	36 ± 3	b
0.078359 ± 0.000026	13.36 ± 0.12	26 ± 10	b
0.036035 ± 0.000004	14.08 ± 1.10	9 ± 6	c
0.078661 ± 0.000012	14.06 ± 0.04	39 ± 5	b
0.039034 ± 0.000010	13.05 ± 0.07	12 ± 6	b
0.080128 ± 0.000030	13.59 ± 0.06	69 ± 12	b
0.039644 ± 0.000006	13.38 ± 0.07	12 ± 4	b
0.082165 ± 0.000018	13.34 ± 0.12	13 ± 9	b
0.039952 ± 0.000024	13.09 ± 0.08	38 ± 12	b
0.082351 ± 0.000010	14.13 ± 0.20	19 ± 5	b
0.041771 ± 0.000008	13.21 ± 0.07	12 ± 4	b
0.083518 ± 0.000021	13.07 ± 0.11	17 ± 10	b
0.042824 ± 0.000014	13.54 ± 0.08	15 ± 5	b
0.084793 ± 0.000017	13.18 ± 0.09	18 ± 8	b

Note. Only a portion of this table is shown. The full table will be available in the online version of the paper. (1) and (5): H I redshift. (2) and (6): H I column density from Voigt profile fitting. (3) and (7): H I Doppler parameter from Voigt profile fitting. (4) and (8): Confidence label: (a) 'secure'; (b) 'probable'; and (c) 'uncertain' (see Section 4.5.3 for definitions). See Section 4.5 for further details.

Table A.8: H I absorption systems in QSO J221806.67+005223.6.

z (1)	$\log(N_{\rm HI}/{\rm cm}^{-2})$ (2)	$b_{\rm HI}$ (km s^{-1}) (3)	label (4)
0.788733 ± 0.000049	14.22 ± 0.02	142 ± 13	b
0.948039 ± 0.000071	14.30 ± 1.07	36 ± 60	c
0.801500 ± 0.000096	14.21 ± 0.06	111 ± 26	b
0.999144 ± 0.000091	13.86 ± 0.04	112 ± 23	b
0.809249 ± 0.000097	14.28 ± 0.04	183 ± 24	b
1.012488 ± 0.000108	13.81 ± 0.06	87 ± 28	b
0.839592 ± 0.000086	14.19 ± 0.13	61 ± 30	b
1.015784 ± 0.000103	14.08 ± 0.04	148 ± 24	b
0.841663 ± 0.000058	15.05 ± 0.61	94 ± 38	c
1.048160 ± 0.000030	14.84 ± 0.08	71 ± 5	a
0.844030 ± 0.000089	14.12 ± 0.09	72 ± 28	b
1.051686 ± 0.000080	14.04 ± 0.04	121 ± 18	a
0.878483 ± 0.000117	15.58 ± 2.87	96 ± 99	c
1.083814 ± 0.000088	14.99 ± 15.67	26 ± 151	c
0.886827 ± 0.006759	14.18 ± 21.99	90 ± 390	c
1.084660 ± 0.000347	14.11 ± 0.14	190 ± 38	a
0.887010 ± 0.003798	14.19 ± 21.18	65 ± 989	c
1.093294 ± 0.000056	14.28 ± 1.64	27 ± 39	c
0.919839 ± 0.000787	13.68 ± 0.42	147 ± 143	b
1.098911 ± 0.000068	14.00 ± 0.35	33 ± 27	a
0.921107 ± 0.000234	13.73 ± 0.36	77 ± 45	b
1.130129 ± 0.000042	15.23 ± 0.56	24 ± 6	c
0.944475 ± 0.000473	14.00 ± 0.13	315 ± 106	b
1.213798 ± 0.000032	14.51 ± 0.30	39 ± 9	a
0.947890 ± 0.000220	14.18 ± 0.13	191 ± 62	b
1.217410 ± 0.000024	15.88 ± 0.59	70 ± 12	c

Note. This is the full table. (1) and (5): H I redshift. (2) and (6): H I column density from Voigt profile fitting. (3) and (7): H I Doppler parameter from Voigt profile fitting. (4) and (8): Confidence label: (a) 'secure'; (b) 'probable'; and (c) 'uncertain' (see Section 4.5.3 for definitions). See Section 4.5 for further details.

Table A.9: Spectroscopic catalog of objects in the Q0107 field.

R.A. (degrees) (1)	Dec. (degrees) (2)	z (3)	z label (4)	Spec. Type (5)	R (mag) (6)	CLASS_STAR (7)	Instrument (8)
17.38011	-2.45953	...	c	none	22.04 ± 0.02	0.89	VIMOS
17.38029	-2.44843	...	c	none	21.58 ± 0.01	0.91	VIMOS
17.38067	-2.39631	...	c	none	22.86 ± 0.03	0.85	VIMOS
17.38092	-2.29300	...	c	none	22.76 ± 0.06	0.01	VIMOS
17.38147	-2.45457	...	c	none	20.60 ± 0.01	0.03	VIMOS
17.38153	-2.28402	0.8206	a	SF	22.62 ± 0.03	0.22	VIMOS
17.38383	-2.30767	...	c	none	22.85 ± 0.05	0.87	VIMOS
17.38384	-2.31244	0.2070	a	SF	21.49 ± 0.01	0.91	VIMOS
17.38433	-2.42912	0.5758	a	SF	21.85 ± 0.02	0.98	VIMOS
17.38459	-2.38049	0.5658	a	SF	21.46 ± 0.01	0.11	VIMOS
17.38593	-2.42506	...	c	none	21.88 ± 0.03	0.72	VIMOS
17.38661	-2.27211	0.1908	a	non-SF	18.48 ± 0.01	0.62	VIMOS
17.38672	-2.43483	0.2604	a	SF	22.25 ± 0.04	0.04	VIMOS
17.38769	-2.39048	0.1898	a	non-SF	18.92 ± 0.01	0.04	VIMOS
17.38899	-2.38348	0.4298	a	non-SF	19.57 ± 0.01	0.04	VIMOS
17.38948	-2.46353	...	c	none	22.78 ± 0.06	0.15	VIMOS
17.38948	-2.28029	...	c	none	22.40 ± 0.02	0.12	VIMOS
17.39174	-2.23779	0.8750	b	SF	21.92 ± 0.02	0.06	VIMOS
17.39238	-2.32387	0.3228	a	SF	19.35 ± 0.01	0.03	VIMOS
17.39346	-2.26905	0.5678	b	SF	22.00 ± 0.02	0.03	VIMOS
17.39372	-2.26188	...	c	none	23.22 ± 0.04	0.92	VIMOS
17.39382	-2.26352	0.1235	a	SF	20.76 ± 0.01	0.03	VIMOS
17.39425	-2.32939	0.1858	a	SF	20.60 ± 0.01	0.03	VIMOS
17.39534	-2.22252	0.4318	b	SF	21.52 ± 0.01	0.34	VIMOS
17.39548	-2.46720	0.4318	a	SF	21.51 ± 0.01	0.57	VIMOS
17.39580	-2.32021	...	c	none	22.84 ± 0.04	0.92	VIMOS
17.39689	-2.32676	...	c	none	22.78 ± 0.05	0.10	VIMOS
17.39936	-2.44477	...	c	none	21.03 ± 0.01	0.03	VIMOS
17.40124	-2.25178	...	c	none	22.86 ± 0.04	0.02	VIMOS
17.40169	-2.41860	...	c	none	22.95 ± 0.04	0.05	VIMOS
17.40238	-2.36791	0.7214	b	SF	22.92 ± 0.05	0.76	VIMOS
17.40259	-2.24309	0.5698	a	SF	22.66 ± 0.03	0.79	VIMOS
17.40325	-2.25935	0.0000	b	star	20.68 ± 0.01	0.03	VIMOS
17.40331	-2.27535	0.7548	a	SF	21.84 ± 0.02	0.03	VIMOS
17.40371	-2.25559	...	c	none	19.75 ± 0.01	0.03	VIMOS
17.40562	-2.22848	...	c	none	21.72 ± 0.01	0.25	VIMOS
17.40621	-2.39338	0.7564	a	SF	22.67 ± 0.04	0.77	VIMOS
17.40647	-2.44007	0.0000	a	star	18.79 ± 0.01	0.98	VIMOS
17.40670	-2.30130	0.5778	b	SF	21.37 ± 0.02	0.02	VIMOS
17.40800	-2.24577	0.5710	a	SF	20.94 ± 0.01	0.87	VIMOS
17.40895	-2.29587	0.4693	a	non-SF	21.61 ± 0.01	0.98	VIMOS
17.40936	-2.40779	0.5125	b	SF	21.93 ± 0.02	0.03	VIMOS

Note. Only a portion of this table is shown. The full table will be available in the online version of the paper. (1) Right ascension (J2000). (2) Declination (J2000). (3) Redshift. (4) Redshift label: secure ('a'), possible ('b'), no idea ('c'), undefined ('n'). (5) Spectral type: star-forming galaxy ('SF'), non-star-forming ('non-SF'), star ('star'), active galactic nuclei ('AGN'), undefined ('none'). (6) R-band magnitude (MAG_AUTO) given by SEXTRACTOR; we note that these uncertainties might be underestimated by a factor of ~ 3. (7) CLASS_STAR given by SEXTRACTOR. (8) Instrument. See Section 4.6 for further details.

Table A.10: Spectroscopic catalog of objects in the J1005 field.

R.A. (degrees)	Dec. (degrees)	z	z label	Spec. Type	R (mag)	CLASS_STAR	Instrument
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
151.20108	1.49272	0.0010	a	star	18.97 ± 0.01	0.98	VIMOS
151.20265	1.61368	1.2043	b	SF	23.28 ± 0.05	0.66	VIMOS
151.20276	1.43851	0.1284	b	SF	22.56 ± 0.03	0.15	VIMOS
151.20418	1.44580	0.9792	b	non-SF	...	...	VIMOS
151.20418	1.44879	...	c	none	23.00 ± 0.04	0.70	VIMOS
151.20469	1.66112	...	c	none	22.13 ± 0.03	0.00	VIMOS
151.20507	1.64756	-0.0001	a	star	21.69 ± 0.01	0.98	VIMOS
151.20593	1.57977	0.5020	a	SF	22.04 ± 0.02	0.57	VIMOS
151.20654	1.46163	...	c	none	21.76 ± 0.01	0.98	VIMOS
151.20737	1.52252	0.3741	a	SF	22.30 ± 0.02	0.90	VIMOS
151.20745	1.44131	...	c	none	21.07 ± 0.01	0.01	VIMOS
151.20786	1.59262	0.6756	a	SF	22.88 ± 0.05	0.00	VIMOS
151.20786	1.65663	0.6171	a	SF	22.01 ± 0.03	0.02	VIMOS
151.20807	1.51942	0.3758	a	SF	21.99 ± 0.02	0.07	VIMOS
151.20824	1.61967	...	c	none	21.16 ± 0.01	0.98	VIMOS
151.20876	1.65437	0.4140	a	SF	21.10 ± 0.01	0.02	VIMOS
151.20898	1.60191	-0.0003	a	star	22.30 ± 0.02	0.96	VIMOS
151.20899	1.60833	0.1833	a	SF	20.15 ± 0.01	0.03	VIMOS
151.21012	1.43022	0.0007	a	star	21.14 ± 0.01	0.98	VIMOS
151.21039	1.49978	0.6186	a	SF	22.18 ± 0.02	0.52	VIMOS
151.21094	1.48094	0.3369	b	SF	20.21 ± 0.01	0.03	VIMOS
151.21137	1.56292	0.4217	a	SF	...	...	VIMOS
151.21501	1.45867	-0.0002	a	star	22.54 ± 0.02	0.95	VIMOS
151.21624	1.66290	0.4349	a	non-SF	19.99 ± 0.01	0.03	VIMOS
151.21690	1.46603	0.0004	b	star	22.21 ± 0.02	0.96	VIMOS
151.21765	1.60560	0.3046	b	SF	20.63 ± 0.01	0.03	VIMOS
151.21766	1.51071	0.2668	a	SF	21.46 ± 0.01	0.03	VIMOS
151.21860	1.62702	0.3607	a	non-SF	20.61 ± 0.01	0.98	VIMOS
151.21912	1.47104	0.2784	a	SF	21.62 ± 0.02	0.10	VIMOS
151.21917	1.46904	0.8439	a	SF	21.37 ± 0.01	0.79	VIMOS
151.22008	1.59780	-0.0006	b	star	19.84 ± 0.01	0.98	VIMOS
151.22048	1.59640	0.3799	a	SF	21.13 ± 0.01	0.37	VIMOS
151.22182	1.63121	0.3408	a	SF	22.21 ± 0.02	0.44	VIMOS
151.22212	1.66830	0.4357	a	non-SF	19.41 ± 0.01	0.03	VIMOS
151.22541	1.62459	0.5973	a	SF	22.29 ± 0.03	0.15	VIMOS
151.22703	1.57367	0.1773	a	SF	20.43 ± 0.01	0.04	VIMOS
151.22793	1.50934	...	c	none	23.07 ± 0.05	0.00	VIMOS
151.22802	1.64551	0.4308	a	SF	22.23 ± 0.02	0.15	VIMOS
151.22852	1.47628	0.4130	a	SF	21.24 ± 0.01	0.97	VIMOS
151.22882	1.48978	1.2499	b	AGN	21.68 ± 0.01	0.98	VIMOS
151.23032	1.67427	0.4658	a	SF	21.16 ± 0.01	0.02	VIMOS
151.23139	1.47889	...	c	none	20.40 ± 0.01	0.98	VIMOS

Note. Only a portion of this table is shown. The full table will be available in the online version of the paper. (1) Right ascension (J2000). (2) Declination (J2000). (3) Redshift. (4) Redshift label: secure ('a'), possible ('b'), no idea ('c'), undefined ('n'). (5) Spectral type: star-forming galaxy ('SF'), non-star-forming ('non-SF'), star ('star'), active galactic nuclei ('AGN'), undefined ('none'). (6) R-band magnitude (MAG_AUTO) given by SEXTRACTOR; we note that these uncertainties might be underestimated by a factor of ~ 3. (7) CLASS_STAR given by SEXTRACTOR. (8) Instrument. See Section 4.6 for further details.

Table A.11: Spectroscopic catalog of objects in the J1022 field.

R.A. (degrees) (1)	Dec. (degrees) (2)	z (3)	z label (4)	Spec. Type (5)	R (mag) (6)	CLASS_STAR (7)	Instrument (8)
155.37715	1.39655	0.8507	a	SF	21.53 ± 0.02	0.00	VIMOS
155.38284	1.41530	1.1483	b	SF	23.39 ± 0.06	0.00	VIMOS
155.38420	1.40041	0.0000	a	star	20.79 ± 0.01	0.98	VIMOS
155.38583	1.47086	0.6002	a	non-SF	21.43 ± 0.01	0.02	VIMOS
155.38706	1.48588	0.5856	b	non-SF	...	...	VIMOS
155.38832	1.46557	0.6024	a	non-SF	21.82 ± 0.02	0.13	VIMOS
155.38857	1.40637	-0.0001	a	star	21.86 ± 0.01	0.98	VIMOS
155.38882	1.45225	0.5859	a	SF	21.92 ± 0.02	0.42	VIMOS
155.38967	1.48727	0.5850	a	SF	...	...	VIMOS
155.39031	1.45981	0.3872	a	non-SF	21.95 ± 0.02	0.19	VIMOS
155.39137	1.44240	0.2793	a	SF	22.00 ± 0.02	0.03	VIMOS
155.39313	1.37581	0.3280	a	SF	22.76 ± 0.03	0.05	VIMOS
155.39317	1.38055	0.5095	a	SF	21.71 ± 0.01	0.14	VIMOS
155.39435	1.43998	...	c	none	21.64 ± 0.01	0.98	VIMOS
155.39496	1.44664	0.8356	a	AGN	20.79 ± 0.01	0.98	VIMOS
155.39525	1.45809	0.5434	a	SF	22.62 ± 0.03	0.01	VIMOS
155.39590	1.41016	0.3792	a	SF	21.85 ± 0.02	0.02	VIMOS
155.39685	1.47371	0.3886	a	non-SF	20.79 ± 0.01	0.04	VIMOS
155.39896	1.39150	0.3822	a	SF	20.52 ± 0.01	0.02	VIMOS
155.40109	1.43295	1.1802	a	SF	22.39 ± 0.03	0.17	VIMOS
155.40404	1.42086	0.0001	a	star	20.06 ± 0.01	0.98	VIMOS
155.40579	1.46164	0.4315	b	non-SF	21.78 ± 0.01	0.40	VIMOS
155.40968	1.48033	0.3387	a	SF	22.10 ± 0.03	0.00	VIMOS
155.40993	1.38644	0.5379	a	SF	21.64 ± 0.01	0.63	VIMOS
155.41473	1.42251	0.2704	a	SF	22.32 ± 0.02	0.96	VIMOS
155.41755	1.48266	...	c	none	20.59 ± 0.01	0.02	VIMOS
155.41755	1.48306	0.6919	a	SF	20.59 ± 0.01	0.02	VIMOS
155.42122	1.41347	0.5490	a	SF	20.92 ± 0.01	0.03	VIMOS
155.42184	1.45377	0.7426	b	non-SF	22.24 ± 0.03	0.00	VIMOS
155.42204	1.50142	0.7131	a	SF	21.65 ± 0.02	0.00	VIMOS
155.42233	1.47884	-0.0001	a	star	20.88 ± 0.01	0.15	VIMOS
155.42239	1.50883	0.2774	a	SF	...	...	VIMOS
155.42247	1.46069	0.6690	a	SF	22.38 ± 0.03	0.00	VIMOS
155.42297	1.49994	0.0001	a	star	21.24 ± 0.01	0.07	VIMOS
155.42303	1.64718	0.0001	a	star	22.18 ± 0.02	0.98	VIMOS
155.42307	1.44942	0.9721	b	AGN	21.17 ± 0.01	0.88	VIMOS
155.42318	1.42447	0.0003	a	star	20.24 ± 0.01	0.94	VIMOS
155.42378	1.43253	...	c	none	...	...	VIMOS
155.42378	1.43483	0.3786	a	SF	22.80 ± 0.03	0.04	VIMOS
155.42421	1.59549	0.2793	b	SF	21.56 ± 0.02	0.01	VIMOS
155.42462	1.65111	...	c	none	22.36 ± 0.02	0.97	VIMOS
155.42480	1.57184	0.3838	a	non-SF	19.88 ± 0.01	0.03	VIMOS

Note. Only a portion of this table is shown. The full table will be available in the online version of the paper. (1) Right ascension (J2000). (2) Declination (J2000). (3) Redshift. (4) Redshift label: secure ('a'), possible ('b'), no idea ('c'), undefined ('n'). (5) Spectral type: star-forming galaxy ('SF'), non-star-forming ('non-SF'), star ('star'), active galactic nuclei ('AGN'), undefined ('none'). (6) R-band magnitude (MAG_AUTO) given by SEXTRACTOR; we note that these uncertainties might be underestimated by a factor of ~ 3. (7) CLASS_STAR given by SEXTRACTOR. (8) Instrument. See Section 4.6 for further details.

Table A.12: Spectroscopic catalog of objects in the J2218 field.

R.A. (degrees) (1)	Dec. (degrees) (2)	z (3)	z label (4)	Spec. Type (5)	R (mag) (6)	CLASS_STAR (7)	Instrument (8)
334.33420	0.88225	...	c	none	22.66 ± 0.06	0.76	VIMOS
334.33427	0.87096	0.7139	a	SF	22.48 ± 0.06	0.06	VIMOS
334.33532	0.76281	-0.0007	a	star	21.45 ± 0.02	0.95	VIMOS
334.33539	0.95526	-0.0006	a	star	20.36 ± 0.01	0.98	VIMOS
334.33540	0.87713	-0.0001	a	star	21.14 ± 0.02	0.98	VIMOS
334.33540	0.87816	...	c	none	21.14 ± 0.02	0.98	VIMOS
334.33553	0.89967	0.2770	a	SF	19.98 ± 0.01	0.02	VIMOS
334.33776	0.94074	-0.0008	a	star	21.05 ± 0.02	0.98	VIMOS
334.33805	0.75379	0.4266	a	non-SF	20.18 ± 0.01	0.06	VIMOS
334.33840	0.89023	-0.0003	a	star	20.48 ± 0.01	0.98	VIMOS
334.33872	0.86856	-0.0002	a	star	22.82 ± 0.06	0.73	VIMOS
334.34238	0.85964	0.0000	a	star	21.57 ± 0.02	0.97	VIMOS
334.34268	0.95394	0.3552	a	SF	20.23 ± 0.01	0.02	VIMOS
334.34329	0.80460	...	c	none	22.33 ± 0.05	0.88	VIMOS
334.34470	0.80777	0.5634	b	SF	22.03 ± 0.04	0.45	VIMOS
334.34497	0.81192	-0.0006	a	star	21.33 ± 0.02	0.98	VIMOS
334.34497	0.81424	...	c	none	22.04 ± 0.04	0.16	VIMOS
334.34521	0.80232	0.2780	b	non-SF	21.17 ± 0.02	0.24	VIMOS
334.34639	0.94785	...	c	none	21.87 ± 0.04	0.16	VIMOS
334.34675	0.72373	...	c	none	21.52 ± 0.04	0.00	VIMOS
334.34679	0.70623	...	c	none	21.64 ± 0.03	0.97	VIMOS
334.34808	0.86693	...	c	none	22.62 ± 0.07	0.09	VIMOS
334.34811	0.72829	...	c	none	21.43 ± 0.03	0.01	VIMOS
334.34837	0.90869	...	c	none	20.85 ± 0.01	0.98	VIMOS
334.34899	0.74437	...	c	none	21.08 ± 0.02	0.98	VIMOS
334.34957	0.71501	2.6775	b	AGN	20.71 ± 0.01	0.98	VIMOS
334.35004	0.73447	0.5102	a	SF	21.49 ± 0.03	0.01	VIMOS
334.35004	0.73733	...	c	none	...	...	VIMOS
334.35029	0.77934	...	c	none	21.11 ± 0.02	0.01	VIMOS
334.35062	0.93009	...	c	none	21.87 ± 0.05	0.00	VIMOS
334.35116	0.89441	0.5197	a	non-SF	20.90 ± 0.02	0.04	VIMOS
334.35170	0.76951	0.2519	a	SF	22.05 ± 0.05	0.01	VIMOS
334.35180	0.79917	0.0001	a	star	22.98 ± 0.07	0.73	VIMOS
334.35212	0.78120	-0.0010	a	star	19.11 ± 0.01	0.90	VIMOS
334.35303	0.96757	-0.0010	a	star	20.74 ± 0.01	0.98	VIMOS
334.35317	0.71804	...	c	none	20.84 ± 0.02	0.42	VIMOS
334.35417	0.91667	...	c	none	20.37 ± 0.02	0.01	VIMOS
334.35472	0.81685	-0.0005	a	star	19.54 ± 0.01	0.98	VIMOS
334.35482	0.85446	...	c	none	20.40 ± 0.01	0.98	VIMOS
334.35491	0.78289	0.4486	b	non-SF	21.54 ± 0.03	0.34	VIMOS
334.35506	0.74805	...	c	none	21.14 ± 0.02	0.97	VIMOS
334.35588	0.90488	...	c	none	20.56 ± 0.01	0.98	VIMOS

Note. Only a portion of this table is shown. The full table will be available in the online version of the paper. (1) Right ascension (J2000). (2) Declination (J2000). (3) Redshift. (4) Redshift label: secure ('a'), possible ('b'), no idea ('c'), undefined ('n'). (5) Spectral type: star-forming galaxy ('SF'), non-star-forming ('non-SF'), star ('star'), active galactic nuclei ('AGN'), undefined ('none'). (6) R-band magnitude (MAG_AUTO) given by SEXTRACTOR; we note that these uncertainties might be underestimated by a factor of ~ 3. (7) CLASS_STAR given by SEXTRACTOR. (8) Instrument. See Section 4.6 for further details.

Bibliography

Abazajian, K. N., Adelman-McCarthy, J. K., Agüeros, M. A., Allam, S. S., Allende Prieto, C., An, D., Anderson, K. S. J., Anderson, S. F., Annis, J., Bahcall, N. A., and et al.: 2009, ApJS **182**, 543

Abraham, R. G., Glazebrook, K., McCarthy, P. J., Crampton, D., Murowinski, R., Jørgensen, I., Roth, K., Hook, I. M., Savaglio, S., Chen, H.-W., Marzke, R. O., and Carlberg, R. G.: 2004, AJ **127**, 2455

Adelberger, K. L., Shapley, A. E., Steidel, C. C., Pettini, M., Erb, D. K., and Reddy, N. A.: 2005, ApJ **629**, 636

Adelberger, K. L., Steidel, C. C., Shapley, A. E., and Pettini, M.: 2003, ApJ **584**, 45

Alfonso-Faus, A.: 2008, Ap&SS **318**, 117

Altay, G. and Theuns, T.: 2013, MNRAS **434**, 748

Altay, G., Theuns, T., Schaye, J., Crighton, N. H. M., and Dalla Vecchia, C.: 2011, ApJ **737**, L37

Aprile, E., Alfonsi, M., Arisaka, K., Arneodo, F., Balan, C., Baudis, L., Bauermeister, B., Behrens, A., Beltrame, P., Bokeloh, K., Brown, E., Bruno, G., Budnik, R., Cardoso, J. M. R., Chen, W.-T., Choi, B., Cline, D., Colijn, A. P., Contreras, H., Cussonneau, J. P., Decowski, M. P., Duchovni, E., Fattori, S., Ferella, A. D., Fulgione, W., Gao, F., Garbini, M., Ghag, C., Giboni, K.-L., Goetzke, L. W., Grignon, C., Gross, E., Hampel, W., Kaether, F., Kish, A., Lamblin, J., Landsman, H., Lang, R. F., Le Calloch, M., Levy, C., Lim, K. E., Lin, Q., Lindemann, S.,

Lindner, M., Lopes, J. A. M., Lung, K., Marrodán Undagoitia, T., Massoli, F. V., Melgarejo Fernandez, A. J., Meng, Y., Molinario, A., Nativ, E., Ni, K., Oberlack, U., Orrigo, S. E. A., Pantic, E., Persiani, R., Plante, G., Priel, N., Rizzo, A., Rosendahl, S., dos Santos, J. M. F., Sartorelli, G., Schreiner, J., Schumann, M., Scotto Lavina, L., Scovell, P. R., Selvi, M., Shagin, P., Simgen, H., Teymourian, A., Thers, D., Vitells, O., Wang, H., Weber, M., and Weinheimer, C.: 2012, *Physical Review Letters* **109(18)**, 181301

Aragón-Calvo, M. A., van de Weygaert, R., and Jones, B. J. T.: 2010, MNRAS **408**, 2163

Babcock, H. W.: 1939, *Lick Observatory Bulletin* **19**, 41

Baugh, C. M.: 2006, *Reports on Progress in Physics* **69**, 3101

Baugh, C. M., Lacey, C. G., Frenk, C. S., Granato, G. L., Silva, L., Bressan, A., Benson, A. J., and Cole, S.: 2005, MNRAS **356**, 1191

Benson, A. J., Hoyle, F., Torres, F., and Vogeley, M. S.: 2003, MNRAS **340**, 160

Bernabei, R., Belli, P., Cappella, F., Cerulli, R., Dai, C. J., D'Angelo, A., He, H. L., Incicchitti, A., Kuang, H. H., Ma, X. H., Montecchia, F., Nozzoli, F., Prosperi, D., Sheng, X. D., Wang, R. G., and Ye, Z. P.: 2010, *European Physical Journal C* **67**, 39

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

Bielby, R. M., Shanks, T., Weilbacher, P. M., Infante, L., Crighton, N. H. M., Bornancini, C., Bouché, N., Héraudeau, P., Lambas, D. G., Lowenthal, J., Minniti, D., Padilla, N., Petitjean, P., and Theuns, T.: 2011, MNRAS **414**, 2

Bond, J. R., Kofman, L., and Pogosyan, D.: 1996, Nature **380**, 603

Bond, N. A., Strauss, M. A., and Cen, R.: 2010, MNRAS **409**, 156

Borgani, S., Governato, F., Wadsley, J., Menci, N., Tozzi, P., Quinn, T., Stadel, J., and Lake, G.: 2002, MNRAS **336**, 409

Bottini, D., Garilli, B., Maccagni, D., Tresse, L., Le Brun, V., Le Fèvre, O., Picat, J. P., Scaramella, R., Scodeggio, M., Vettolani, G., Zanichelli, A., Adami, C., Arnaboldi, M., Arnouts, S., Bardelli, S., Bolzonella, M., Cappi, A., Charlot, S., Ciliegi, P., Contini, T., Foucaud, S., Franzetti, P., Guzzo, L., Ilbert, O., Iovino, A., McCracken, H. J., Marano, B., Marinoni, C., Mathez, G., Mazure, A., Meneux, B., Merighi, R., Paltani, S., Pollo, A., Pozzetti, L., Radovich, M., Zamorani, G., and Zucca, E.: 2005, PASP **117**, 996

Bowen, D. V., Pettini, M., and Blades, J. C.: 2002, ApJ **580**, 169

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

Burbidge, E. M., Lynds, C. R., and Stockton, A. N.: 1968, ApJ **152**, 1077

Carroll, S. M.: 2001, *Living Reviews in Relativity* **4**, 1

Carswell, R. F. and Rees, M. J.: 1987, MNRAS **224**, 13P

Ceccarelli, L., Padilla, N. D., Valotto, C., and Lambas, D. G.: 2006, MNRAS **373**, 1440

Cen, R. and Ostriker, J. P.: 1999, ApJ **514**, 1

Chen, H.-W., Lanzetta, K. M., Webb, J. K., and Barcons, X.: 1998, ApJ **498**, 77

Chen, H.-W. and Mulchaey, J. S.: 2009, ApJ **701**, 1219

Chen, H.-W., Prochaska, J. X., Weiner, B. J., Mulchaey, J. S., and Williger, G. M.: 2005, ApJ **629**, L25

Clowe, D., Bradač, M., Gonzalez, A. H., Markevitch, M., Randall, S. W., Jones, C., and Zaritsky, D.: 2006, ApJ **648**, L109

Coil, A. L., Davis, M., Madgwick, D. S., Newman, J. A., Conselice, C. J., Cooper, M., Ellis, R. S., Faber, S. M., Finkbeiner, D. P., Guhathakurta, P., Kaiser, N., Koo,

D. C., Phillips, A. C., Steidel, C. C., Weiner, B. J., Willmer, C. N. A., and Yan, R.: 2004, ApJ **609**, 525

Colberg, J. M., Pearce, F., Foster, C., Platen, E., Brunino, R., Neyrinck, M., Basilakos, S., Fairall, A., Feldman, H., Gottlöber, S., Hahn, O., Hoyle, F., Müller, V., Nelson, L., Plionis, M., Porciani, C., Shandarin, S., Vogeley, M. S., and van de Weygaert, R.: 2008, MNRAS **387**, 933

Colberg, J. M., Sheth, R. K., Diaferio, A., Gao, L., and Yoshida, N.: 2005, MNRAS **360**, 216

Colberg, J. M., White, S. D. M., Yoshida, N., MacFarland, T. J., Jenkins, A., Frenk, C. S., Pearce, F. R., Evrard, A. E., Couchman, H. M. P., Efstathiou, G., Peacock, J. A., Thomas, P. A., and Virgo Consortium: 2000, MNRAS **319**, 209

Cole, S. and Lacey, C.: 1996, MNRAS **281**, 716

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., De Propris, R., 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 Taylor, K.: 2001, MNRAS **328**, 1039

Crain, R. A., Theuns, T., Dalla Vecchia, C., Eke, V. R., Frenk, C. S., Jenkins, A., Kay, S. T., Peacock, J. A., Pearce, F. R., Schaye, J., Springel, V., Thomas, P. A., White, S. D. M., and Wiersma, R. P. C.: 2009, MNRAS **399**, 1773

Crighton, N. H. M., Bechtold, J., Carswell, R. F., Davé, R., Foltz, C. B., Jannuzi, B. T., Morris, S. L., O'Meara, J. M., Prochaska, J. X., Schaye, J., and Tejos, N.: 2013, MNRAS **433**, 178

Crighton, N. H. M., Bielby, R., Shanks, T., Infante, L., Bornancini, C. G., Bouché, N., Lambas, D. G., Lowenthal, J. D., Minniti, D., Morris, S. L., Padilla, N., Péroux,

C., Petitjean, P., Theuns, T., Tummuangpak, P., Weilbacher, P. M., Wisotzki, L., and Worseck, G.: 2011, MNRAS **414**, 28

Crighton, N. H. M., Morris, S. L., Bechtold, J., Crain, R. A., Jannuzi, B. T., Shone, A., and Theuns, T.: 2010, MNRAS **402**, 1273

Croft, R. A. C., Weinberg, D. H., Katz, N., and Hernquist, L.: 1998, ApJ **495**, 44

Dalla Vecchia, C. and Schaye, J.: 2008, MNRAS **387**, 1431

Danforth, C. W. and Shull, J. M.: 2008, ApJ **679**, 194

Danforth, C. W., Shull, J. M., Rosenberg, J. L., and Stocke, J. T.: 2006, ApJ **640**, 716

Davé, R., Hernquist, L., Katz, N., and Weinberg, D. H.: 1999, ApJ **511**, 521

Davé, R., Oppenheimer, B. D., Katz, N., Kollmeier, J. A., and Weinberg, D. H.: 2010, MNRAS **408**, 2051

Davies, R. L., Allington-Smith, J. R., Bettess, P., Chadwick, E., Content, R., Dodsworth, G. N., Haynes, R., Lee, D., Lewis, I. J., Webster, J., Atad, E., Beard, S. M., Ellis, M., Hastings, P. R., Williams, P. R., Bond, T., Crampton, D., Davidge, T. J., Fletcher, M., Leckie, B., Morbey, C. L., Murowinski, R. G., Roberts, S., Saddlemyer, L. K., Sebesta, J., Stilburn, J. R., and Szeto, K.: 1997, in A. L. Ardeberg (ed.), *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Vol. 2871 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, pp 1099–1106

Davis, M. and Peebles, P. J. E.: 1983, ApJ **267**, 465

Dicke, R. H., Peebles, P. J. E., Roll, P. G., and Wilkinson, D. T.: 1965, ApJ **142**, 414

Doyle, M. T., Drinkwater, M. J., Rohde, D. J., Pimbblet, K. A., Read, M., Meyer, M. J., Zwaan, M. A., Ryan-Weber, E., Stevens, J., Koribalski, B. S., Webster, R. L., Staveley-Smith, L., Barnes, D. G., Howlett, M., Kilborn, V. A., Waugh, M., Pierce, M. J., Bhathal, R., de Blok, W. J. G., Disney, M. J., Ekers, R. D., Freeman, K. C.,

Garcia, D. A., Gibson, B. K., Harnett, J., Henning, P. A., Jerjen, H., Kesteven, M. J., Knezek, P. M., Mader, S., Marquarding, M., Minchin, R. F., O'Brien, J., Oosterloo, T., Price, R. M., Putman, M. E., Ryder, S. D., Sadler, E. M., Stewart, I. M., Stootman, F., and Wright, A. E.: 2005, MNRAS **361**, 34

Draine, B. T.: 2011, *Physics of the Interstellar and Intergalactic Medium*

Einstein, A.: 1905, *Annalen der Physik* **322**, 891

Einstein, A.: 1916, *Annalen der Physik* **354**, 769

Estrada, J., Sefusatti, E., and Frieman, J. A.: 2009, ApJ **692**, 265

Faber, S. M., Phillips, A. C., Kibrick, R. I., Alcott, B., Allen, S. L., Burrous, J., Cantrall, T., Clarke, D., Coil, A. L., Cowley, D. J., Davis, M., Deich, W. T. S., Dietsch, K., Gilmore, D. K., Harper, C. A., Hilyard, D. F., Lewis, J. P., McVeigh, M., Newman, J., Osborne, J., Schiavon, R., Stover, R. J., Tucker, D., Wallace, V., Wei, M., Wirth, G., and Wright, C. A.: 2003, in M. Iye and A. F. M. Moorwood (eds.), *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Vol. 4841 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, pp 1657–1669

Fan, X., Narayanan, V. K., Lupton, R. H., Strauss, M. A., Knapp, G. R., Becker, R. H., White, R. L., Pentericci, L., Leggett, S. K., Haiman, Z., Gunn, J. E., Ivezić, Ž., Schneider, D. P., Anderson, S. F., Brinkmann, J., Bahcall, N. A., Connolly, A. J., Csabai, I., Doi, M., Fukugita, M., Geballe, T., Grebel, E. K., Harbeck, D., Hennessy, G., Lamb, D. Q., Miknaitis, G., Munn, J. A., Nichol, R., Okamura, S., Pier, J. R., Prada, F., Richards, G. T., Szalay, A., and York, D. G.: 2001, AJ **122**, 2833

Fan, X., Strauss, M. A., Becker, R. H., White, R. L., Gunn, J. E., Knapp, G. R., Richards, G. T., Schneider, D. P., Brinkmann, J., and Fukugita, M.: 2006, AJ **132**, 117

Ferland, G. J., Korista, K. T., Verner, D. A., Ferguson, J. W., Kingdon, J. B., and Verner, E. M.: 1998, PASP **110**, 761

Forbes, D. A. and Kroupa, P.: 2011, PASA **28**, 77

Ford, A. B., Oppenheimer, B. D., Davé, R., Katz, N., Kollmeier, J. A., and Weinberg, D. H.: 2013, MNRAS **432**, 89

Friedman, A.: 1922, *Zeitschrift fur Physik* **10**, 377

Fukugita, M., Hogan, C. J., and Peebles, P. J. E.: 1998, ApJ **503**, 518

Fukugita, M. and Peebles, P. J. E.: 2004, ApJ **616**, 643

Fumagalli, M., Prochaska, J. X., Kasen, D., Dekel, A., Ceverino, D., and Primack, J. R.: 2011, MNRAS **418**, 1796

Gamow, G.: 1948, *Physical Review* **74**, 505

Gehrels, N.: 1986, ApJ **303**, 336

González, R. E. and Padilla, N. D.: 2010, MNRAS **407**, 1449

Green, J. C., Froning, C. S., Osterman, S., Ebbets, D., Heap, S. H., Leitherer, C., Linsky, J. L., Savage, B. D., Sembach, K., Shull, J. M., Siegmund, O. H. W., Snow, T. P., Spencer, J., Stern, S. A., Stocke, J., Welsh, B., Béland, S., Burgh, E. B., Danforth, C., France, K., Keeney, B., McPhate, J., Penton, S. V., Andrews, J., Brownsberger, K., Morse, J., and Wilkinson, E.: 2012, ApJ **744**, 60

Greenstein, J. L. and Schmidt, M.: 1967, ApJ **148**, L13

Grogin, N. A. and Geller, M. J.: 1998, ApJ **505**, 506

Gunn, J. E. and Peterson, B. A.: 1965, ApJ **142**, 1633

Haardt, F. and Madau, P.: 1996, ApJ **461**, 20

Hajdukovic, D. S.: 2011, Ap&SS **334**, 215

Hajdukovic, D. S.: 2012, Ap&SS **339**, 1

Hamilton, A. J. S.: 1993, ApJ **417**, 19

Hamuy, M., Suntzeff, N. B., Heathcote, S. R., Walker, A. R., Gigoux, P., and Phillips, M. M.: 1994, PASP **106**, 566

Hamuy, M., Walker, A. R., Suntzeff, N. B., Gigoux, P., Heathcote, S. R., and Phillips, M. M.: 1992, PASP **104**, 533

Hewett, P. C.: 1982, MNRAS **201**, 867

Hong, T., Han, J. L., Wen, Z. L., Sun, L., and Zhan, H.: 2012, ApJ **749**, 81

Hopkins, A. M. and Beacom, J. F.: 2006, ApJ **651**, 142

Horne, K.: 1986, PASP **98**, 609

Hoyle, F. and Vogeley, M. S.: 2002, ApJ **566**, 641

Hubble, E.: 1929, *Proceedings of the National Academy of Science* **15**, 168

Hui, L. and Gnedin, N. Y.: 1997, MNRAS **292**, 27

Hui, L. and Rutledge, R. E.: 1999, ApJ **517**, 541

Hummels, C. B., Bryan, G. L., Smith, B. D., and Turk, M. J.: 2013, MNRAS **430**, 1548

Icke, V.: 1984, MNRAS **206**, 1P

Impey, C. D., Petry, C. E., and Flint, K. P.: 1999, ApJ **524**, 536

Kaiser, N.: 1987, MNRAS **227**, 1

Keeney, B. A., Danforth, C. W., Stocke, J. T., France, K., and Green, J. C.: 2012, PASP **124**, 830

Keeney, B. A., Stocke, J. T., Rosenberg, J. L., Danforth, C. W., Ryan-Weber, E. V., Shull, J. M., Savage, B. D., and Green, J. C.: 2013, ApJ **765**, 27

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

Kerscher, M.: 1999, A&A **343**, 333

Kerscher, M., Szapudi, I., and Szalay, A. S.: 2000, ApJ **535**, L13

Keyes, C. D., Koratkar, A. P., Dahlem, M., Hayes, J., Christensen, J., and Martin, S.: 1995, *Faint Object Spectrograph Instrument Handbook v. 6.0*

Kirkman, D., Tytler, D., Lubin, D., and Charlton, J.: 2007, MNRAS **376**, 1227

Komatsu, E., Smith, K. M., Dunkley, J., Bennett, C. L., Gold, B., Hinshaw, G., Jarosik, N., Larson, D., Nolta, M. R., Page, L., Spergel, D. N., Halpern, M., Hill, R. S., Kogut, A., Limon, M., Meyer, S. S., Odegard, N., Tucker, G. S., Weiland, J. L., Wollack, E., and Wright, E. L.: 2011, ApJS **192**, 18

Kreckel, K., Platen, E., Aragón-Calvo, M. A., van Gorkom, J. H., van de Weygaert, R., van der Hulst, J. M., Kovač, K., Yip, C.-W., and Peebles, P. J. E.: 2011, AJ **141**, 4

Lagos, C. D. P., Cora, S. A., and Padilla, N. D.: 2008, MNRAS **388**, 587

Landy, S. D. and Szalay, A. S.: 1993, ApJ **412**, 64

Lanzetta, K. M., Bowen, D. V., Tytler, D., and Webb, J. K.: 1995, ApJ **442**, 538

Le Fèvre, O., Saisse, M., Mancini, D., Brau-Nogue, S., Caputi, O., Castinel, L., D'Odorico, S., Garilli, B., Kissler-Patig, M., Lucuix, C., Mancini, G., Pauget, A., Sciarretta, G., Scodeggio, M., Tresse, L., and Vettolani, G.: 2003, in M. Iye and A. F. M. Moorwood (eds.), *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Vol. 4841 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, pp 1670–1681

Le Fèvre, O., Vettolani, G., Garilli, B., Tresse, L., Bottini, D., Le Brun, V., Maccagni, D., Picat, J. P., Scaramella, R., Scodeggio, M., Zanichelli, A., Adami, C., Arnaboldi, M., Arnouts, S., Bardelli, S., Bolzonella, M., Cappi, A., Charlot, S., Ciliegi,

P., Contini, T., Foucaud, S., Franzetti, P., Gavignaud, I., Guzzo, L., Ilbert, O., Iovino, A., McCracken, H. J., Marano, B., Marinoni, C., Mathez, G., Mazure, A., Meneux, B., Merighi, R., Paltani, S., Pellò, R., Pollo, A., Pozzetti, L., Radovich, M., Zamorani, G., Zucca, E., Bondi, M., Bongiorno, A., Busarello, G., Lamareille, F., Mellier, Y., Merluzzi, P., Ripepi, V., and Rizzo, D.: 2005, A&A **439**, 845

Lehner, N., Howk, J. C., Tripp, T. M., Tumlinson, J., Prochaska, J. X., O'Meara, J. M., Thom, C., Werk, J. K., Fox, A. J., and Ribaudo, J.: 2013, ApJ **770**, 138

Lehner, N., Savage, B. D., Richter, P., Sembach, K. R., Tripp, T. M., and Wakker, B. P.: 2007, ApJ **658**, 680

Lemaître, G.: 1927, *Annales de la Societe Scietifique de Bruxelles* **47**, 49

Lemaître, G.: 1931, Nature **127**, 706

Lewis, A., Challinor, A., and Lasenby, A.: 2000, ApJ **538**, 473

Lewis, I., Balogh, M., De Propris, R., Couch, W., Bower, R., Offer, A., Bland-Hawthorn, J., Baldry, I. K., Baugh, C., Bridges, T., Cannon, R., Cole, S., Colless, M., Collins, C., Cross, N., Dalton, G., Driver, S. P., Efstathiou, G., Ellis, R. S., Frenk, C. S., Glazebrook, K., Hawkins, E., Jackson, C., Lahav, O., Lumsden, S., Maddox, S., Madgwick, D., Norberg, P., Peacock, J. A., Percival, W., Peterson, B. A., Sutherland, W., and Taylor, K.: 2002, MNRAS **334**, 673

Lopez, S., Barrientos, L. F., Lira, P., Padilla, N., Gilbank, D. G., Gladders, M. D., Maza, J., Tejos, N., Vidal, M., and Yee, H. K. C.: 2008, ApJ **679**, 1144

Manning, C. V.: 2002, ApJ **574**, 599

Manning, C. V.: 2003, ApJ **591**, 79

Meiksin, A. A.: 2009, *Reviews of Modern Physics* **81**, 1405

Michelson, A. A.: 1925, ApJ **61**, 137

Michelson, A. A. and Gale, H. G.: 1925, ApJ **61**, 140

Michelson, A. A. and Morley, E. W.: 1887, *Sidereal Messenger, vol. 6, pp.306-310* **6**, 306

Michelson, A. A., Pease, F. G., and Pearson, F.: 1935, ApJ **82**, 26

Milgrom, M.: 1983, ApJ **270**, 365

Mo, H., van den Bosch, F. C., and White, S.: 2010, *Galaxy Formation and Evolution*

Mo, H. J.: 1994, MNRAS **269**, L49

Mo, H. J. and Morris, S. L.: 1994, MNRAS **269**, 52

Mo, H. J., Peacock, J. A., and Xia, X. Y.: 1993, MNRAS **260**, 121

Moffat, J. W.: 2006, JCAP **3**, 4

Moos, H. W., Cash, W. C., Cowie, L. L., Davidsen, A. F., Dupree, A. K., Feldman, P. D., Friedman, S. D., Green, J. C., Green, R. F., Gry, C., Hutchings, J. B., Jenkins, E. B., Linsky, J. L., Malina, R. F., Michalitsianos, A. G., Savage, B. D., Shull, J. M., Siegmund, O. H. W., Snow, T. P., Sonneborn, G., Vidal-Madjar, A., Willis, A. J., Woodgate, B. E., York, D. G., Ake, T. B., Andersson, B.-G., Andrews, J. P., Barkhouser, R. H., Bianchi, L., Blair, W. P., Brownsberger, K. R., Cha, A. N., Chayer, P., Conard, S. J., Fullerton, A. W., Gaines, G. A., Grange, R., Gummin, M. A., Hebrard, G., Kriss, G. A., Kruk, J. W., Mark, D., McCarthy, D. K., Morbey, C. L., Murowinski, R., Murphy, E. M., Oegerle, W. R., Ohl, R. G., Oliveira, C., Osterman, S. N., Sahnow, D. J., Saisse, M., Sembach, K. R., Weaver, H. A., Welsh, B. Y., Wilkinson, E., and Zheng, W.: 2000, ApJ **538**, L1

Morris, S. L. and Jannuzi, B. T.: 2006, MNRAS **367**, 1261

Morris, S. L. and van den Bergh, S.: 1994, ApJ **427**, 696

Morris, S. L., Weymann, R. J., Dressler, A., McCarthy, P. J., Smith, B. A., Terrile, R. J., Giovanelli, R., and Irwin, M.: 1993, ApJ **419**, 524

Morris, S. L., Weymann, R. J., Savage, B. D., and Gilliland, R. L.: 1991, ApJ **377**, L21

Navarro, J. F., Frenk, C. S., and White, S. D. M.: 1996, ApJ **462**, 563

Navarro, J. F., Frenk, C. S., and White, S. D. M.: 1997, ApJ **490**, 493

Newman, J. A., Cooper, M. C., Davis, M., Faber, S. M., Coil, A. L., Guhathakurta, P., Koo, D. C., Phillips, A. C., Conroy, C., Dutton, A. A., Finkbeiner, D. P., Gerke, B. F., Rosario, D. J., Weiner, B. J., Willmer, C. N. A., Yan, R., Harker, J. J., Kassin, S. A., Konidaris, N. P., Lai, K., Madgwick, D. S., Noeske, K. G., Wirth, G. D., Connolly, A. J., Kaiser, N., Kirby, E. N., Lemaux, B. C., Lin, L., Lotz, J. M., Luppino, G. A., Marinoni, C., Matthews, D. J., Metevier, A., and Schiavon, R. P.: 2013, ApJS **208**, 5

Norberg, P., Baugh, C. M., Gaztañaga, E., and Croton, D. J.: 2009, MNRAS **396**, 19

Oke, J. B.: 1990, AJ **99**, 1621

O'Meara, J. M., Prochaska, J. X., Burles, S., Prochter, G., Bernstein, R. A., and Burgess, K. M.: 2007, ApJ **656**, 666

Oppenheimer, B. D., Davé, R., Katz, N., Kollmeier, J. A., and Weinberg, D. H.: 2012, MNRAS **420**, 829

Osterbrock, D. E., Fulbright, J. P., and Bida, T. A.: 1997, PASP **109**, 614

Osterbrock, D. E., Fulbright, J. P., Martel, A. R., Keane, M. J., Trager, S. C., and Basri, G.: 1996, PASP **108**, 277

Padilla, N., Lacerna, I., Lopez, S., Barrientos, L. F., Lira, P., Andrews, H., and Tejos, N.: 2009, MNRAS **395**, 1135

Padilla, N., Lambas, D. G., and González, R.: 2010, MNRAS **409**, 936

Pan, D. C., Vogeley, M. S., Hoyle, F., Choi, Y.-Y., and Park, C.: 2012, MNRAS **421**, 926

Park, C., Choi, Y.-Y., Vogeley, M. S., Gott, III, J. R., Blanton, M. R., and SDSS Collaboration: 2007, ApJ **658**, 898

Paschos, P., Jena, T., Tytler, D., Kirkman, D., and Norman, M. L.: 2009, MNRAS **399**, 1934

Peebles, P. J. E.: 1980, *The large-scale structure of the universe*

Peebles, P. J. E.: 2001, ApJ **557**, 495

Penton, S. V., Shull, J. M., and Stocke, J. T.: 2000, ApJ **544**, 150

Penton, S. V., Stocke, J. T., and Shull, J. M.: 2002, ApJ **565**, 720

Penton, S. V., Stocke, J. T., and Shull, J. M.: 2004, ApJS **152**, 29

Penzias, A. A. and Wilson, R. W.: 1965, ApJ **142**, 419

Perlmutter, S., Aldering, G., Goldhaber, G., Knop, R. A., Nugent, P., Castro, P. G., Deustua, S., Fabbro, S., Goobar, A., Groom, D. E., Hook, I. M., Kim, A. G., Kim, M. Y., Lee, J. C., Nunes, N. J., Pain, R., Pennypacker, C. R., Quimby, R., Lidman, C., Ellis, R. S., Irwin, M., McMahon, R. G., Ruiz-Lapuente, P., Walton, N., Schaefer, B., Boyle, B. J., Filippenko, A. V., Matheson, T., Fruchter, A. S., Panagia, N., Newberg, H. J. M., Couch, W. J., and Supernova Cosmology Project: 1999, ApJ **517**, 565

Petry, C. E., Impey, C. D., Fenton, J. L., and Foltz, C. B.: 2006, AJ **132**, 2046

Pettini, M. and Cooke, R.: 2012, MNRAS **425**, 2477

Pierleoni, M., Branchini, E., and Viel, M.: 2008, MNRAS **388**, 282

Planck Collaboration, Ade, P. A. R., Aghanim, N., Armitage-Caplan, C., Arnaud, M., Ashdown, M., Atrio-Barandela, F., Aumont, J., Baccigalupi, C., Banday, A. J., and et al.: 2013, *ArXiv e-prints:1303.5076*

Press, W. H. and Schechter, P.: 1974, ApJ **187**, 425

Prochaska, J. X. and Tumlinson, J.: 2009, *Baryons: What,When and Where?*, p. 419

Prochaska, J. X., Weiner, B., Chen, H.-W., Cooksey, K. L., and Mulchaey, J. S.: 2011a, ApJS **193**, 28

Prochaska, J. X., Weiner, B., Chen, H.-W., Mulchaey, J., and Cooksey, K.: 2011b, ApJ **740**, 91

Rakic, O., Schaye, J., Steidel, C. C., and Rudie, G. C.: 2011, MNRAS **414**, 3265

Rakic, O., Schaye, J., Steidel, C. C., and Rudie, G. C.: 2012, ApJ **751**, 94

Rauch, M.: 1998, ARA&A **36**, 267

Rees, M. J. and Ostriker, J. P.: 1977, MNRAS **179**, 541

Regos, E. and Geller, M. J.: 1991, ApJ **377**, 14

Richter, P.: 2012, ApJ **750**, 165

Riess, A. G., Filippenko, A. V., Challis, P., Clocchiatti, A., Diercks, A., Garnavich, P. M., Gilliland, R. L., Hogan, C. J., Jha, S., Kirshner, R. P., Leibundgut, B., Phillips, M. M., Reiss, D., Schmidt, B. P., Schommer, R. A., Smith, R. C., Spyromilio, J., Stubbs, C., Suntzeff, N. B., and Tonry, J.: 1998, AJ **116**, 1009

Robertson, H. P.: 1935, ApJ **82**, 284

Rojas, R. R., Vogeley, M. S., Hoyle, F., and Brinkmann, J.: 2005, ApJ **624**, 571

Rollinde, E., Petitjean, P., Pichon, C., Colombi, S., Aracil, B., D'Odorico, V., and Haehnelt, M. G.: 2003, MNRAS **341**, 1279

Rubin, V. C., Ford, W. K. J., and . Thonnard, N.: 1980, ApJ **238**, 471

Rudie, G. C., Steidel, C. C., Trainor, R. F., Rakic, O., Bogosavljević, M., Pettini, M., Reddy, N., Shapley, A. E., Erb, D. K., and Law, D. R.: 2012, ApJ **750**, 67

Rutledge, R. E.: 1998, ApJ **502**, 130

Ryan-Weber, E. V.: 2006, MNRAS **367**, 1251

Ryden, B.: 2003, *Introduction to cosmology*

Savage, B. D., Narayanan, A., Wakker, B. P., Stocke, J. T., Keeney, B. A., Shull, J. M., Sembach, K. R., Yao, Y., and Green, J. C.: 2010, ApJ **719**, 1526

Savage, B. D. and Sembach, K. R.: 1991, ApJ **379**, 245

Scannapieco, C., Wadepuhl, M., Parry, O. H., Navarro, J. F., Jenkins, A., Springel, V., Teyssier, R., Carlson, E., Couchman, H. M. P., Crain, R. A., Dalla Vecchia, C., Frenk, C. S., Kobayashi, C., Monaco, P., Murante, G., Okamoto, T., Quinn, T., Schaye, J., Stinson, G. S., Theuns, T., Wadsley, J., White, S. D. M., and Woods, R.: 2012, MNRAS **423**, 1726

Schaye, J.: 2001, ApJ **559**, 507

Schaye, J. and Dalla Vecchia, C.: 2008, MNRAS **383**, 1210

Schaye, J., Dalla Vecchia, C., Booth, C. M., Wiersma, R. P. C., Theuns, T., Haas, M. R., Bertone, S., Duffy, A. R., McCarthy, I. G., and van de Voort, F.: 2010, MNRAS **402**, 1536

Schaye, J., Theuns, T., Leonard, A., and Efstathiou, G.: 1999, MNRAS **310**, 57

Scodeggio, M., Franzetti, P., Garilli, B., Zanichelli, A., Paltani, S., Maccagni, D., Bottini, D., Le Brun, V., Contini, T., Scaramella, R., Adami, C., Bardelli, S., Zucca, E., Tresse, L., Ilbert, O., Foucaud, S., Iovino, A., Merighi, R., Zamorani, G., Gavignaud, I., Rizzo, D., McCracken, H. J., Le Fèvre, O., Picat, J. P., Vettolani, G., Arnaboldi, M., Arnouts, S., Bolzonella, M., Cappi, A., Charlot, S., Ciliegi, P., Guzzo, L., Marano, B., Marinoni, C., Mathez, G., Mazure, A., Meneux, B., Pellò, R., Pollo, A., Pozzetti, L., and Radovich, M.: 2005, PASP **117**, 1284

Sérsic, J. L.: 1963, *Boletin de la Asociacion Argentina de Astronomia La Plata Argentina* **6**, 41

Sharp, N. A.: 1979, A&A **74**, 308

Sheth, R. K. and van de Weygaert, R.: 2004, MNRAS **350**, 517

Shone, A. M., Morris, S. L., Crighton, N., and Wilman, R. J.: 2010, MNRAS **402**, 2520

Shull, J. M., Giroux, M. L., Penton, S. V., Tumlinson, J., Stocke, J. T., Jenkins, E. B., Moos, H. W., Oegerle, W. R., Savage, B. D., Sembach, K. R., York, D. G., Green, J. C., and Woodgate, B. E.: 2000, ApJ **538**, L13

Shull, J. M., Smith, B. D., and Danforth, C. W.: 2012, ApJ **759**, 23

Smith, R. E., Peacock, J. A., Jenkins, A., White, S. D. M., Frenk, C. S., Pearce, F. R., Thomas, P. A., Efstathiou, G., and Couchman, H. M. P.: 2003, MNRAS **341**, 1311

Spinrad, H., Filippenko, A. V., Yee, H. K., Ellingson, E., Blades, J. C., Bahcall, J. N., Jannuzi, B. T., Bechtold, J., and Dobrzycki, A.: 1993, AJ **106**, 1

Springel, V., White, S. D. M., Jenkins, A., Frenk, C. S., Yoshida, N., Gao, L., Navarro, J., Thacker, R., Croton, D., Helly, J., Peacock, J. A., Cole, S., Thomas, P., Couchman, H., Evrard, A., Colberg, J., and Pearce, F.: 2005, Nature **435**, 629

Stinson, G. S., Brook, C., Prochaska, J. X., Hennawi, J., Shen, S., Wadsley, J., Pontzen, A., Couchman, H. M. P., Quinn, T., Macciò, A. V., and Gibson, B. K.: 2012, MNRAS **425**, 1270

Stocke, J. T., Danforth, C. W., Shull, J. M., Penton, S. V., and Giroux, M. L.: 2007, ApJ **671**, 146

Stocke, J. T., Keeney, B. A., Danforth, C. W., Shull, J. M., Froning, C. S., Green, J. C., Penton, S. V., and Savage, B. D.: 2013, ApJ **763**, 148

Stocke, J. T., Penton, S. V., Danforth, C. W., Shull, J. M., Tumlinson, J., and McLin, K. M.: 2006, ApJ **641**, 217

Stocke, J. T., Shull, J. M., Penton, S., Donahue, M., and Carilli, C.: 1995, ApJ **451**, 24

Tejos, N., Lopez, S., Prochaska, J. X., Bloom, J. S., Chen, H.-W., Dessauges-Zavadsky, M., and Maureira, M. J.: 2009, ApJ **706**, 1309

Tejos, N., Lopez, S., Prochaska, J. X., Chen, H.-W., and Dessauges-Zavadsky, M.: 2007, ApJ **671**, 622

Tejos, N., Morris, S. L., Crighton, N. H. M., Theuns, T., Altay, G., and Finn, C. W.: 2012, MNRAS **425**, 245

Tejos, N., Morris, S. L., Finn, C. W., Crighton, N. H. M., Bechtold, J., Jannuzi, B. T., Schaye, J., Theuns, T., Altay, G., Le Fevre, O., Ryan-Weber, E., and Dave, R.: 2013, *ArXiv e-prints:1309.5950*

Tepper-García, T., Richter, P., Schaye, J., Booth, C. M., Dalla Vecchia, C., and Theuns, T.: 2012, MNRAS **425**, 1640

Theuns, T., Leonard, A., and Efstathiou, G.: 1998a, MNRAS **297**, L49

Theuns, T., Leonard, A., Efstathiou, G., Pearce, F. R., and Thomas, P. A.: 1998b, MNRAS **301**, 478

Theuns, T., Leonard, A., Schaye, J., and Efstathiou, G.: 1999, MNRAS **303**, L58

Theuns, T., Viel, M., Kay, S., Schaye, J., Carswell, R. F., and Tzanavaris, P.: 2002, ApJ **578**, L5

Thom, C., Tumlinson, J., Werk, J. K., Prochaska, J. X., Oppenheimer, B. D., Peeples, M. S., Tripp, T. M., Katz, N. S., O'Meara, J. M., Brady Ford, A., Davé, R., Sembach, K. R., and Weinberg, D. H.: 2012, ApJ **758**, L41

Tikhonov, A. V. and Klypin, A.: 2009, MNRAS **395**, 1915

Tilton, E. M., Danforth, C. W., Shull, J. M., and Ross, T. L.: 2012, ApJ **759**, 112

Tripp, T. M., Lu, L., and Savage, B. D.: 1998, ApJ **508**, 200

Tumlinson, J., Thom, C., Werk, J. K., Prochaska, J. X., Tripp, T. M., Weinberg, D. H., Peeples, M. S., O'Meara, J. M., Oppenheimer, B. D., Meiring, J. D., Katz, N. S., Davé, R., Ford, A. B., and Sembach, K. R.: 2011, *Science* **334**, 948

Tummuangpak, P., Shanks, T., Bielby, R., Crighton, N. H. M., Francke, H., Infante, L., and Theuns, T.: 2013, *ArXiv e-prints:1304.4598*

van de Hulst, H. C. and Reesinck, J. J. M.: 1947, ApJ **106**, 121

van de Voort, F., Schaye, J., Booth, C. M., and Dalla Vecchia, C.: 2011, MNRAS **415**, 2782

van de Weygaert, R. and van Kampen, E.: 1993, MNRAS **263**, 481

Wakker, B. P. and Savage, B. D.: 2009, ApJS **182**, 378

Walker, A. G.: 1935, MNRAS **95**, 263

Weinberg, S.: 1989, *Reviews of Modern Physics* **61**, 1

Werk, J. K., Prochaska, J. X., Thom, C., Tumlinson, J., Tripp, T. M., O'Meara, J. M., and Peeples, M. S.: 2013, ApJS **204**, 17

White, S. D. M. and Frenk, C. S.: 1991, ApJ **379**, 52

White, S. D. M., Frenk, C. S., and Davis, M.: 1983, ApJ **274**, L1

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

Wiersma, R. P. C., Schaye, J., and Smith, B. D.: 2009a, MNRAS **393**, 99

Wiersma, R. P. C., Schaye, J., Theuns, T., Dalla Vecchia, C., and Tornatore, L.: 2009b, MNRAS **399**, 574

Williger, G. M., Carswell, R. F., Weymann, R. J., Jenkins, E. B., Sembach, K. R., Tripp, T. M., Davé, R., Haberzettl, L., and Heap, S. R.: 2010, MNRAS **405**, 1736

Williger, G. M., Heap, S. R., Weymann, R. J., Davé, R., Ellingson, E., Carswell, R. F., Tripp, T. M., and Jenkins, E. B.: 2006, ApJ **636**, 631

Williger, G. M., Smette, A., Hazard, C., Baldwin, J. A., and McMahon, R. G.: 2000, ApJ **532**, 77

Willman, B. and Strader, J.: 2012, AJ **144**, 76

Wilman, R. J., Morris, S. L., Jannuzi, B. T., Davé, R., and Shone, A. M.: 2007, MNRAS **375**, 735

Wolfe, A. M., Gawiser, E., and Prochaska, J. X.: 2005, ARA&A **43**, 861

Woodgate, B. E., Kimble, R. A., Bowers, C. W., Kraemer, S., Kaiser, M. E., Danks, A. C., Grady, J. F., Loiacono, J. J., Brumfield, M., Feinberg, L., Gull, T. R., Heap, S. R., Maran, S. P., Lindler, D., Hood, D., Meyer, W., Vanhouten, C., Argabright, V., Franka, S., Bybee, R., Dorn, D., Bottema, M., Woodruff, R., Michika, D., Sullivan, J., Hetlinger, J., Ludtke, C., Stocker, R., Delamere, A., Rose, D., Becker, I., Garner, H., Timothy, J. G., Blouke, M., Joseph, C. L., Hartig, G., Green, R. F., Jenkins, E. B., Linsky, J. L., Hutchings, J. B., Moos, H. W., Boggess, A., Roesler, F., and Weistrop, D.: 1998, PASP **110**, 1183

Yoon, J. H., Putman, M. E., Thom, C., Chen, H.-W., and Bryan, G. L.: 2012, ApJ **754**, 84

Zheng, Z. and Miralda-Escudé, J.: 2002, ApJ **568**, L71

Zwicky, F.: 1933, *Helvetica Physica Acta* **6**, 110

Zwicky, F.: 1937, ApJ **86**, 217