On the connection between the metal-enriched intergalactic medium and galaxies: an O VI-galaxy cross-correlation study at z < 1

Finn, C.W.; Morris, S.L.; Tejos, N.; Crighton, N.H.M.; Perry, R; Fumagalli, M.; ... ; Bartle, S.

Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **460,** 590–616 (2016) doi:10.1093/mnras/stw918
Advance Access publication 2016 April 21

On the connection between the metal-enriched intergalactic medium and galaxies: an O VI–galaxy cross-correlation study at $z < 1$

Charles W. Finn,[1,2★] Simon L. Morris,[1★] Nicolas Tejos,[3] Neil H. M. Crighton,[4] Robert Perry,[2] Michele Fumagalli,[1,2] Rich Bielby,[1] Tom Theuns,[2] Joop Schaye,[5] Tom Shanks,[1] Jochen Liske,[6] Madusha L. P. Gunawardhana[1,2] and Stephanie Bartle[1]

[1]*Department of Physics, Durham University, South Road, Durham DH1 3LE, UK*
[2]*Department of Physics, Institute for Computational Cosmology, Durham University, South Road, Durham DH1 3LE, UK*
[3]*University of California Observatories-Lick Observatory, University of California, Santa Cruz, CA 95064, USA*
[4]*Centre for Astrophysics and Supercomputing, Swinburne University of Technology, PO Box 218, VIC 3122, Australia*
[5]*Leiden Observatory, Leiden University, PO Box 9513, NL-2300 RA Leiden, the Netherlands*
[6]*Hamburger Sternwarte, Universität Hamburg, Gojenbergsweg 112, D-21029 Hamburg, Germany*

Accepted 2016 April 18. Received 2016 April 17; in original form 2015 November 5

ABSTRACT

We present new results on the auto- and cross-correlation functions of galaxies and O VI absorbers in a $\sim$18 Gpc^3 comoving volume at $z < 1$. We use a sample of 51 296 galaxies and 140 O VI absorbers in the column density range $13 \lesssim \log N \lesssim 15$ to measure two-point correlation functions in the two dimensions transverse and orthogonal to the line of sight $\xi(r_\perp, r_\parallel)$. We furthermore infer the corresponding 'real-space' correlation functions, $\xi(r)$, by projecting $\xi(r_\perp, r_\parallel)$ along $r_\parallel$, and assuming a power-law form, $\xi(r) = (r/r_0)^{-\gamma}$. Comparing the results from the absorber-galaxy cross-correlation function, $\xi_{\rm ag}$, the galaxy autocorrelation function, $\xi_{\rm gg}$, and the absorber autocorrelation function, $\xi_{\rm aa}$, we constrain the statistical connection between galaxies and the metal-enriched intergalactic medium as a function of star formation activity. We also compare these results to predictions from the EAGLE cosmological hydrodynamical simulation and find a reasonable agreement. We find that: (i) O VI absorbers show very little velocity dispersion with respect to galaxies on $\sim$ Mpc scales, likely $\lesssim$100 km s^{-1}; (ii) O VI absorbers are less clustered, and potentially more extended around galaxies than galaxies are around themselves; (iii) on $\gtrsim$100 kpc scales, the likelihood of finding O VI absorbers around star-forming galaxies is similar to the likelihood of finding O VI absorbers around non-star-forming galaxies; and (iv) O VI absorbers are either not ubiquitous to galaxies in our sample, or their distribution around them is patchy on scales $\gtrsim$100 kpc (or both), at least for the column densities at which most are currently detected.

Key words: galaxies: formation – intergalactic medium – quasars: absorption lines – large-scale structure of Universe.

1 INTRODUCTION

The connection between the intergalactic medium (IGM) and galaxies is fundamental to our understanding of the formation and evolution of galaxies and the large-scale structure of the Universe. This is because there exists a continuous interplay between galaxies and the plasma around them, which fuels the formation of stars and the hierarchical assembly of cosmic structures. In the theoretical Λ cold dark matter (ΛCDM) paradigm, the two main physical processes that drive this assembly are: (i) the accretion of intergalactic matter in 'hot' and 'cold' modes (e.g. Rees & Ostriker 1977; White & Rees 1978; White & Frenk 1991; Kereš et al. 2005; Dekel & Birnboim 2006; Dekel, Sari & Ceverino 2009; van de Voort et al. 2011); and (ii) winds emanating from galaxies generated mostly by supernova (SN) explosions and active galactic nuclei (AGN; e.g. Baugh et al. 2005; Veilleux, Cecil & Bland-Hawthorn 2005; Bower et al. 2006; Lagos, Cora & Padilla 2008; Creasey, Theuns & Bower 2013). These winds are also thought to be responsible for enriching the IGM with metals (e.g. Schaye 2001; Simcoe et al. 2012). Observational studies are producing results largely consistent with this picture, but better constraints are nevertheless required if we are to understand these processes in detail.

To achieve a thorough understanding of the formation of galaxies and cosmic structure, we must correctly describe the behaviour and

★E-mail: c.w.finn2301@gmail.com (CWF); simon.morris@durham.ac.uk (SLM)

evolution of the baryonic matter in the Universe. For this, we require hydrodynamical simulations following the evolution of baryons and dark matter together within a cosmological volume (e.g. Crain et al. 2009; Davé et al. 2010; Vogelsberger et al. 2014; Schaye et al. 2015). Large volumes are fundamentally important, since the simulations must be able to reproduce the statistics of the present-day galaxy population. Unfortunately, due to the computational cost, there is a fundamental reliance on uncertain 'subgrid' prescriptions to capture the relevant physics on scales smaller than the resolution limit (e.g. Schaye et al. 2010; Scannapieco et al. 2012). To glean trust-worthy physical insight from these simulations, we must therefore place tight constraints on the subgrid physics. Observations of the gaseous environments around galaxies play a major role in this goal, as the simulations are not typically calibrated to match these observations. They therefore provide an important test that is independent of any 'fine tuning'. In particular, observations of the metal-enriched components of the IGM are expected to provide especially robust constraints, since their distribution and dynamics are found to be sensitive to details of the typically implemented subgrid feedback prescriptions (e.g. Wiersma, Schaye & Theuns 2011; Ford et al. 2013; Hummels et al. 2013; Suresh et al. 2015).

Unfortunately, despite being the main reservoir of baryons at all epochs, the extremely low density of the IGM makes it difficult to observe. The best method at present is through the analysis of absorption lines in quasar (QSO) spectra. These lines appear due to the scattering of ultraviolet (UV) photons by intervening gas along the line of sight (LOS). The resulting characterization of the IGM is therefore often limited to being one-dimensional. Nevertheless, by combining information from multiple LOS, we are able to construct a statistical picture of the distribution and dynamics of gas in the Universe (e.g. Chen & Mulchaey 2009; Tejos et al. 2014).

Observations of the IGM at low redshifts ($z < 1$) have improved dramatically over the last few years with the advent of the Cosmic Origins Spectrograph (COS) on the *Hubble Space telescope* (*HST*; Green et al. 2012). With a sensitivity more than 10 times that of its predecessor, COS has provided observations of hundreds of QSOs in the far ultraviolet (FUV). Observations at these wavelengths are fundamentally important, as they allow for a mapping of the H I and metal content of the IGM to $z = 0$. The capabilities of COS have been exploited extensively to probe both cool ($T \sim 10^4$ K) gas, traced mostly by the Lyα forest, and warmer ($\sim 10^5$–10^6 K) gas, traced by broad Lyα, O VI and Ne VIII absorption (e.g. Savage, Lehner & Narayanan 2011; Tumlinson et al. 2011; Lehner et al. 2013; Meiring et al. 2013; Liang & Chen 2014; Shull, Danforth & Tilton 2014; Stocke et al. 2014; Hussain et al. 2015; Danforth et al. 2016). These ions probably trace up to ∼60 per cent of all the baryons, with only ∼10 per cent in the luminous constituents of the Universe (stars and galaxies), and the rest in an even hotter plasma at $T > 10^6$ K (Fukugita & Peebles 2004).

To date, much of the work on the low-redshift IGM in relation to galaxies has taken a 'galaxy-centric' approach, with a primary focus on the properties of the so-called circumgalactic medium (CGM). A number of successful programmes have been designed with this goal in mind, notably the 'COS-Haloes' survey (Tumlinson et al. 2013), and various programmes by the COS guaranteed time observations (GTO) team (e.g. Keeney et al. 2013; Stocke et al. 2013). These studies implicitly assume a one-to-one correspondence between absorption systems and the closest observed galaxy, which is problematic due to the incomplete sampling of galaxies in any galaxy survey. Despite this shortcoming, it is clear from these studies that there is a nearly ubiquitous presence of cool ($T \approx 10^4$–10^5 K) metal-enriched gas surrounding galaxies to impact parameters of ∼150 kpc (see e.g. Prochaska et al. 2011b; Werk et al. 2013). Ionization models suggest this cool CGM, combined with an additional hotter component traced by collisionally ionized O VI, can account for at least half of the baryons expected from big bang nucleosynthesis that were originally unaccounted for (Fukugita, Hogan & Peebles 1998; McGaugh et al. 2010; Shull, Smith & Danforth 2012; Werk et al. 2014), although see Oppenheimer & Schaye (2013b) and Vasiliev, Ryabova & Shchekinov (2015) for important caveats. Nevertheless, 30–40 per cent of the baryons may still be unaccounted for, residing in the so-called warm-hot intergalactic medium (WHIM) predicted by cosmological hydrodynamical simulations (e.g. Cen & Ostriker 1999; Davé et al. 2001). An unambiguous detection of the WHIM is needed if we are to validate the predictions of these simulations (see Tejos et al. 2016, for a recent attempt at addressing this problem).

In this paper, we address the connection between the metal-enriched IGM and galaxies at $z < 1$ via an analysis of the two-point cross- and autocorrelation functions of galaxies and O VI absorbers. The advantage of this approach is two-fold: (i) we do not rely on associating a particular intergalactic absorber with a particular galaxy (or set of galaxies), which in many instances is ambiguous; and (ii) we are robust to galaxy/absorber completeness variations, since we are measuring a clustering excess as a function of scale relative to a random expectation that takes into account the relevant selection functions (Tejos et al. 2014). Using these measurements, we are therefore able to investigate the distribution and dynamics of the metal-enriched IGM around galaxies on both the CGM scale ($\lesssim$300 kpc), and to much larger scales ($\gg$1 Mpc). We use the O VI $\lambda\lambda$1031, 1037 doublet largely because the component transitions have high oscillator strengths, and possess rest-frame wavelengths that make them accessible in the redshift range $0.1 \lesssim z \lesssim 0.7$ with current FUV instrumentation. O VI absorbers are thus a convenient tracer of the metal-enriched gas in the IGM. In addition, they are thought to trace both cool, photoionized plasmas in the temperature range $10^4 \lesssim T \lesssim 10^5$ K, and hotter, collisionally ionized gas at temperatures $10^5 \lesssim T \lesssim 10^7$ K (e.g. Tripp et al. 2001; Danforth & Shull 2005, 2008; Thom & Chen 2008; Tepper-García et al. 2011; Oppenheimer et al. 2012; Savage et al. 2014; Stocke et al. 2014), the latter of which is commonly referred to as the WHIM (e.g. Cen & Ostriker 1999; Davé et al. 2001; Fukugita & Peebles 2004). They may also form in more complicated scenarios, e.g. in conductive or turbulent interfaces between gaseous components at multiple temperatures (e.g. Borkowski, Balbus & Fristrom 1990; Kwak & Shelton 2010). Furthermore, the ionization fractions of O VI absorbers may be high in environments close to star-forming or post-starburst galaxies, or any galaxy where there has been recent or ongoing active galactic nuclei (AGN) activity due to non-equilibrium effects and long recombination time-scales (e.g. Oppenheimer & Schaye 2013a,b; Vasiliev et al. 2015). This makes O VI absorbers effective tracers of metal-enriched gas in environments like these, even for low metallicities. We bear in mind the many, potentially complex formation scenarios for O VI absorbers in the interpretation of our results.

This paper is structured as follows. In Section 2, we describe the observational data sets used in this work. In Sections 3 and 4, we describe the data analysis relating to IGM absorption systems and galaxies, respectively. In Section 5, we describe the creation of a set of comparison data drawn from the Evolution and Assembly of GaLaxies and their Environments (EAGLE) cosmological hydrodynamical simulation. In Section 6, we describe the mathematical formalisms used to compute the auto- and cross-correlation functions of galaxies and absorbers, and describe the creation of the random samples that are crucial to this analysis. In Section 7, we

present the results of our correlation function analysis. In Section 8, we present a discussion of these results and a comparison to the literature. In Section 9, we summarize our findings and outline the main conclusions of this work.

All distances are in comoving coordinates unless otherwise stated. We assume a ΛCDM cosmology, with parameters set to the best-fitting values determined from the 2013 analysis of data from the *Planck* satellite (Planck Collaboration XVI 2014).

2 OBSERVATIONS

The observational data consists of 50 independent fields with small angular coverage in each of which we have at least one QSO spectrum obtained from Hubble Space Telescope (*HST*)/COS and a large number of spectra, or spectroscopic measurements of galaxies at $z \lesssim 1$. There are 60 QSOs in the sample. Out of 50 fields, 6 are inherited from our previous work on the H I-galaxy cross-correlation (Tejos et al. 2014). We have greatly expanded on this original sample by incorporating a large number of publicly available data sets. We describe in detail the sample of galaxies and QSOs and summarize the data reduction procedures in the following subsections.

2.1 QSOs

We have used *HST*/COS and Faint Object Spectrograph (FOS) spectroscopy of 60 QSOs to characterize the diffuse IGM through analysis of intervening H I and metal absorption systems. Of these, seven have had COS spectroscopy obtained and presented by our collaboration in previous work (Crighton et al. 2013; Finn et al. 2014; Tejos et al. 2014). Details of the QSO observations are summarized in Table 1.

2.1.1 The sample

The QSOs used in this work were selected to lie in fields well surveyed for their galaxy content, and having spectroscopy with good signal-to-noise ratio (signal-to-noise ratio (SNR), $\gtrsim$10). Their spectra have been obtained by a number of collaborations, including our own, for a variety of specific programmes. In Table 1, we list the *HST* proposal ID(s) and principal investigator(s) associated with the data obtained for each QSO, and we refer the reader to the proposal abstracts for details on the associated science cases.

To obtain high SNR observations in a reasonable amount of observing time with *HST*, most of the QSOs in this sample were selected on the basis of having FUV fluxes $\gtrsim$100 μJ. This biases the QSO sample to be of high luminosity, which potentially has implications for their local environments. However, we note that the regions of the Universe along the LOS to most of these QSOs are effectively random, and we proceed with this assumption throughout the forthcoming analysis.

2.1.2 Data reduction

All of the COS data were reduced with the CALCOS pipeline. In particular, the COS QSO spectra obtained by our collaboration, namely, LBQS J0107−0235A, LBQS J0107−0235B, LBQS J0107−0232, FBQS J0209−0438, HE 1003+0149, SDSS J135726.27+043541.4 and FBQS J2218+0052, and additionally LBQS 1019+0147 and LBQS 1435−067, were reduced using v2.18.5 of the pipeline in combination with PYTHON routines developed by the authors,[1] which are based loosely on IDL routines developed by the COS GTO team.[2] For full details, see Tejos et al. (2014) and Finn et al. (2014). All of the other COS spectra were reduced as described in Danforth et al. (2014), using CALCOS versions contemporary with their observation epoch.

FOS data were reduced using the CALFOS pipeline. We refer the reader to Tejos et al. (2014) for full details.

2.2 Galaxies

The galaxy data is obtained from a number of different instruments and surveys. We include data collected by our own collaboration from the Deep Imaging Multi-Object Spectrograph (DEIMOS), Gemini Multi-Object Spectrograph (GMOS), Canada–France–Hawaii Telescope (CFHT) multi-object spectrograph and Very Large Telescope (VLT) Visible Multi-Object Spectrograph (VIMOS; Morris & Jannuzi 2006; Tejos et al. 2014, hereafter T14 and T14-Q0107). We make use of the Sloan Digital Sky Survey (SDSS; Abazajian et al. 2009), 2dF Galaxy Redshift Survey (2dFGRS; Colless et al. 2001), Galaxy and Mass Assembly (GAMA) survey (Driver et al. 2011), VLT VIMOS Deep Survey (VVDS; Le Fèvre et al. 2005) and VIMOS Public Extragalactic Redshift Survey (VIPERS; Guzzo et al. 2014). We also include data from the Las Campanas Observatory (LCO)/Wide Field Reimaging CCD Camera (WFCCD) galaxy survey of 20 fields surrounding UV-bright QSOs (Prochaska et al. 2011a, hereafter P11), galaxy data around PKS 0405−123 presented in Johnson, Chen & Mulchaey (2013, hereafter J13); and galaxy data around HE 0226−4110 and PG 1216+069 presented in Chen & Mulchaey (2009, hereafter C09). The latter two surveys made use of the Inamori-Magellan Areal Camera & Spectrograph (IMACS) and Low Dispersion Survey Spectrograph 3 (LDSS3) at LCO.

2.2.1 The sample

The surveys that make up our galaxy sample cover regions close to all of the QSO sight-lines used to characterize the IGM. Some were conducted for the primary purpose of mapping galaxies close to a particular QSO sight-line, while others serendipitously cover regions where there are bright QSOs with *HST* spectroscopy. Those that fall in the latter category are the large Sloan Digital Sky Survey (SDSS), 2dF Galaxy Redshift Survey (2dFGRS), GAMA, Very Large Telescope (VLT) VLT Visible Multi-Object Spectrograph (VIMOS) Deep Survey (VVDS) and VIMOS Public Extragalactic Redshift Survey (VIPERS) surveys. For SDSS, we adopt just those galaxies in the main sample, i.e. SDSS-I/II (see Abazajian et al. 2009, for details). We restrict our combined galaxy sample to 4 × 4 square degree fields centred on each QSO.[3] This means that we can sample galaxy-absorber pairs to transverse separations of ∼15 comoving Mpc at the median redshift of our sample ($z_{\rm median} = 0.19$), ∼10 comoving Mpc at $z \sim 0.07$ and ∼1 comoving Mpc at $z \sim 0.005$. We discard all objects with $z < 0.005$, regardless of their classification, on the basis that they may be stars. Some fields are made larger by virtue of there being more than one QSO that inhabits a particular 4 × 4 square degree region.

We summarize our combined galaxy sample in Table 2. As an indication of survey depth, we list the median redshift for each survey, and the 95th percentile of the redshift distribution, which we

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

[2] http://casa.colorado.edu/danforth/science/cos/costools.html

[3] Note that a number of surveys cover areas of sky that are smaller than this.

Table 1. The QSO sample.

QSO name	G130M		G160M		NUV[c]		Programme ID(s)	P.I.(s)
	t_{exp} (ks)[a]	S/N[b]	t_{exp} (ks)[a]	S/N[b]	t_{exp} (ks)[a]	S/N[b]		
PG 0003+158	10.4	22	10.9	19	–	–	12038	Green
PG 0026+129	1.9	18	–	–	–	–	12569	Veilleux
HE 0056−3622	7.8	25	5.7	15	–	–	12604	Fox
LBQS 0107−0235A	28.1	9	44.3	8	35.6	30	11585, 6592, 6100, 5320	Crighton, Foltz
LBQS 0107−0235B	21.2	9	21.2	7	13.0	30	11585, 6592, 6100, 5320	Crighton, Foltz
LBQS 0107−0232	–	–	83.5	7	32.8	18	11585, 6592, 6100	Crighton, Foltz
B0117−2837	5.2	24	8.5	19	–	–	12204	Thom
Ton S210	5.0	41	5.5	26	–	–	12204	Thom
PG 0157+001	1.8	16	–	–	–	–	12569	Veilleux
FBQS J0209−0438	14.0	12	28.1	10	14.4	12	12264	Morris
HE 0226−4110	8.8	34	7.8	24	–	–	11541	Green
PKS 0405−123	24.2	59	11.1	30	–	–	11508, 11541	Noll, Green
RBS 542	23.2	61	15.2	35	–	–	11686	Arav
PKS 0558−504	1.8	19	0.7	10	–	–	11692	Howk
SDSS J080908.13+461925.6	5.7	15	5.0	13	–	–	12248	Tumlinson
PG 0832+251	8.8	14	6.8	12	–	–	12025	Green
PG 0844+349	1.9	18	–	–	–	–	12569	Veilleux
Mrk 106	9.3	28	7.6	18	–	–	12029	Green
RXS J09565−0452	7.7	16	–	–	–	–	12275	Wakker
PG 0953+414	6.2	38	5.6	26	–	–	12038	Green
PG 1001+291	7.1	21	6.8	17	–	–	12038	Green
HE 1003+0149	14.0	9	22.3	9	–	–	12264	Morris
FBQS J1010+3003	12.8	17	10.8	10	–	–	12025	Green
Ton 1187	2.0	16	–	–	–	–	12275	Wakker
PG 1011−040	6.7	29	4.7	18	–	–	11524	Green
LBQS 1019+0147	2.2	6	2.9	5	–	–	11598	Tumlinson
1ES 1028+511	20.0	20	14.6	13	–	–	12025	Green
1SAX J1032.3+5051	13.5	12	11.3	6	–	–	12025	Green
PG 1048+342	7.8	23	11.0	16	–	–	12024	Green
PG 1049−005	3.6	14	2.8	12	–	–	12248	Tumlinson
HS 1102+3441	11.3	17	11.3	13	–	–	11541	Green
SBS 1108+560	10.6	4	8.9	14	–	–	12025	Green
PG 1115+407	5.1	23	5.7	15	–	–	11519	Green
PG 1116+215	6.1	39	5.5	28	–	–	12038	Green
PG 1121+422	5.0	21	5.8	13	–	–	12604	Fox
SBS 1122+594	9.9	14	10.5	12	–	–	11520	Green
Ton 580	4.9	21	5.6	16	–	–	11519	Green
3C 263	15.4	34	18.0	23	–	–	11541	Green
PG 1216+069	5.1	24	5.6	16	–	–	12025	Green
3C 273	4.0	73	–	–	–	–	12038	Green
HE 1128+0131	13.2	44	11.0	36	–	–	11686	Arav
PG 1229+204	1.9	17	–	–	–	–	12569	Veilleux
PG 1259+593	12.0	32	11.2	24	–	–	11541	Green
PKS 1302−102	7.4	26	6.9	20	–	–	12038	Green
PG 1307+085	1.8	18	–	–	–	–	12569	Veilleux
SDSS J135726.27+043541.4	14.0	9	28.1	7	14.4	11	12264	Morris
PG 1424+240	6.4	21	7.9	21	–	–	12612	Stocke
PG 1435−067	1.9	14	–	–	–	–	12569	Veilleux
LBQS 1435−0134	22.3	23	34.2	16	–	–	11741	Tripp
Mrk 478	1.9	18	–	–	–	–	12569	Veilleux
Ton 236	8.3	18	9.4	15	–	–	12038	Green
1ES 1553+113	10.8	33	12.3	26	–	–	11520, 12025	Green
Mrk 877	1.8	18	–	–	–	–	12569	Veilleux
PKS 2005−489	2.5	24	1.9	15	–	–	11520	Green
Mrk 1513	6.9	32	4.8	20	–	–	11524	Green
PHL 1811	3.9	36	3.1	24	–	–	12038	Green
PKS 2155−304	4.6	45	–	–	–	–	12038	Green
FBQS J2218+0052	–	–	–	–	20.2	10	12264	Morris
MR 2251−178	5.6	38	5.4	30	–	–	12029	Green
4C 01.61	1.8	20	–	–	–	–	12569	Veilleux

[a]Total exposure time in ks.
[b]Median SNR per resolution element.
[c]FOS gratings G270H and/or G190H for LBQS 0107−0235A, LBQS 0107−0235B and LBQS 0107−0232, COS G230L grating otherwise.

Table 2. The galaxy sample.

Survey	$N_{\rm galaxies}$[a]	$z_{\rm median}$[b]	z_{95}[c]	$m_{\rm limit}$[d]	Reference
SDSS	41 342	0.10	0.19	$r < 17.77$	Abazajian et al. (2009)
2dFGRS	10 643	0.11	0.22	$b_{\rm J} < 19.45$	Colless et al. (2001)
GAMA	8636	0.22	0.40	$r < 19.8$	Driver et al. (2011)
VVDS	18 181	0.58	1.07	$I < 22.5$	Le Fèvre et al. (2005)
VIPERS	24 183	0.70	1.06	$I < 22.5$[e]	Guzzo et al. (2014)
T14	1049	0.43	0.93	$R < 23.5$[f]	Tejos et al. (2014)
T14-Q0107	962	0.55	1.07	Various[g]	Tejos et al. (2014)
P11	900	0.16	0.36	$R < 20$[h]	Prochaska et al. (2011a)
C09	810	0.36	0.64	$R < 22$	Chen & Mulchaey (2009)
J13	443	0.41	0.81	$R < 23$	Johnson et al. (2013)
ALL	**107 149**	**0.19**	**1.00**	–	–

[a]Number of galaxies with spectroscopically confirmed redshifts (not labelled 'c' – see Section 4.1).
[b]Median redshift for the survey.
[c]The 95th percentile of the redshift distribution.
[d]Magnitude limit for the survey.
[e]Colour cuts also applied.
[f]Priority given to objects with $R < 22.5$.
[g]VIMOS: $R < 23$, priority given to objects with $R < 22$. DEIMOS: $R < 24.5$, priority given to brighter objects, colour cuts also applied. GMOS: top priority given to objects with $R < 22$, second priority given to objects with $22 < R < 23$, last priority given to objects with $23 < R < 24$. CFHT: $R < 23.5$ (indicative only).
[h]$R < 19.5$ for some fields.

denote z_{95}. This is more informative than the maximum of the redshift distribution, as many surveys show long tails to high redshift due to the presence of luminous AGN. We also list the magnitude limit for each survey, which in many cases is only indicative (see the table footnotes for more details). There are 107 149 galaxies in our combined sample, which has a median redshift of 0.19. In Table 3, we summarize the QSO sight-line fields. We list the number of QSOs in each field and give an indication of the area and comoving volume covered by each. For the latter, we define the edge of the volume by the minimum of ($z_{\rm QSO}$, z_{95}), where $z_{\rm QSO}$ denotes the maximum QSO redshift for the field.

2.2.2 *Data reduction*

Galaxy data obtained with VIMOS pre-2011 were reduced using VIPGI pipeline (Scodeggio et al. 2005), and after this time using the ESOREX pipeline with the exception of that taken for VIPERS, which has its own dedicated pipeline (Guzzo et al. 2014). For full details, see Le Fèvre et al. (2005) and Tejos et al. (2014). Data from DEIMOS was reduced using the DEEP2 DEIMOS Data Pipeline (Newman et al. 2013).[4] GMOS data was reduced using the Gemini Image Reduction and Analysis Facility (IRAF; see Tejos et al. 2014, for details). The reduction of the CFHT data is described in Morris & Jannuzi (2006). SDSS, GAMA and 2dFGRS data reduction procedures are described in Stoughton et al. (2002), Hopkins et al. (2013) and Colless et al. (2001), respectively. Details on the reduction of the LCO/WFCCD data can be found in Prochaska et al. (2011a). The reduction of the LCO/IMACS and LDSS3 data is described fully in Chen & Mulchaey (2009) and Johnson et al. (2013).

3 ANALYSIS OF THE IGM DATA

The following sections briefly describe the processes involved in creating absorption line lists from the reduced COS and FOS data obtained and/or analysed by our collaboration. For more description, see Tejos et al. (2014) and Finn et al. (2014). For the majority of the QSOs, we obtained absorption line lists directly as a result of the analysis in Danforth et al. (2014), which was downloaded as a high-level science product from the Mikulski Archive for Space Telescopes (MAST).[5] These lists were assembled using an automated line identification and fitting algorithm, with subsequent human verification. We refer the reader to Danforth et al. (2014) for a full description of the analysis and line list creation for these spectra. It is important to note that the absorption line lists presented by Danforth et al. (2014) have been updated since the analysis conducted in this paper. These new results are presented in Danforth et al. (2014). They are based on the analysis of a further seven AGN sight-lines, and have better detection statistics owing to improved spectrum extraction and background subtraction. Although a refreshed analysis using this new data set is desirable, the general conclusions in this paper are likely to remain valid. In particular, any spurious line detections in the older analysis of Danforth et al. (2014) should only act to decrease the overall statistical significance of our results, rather than changing their implications.

3.1 Continuum fitting

Before line identification and fitting, the reduced QSO spectra are normalized by an estimate of the pseudo-continuum (continuum emission + line emission). We estimate this using a technique similar to that described in Young et al. (1979), Carswell et al. (1982) and Aguirre, Schaye & Theuns (2002). Each spectrum is split up into an arbitrary number of wavelength intervals, and a cubic spline fit through the set of points defined by the median flux in each interval. Pixels falling an arbitrary $n\sigma$ below the continuum are rejected, the median flux is recalculated, and the fit performed again. Here, σ is the standard deviation of the flux in each wavelength interval. We iterate over this process until the fit converges with an

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

[5] http://archive.stsci.edu/prepds/igm/

Table 3. QSO sight-line fields.

Field name	N_{QSO}	Area (sr)[a]	V_c (Gpc^3)[b]	Instrument/survey
J0005+1609	1	0.002 21	0.181	SDSS
J0029+1316	1	0.000 03	0.002	WFCCD
J0058−3606	1	0.002 18	0.153	2dFGRS
J0110−0218	3	0.000 04	0.049	CFHT, VIMOS, DEIMOS, GMOS
J0120−2821	2	0.004 87	0.623	2dFGRS
J0159+0023	1	0.003 09	0.215	SDSS
J0209−0438	1	0.003 14	4.700	VIPERS
J0228−1904	1	0.000 05	0.024	IMACS, LDSS3
J0407−1211	1	0.000 29	0.176	WFCCD, IMACS, LDSS3
J0426−5712	1	0.001 35	0.041	2dFGRS
J0559−5026	1	0.000 04	0.002	WFCCD
J0809+4619	1	0.004 87	0.436	SDSS
J0835+2459	1	0.004 87	0.369	SDSS
J0847+3445	1	0.004 87	0.059	SDSS
J0919+5521	1	0.004 87	0.201	SDSS
J0956−0453	1	0.003 32	0.210	2dFGRS
J0956+4115	2	0.004 87	0.452	SDSS
J1005+0134	1	0.000 82	1.082	SDSS, VVDS, VIMOS
J1007+2929	2	0.010 97	1.042	SDSS
J1013+3551	1	0.004 87	0.088	SDSS
J1014−0418	1	0.004 87	0.049	2dFGRS
J1022+0132	1	0.000 04	0.027	SDSS, VIMOS
J1031+5052	1	0.004 87	0.467	SDSS
J1051−0051	1	0.002 89	0.213	SDSS
J1058+3412	2	0.010 97	1.010	SDSS
J1118+5728	2	0.019 50	1.697	SDSS
J1119+2119	1	0.003 63	0.291	SDSS, WFCCD
J1121+4113	1	0.010 97	1.060	SDSS
J1131+3114	1	0.004 87	0.502	SDSS
J1139+6547	1	0.004 87	0.378	SDSS
J1226+0319	3	0.019 50	3.638	SDSS, IMACS, LDSS3, WFCCD
J1232+2009	1	0.004 87	0.057	SDSS
J1301+5902	2	0.004 87	0.488	SDSS
J1305−1033	1	0.000 04	0.007	WFCCD
J1309+0819	1	0.004 87	0.309	SDSS, WFCCD
J1357+0435	1	0.000 91	1.225	SDSS, VVDS
J1427+2348	1	0.004 87	0.372	SDSS
J1437−0147	1	0.004 87	1.646	GAMA
J1438−0658	1	0.001 82	0.078	2dFGRS
J1442+3526	1	0.004 87	0.088	SDSS
J1528+2825	1	0.004 87	0.382	SDSS
J1555+1111	1	0.004 87	0.577	SDSS, WFCCD
J1620+1724	1	0.004 87	0.169	SDSS
J2009−4849	1	0.000 04	0.001	WFCCD
J2132+1008	1	0.002 96	0.035	SDSS
J2155−0922	1	0.000 03	0.003	WFCCD
J2158−3013	1	0.004 87	0.183	2dFGRS, WFCCD
J2218+0052	1	0.001 09	1.464	SDSS, VVDS, VIMOS
J2254−1734	1	0.001 22	0.015	2dFGRS
J2351−0109	1	0.002 56	0.201	SDSS

[a]Field area, approximating the survey region as a rectangle.
[b]Comoving volume covered by the survey up to the minimum of (z_{QSO}, z_{95}).

approximately Gaussian distribution of flux values above the continuum. The appropriate value of n is found to vary from spectrum to spectrum, with values adopted in the range 1.5–3. From trial and error, the best value depends on SNR and location either within, or outside of the Lyα forest.

The continuum fitting process described above generally works well in regions of the spectra where the continuum varies smoothly. For regions where it fails, we adjust the continuum manually by hand. This is typically at the cusps of emission lines, in the Galactic Lyα absorption trough, at the absorption edge of Lyman limit systems and at the detector edges.

3.2 Absorption line identification

We identified absorption lines attributable to a particular ion and transition by performing a manual search through each QSO spectrum. We begin by searching for Galactic absorption lines at $z = 0$ and associated absorption lines at z_{QSO}. We then work systematically from z_{QSO} to $z = 0$, identifying H I absorbers on the basis of

there being at least two clearly detected Lyman series transitions at a given redshift. We simultaneously identify any metal absorbers coincident with the redshift of these H I absorbers.[6] Next, we scan through each spectrum again, identifying any 'high-ionization' doublets (namely Ne VIII, O VI, N V, C IV and Si IV) that may appear independently of any H I absorption. Finally, we assume lines in short wavelength regions of the spectra where there is no Lyβ coverage to be attributable to Lyα, and again look for coincident metal absorbers. For all identified ions, we set an initial guess for the number of velocity components, and for each component a column density and Doppler broadening parameter. This process typically accounts for >95 per cent of all absorption lines with equivalent widths at the $>3\sigma$ significance level.[7]

3.3 Voigt profile fitting

We fit Voigt profiles to the identified absorption-line systems with VPFIT,[8] accounting for the non-Gaussian COS line-spread function (LSF) at each wavelength by interpolating between the tables provided by the Space Telescope Science Institute (STScI).[9] In VPFIT, a χ^2 minimization is performed to fit Voigt profiles that are first convolved with the wavelength dependent COS LSF.

We begin with an initial list of guesses provided by the identification algorithm described in the previous section, and give these as input to VPFIT. All transitions of a given ion in a given system are fitted simultaneously, such that for each system, every transition of that ion shares the same redshift. Here, by 'system' we refer to the set of transitions belonging to a single ion at a single redshift, and we shall hereafter refer to these 'systems' as 'absorbers'. In general, we do not make any assumption as to whether different ions belong together at the same redshift in the same physical absorption complex (although see Finn et al. 2014, for a special case). Therefore, the redshifts for coincident ions are free to vary, consistent with the observation that some coincident ions show small velocity offsets. Fitted profiles are visually inspected, and initial guesses tweaked in rare cases where VPFIT fails to converge on a sensible result. We adopt only the minimum number of velocity components needed to minimize the reduced χ^2 value on the fit.

3.4 The absorption line catalogues

For each QSO spectrum, we compile an absorption line list based on the identification and Voigt profile fitting procedures just described. These are lists of absorbers (in a given ion), where each has a redshift (z), log column density ($\log N$) and Doppler broadening parameter (b), together with the associated 1σ uncertainties calculated during the fitting process. We also assign each absorber the right-ascension and declination of the QSO, such that it can be assigned a unique position in redshift space for cross-correlation. Additionally, a flagging system is employed to categorize the reliability of each absorber identification/fit. This scheme is similar to that in Tejos et al. (2014) and is defined as follows.

[6] Coincident here loosely means at $\Delta v \lesssim 50\,\mathrm{km\ s^{-1}}$. We are empirically motivated to search for metal absorbers at small (or zero) velocity separations from high column density H I absorbers ($\log N > 10^{15}\,\mathrm{cm^{-2}}$), but we do not make any prior assumption on the physical mechanisms that give rise to these coincidences.

[7] See Keeney et al. (2012) for a detailed discussion on the significance of absorption lines in *HST*/COS spectra.

[8] http://www.ast.cam.ac.uk/~rfc/vpfit.html

[9] http://www.stsci.edu/hst/cos/performance/spectral_resolution

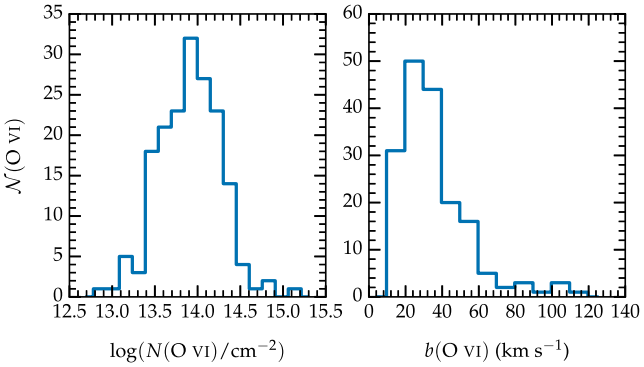

Figure 1. Statistics of O VI absorbers in our survey. The left-hand panel shows the histogram of column densities, and the right shows the histogram of Doppler broadening parameters.

(i) *Secure* ('a'): systems that are detected on the basis of at least two transitions (in the same ion), with $\log N/\Delta(\log N) > 30$, and each transition having an equivalent width significant at the $>3\sigma$ level.

(ii) *Probable* ('b'): H I systems detected on the basis of Lyα only (after ruling out all other possibilities and with equivalent widths significant at the $>4\sigma$ level), or metal-line systems detected on the basis of one transition with equivalent widths significant at the $>3\sigma$ level and with one or more accompanying Lyman series transitions. Both possibilities also with the requirement $\log N/\Delta(\log N) > 30$.

(iii) *Uncertain* ('c'): systems for which $\log N/\Delta(\log N) < 30$ and/or equivalent widths detected at the $<3\sigma$ level.

Absorbers in category 'c' are excluded from scientific analysis. This scheme is also applied to the measurements presented in Danforth et al. (2014). The scheme ensures that we only include absorbers in our sample that are both well constrained and statistically significant. The requirement that H I absorbers detected on the basis of Lyα only must have equivalent widths significant at the $>4\sigma$ level is motivated in Danforth et al. (2014), and is estimated to reduce the number of spurious detections to ~3 per spectrum.

For the analysis performed in this paper, we consider just the O VI samples. In Fig. 1, we show histograms of column density and Doppler broadening parameter for our O VI sample. There are a total of 181 O VI systems that possess reliability flags 'a' or 'b'. These absorption systems range over a factor of 100 in column density down to our detection limit (N(O VI) $\approx 10^{13}$ cm^{-2}), in marked contrast to H I absorbers that are observed to span ~10 orders of magnitude in column density. The number of O VI absorbers drops off fairly rapidly below $10^{13.5}$ cm^{-2}, and we are typically 100 per cent complete at N(O VI) $> 10^{14}$ cm^{-2}. Doppler broadening parameters show a long tail to high values, and a sharp cut off at ~10 km s^{-1}, which roughly corresponds to the spectral resolution of COS. This distribution is similar to that presented in Danforth et al. (2016). There may be a population of very narrow O VI absorbers, but we are not sensitive to them here.

4 ANALYSIS OF THE GALAXY DATA

The following sections describe the analyses performed on the 1D extracted galaxy spectra and photometric parent samples. Much of this analysis builds on that already presented in Tejos et al. (2014). For the SDSS, 2dFGRS, VVDS and VIPERS surveys, and all galaxy data presented in Chen & Mulchaey (2009), Prochaska et al. (2011a) and Johnson et al. (2013), we work from the catalogued

magnitudes, redshifts and spectral line measurements (where available).[10] GAMA galaxy catalogues used in this study are from phase II of the survey, and are not publicly available at the time of writing. For SDSS, we make use of the spectral line measurements presented in (Brinchmann et al. 2004, see Section 4.4 for details).

4.1 Redshift determination

The majority of the galaxy redshifts in our VIMOS, DEIMOS, GMOS and CFHT samples were obtained by cross-correlating galaxy, star and QSO templates from SDSS[11] with each observed spectrum (see Morris & Jannuzi 2006; Tejos et al. 2014, for a full description). Each galaxy was then assigned a quality flag to indicate the reliability of the assigned redshift. The scheme is designed as follows.

(i) *Secure* ('a'): at least three well-identified spectral features (emission or absorption lines) or two well-identified emission lines.
(ii) *Possible* ('b'): only one or two spectral features.
(iii) *Uncertain* ('c'): no clear spectral features.

(Tejos et al. 2014). Flag 'c' is typically raised for spectra with low SNR, or due to an intrinsic lack of observable features at the instrumental resolution. We do not use these redshifts in any of the forthcoming analysis. For all other galaxy redshifts, we map the corresponding quality flags on to our scheme to ensure a unified definition for 'secure', 'possible', or 'uncertain' as follows.

In SDSS, we simply adopt all galaxies with a warning flag of 0 (indicating no warnings) as being secure (label 'a'), and flag all other redshifts as 'c' (see Stoughton et al. 2002, for details on SDSS flags).

The 2dFGRS scheme is defined in terms of absorption redshifts and emission redshifts separately. In brief, for absorption redshifts, a quality parameter Q_a is defined in terms of a variable R, being the ratio of peak to noise in the cross-correlation with the best-fitting template, as follows:

$$Q_a = \begin{cases} 4 & R > 5.0, \\ 3 & R > 4.5, \\ 2 & R > 4.0, \\ 1 & R > 3.5, \\ 0 & \text{otherwise}, \end{cases}$$

with a further requirement that $Q_a = 3$ and 4 redshifts are obtained to within 600 km s^{-1} across four and six of the eight spectral templates, respectively. For emission redshifts, the parameter Q_e is defined as

$$Q_e = \begin{cases} 4 & \text{three or more detected lines}, \\ 2 & \text{two lines, or one strong line}, \\ 1 & \text{one weak line}, \\ 0 & \text{no lines}. \end{cases}$$

The combined redshift quality flag, Q_b, is then determined as $Q_b = \max(Q_a, Q_e)$, unless the difference between the absorption and emission redshifts is $<$600 km s^{-1}, in which case, $Q_b = \max(Q_a, Q_e, 3)$, or if $Q_a \geq 2$ and $Q_e \geq 2$ and the difference between absorption and emission redshifts is $>$600 km s^{-1}, in which case $Q_b = 1$. An overall redshift flag, Q, is determined via human verification of the automated redshift measurement, with the option of manually fitting Gaussian lines to spectral features as a means to obtain the redshift. The scheme is then as follows:

$$Q = \begin{cases} 5 & \text{reliable redshift, high}-\text{quality spectrum}, \\ 4 & \text{reliable redshift}, \\ 3 & \text{probable redshift}, \\ 2 & \text{possible, but doubtful redshift}, \\ 1 & \text{no redshift could be estimated} \end{cases}$$

(see Colless et al. 2001, for a detailed description). We perform the mapping $Q \geq 4 \rightarrow$ 'a', $Q = 3 \rightarrow$ 'b', $Q < 3 \rightarrow$ 'c'.

GAMA redshifts are derived using the AUTOZ code (Baldry et al. 2014) and assigned a quality parameter, nQ, in the range 1–4 based on quantitative estimates of their reliability (Baldry et al. 2014; Liske et al. 2015). We use the same mapping as above to translate nQ to our quality flag.

The VVDS scheme is defined according to the following numbering scheme: (0) no redshift (no features); (1) tentative redshift (weak features, continuum shape); (2) secure redshift (several features); (3) very secure redshift (strong spectral features); (4) completely secure redshift (obvious spectral features); (9) redshift based on single secure feature. Added to this are the prefixes 1 and 2, to mean broad line AGN and secondary target, respectively (see Le Fèvre et al. 2005, for more details). We perform the following mapping:

(i) {4, 14, 24, 214, 3, 13, 23, 213} $\rightarrow$ 'a',
(ii) {2, 12, 22, 212, 9, 19, 29, 219} $\rightarrow$ 'b',
(iii) {1, 11, 21, 211, 0} $\rightarrow$ 'c'.

The VIPERS scheme is identical to that of VVDS, but with the addition of a decimal fraction to each flag depending on the photometric redshift from the accompanying 5-band CFHT Legacy Survey (CFHTLS) photometry. If the spectroscopic redshift falls within the 1σ confidence interval on the photometric redshift, a value of 0.5 is added. If it falls within the 2σ confidence interval, a value of 0.4 is added. If it falls outside the 2σ confidence interval, a value of 0.2 is added, and when there is no photometric redshift, a value of 0.1 is added (see Guzzo et al. 2014, for more details). We adopt the same mapping as for VVDS, regardless of the added decimal fraction.

Redshifts for galaxies presented in Chen & Mulchaey (2009) and Prochaska et al. (2011a) are only provided where they are deemed reliable. We therefore label objects having an assigned redshift with flag 'a', and all other objects flag 'c'.

In Johnson et al. (2013), the redshift flagging scheme is defined (A) secure ($\geq$2 features); (B) 1 feature; (C) observed but no features; and (N) not observed. We perform the mapping A $\rightarrow$ 'a', B $\rightarrow$ 'b', {C, N} $\rightarrow$ 'c'.

4.2 Global astrometry/photometry solutions

In two fields, J1005−0134 and J2218+0052, VIMOS observations obtained by our collaboration supplement VVDS galaxy data at small angular separations from each QSO (Tejos et al. 2014). We have improved the photometric and astrometric calibration for these data as follows.

[10] Catalogues for 2dFGRS, SDSS, VVDS and VIPERS galaxies are obtained from http://www.2dfgrs.net, http://skyserver.sdss.org/casjobs, http://cesam.oamp.fr/vvdsproject/vvds.htm and http://vipers.inaf.it/rel-pdr1.html, respectively. Catalogues from the analysis in Chen & Mulchaey (2009), Prochaska et al. (2011a) and Johnson et al. (2013) were obtained from http://vizier.cfa.harvard.edu/viz-bin/VizieR.
[11] http://www.sdss.org/dr7/algorithms/spectemplates/

Table 4. Spectral index definitions.

Index	Blue continuum (Å)	Line (Å)	Red continuum (Å)
[O II]	3655–3705	3708.5–3748.5	3750–3800
Hδ	4030–4080	4082.0–4122.0	4125–4170
Hγ	4230–4270	4321.5–4361.5	4365–4400
Hβ	4785–4820	4842.5–4882.5	5030–5100
[O III]	4785–4820	4988.0–5028.0	5030–5100
H α + [N II]	6460–6520	6544.5–6584.5	6610–6670
[S II]	6640–6700	6713.0–6753.0	6760–6810

Astrometry and *R*-band photometry was originally obtained from the VIMOS pre-imaging data. However, these fields overlap with the VIRMOS deep imaging survey (Le Fèvre et al. 2004), for which the astrometric and photometric calibration is superior. This photometric data set extends coverage to the *B*, *V* and *I* bands, and forms the basis for target selection in the VVDS. We therefore set about matching these data sets to ensure global astrometric and photometric consistency across these fields. We made use of the SEXTRACTOR (Bertin & Arnouts 1996) and SCAMP (Bertin 2006) software packages to automatically map galaxy positions from the VIMOS detector plane to world coordinates using sources detected in SDSS as a reference. This brought the astrometric solution to within 1 arcsec of the VIRMOS deep imaging survey, which is below the typical seeing level. We then cross-matched the photometric catalogues, and calculated the mean *R*-band magnitude offset needed to bring the two into statistical agreement. We did this for each VIMOS quadrant separately. The typical offset was ∼0.4 mag. Not all sources could be matched to those from the VIRMOS deep imaging survey due to regions of the imaging for that survey that are poorly calibrated. For these sources, we keep offset VIMOS *R*-band magnitudes, whereas elsewhere we assign the appropriate matched *BVRI* photometry.

4.3 Spectral line measurements

For the VVDS and VIPERS surveys, no spectral line measurements or indications of spectral type are made available. We therefore performed our own analysis where possible, as a means to estimate the star formation activity for the galaxies in these surveys (see Section 4.4 for details). We also performed this analysis on the VIMOS, GMOS and CFHT data collected by our collaboration (Morris & Jannuzi 2006; Tejos et al. 2014). Originally, the spectral types for these galaxies were determined by assigning the spectral type of the best-fitting template as part of the redshift determination process. We improve upon this by measuring spectral line fluxes, as described below.

For each galaxy spectrum, where spectral coverage, resolution and SNR allowed, we estimated the integrated fluxes and local continuum level around the [O II], Hδ, Hγ, Hβ, [O III], Hα, [N II] and [S II] emission lines. For this, we used the spectral line indices defined in Table 4, and a direct integration over spectral pixels. The continuum level is obtained by iteratively clipping points 1.5σ below the estimated continuum in a manner similar to that described in Section 3.1, taking a mean of the 'unclipped' pixels either side of the line, and linearly interpolating between the points defined by these means. This makes the continuum estimate reasonably robust to underlying absorption, but it occasionally fails at the edges of some spectra where there is a loss in sensitivity, leading to poor flux-calibration, and a rapid fall-off in the continuum. The other main cause for continuum misplacement is the presence of occasional contaminating zero-orders[12] lying on top of the galaxy spectra, or regions of bad sky subtraction. Note that we do not accurately remove the stellar continuum (including the underlying stellar absorption) in our procedure. This inevitably affects the reliability of the inferred emission line fluxes (the Balmer emission lines in particular), however our approach suffices for the purposes of splitting the galaxy sample into star-forming and non-star-forming populations (see Section 4.4 for details).

The line indices in Table 4 are optimized for spectra taken with VIMOS at a spectral resolution $R = 200$, appropriate for the VVDS and VIPERS surveys, and for the VIMOS data presented in Tejos et al. (2014). For the GMOS and CFHT data, we narrowed the line indices to reflect the higher spectral resolution obtained by these instruments (see Tejos et al. 2014, and references therein for details). For integrated line fluxes detected above a 3σ significance threshold, we also attempt to fit Gaussian profiles, which are usually adopted in preference to the pixel measurements. We revert to the pixel measurements when the fitting routine returns a Gaussian with zero amplitude, indicating that the fit has failed. All Gaussian fits are performed with a χ^2 minimization employing the Levenberg–Marquardt algorithm. The rest-frame standard deviation of each Gaussian line is bounded between $0.5\sigma_{\rm LSF}$ Å and $\sqrt{10^2 + \sigma^2_{\rm LSF}}$ Å, where $\sigma_{\rm LSF}$ is the standard deviation of the (assumed Gaussian) LSF, which helps to identify broad emission lines that are likely of AGN origin, and contaminating sky lines. We fit the [O II] doublet, Hδ, Hγ and [S II] doublet lines separately. The [O II] doublet is not resolved by our spectra, so we fit it as a single line. For the [S II] doublet, which is marginally resolved, we tie the Gaussian standard deviations in the fit, and fix the line ratio [S II] λ6716/[S II] λ6731 to the expected (but not fixed)[13] ratio of 1/1.4 (Osterbrock 1989). We fit the Hβ/[O III] line complex simultaneously, tying together the Gaussian standard deviations, and fixing the [O III] λ4956/[O III] λ5007 ratio to the expected value of 1/2.98 (Storey & Zeippen 2000). We also fit the Hα/[N II] complex simultaneously, tying together the Gaussian standard deviations, and fixing the [N II] λ6548/[N II] λ6583 ratio to the expected value of 1/2.95 (Osterbrock 1989). The [N II] lines are barely resolved from the Hα line in our spectra, so we perform an alternative, single Gaussian fit to just the Hα line in every case. If this fit gives a smaller χ^2 value than the three-component fit, we assign the resulting Gaussian parameters to the Hα line, and report no [N II] measurements in these instances. Despite not having sufficient spectral resolution to properly resolve the [N II] components, in instances where these lines are strong, the resulting line profile has definite asymmetry, which motivates us to decompose the line profile. Uncertainties on the fitted Gaussian parameters are estimated by generating 100 Monte Carlo realizations of the data. For each realization, we add a number to every pixel flux, randomly generated from a Gaussian distribution of values with standard deviation equal to its 1σ uncertainty. Each of these 100 realizations are then fit using the same procedure as in the nominal case, and the standard deviation over the resulting best-fitting parameter values are taken as the 1σ uncertainty on the measurement.

To identify bad measurements in our galaxy spectra, we have devised a flagging scheme as follows.

(i) Flag (0): no warnings.

(ii) Flag (1): measurement may be affected by the OH forest between 8600 and 8700 Å.

[12] Crowding on the detector can often lead to zero-order spectra landing on regions inhabited by first-order spectra.

[13] Assumes a gas density and temperature.

(iii) Flag (2): line was fit with the maximum/minimum allowed Gaussian standard deviation.

(iv) Flag (3): line coincides with a region above a user-specified sky spectrum threshold.

(v) Flag (4): line may be affected by the O_2 telluric absorption at ∼7600 Å.

(vi) Flag (5): bad continuum reduced χ^2 (> 10).

(vii) Flag (6): no spectral coverage.

The quality of the sky subtraction in our spectra makes the line measurements reasonably robust to flag (1). Flag (2) is implemented to identify potential broad-line AGN and contaminating sky lines. Flag (3) is mainly implemented to eliminate contaminating zero orders. We found that the VIMOS ESOREX pipeline reduction software often incorrectly identifies zero orders offset from the galaxy spectrum of interest and tries to correct for them, leaving deep, artificial absorption features in the extracted 1D galaxy spectra. Nevertheless, these appear as broad spikes in the extracted 1D sky spectra, and can be identified by adopting a threshold sky value. Flag (4) is implemented because most of our spectra are not corrected for the O_2 telluric, and even in spectra that are corrected for this contaminating feature, the correction is highly uncertain due to the narrow 'picket-fence' nature of the absorption. Flag (5) is implemented to identify line measurements that are marred by bad continuum estimation. We allow for relatively high reduced χ^2 values in view of occasional absorption that raises the value of this statistic even for reasonable continuum estimations. Flag (6) is implemented to identify lines not measured due to insufficient spectral coverage. We enforce coverage across the entire region defined by each of the line indices in Table 4. Line measurements that do not raise any of the aforementioned flags are assigned flag (0), to indicate that there are no warnings.

In all of our spectra, we reject measurements that raise flags (2), (4), (5) and (6). For VIMOS spectra reduced with ESOREX, we additionally reject measurements that raised flag (3). In practice, this flag is reserved for those spectra only.

4.4 Star formation activity

For the VVDS and VIPERS surveys, and our VIMOS, GMOS and CFHT data, we use the spectral line measurements described in the previous section to split the sample of galaxies in terms of their star formation activity. For GAMA, we also do the same using a very similar set of line measurements provided by the GAMA survey team (see Hopkins et al. 2013, for a description). We aim simply to define galaxies as 'star-forming', 'non-star-forming', 'AGN-dominated', or 'unclassified'. Although we could have calculated star formation rates for many of our galaxies using standard procedures (e.g. Kennicutt 1998; Moustakas, Kennicutt & Tremonti 2006), estimating *K*-corrections at redshifts $\gtrsim$0.5 becomes increasingly uncertain, and in general we lack a homogeneous set of multi-band photometric measurements across our sample to allow for a consistent approach. In any case, simply splitting our sample purely on the basis of spectral line measurements suffices for our requirements. We adopt a very similar prescription to that outlined in Brinchmann et al. (2004), whose classification scheme was applied to the SDSS galaxies in our sample.

First, we attempt to identify galaxies whose spectra are dominated by an AGN component. A number of broad-line AGN are already identified on the basis of their redshift determination via cross-correlation with AGN templates (Tejos et al. 2014). For the remaining galaxies, we perform a classification on the basis of a Baldwin, Phillips & Terlevich (1981, hereafter BPT) diagram, shown in Fig. 2. Kewley et al. (2001) performed spectral energy distribution (SED) and photoionization modelling to find a theoretical discriminating line between star-forming galaxies and AGN on a BPT diagram of $\log([\mathrm{N\,II}]/\mathrm{H}\,\alpha)$ versus $\log([\mathrm{O\,III}]/\mathrm{H}\,\beta)$. This discriminating line is shown in Fig. 2, along with the subset of our galaxies that have SNR > 3 in each of the measured Hα, Hβ, [N II] and [O III] line fluxes, and no severe warning flags associated with these (we accept flags 0 and 1 in the scheme devised above). High values of $\log([\mathrm{N\,II}]/\mathrm{H}\,\alpha)$ and $\log([\mathrm{O\,III}]/\mathrm{H}\,\beta)$ are driven by a hard extreme ultraviolet (EUV) spectrum attributable to AGN activity in the host galaxies, pushing these galaxies to the top-right corner of the diagram. Although we can make an AGN classification on the basis of this diagram regardless of SNR, we use this technique only for SNR > 3, since below this, an increasing fraction of galaxies have measured line fluxes that are negative, and the non-symmetric distribution of galaxies on this diagram leads to classification biases. Typically, only a very small subset of our galaxies for which we performed spectral line measurements have all the required lines measured above our SNR criterion, but we are able to expand the classification by using only single line ratios. These are indicated by the dashed lines in Fig. 2, and they correspond to $\log([\mathrm{N\,II}]/\mathrm{H}\,\alpha) > 0.3$ and $\log([\mathrm{O\,III}]/\mathrm{H}\,\beta) > 1$. Clearly these classifiers are less effective than that using both line ratios. However, this does allow us to classify AGN according to their line ratios over a larger number of galaxies. We find that only around 5 per cent of our galaxies are classified as AGN, but we cannot rule out a small additional population that could not be identified in the manner just described.

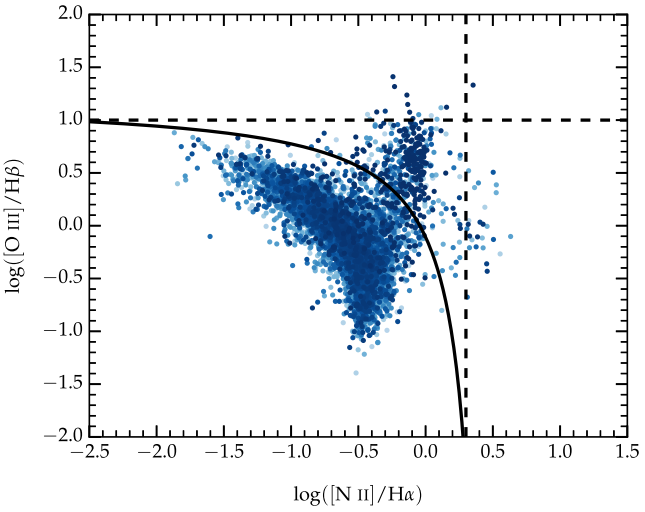

Figure 2. The distribution of galaxies in our sample on a BPT line-ratio diagram. The solid line indicates the discrimination line from Kewley et al. (2001), separating star-forming galaxies from AGN. The dashed lines indicate the single-line ratio diagnostics we also employ to identify AGN when two line ratios are not available.

After identifying AGN, we assume that the rest of the galaxies are regular star-forming or non-star-forming galaxies. We identify star-forming galaxies as those that show measured fluxes that are positive and with SNR > 2 in any one of the Hα, Hβ, or [O II] lines. Those that do not meet this criterion are identified as non-star-forming galaxies. Those galaxies that do not have good measurements of any of the Hα, Hβ, or [O II] emission lines (due to bad flags and/or lack of spectral coverage) are marked as 'unclassified'. These galaxies nevertheless have redshift measurements from e.g. the Ca II H and K, CH molecule *G* band, Mg λ5175 and Na λ5894 lines.

For all other galaxies in our sample, we obtain their spectral classifications from the literature, typically found from principle component analyses (see Chen & Mulchaey 2009; Prochaska et al. 2011a; Johnson et al. 2013; Tejos et al. 2014, for details). Overall, we find that ∼55 per cent of the galaxies in our sample are classified as star forming, ∼35 per cent are classified as non-star forming, ∼5 per cent are classified as AGN and ∼5 per cent are unclassified.

5 GAS AND GALAXIES IN THE EAGLE HYDRODYNAMICAL SIMULATION

In the following sections, we present the methods used to extract a comparison data set from the EAGLE project, which is a suite of hydrodynamical simulations that follow the formation and evolution of galaxies and supermassive black holes in volumes representative of a ΛCDM Universe. We begin by briefly describing the pertinent aspects of the simulation, and discuss its key advantages and limitations. We then follow with a detailed description of the processes involved in generating mock catalogues of galaxies and absorption systems, designed to mimic as closely as possible the observations.

5.1 The EAGLE simulations

The EAGLE project (Crain et al. 2015; Schaye et al. 2015) is a suite of cosmological hydrodynamical simulations representative of a ΛCDM Universe. The simulations were run with the smoothed particle hydrodynamics (SPH) code GADGET3 in cubic volumes 12.5, 25, 50 and 100 comoving Mpc on a side.[14] State-of-the-art numerical techniques and subgrid models are used to capture various physical processes important to galaxy formation and evolution. These include radiative gas cooling, star formation, mass-loss from stars, metal enrichment, energy feedback from star formation and AGN and gas accretion on to, and mergers of, supermassive black holes. The efficiency of stellar feedback and the mass accretion on to black holes is calibrated to match the present-day stellar mass function of galaxies (subject to the additional constraint that the galaxies sizes need to be reasonable), and the efficiency of AGN feedback is calibrated to match the observed relation between stellar mass and black hole mass. Calibrations such as these are necessary, since the underlying physics behind galaxy feedback is neither well understood, nor well constrained observationally, and the resolution of the simulations is insufficient for ab initio predictions of the feedback efficiency.

In this work, we use only the largest simulation volume (100 Mpc^3), referred to as L100N1504, containing 1504^3 SPH particles. For this, the initial baryonic particle mass is $m_g = 1.81 \times 10^6$ M$_\odot$, the dark matter particle mass is $m_{dm} = 9.70 \times 10^6$ M$_\odot$, and the comoving, Plummer-equivalent gravitational softening length is 2.66 kpc. At this resolution, EAGLE is marginally sufficient to resolve the Jeans scales in the warm interstellar medium (ISM). We shall refer to three particle types in EAGLE: dark matter particles, star particles and gas particles. Dark matter particles are evolved using the N-body part of the code, which simulates just the gravitational interactions between particles, while gas particles are also subject to hydrodynamical forces. Gas particles above a metallicity-dependent density threshold are converted to star particles stochastically. More details can be found in Schaye et al. (2015).

[14] There is also a set of high-resolution 'zoom' simulations (see Sawala et al. 2015, for details).

We will test the feedback prescriptions in EAGLE by examining its predictions for the distribution and dynamics of O VI absorbers around galaxies, which we can then compare to our observational sample. A test such as this has considerable diagnostic power, since the simulation was not calibrated to match observations such as these. Even though we cannot hope to learn much about the detailed physics governing SN and AGN feedback, a simulation that matches these observations should nevertheless provide important insights on the gas flows around galaxies, which are responsible for driving the evolution of key galaxy properties, such as their star formation rates, and the mass–metallicity relationship. To perform this test, we need to generate mock catalogues of galaxies and absorbers from the simulation, which is the subject of the following sections.

5.2 Intergalactic gas

We characterize the IGM in EAGLE in a very similar manner to observations by drawing synthetic QSO sight-lines through the simulation volume. We follow the procedure outlined in (Theuns et al. 1998, see their appendix A4), using a modified version of the artificial transmission spectra code, SPECWIZARD. This works as follows. For a given (x, y, z) coordinate and orientation in the simulation volume at a given redshift snapshot, specifying a one-dimensional sight-line, SPECWIZARD first extracts all SPH gas particles that intersect that sight-line. The sight-line is then divided into an arbitrarily high number of bins of width Δ in real space. These bins are labelled from zero to $\dot{a}L$ in velocity space, each having a width of 1 km s^{-1}, where $a(z)$ is the dimensionless scale-factor and L is the box size in comoving coordinates. For each bin, the code then calculates the local physical density, ρ_X, and temperature, T_X, for an ionic species X, weighted by the SPH smoothing kernel and abundance of species X, assuming ionization equilibrium in the presence of a Haardt & Madau (2001) UV background radiation field. Then, for a given atomic transition, i, of species X, assuming only thermal line broadening, a bin k, corresponding to a velocity $v(k)$, will suffer absorption due to material in bin j, at velocity $v(j)$, by an amount $\mathrm{e}^{-\tau(k)}$, where

$$\tau(k) = \sigma_{X_i} \frac{1}{\sqrt{\pi}} \frac{c}{V_X(j)} \rho_X(j) a \Delta \exp\left[-\left(\frac{v(k) - v(j)}{V_X(j)} \right)^2 \right], \quad (1)$$

and

$$V_X^2(j) = \frac{2kT_X(j)}{m_X} \quad (2)$$

(Theuns et al. 1998). Here, σ_{X_i} is the absorption cross-section of the transition, c is the speed of light and $V_X(j)$ is the Doppler width of species X with mass m_X. For the vast majority of the absorption along these sight-lines, the physical densities are small enough that a purely thermally broadened line-profile is a good approximation to the real one.

To create mock catalogues of O VI absorbers, we use the method above to calculate the optical depth, $\tau(v)$, in O VI λ1031 along 25 000 randomly drawn sight-lines parallel to the z-axis (our pseudo-redshift axis) through each of seven different redshift snapshots over the range $0.1 \lesssim z \lesssim 0.7$ (the dominant range covered by our observational sample). We then take peaks in the τ distribution above a threshold value of 0.0005 (arbitrary) along each sight-line to correspond to the optical depth at the absorption line centres, and calculate the absorbing column density assuming a Doppler broadening parameter equivalent to V_X in the equations above. Each absorber is then assigned the (x, y) coordinate of the sight-line it was extracted from, and the velocity, v, at which it was

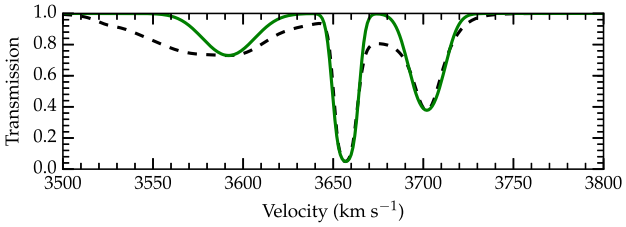

Figure 3. Voigt profiles derived from peaks in the optical depth distribution (plotted in green) compared to the $e^{-\tau}$ spectrum. The predicted Voigt profiles show simpler velocity structure than is apparent in the spectrum, highlighting an important caveat to our approach.

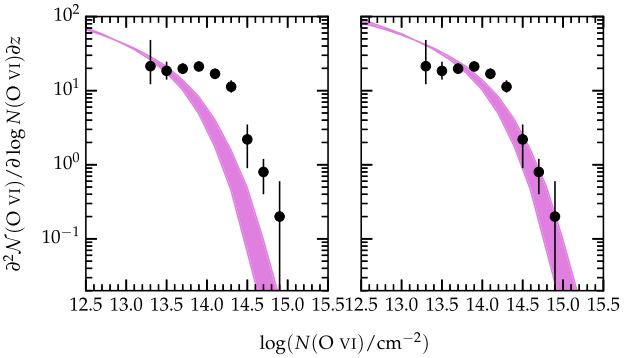

Figure 4. The column density distribution function of O VI absorbers from (Danforth et al. 2014, data points) compared to predictions from EAGLE snapshots over the redshift range $0.1 \lesssim z \lesssim 0.7$ (shaded magenta region). The shape of the column density distribution function predicted from EAGLE is not far from that in the data (left-hand panel), but a shift of 0.3 dex in column density (right-hand panel) is necessary to obtain the required number of high column density absorbers from the simulation (see the text for justification).

extracted, which we convert to a position r_z along the z-axis via $r_z = v(1+z)/H(z)$, where z is the redshift of the simulation snapshot, and $H(z)$ is the value of the Hubble parameter at that redshift.

It is important to note that the above procedure differs from the observational one, whereby Voigt profiles are fit to the absorption features in transmission spectra ($e^{-\tau}$) in order to extract column densities and Doppler broadening parameters. This approach enables deblending of multiple-component absorption, and takes into account the broadening of the lines due to the instrumental profile, enabling accurate recovery of their column densities. The technique was originally devised under the premise that the intervening absorption lines in QSO spectra arose from discrete absorbing clouds, but this picture was challenged early by the smoothly distributed IGM captured in cosmological hydrodynamical simulations (e.g. Cen et al. 1994; Hernquist et al. 1996; Miralda-Escudé et al. 1996; Theuns et al. 1998; Davé et al. 1999). Absorbers that have a large spatial extent take part in the Hubble expansion, which leads to a line profile that deviates from a Voigt profile. Voigt profile fitting the transmission spectra will therefore glean slightly different results to simply taking peaks in the τ distribution. In particular, we might expect a larger number of absorbers, and some differences in the derived Doppler broadening parameters and column densities. As a check, we derived Voigt profiles for lines recovered in some of the sight-lines extracted from the simulation using our τ peak method, and plotted these on top of the transmission spectrum. An example is shown in Fig. 3. The transmission spectrum is shown as the black dashed line, and the predicted Voigt profiles are shown in green. More structure is apparent in the real spectrum, which would yield a larger number of Voigt components in Voigt profile fitting, and column densities and Doppler broadening parameters that differ slightly from those we have recovered via the τ peak method. This is an obvious caveat to our approach, which we bear in mind when interpreting our results later on.

Having created a mock catalogue of absorbers from the EAGLE simulation, we now examine whether the global statistics of our mock population matches the observed one. To do so, we investigate the column density distribution function of O VI absorbers. This is defined as the number, $\mathcal{N}$, of absorption lines per unit column density, $\mathrm{d}N_{\mathrm{O\ VI}}$, per unit redshift, $\mathrm{d}z$:

$$f(N_{\mathrm{O\ VI}}) = \frac{\mathrm{d}^2\mathcal{N}}{\mathrm{d}N_{\mathrm{O\ VI}}\mathrm{d}z}. \quad (3)$$

It is also sometimes defined in terms of the absorption distance $\mathrm{d}X$, which is related to $\mathrm{d}z$ via $\mathrm{d}X/\mathrm{d}z = H_0(1+z)^2/H(z)$, where $H(z)$ is the Hubble parameter. We compare to the measurement presented in (Danforth et al. 2014, data points in Fig. 4), whose data form the majority of our O VI sample, although we remind the reader that this measurement was recently updated in Danforth et al. (2016). To calculate this for the O VI absorbers extracted from EAGLE, we adopt the same bin size as was used for the data, $\Delta(\log N_{\mathrm{O\ VI}}) = 0.2$, and note that we have $n_{\mathrm{LOS}} = 25\,000$ sight-lines through the $L = 100$ Mpc simulation volume, then calculate $\Delta z = n_{\mathrm{LOS}}LH(z)/c$, where c is the speed of light. We do this for each of the redshift snapshots in the range $0.1 \lesssim z \lesssim 0.7$, and plot the resulting curves as the shaded region in the left-hand panel of Fig. 4. We note that a similar calculation has been performed by projecting the simulation volume along one axis, and taking the summed O VI column densities in each two-dimensional grid cell, yielding similar results (see Schaye et al. 2015, for details). The shape of the column density distribution function from EAGLE is not far from that of the observed one, although the 'knee' seen in the observed distribution is not so well pronounced in the simulation. We note also that the distribution of simulated O VI absorbers falls off at lower column densities than in the data. Uncertainties in the shape and normalization of the UV background may be enough to account for this difference, along with uncertainties in the metal yields from star formation (Schaye et al. 2015). Due to these differences, it is necessary to shift the simulated O VI absorption column densities by 0.3 dex towards higher values in order to draw a subsample of O VI absorbers from the simulation that is matched to the selection functions in the data (right-hand panel in Fig. 4). We therefore proceed from this point on with simulated O VI absorbers whose column densities are rescaled by an extra 0.3 dex in column density.

To create a matched sample of O VI absorption systems from the EAGLE simulation, we begin by optimally binning the column density histogram of the observed sample using the method presented in Knuth (2006; see Section 6.2 for a description and motivation), then smooth the histogram with a Gaussian having standard deviation equal to the bin size. We then interpolate a cubic spline over the smoothed histogram, and normalize by the area underneath the curve to produce a probability density function (PDF) for the observed column densities. We also do the same on the column density distribution of simulated samples, which exhibits a turn over at very low column densities ($N(\mathrm{O\ VI}) < 10^{11}$ cm^{-2}) due to the arbitrarily imposed optical depth cut. We then randomly draw a 10 per cent subset of the simulated O VI absorbers from the probability density distribution of the data, inverse weighted by that of the simulation. We illustrate this in Fig. 5. The observed PDF is shown in blue, the simulated one in red, and that of the 10 per cent subset in green. As

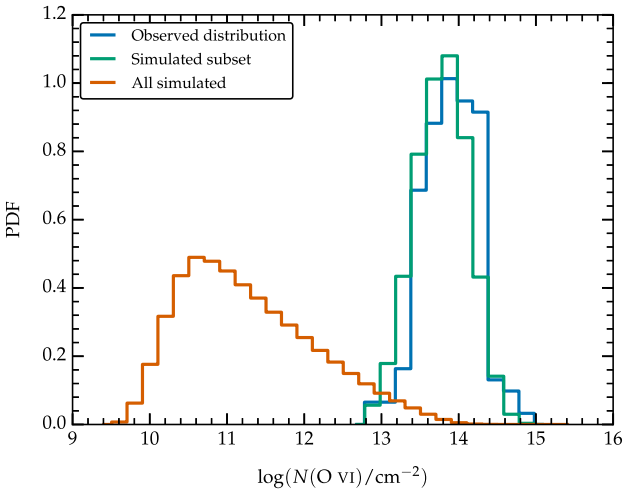

Figure 5. PDF for the observed O VI column densities (blue), the simulated ones (red) and the simulated subset (green). See the text for a full description.

is clearly evident from the figure, we are able to extract a subsample of O VI absorbers from the EAGLE simulation that reproduces well the selection bias towards high column densities in the data.

5.3 Galaxies

Galaxies in EAGLE are defined by the minimum of the gravitational potential in 'subhaloes' identified by the SUBFIND algorithm (Springel et al. 2001; Dolag et al. 2009). A 'subhalo' is defined in this algorithm as a gravitationally bound, locally overdense group of particles, identified in the following fashion. First, dark matter haloes are found by implementing the friends-of-friends (FoF; Davis et al. 1985) algorithm on the dark matter particles. Gas and star particles in the simulation are then assigned to the same (if any) FoF halo as their nearest dark matter particles. Next, substructure candidates are identified by finding overdense regions within each FoF halo that are bounded by saddle points in the density distribution. Finally, particles not gravitationally bound to the substructure are discarded, and the resulting substructures are termed subhaloes. For EAGLE, it was also necessary to merge subhaloes separated by less than the minimum of 3 kpc and the stellar half-mass radius, to remove a small number of very low mass subhaloes whose mass was dominated by a single particle (Schaye et al. 2015).

The stellar mass of each galaxy in EAGLE is defined to be the sum of the masses of all the star particles that belong to the corresponding subhalo, and that are within a 3D aperture of radius 30 kpc. The choice of aperture was motivated by the fact that this gives a nearly identical galaxy stellar mass function as applying the Petrosian apertures used in the observations (Schaye et al. 2015). All galaxy properties in EAGLE, including the star formation rates, are evaluated in this aperture. For the purpose of matching to the observational sample, we limit the EAGLE sample to galaxies with stellar masses $\log(M_\star/\mathrm{M}_\odot) > 8$. We perform this mass cut for two reasons: (i) to eliminate galaxies consisting of $\lesssim$100 star particles, as these galaxies will have star formation rates that are not robust; and (ii) to reflect our observed sample, for which stellar masses $\log(M_\star/\mathrm{M}_\odot) < 8$ are comparatively rare due to the limiting magnitudes of the surveys (see, for example, Li & White 2009; Baldry et al. 2012).

To split the EAGLE sample of galaxies in terms of their star formation activity, we adopt the same scheme as in Schaye et al. (2015), whereby galaxies with specific star formation rates $\dot{M}_\star/M_\star > 0.1\,\mathrm{Gyr}^{-1}$ are defined to be star-forming galaxies, and all others are defined to be non-star-forming galaxies. We do not attempt to find a cut in star formation rate (or specific star formation rate) that will match the relative proportions of star-forming and non-star-forming galaxies in the observational sample, since the observational definition is subject to signal-to-noise constraints not present in the simulation. We leave a more precise comparison for future work.

Finally, we define the z-axis of the simulation volume to be the pseudo-redshift axis, and replace the z-coordinate of each galaxy, which we shall denote r_z, with $r_{z,\,\mathrm{pec}} = r_z + v_{\mathrm{pec}}(1+z)/H(z)$, where v_{pec} is the peculiar velocity of the galaxy, and $H(z)$ the Hubble parameter for a redshift z corresponding to the snapshot redshift. We note that the simulation volume is periodic, and we preserve this feature when modifying the z-coordinates of the galaxies. This then mimics the distortions introduced by peculiar velocities along the LOS for galaxy surveys constructed in redshift space.

6 IGM/GALAXY CLUSTERING ANALYSIS

We now make use of our samples of O VI absorbers and galaxies, together with the data sets drawn from the EAGLE cosmological hydrodynamical simulation, to examine their two-point correlation functions. We limit the QSO sight-line fields listed in Table 3 to just those where we found O VI absorption in at least one QSO spectrum with confidence label 'a' or 'b'. We also exclude O VI absorbers within 5000 km s^{-1} of the QSO redshift, on the conservative assessment that these may be intrinsic to the QSO host galaxy. Galaxies are limited to those with redshift confidence labels 'a' or 'b'. The resulting sample for this study then consists of 27 fields and 32 QSO sight-lines, which contain 51 296 galaxies and 140 O VI absorbers, respectively. Our full sample of galaxies and QSOs will be used to investigate the H I-galaxy cross-correlation function in future work. This will serve as an extension to the work presented in Tejos et al. (2014).

6.1 The two-point correlation function

To address the statistical connection between the metal-enriched IGM (traced by O VI absorbers) and galaxies, we focus on a two-point correlation analysis. The advantages of doing so are two-fold: (i) we do not rely on associating a particular intergalactic absorber with a particular galaxy (or set of galaxies), which in many instances is ambiguous; and (ii) we are robust to galaxy/absorber completeness variations in our survey, since we are measuring a clustering excess as a function of scale relative to a random expectation that takes into account the survey selection function. For the latter, it is important to note that a deficit of one population relative to another at a given scale due to a selection bias *will* bias the measurement at those scales if it is not properly modelled and corrected for.

The two-point correlation function, $\xi(r)$, is defined as the probability *excess* of finding two points separated by a distance r with respect to the random expectation.[15] We shall use the correlation function between pairs of galaxies, ξ_{gg} (the galaxy autocorrelation function), between pairs of absorbers, ξ_{aa} (the absorber autocorrelation function) and between absorber-galaxy pairs, ξ_{ag} (the absorber-galaxy cross-correlation function) to gain insights on the relationship between them.

[15] Assumes isotropy, since r in this case is a scalar quantity.

For each field, we assign a central coordinate in right-ascension and declination (α_0, δ_0). Then, for every object in that field with a spectroscopic redshift z, we calculate its position in Cartesian (x, y, z) coordinates as follows:

$$\begin{aligned} x &\equiv r(z)\cos(\Delta\delta)\cos(\Delta\alpha), \\ y &\equiv r(z)\cos(\Delta\delta)\sin(\Delta\alpha), \\ z &\equiv r(z)\sin(\Delta\delta), \end{aligned} \tag{4}$$

where $r(z)$ is the comoving distance, $\Delta\delta \equiv (\delta - \delta_0)$ and $\Delta\alpha \equiv (\alpha - \alpha_0)\cos(\delta_0)$. All of our fields are away from the celestial poles and have small angular coverage, making this transformation accurate. The x coordinate is parallel to the LOS, while the y and z coordinates are perpendicular (transverse) to it, so given that peculiar velocities contribute to the redshifts of objects in addition to cosmological expansion, our coordinate system is subject to distortions, often termed 'redshift–space distortions' (e.g. Kaiser 1987). We therefore measure correlation functions parallel and transverse to the LOS independently, i.e. we measure the two-dimensional two-point correlation function $\xi(r_\perp, r_\parallel)$, where for a given pair of objects denoted i and j, we have

$$\begin{aligned} r_\perp &\equiv \sqrt{|y_i - y_j|^2 + |z_i - z_j|^2}, \\ r_\parallel &\equiv |x_i - x_j|. \end{aligned} \tag{5}$$

Deviations from an isotropic signal in these coordinates can then be attributed to peculiar velocities along the LOS and/or large-scale bulk motions between objects in our sample.

We use the (Landy & Szalay 1993, LS) estimator to calculate the autocorrelation functions:

$$\xi(r_\perp, r_\parallel) = \frac{DD/n_{\rm DD} - 2DR/n_{\rm DR}}{RR/n_{\rm RR}} + 1, \tag{6}$$

where DD, DR and RR are the number of data–data, data–random and random–random pairs, respectively, at a given $r_\perp$ and $r_\parallel$ (or at a given s for the redshift–space correlation function), and the values of n correspond to their normalization factors:

$$\begin{aligned} n_{\rm DD} &= N(N-1)/2, \\ n_{\rm DR} &= \vartheta N^2, \\ n_{\rm RR} &= \vartheta N(\vartheta N - 1)/2, \end{aligned} \tag{7}$$

where N is the total number of real objects, and ϑN is the total number of random ones. We write the normalization factors in this way to highlight that the random samples always have an integer number ϑ times as many objects as the real ones. For the absorber-galaxy cross-correlation function, we adopt the following, generalized form of the LS estimator:

$$\xi_{\rm ag}(r_\perp, r_\parallel) = \frac{D_{\rm a}D_{\rm g}/n_{\rm ag}^{\rm DD} - D_{\rm a}R_{\rm g}/n_{\rm ag}^{\rm DR} - R_{\rm a}D_{\rm g}/n_{\rm ag}^{\rm RD}}{R_{\rm a}R_{\rm g}/n_{\rm ag}^{\rm RR}} + 1 \tag{8}$$

(e.g. Adelberger et al. 2003; Tejos et al. 2014), where $D_{\rm a}D_{\rm g}$, $D_{\rm a}R_{\rm g}$, $R_{\rm a}D_{\rm g}$ and $R_{\rm a}R_{\rm g}$ are the data–data, data–random, random–data and random–random absorber-galaxy pairs, respectively, and their normalization factors, n, are

$$\begin{aligned} n_{\rm ag}^{\rm DD} &= N_{\rm a}N_{\rm g}, \\ n_{\rm ag}^{\rm DR} &= \vartheta_{\rm g}N_{\rm a}N_{\rm g}, \\ n_{\rm ag}^{\rm RD} &= \vartheta_{\rm a}N_{\rm a}N_{\rm g}, \\ n_{\rm ag}^{\rm RR} &= \vartheta_{\rm a}\vartheta_{\rm g}N_{\rm a}N_{\rm g}, \end{aligned} \tag{9}$$

where $N_{\rm a}$ and $N_{\rm g}$ are the total number of absorbers and galaxies, respectively, and $\vartheta_{\rm a}N_{\rm a}$ and $\vartheta_{\rm g}N_{\rm g}$ are the total number of random absorbers and random galaxies, respectively. Landy & Szalay (1993) have shown that these estimators minimize the variance in the correlation function, and so are preferable to other proposed estimators.

A useful quantity, which we also compute, is the projection of the two-dimensional two-point correlation function along the LOS:

$$\Xi(r_\perp) = 2\int_0^\infty {\rm d}r_\parallel\, \xi(r_\perp, r_\parallel). \tag{10}$$

In reality, one only integrates $\xi(r_\perp, r_\parallel)$ up to a finite $r_\parallel$ where the correlation function ceases to be well measured, or where it is consistent with zero. The advantage of calculating this quantity is that it integrates over correlations smeared along the LOS due to peculiar motions, and is insensitive to redshift–space distortions on small transverse scales where bulk flows are not important (Davis & Peebles 1983). We can therefore find a relation between the 'real-space' correlation function (free of distortions), $\xi(r)$, and $\Xi(r_\perp)$ as

$$\begin{aligned} \Xi(r_\perp) &= 2\int_0^\infty {\rm d}r_\parallel\, \xi(r) \\ &= 2\int_{r_\perp}^\infty {\rm d}r\, \xi(r)\frac{r}{\sqrt{r^2 - r_\perp^2}}, \end{aligned} \tag{11}$$

which gives $\xi(r)$ as the inverse Abel transform

$$\xi(r) = -\frac{1}{\pi}\int_r^\infty \frac{{\rm d}\Xi(r_\perp)}{{\rm d}r_\perp}\frac{{\rm d}r_\perp}{\sqrt{r_\perp^2 - r^2}}. \tag{12}$$

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

$$\xi(r) = \left(\frac{r}{r_0}\right)^{-\gamma}, \tag{13}$$

equation (11) yields

$$\Xi(r_\perp) = A(r_0, \gamma)r_\perp^{1-\gamma}, \tag{14}$$

where $A(r_0, \gamma) = r_0^\gamma \Gamma(1/2)\Gamma[(\gamma - 1)/2]/\Gamma(\gamma/2)$, and Γ is the Gamma function. Fitting a power-law form to $\Xi(r_\perp)$ therefore allows determination of r_0 and γ, and hence $\xi(r)$, for $\gamma > 1$. Here, r_0 is usually referred to as the 'correlation length', and γ is the slope of the correlation function.

Given the volume-limited nature of any survey, all estimators are biased towards correlation amplitudes that are lower than the real ones. This arises because the mean density of objects is estimated from the survey itself, and is a well-known bias commonly referred to as the 'integral constraint' (Groth & Peebles 1977). Landy & Szalay (1993) showed that the ξ measured using equation (6) or equation (8) and the real one, $\xi_{\rm real}$, are related as

$$1 + \xi = \frac{1 + \xi_{\rm real}}{1 + \xi_V}, \tag{15}$$

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

$$\xi_V \equiv \int_V {\rm d}^2V\; G(r)\xi_{\rm real}(r). \tag{16}$$

Here, $G(r)$ is the normalized geometric window function, which gives the probability of having two volume elements separated by a distance r for a given survey geometry. For a large enough random catalogue, this is accurately approximated as $G(r) \approx {\rm RR}/n_{\rm RR}$.

Although we cannot know $\xi_{\rm real}$ a priori, we can still estimate the integral constraint by obtaining r_0 and γ from a power-law fit to

$\Xi(r_\perp)$, and taking the mean of the random–random pair counts over all $r_\parallel$ bins at each $r_\perp$ bin to give a proxy for ξ_V as

$$\tilde{\xi}_V = \sum_{r_\perp} \langle \mathrm{RR}/n_{\mathrm{RR}} \rangle A(r_0, \gamma) r_\perp^{1-\gamma}. \quad (17)$$

We can then make a (small) correction to our measured correlation function, ξ', to obtain ξ as follows:

$$\xi = \xi' + \tilde{\xi}_V(1 + \xi'). \quad (18)$$

All of the correlation function measurements that follow have this correction applied.

To interpret our correlation function measurements, we follow Adelberger et al. (2003) and Tejos et al. (2014) in using the Cauchy–Schwarz inequality:

$$\xi_{\mathrm{ag}}^2 \leq \xi_{\mathrm{gg}} \xi_{\mathrm{aa}}. \quad (19)$$

The equality can only hold at any given scale when the density fluctuations that give rise to absorbers and galaxies are linearly dependent. In other words, both populations must trace the same underlying distribution of matter with a linear bias (independent of the scale) to achieve $\xi_{\mathrm{ag}}^2 = \xi_{\mathrm{gg}} \xi_{\mathrm{aa}}$ in general.

We estimate the uncertainty in our correlation function measurements using the bootstrap method, which in our experience provides the most conservative measure of the uncertainty (see Tejos et al. 2014, for a discussion). We do this by creating $N_{\mathrm{bs}} = 1000$ sets of 27 fields, randomly chosen (with replacement) from our set of 27 fields, and compute the uncertainty as

$$\Delta^2(\xi) = \frac{1}{N_{\mathrm{bs}}} \sum_i^{N_{\mathrm{bs}}} (\xi_i - \bar{\xi})^2, \quad (20)$$

where ξ_i is the correlation function measured from the ith random set of fields, and $\bar{\xi}$ is the mean of these measurements. Performing bootstrap realizations over fields ensures that we capture the sample variance, as well as the statistical uncertainty in the measurement.

The bootstrap uncertainty estimation is clearly appropriate for our observational sample, for which we have a large number of independent fields, but is not so easily applied to the simulated sample we have assembled, which is drawn from a single cubic volume, 100 comoving Mpc on a side (see Section 5 for details). To apply this uncertainty estimator, we would have to break the simulation down into subvolumes, which limits the scales on which we can measure the correlation functions. For the simulated samples, we therefore quantify the statistical uncertainty on the measurement via the approximate estimator presented in Landy & Szalay (1993):

$$\Delta_{\mathrm{LS}}^2(\xi) \approx \frac{(1+\xi)^2}{n_{\mathrm{DD}}(RR/n_{\mathrm{RR}})} \approx \frac{(1+\xi)^3}{DD}. \quad (21)$$

Note that this is greater than the commonly used Poissonian estimator, $\Delta_{\mathrm{DD}}^2(\xi) = (1+\xi)/\mathrm{DD}$, by a factor of $\sim(1+\xi)^2$, since it takes into account correlations introduced by non-independent cross-pairs. In practice, we are able to achieve negligibly small statistical uncertainties on the measurements from the simulated data, but we note that the sample variance (which may be much larger than the statistical one) is not taken into account.

6.2 Random samples

As is clear from the previous section, the construction of random samples is a crucial part of any correlation function analysis. These random samples need to capture the selection functions in the data, so as not to bias the measurement. We present in the following sections a detailed description of the method for generating random samples of observed and simulated O VI absorbers and galaxies.

6.2.1 *Random galaxy catalogues*

For our observational sample, we create random galaxies for each field and survey independently. This means that individual fields containing galaxies from multiple surveys have separate random catalogues constructed for each of those surveys that are then combined at the end to form the random sample for that field. We do this since the different surveys in our sample have different selection functions (see Table 2), and it is easier to model these separately, rather than attempting to model the combined selection functions. We note that the T14-Q0107 survey listed in Table 2 has a complex selection function, since it combines four subsurveys constructed with different instruments, each with their own specific selection biases. Ideally, we would further split this survey down into its constituent parts, but the number of galaxies attributable to each of the different instruments is too small to reliably model their individual selection functions.

Our process for creating random galaxy catalogues for the observational sample expands upon that described in Tejos et al. (2014), and works as follows. For a given galaxy, in a given field, and from a given survey, we create $\vartheta_{\mathrm{g}} = 10$ random ones, varying the redshift of the galaxy, but preserving its position on the sky and all of its other properties. The random redshifts are drawn from a PDF that is modelled on the observed redshift histogram for galaxies with matching properties in the survey from which the real redshift was obtained. We take into account the observed magnitude of the galaxy (in the broad-band filter that defines the magnitude limit for the survey), and whether it is a star-forming galaxy, a non-star-forming galaxy, or neither of these. For example, if a galaxy in a particular field is drawn from the SDSS, and is a star-forming galaxy with an r-band magnitude $r = 16$, we randomly draw $\vartheta_{\mathrm{g}} = 10$ redshifts from a PDF that is modelled on the redshift histogram in SDSS for star-forming galaxies with that magnitude. In this way, our random samples reflect the individual survey sensitivity functions for galaxies of a given magnitude and spectral type, and also the evolution in the star-forming fraction. For galaxies that have no classification of spectral type, or are classified as AGN, we use the redshift histogram of *all* galaxies from the same survey, and with the same magnitude as that galaxy. Our approach guarantees that we take into account the survey incompleteness in the construction of the random catalogues.

The probability density distributions described above are constructed in the following way. First, we optimally bin redshift distributions using the algorithm presented by Knuth (2006). This is a maximum likelihood method for determining the optimum number of bins needed to both capture the dominant features in the data and minimize the number of random sampling fluctuations. We create histograms in this way for star-forming and non-star-forming galaxies separately, in magnitude bins of size 1, shifted by 0.5 mag, over the range $r = 13$–25. We iteratively increase the magnitude bin sizes at the bright and faint ends of the magnitude distributions to ensure that there are a minimum of 20 galaxies in each redshift histogram. We then smooth the histograms with a Gaussian smoothing kernel having a standard deviation equal to the bin size, to remove spikes in the redshift distributions attributable to large-scale structure (galaxy clusters, filaments, sheets and voids). The probability density distributions are then obtained by interpolating a cubic spline over the

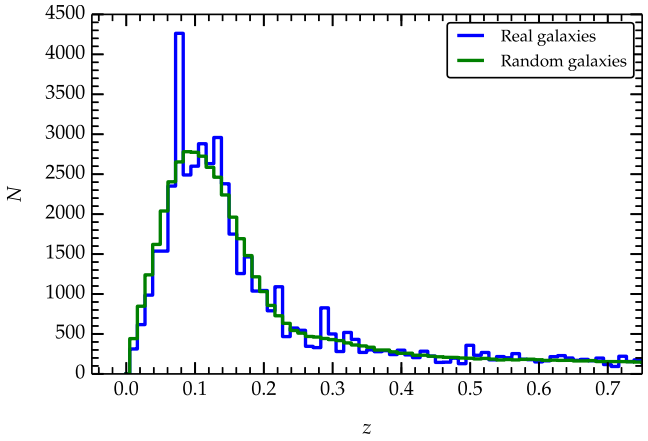

Figure 6. The real galaxy redshift distribution for our total sample (blue), and the random galaxy redshift distribution divided by ϑ_g (green). The smooth shape of the random distribution, following the broad shape of the real distribution, suggests that our approach to generating random catalogues of galaxies is robust.

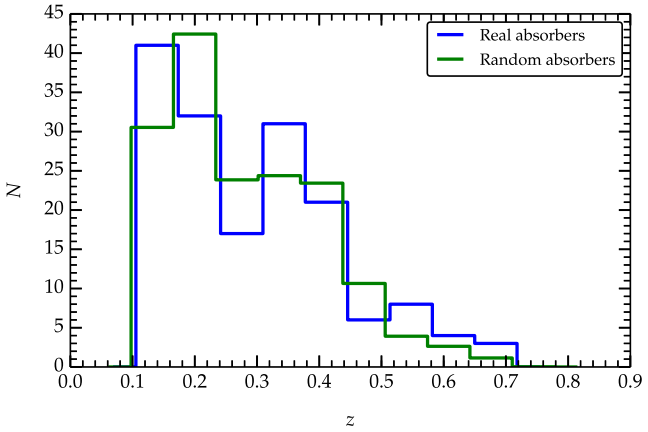

Figure 7. The real O VI redshift distribution for our total sample (blue), and the random O VI redshift distribution divided by ϑ_a (green). The random redshift distribution follows the same selection bias as the data, which suggests our procedure for generating random absorber catalogues is robust.

smoothed histograms, and by normalizing to the area underneath the resulting curve.

In Fig. 6, we plot a histogram showing the redshift distribution of our total galaxy sample in blue, and the random sample divided by ϑ_g in green. The green histogram shows a smooth distribution, reflective of the overall selection function for our sample. Spikes are apparent in the real histogram attributable to large-scale structure. The smooth redshift distribution apparent in our random sample, following the overall shape of the real histogram, suggests that our approach is robust.

Random galaxies for our sample extracted from EAGLE simulation volume are distributed uniformly for each of the redshift snapshots. We created $\vartheta_g = 10$ times as many random galaxies as there are real ones.

6.2.2 *Random absorber catalogues*

For our observed O VI sample, which derives entirely from COS FUV spectra, we created $\vartheta_a = 1000$ random absorbers for every real one, varying the redshift, but preserving all other parameters. In this way, we randomize absorbers only along the QSO sight-lines, not transverse to them, so as to preserve the geometry of our survey. Random redshifts were chosen on the basis of an equivalent width threshold. For every real O VI absorber with Doppler broadening parameter b, and equivalent width W, we calculated the minimum equivalent width, $W_{3\sigma}$, at which the weaker transition in the doublet (O VI λ1037) for that absorber could still be observed above the required 3σ significance threshold as a function of wavelength in the spectrum from which that absorber was obtained. The significance of absorption features in COS spectra cannot be estimated in the usual way due to the non-random noise properties of the instrument (Keeney et al. 2012). Nevertheless, Keeney et al. (2012) provide a formalism for doing this, which we adopt (see their equations 4, 5, 7, 9 and 10). We then transformed wavelength coordinates to redshift coordinates to obtain $W_{3\sigma}$ as a function of z, and distributed random absorbers in z where the condition $W > W_{3\sigma}(z)$ was satisfied. We enforced a maximum redshift equivalent to a -5000 km s^{-1} offset from the QSO redshift, as was done in the data.

In Fig. 7, we show the redshift distribution of O VI absorbers for our full sample in blue in comparison to the random sample divided by ϑ_a in green. A larger number of O VI absorbers are found at lower redshifts, where they are detected in the wavelength range covered by the COS G130M grating. COS has greater sensitivity over this wavelength range, and the spectra from this grating therefore typically have a higher SNR. The shape of the random distribution follows this closely, which indicates our procedure is robust.

For our simulated set of O VI absorbers extracted from the EAGLE simulation, we randomize absorber velocities uniformly along the sight-lines from which they are extracted. We created $\vartheta_a = 10$ times as many random absorbers as there are real ones.

7 RESULTS

In this section, we present the results of our two-point correlation analysis following the mathematical formalism outlined in Section 6.1. The results that follow were computed using the random samples described in the previous section. We present results for our full sample of O VI absorbers and galaxies, and also for the subsamples containing only star-forming and non-star-forming galaxies.

7.1 2D two-point correlation functions

In Fig. 8, we show the two-dimensional correlation functions (top panels) and their uncertainties (bottom panels) for our full sample of O VI absorbers and galaxies. The results are shown in bins of 1 Mpc (comoving) and are derived from pair counts smoothed with a Gaussian kernel having a standard deviation of 1 Mpc. The use of a smoothing kernel strikes a compromise between lowering the shot noise in the measurement, whilst keeping a relatively small bin size. From left to right, the panels show the O VI-galaxy cross-correlation function (ξ_{ag}), the galaxy autocorrelation function (ξ_{gg}), the O VI autocorrelation function (ξ_{aa}) and the ratio $\xi_{ag}^2/\xi_{gg}\xi_{aa}$. Throughout this section, it is important to bear in mind that correlations along the LOS in ξ_{aa} are subject to the often somewhat subjective decomposition of O VI absorption complexes into multiple absorption 'components' (see Section 3.3 for a description of our line-fitting approach).

On inspection of the figure, we see that on small scales, the amplitudes of ξ_{ag}, ξ_{gg} and ξ_{aa} are comparable within the uncertainties. Note that we are unable to probe ξ_{aa} on transverse scales >2 Mpc with our data. The ratio $\xi_{ag}^2/\xi_{gg}\xi_{aa}$ is close to 1 on scales $\lesssim$1 Mpc,

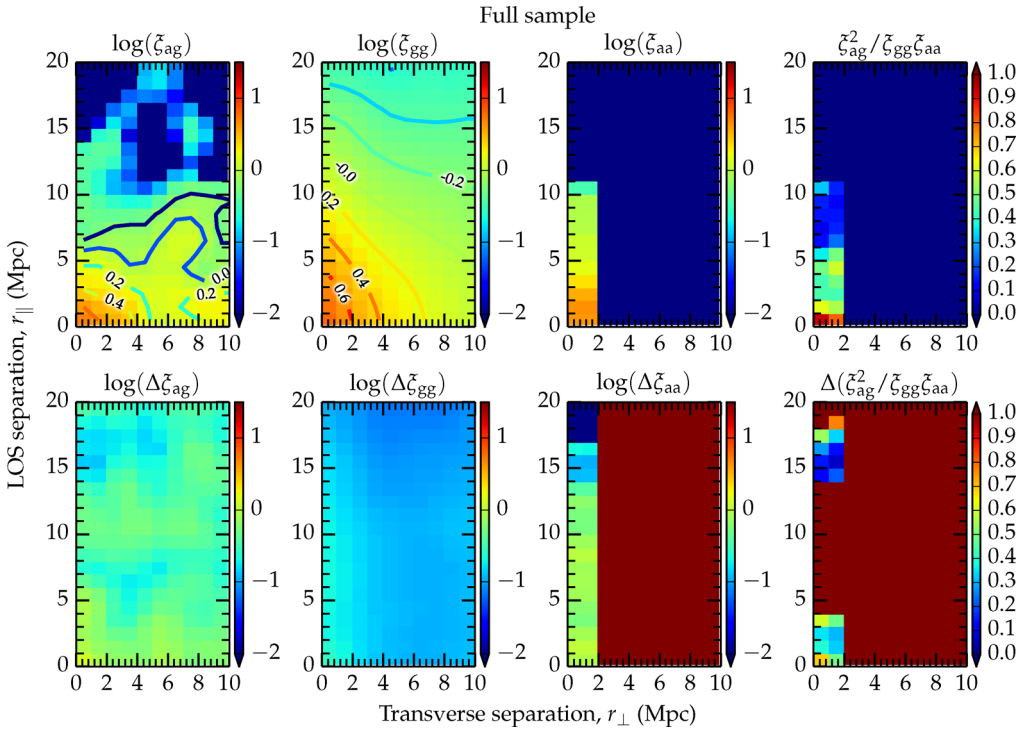

Figure 8. Two-dimensional correlation functions for galaxies and O VI absorption systems (top panels) and their respective uncertainties (bottom panels), as a function of comoving separation parallel ($r_\parallel$) and perpendicular ($r_\perp$) to the LOS. From left to right: the galaxy-O VI cross-correlation function ($\xi_{\rm ag}$), the galaxy autocorrelation function ($\xi_{\rm gg}$), the O VI autocorrelation function ($\xi_{\rm aa}$), and the ratio $\xi^2_{\rm ag}/\xi_{\rm gg}\xi_{\rm aa}$. Note that our data are not suitable for measuring $\xi_{\rm aa}$ and $\xi^2_{\rm ag}/\xi_{\rm gg}\xi_{\rm aa}$ on transverse scales >2 Mpc. The correlation functions are calculated using a bin size of 1 Mpc, and the pair-counts are smoothed with a Gaussian kernel of standard deviation 1 Mpc in both directions.

which suggests that O VI absorbers and galaxies are in close correspondence with one another on these scales. Our data do not have sufficient statistical power to quantify the presence of anisotropies in $\xi_{\rm ag}$ or $\xi_{\rm aa}$, but for $\xi_{\rm ag}$ we can nevertheless examine the 'isocorrelation' contours shown in these plots. We see a reasonably isotropic signal in $\xi_{\rm ag}$, at least on the small scales where our measurement is stronger, which indicates that O VI absorbers show little velocity dispersion with respect to the galaxies. There is a hint of some compression in the signal along the LOS on large scales, indicative of large-scale bulk motions (e.g. Kaiser 1987), but not at a statistically significant level. Deviations from an isotropic signal are present in $\xi_{\rm gg}$, which we expect for a galaxy sample of the size presented here, where a significant fraction of the galaxies reside in groups and clusters, with velocity dispersions of several 100 km s^{-1}. We note that the deviation from isotropy in $\xi_{\rm gg}$ complicates the comparison of clustering amplitudes between $\xi_{\rm ag}$ and $\xi_{\rm gg}$ (and our inferences based on $\xi^2_{\rm ag}/\xi_{\rm gg}\xi_{\rm aa}$ above), since the effect of the redshift–space distortions is to 'smear' the total correlation amplitude at a given $r_\perp$ over a range in $r_\parallel$. The total correlation amplitudes at a given $r_\perp$ are therefore not necessarily as comparable as they appear in Fig. 8. We investigate this further in the next section.

In Fig. 9, we present the same calculation as was illustrated in Fig. 8, but this time for the samples extracted from the EAGLE simulation volume. For this, and all the comparisons that follow, we present results from the $z = 0.271$ EAGLE snapshot, which is roughly the median redshift of our O VI sample. Note that for the EAGLE calculations we do not apply a Gaussian smoothing kernel. Much like in the real case, the amplitudes of $\xi_{\rm ag}$, $\xi_{\rm gg}$ and $\xi_{\rm aa}$ are all very similar at small scales within the uncertainties. We note that the correlation amplitudes in EAGLE are somewhat higher than in the data at these scales, although the Gaussian smoothing kernel employed in the latter does act to lower the correlation amplitude at small separations where they are intrinsically peaked. Inspection of an unsmoothed version of Fig. 8 reveals that the correlation amplitudes on the smallest scales are in fact comparable to those in EAGLE within the uncertainties. Also, much like in the data, there is very little anisotropy on small scales in $\xi_{\rm ag}$. Even without model-fitting, it is clear that the 'anisotropy ratio' along the LOS on <4 Mpc scales is no more than 2:1, which limits the velocity dispersion of O VI around galaxies to $\lesssim$100 km s^{-1}. A highly isotropic signal is seen in $\xi_{\rm aa}$ on <4 Mpc scales as well, which suggests that the O VI absorbers are virtually static with respect to one another on these scales. Intriguingly, the hint of a compression along the LOS in $\xi_{\rm ag}$, seen in the real data on large scales, appears in the simulated sample with high significance. The same is seen in $\xi_{\rm aa}$. This then points to a picture in which O VI absorbers show bulk motions towards both galaxies and themselves on ∼5–10 Mpc scales. Given the low significance of this result in the real data, we caution that this finding is far from conclusive. We note that there are some differences in $\xi_{\rm gg}$ between the data and the simulation. In particular, there is a larger anisotropy in the signal along the LOS in EAGLE compared to the data. However, we note that a good agreement between observations and simulations is not necessarily expected for $\xi_{\rm gg}$, since it depends on the fraction of passive galaxies. Observational studies are biased against selecting low-mass passive galaxies due to the difficulty in assigning a redshift, whereas galaxies of this type will always be present in the simulated samples. Since the galaxy autocorrelation functions are not the primary focus of this study, and we shall leave a more detailed comparison to future work.

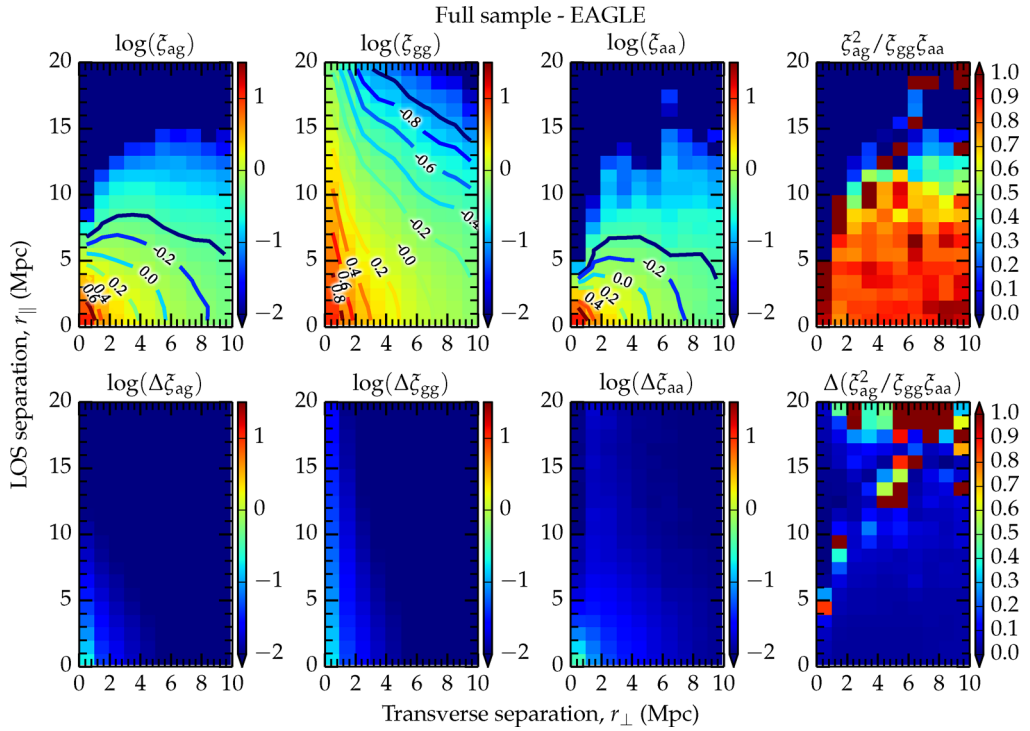

Figure 9. Same as Fig. 8, but for the simulated samples extracted from the EAGLE simulation. Note that for the simulated samples we do not apply a Gaussian smoothing kernel.

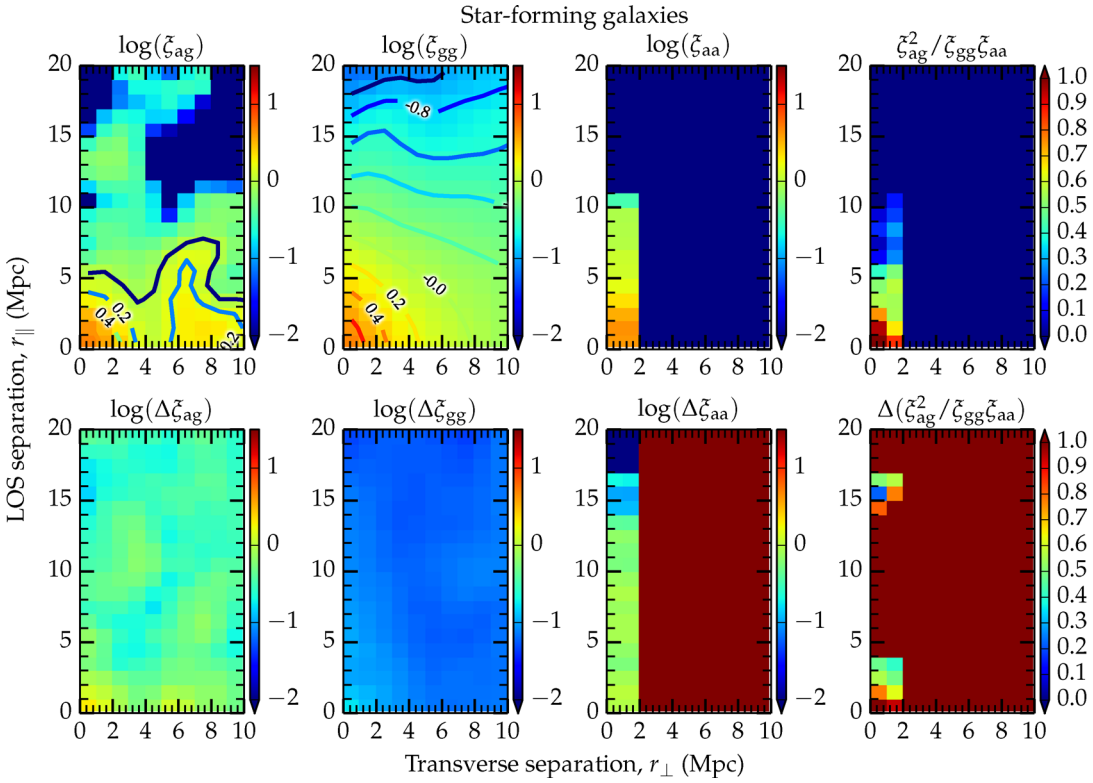

Figure 10. Same as Fig. 8, but for star-forming galaxies only.

Next, we examine the 2D two-point correlation functions for O VI absorbers and star-forming galaxies only. For the real data, this calculation is shown in Fig. 10. Again, we see that ξ_{ag}, ξ_{gg} and ξ_{aa} are all comparable within the uncertainties. This leads to a ratio $\xi^2_{ag}/\xi_{gg}\xi_{aa}$ that is consistent with 1 on small scales. We see no significant anisotropies in ξ_{ag} either, which suggests that on small scales, O VI absorbers and star-forming galaxies trace the same underlying distribution of matter, and show very little velocity

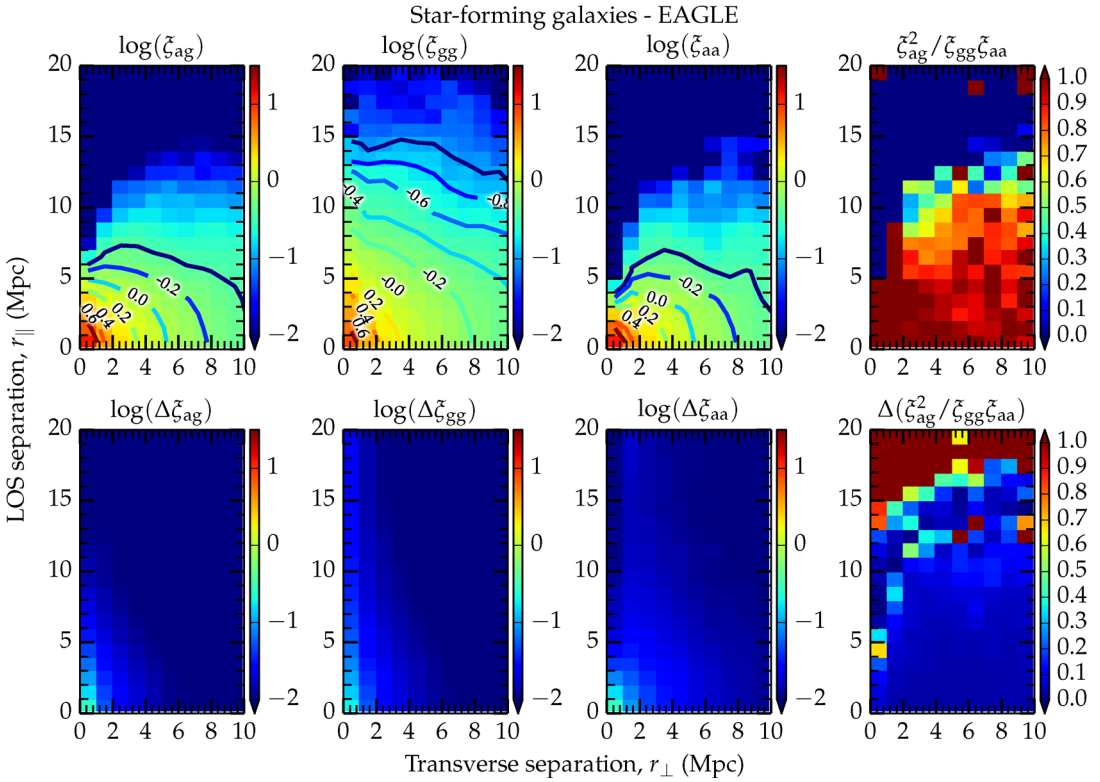

Figure 11. Same as Fig. 10, but for the simulated samples extracted from the EAGLE simulation.

dispersion with respect to one another (certainly $\lesssim$100 km s^{-1}). Redshift–space distortions in ξ_{gg} are less evident for star-forming galaxies, and are qualitatively similar to those in ξ_{ag} and ξ_{aa}, which makes our findings based on $\xi^2_{ag}/\xi_{gg}\xi_{aa}$ somewhat more robust than those from the full sample.

In Fig. 11, we show the same calculation for the EAGLE simulation, again with no Gaussian smoothing kernel applied. In agreement with the real data, ξ_{ag}, ξ_{gg} and ξ_{aa} are all consistent within the uncertainties, and the ratio $\xi^2_{ag}/\xi_{gg}\xi_{aa}$ is consistent with 1. We see the same compression in ξ_{ag} and ξ_{aa} along the LOS on large scales as was seen for the full sample. However, this is not detected in the real data with significance. A small anisotropy is seen in ξ_{ag}, which amounts to a velocity dispersion between O VI absorbers and star-forming galaxies of no more than $\sim$100 km s^{-1}.

Finally, we examine the 2D two-point correlation functions for O VI absorbers and non-star-forming galaxies only. We show this calculation for the real data in Fig. 12. Now we see that the correlation amplitudes of ξ_{ag} and ξ_{aa} are both comparable within the uncertainties, but the amplitude of ξ_{gg} is significantly higher than both of these on small scales. As a result, we see the ratio $\xi^2_{ag}/\xi_{gg}\xi_{aa}$ is nearly consistent with zero. This suggests that O VI absorbers and non-star-forming galaxies trace the underlying distribution of matter differently, although in this case we note that the interpretation of $\xi^2_{ag}/\xi_{gg}\xi_{aa}$ is complicated by redshift–space distortions that are highly prominent in ξ_{gg}. Again, we see no evidence for any anisotropy in ξ_{ag} along the LOS, which implies that O VI absorbers show very little velocity dispersion with respect to non-star-forming galaxies.

The corresponding calculation for the EAGLE simulation is shown in Fig. 13. We see a very similar situation to the data, whereby ξ_{ag} and ξ_{aa} are both similar within the uncertainties, but the amplitude of ξ_{gg} is significantly higher than both of these. This leads to a ratio $\xi^2_{ag}/\xi_{gg}\xi_{aa}$ that is nearly zero on the smallest scales. However, we note that in both the simulation and the data, the clustering amplitudes of O VI absorbers around non-star-forming galaxies are highly comparable to those around star-forming galaxies. This suggests that the likelihood of finding an O VI absorber close to a non-star-forming galaxy should be similar to the likelihood of finding an O VI absorber close to a star-forming galaxy over the scales probed. As is consistently seen in the simulation, there does exist a small anisotropy in ξ_{ag} along the LOS, but only at the level whereby the velocity dispersion between O VI absorbers and non-star-forming galaxies is $\lesssim$100 km s^{-1}, which is broadly consistent with the real data within the uncertainties.

7.2 Correlation functions projected along the LOS

Comparisons of the clustering amplitudes between ξ_{ag}, ξ_{gg} and ξ_{aa} for the 2D two-point correlation functions in the previous section are complicated by the redshift–space distortions that lead to anisotropies in the signal. To better-compare the clustering amplitudes as a function of scale, we now examine the correlation functions that are projected along the LOS, as in equation (11).

In Figs 14–16, we show the correlation functions of O VI absorbers and galaxies projected along the LOS and divided by $r_\perp$, $\Xi(r_\perp)/r_\perp$, for our full sample, and for star-forming and non-star-forming galaxies only. We show the projected O VI–galaxy cross-correlation function, Ξ_{ag}, in blue data points, the projected galaxy autocorrelation function, Ξ_{gg}, in green data points and the projected O VI autocorrelation function, Ξ_{aa}, in red data points. For Ξ_{gg} we integrate to $r_\parallel = 45$ Mpc, and for Ξ_{ag} and Ξ_{aa} we integrate to $r_\parallel = 13$ Mpc. These integration limits are the minimum for which the data points had converged to stable values, indicating that we are fully integrating the reliably measured signal, and minimizing the

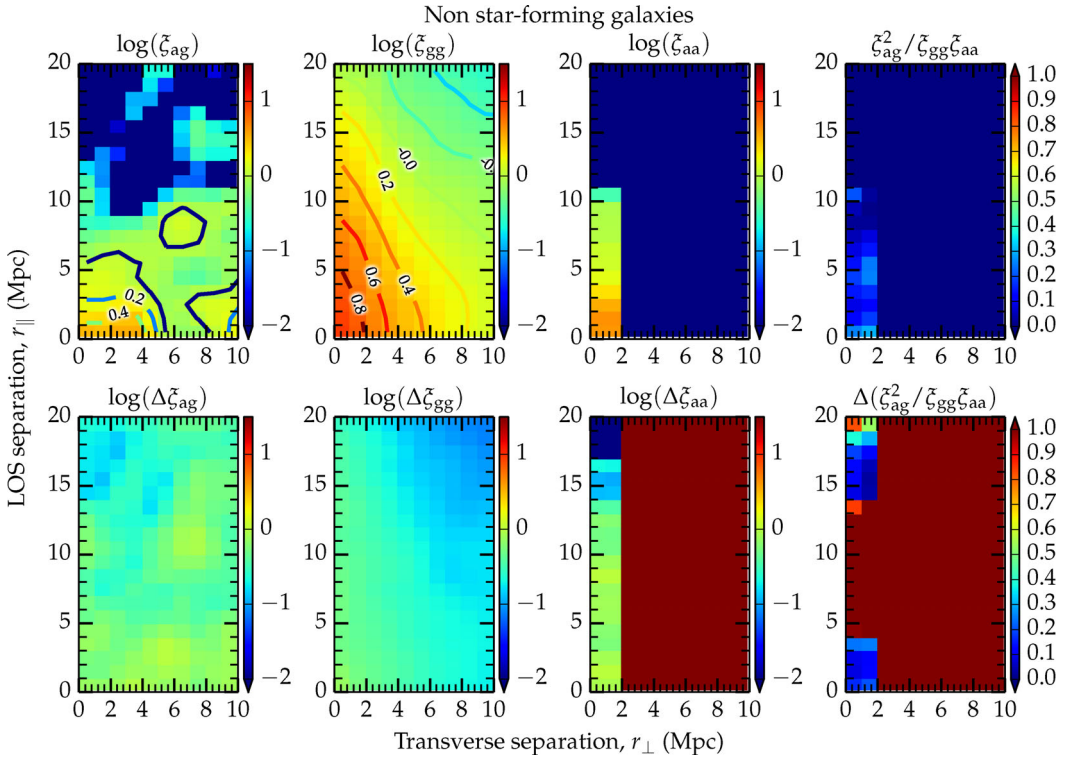

Figure 12. Same as Fig. 8, but for non-star-forming galaxies only.

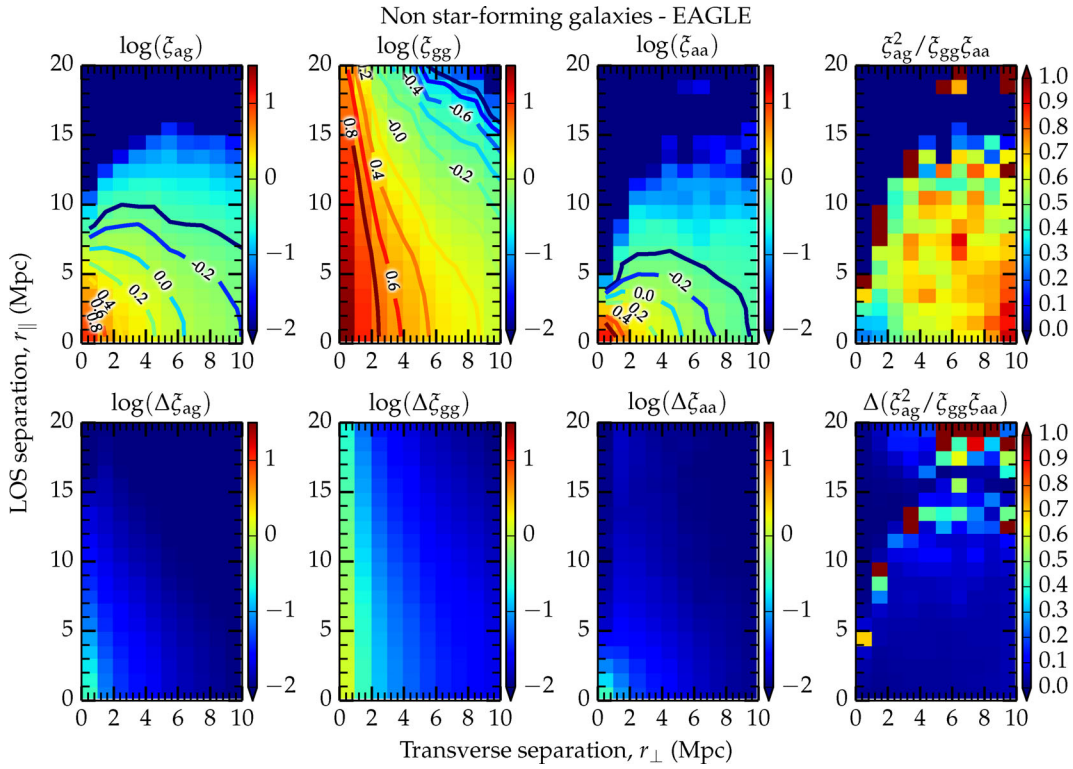

Figure 13. Same as Fig. 12, but for the simulated samples extracted from the EAGLE simulation.

addition of shot noise. We note that the data points are correlated, and that uncertainties smaller than the data points are not shown. The error bars show 1σ bootstrap uncertainties that include both the variance and covariance in the measurement. The lines show the best-fitting power laws to the data (equation 14) using the same colour scheme. We only fit these power laws to the data at $r_\perp < 10$ Mpc, as deviations from a power law are typically seen at $r_\perp > 10$ Mpc. We do not attempt to fit a power-law slope to Ξ_{aa},

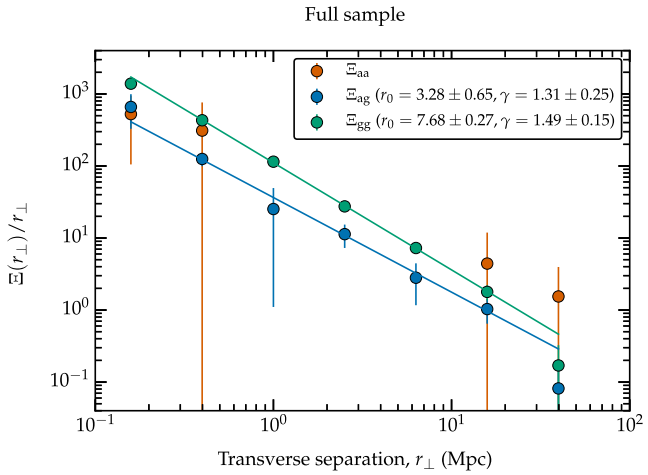

Figure 14. Correlation functions projected along the LOS and divided by $r_\perp$, $\Xi(r_\perp)/r_\perp$, for our full sample of O VI absorbers and galaxies. Our measurement for the projected O VI–galaxy cross-correlation function, Ξ_{ag}, is shown by the blue data points. Green data points show the measurement for the projected galaxy autocorrelation function, Ξ_{gg}, and red data points show the measurement for the projected O VI autocorrelation function, Ξ_{aa}. The lines correspond to the best power-law fits (equation 14) to Ξ_{ag}, Ξ_{gg} and Ξ_{aa} in blue, green and red, respectively. Note that the data points and their uncertainties are correlated. Uncertainties smaller than the symbols are not shown.

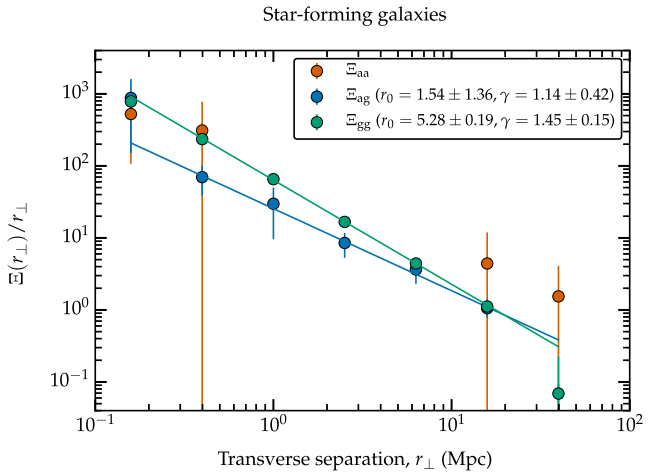

Figure 15. Same as Fig. 14, but for star-forming galaxies only.

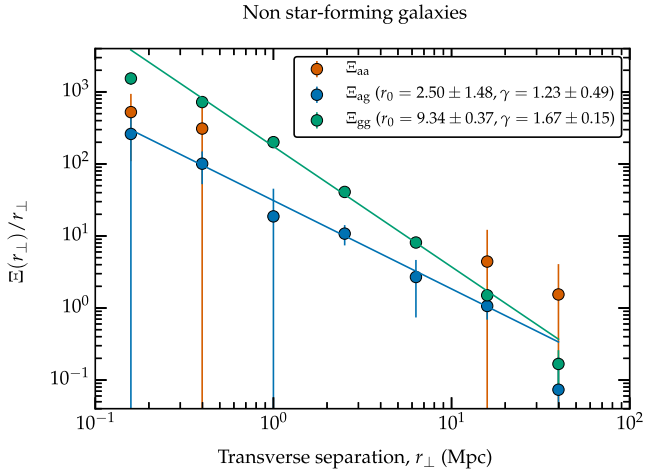

Figure 16. Same as Fig. 14, but for non-star-forming galaxies only.

as these data show very low statistical significance and are likely not robust. It is important to note that while the measurement of ξ_{aa} in the 2D two-point correlation function is dominated by O VI pairs along the LOS, these pairs do not contribute to the measurements in Figs 14–16, which instead come from transverse pairs in the very few closely separated QSO sight-line pairs (and triplets) in our sample. The best-fitting parameters for all of the projected correlation functions are summarized in Table 5.

Starting with Fig. 14, which shows projected correlation functions for our full sample, we note that a power law is a good description of the data for Ξ_{ag} and Ξ_{gg} at $r_\perp < 10$ Mpc. We find that $\xi_{ag}(r)$ has a correlation length of $r_0^{ag} = 3.28 \pm 0.65$ Mpc, and a slope of $\gamma^{ag} = 1.31 \pm 0.24$, whereas $\xi_{gg}(r)$ has a correlation length of $r_0^{gg} = 7.68 \pm 0.27$ Mpc, and a slope of $\gamma^{gg} = 1.49 \pm 0.15$. We therefore find that the correlation length of O VI absorbers around galaxies is less than the correlation length of galaxies with themselves, and this finding is significant at a $>4\sigma$ confidence level. The data also suggests that the slope of Ξ_{ag} is shallower than that of Ξ_{gg}, although they are consistent with one another within the 1σ uncertainties. The amplitude of Ξ_{aa}, and the apparent slope, is consistent with both Ξ_{ag} and Ξ_{gg}. The differing slopes between Ξ_{ag} and Ξ_{gg}, combined with the difference in their correlation lengths, suggests that O VI absorbers and galaxies may not linearly trace the same underlying distribution of matter in general, however this is by no means definitive given the uncertainties on the data, in particular for Ξ_{aa}.

As we did for the 2D two-point correlation functions, we now examine the correlation functions with star-forming and non-star-forming galaxies separately. We show the projected correlation functions with star-forming galaxies only in Fig. 15. Here, we find that $\xi_{ag}(r)$ has a correlation length of $r_0^{ag} = 1.55 \pm 0.35$ Mpc, and a slope of $\gamma^{ag} = 1.14 \pm 0.41$, whereas $\xi_{gg}(r)$ has a correlation length of $r_0^{gg} = 5.28 \pm 0.21$ Mpc, and a slope of $\gamma^{gg} = 1.45 \pm 0.15$. The correlation length of O VI absorbers around star-forming galaxies is therefore still less than the correlation length of star-forming galaxies with themselves, although at a lower significance level ($>2\sigma$), and there remains marginal evidence for a difference in slope between Ξ_{ag} and Ξ_{gg}, although this result is not significant ($<1\sigma$). We therefore find indications that O VI absorbers and star-forming galaxies may not linearly trace the same underlying distribution of matter, similar to the full sample, even though the clustering amplitudes in the cross-correlation function do indicate that O VI absorbers have a strong association with star-forming galaxies in general.

Finally, in Fig. 16, we show the projected correlation functions with non-star-forming galaxies only. Here, we find that $\xi_{ag}(r)$ has a correlation length of $r_0^{ag} = 2.63 \pm 1.56$ Mpc, and a slope of $\gamma^{ag} = 1.22 \pm 0.51$, whereas $\xi_{gg}(r)$ has a correlation length of $r_0^{gg} = 9.35 \pm 0.38$ Mpc, and a slope of $\gamma^{gg} = 1.67 \pm 0.15$. We therefore find that the correlation length of O VI absorbers around non-star-forming galaxies is less than the correlation length of non-star-forming galaxies with themselves at a $>3\sigma$ significance level. The slopes of the correlation functions are still consistent with each other within the uncertainties, but the data nevertheless suggests that the slope in Ξ_{gg} is steeper than that of Ξ_{ag}. This supports indications from Fig. 12 that O VI absorbers and non-star-forming galaxies do not linearly trace the same underlying distribution of matter.

It is well known that non-star-forming galaxies are more biased tracers of matter than are non-star-forming galaxies, as is clearly seen from the difference in the correlation lengths and slopes of their autocorrelation functions. However, the most striking result is the similarity in the correlation lengths and slopes of the cross-correlation functions of O VI absorbers with star-forming and

Table 5. Best-fitting projected correlation function parameters.

Sample	$r_0^{\rm ag}$ (Mpc)	$\gamma^{\rm ag}$	$r_0^{\rm gg}$ (Mpc)	$\gamma^{\rm gg}$
Full sample	3.28 ± 0.65	1.31 ± 0.25	7.68 ± 0.27	1.49 ± 0.15
Star-forming galaxies	1.54 ± 1.36	1.14 ± 0.42	5.28 ± 0.19	1.45 ± 0.15
Non-star-forming galaxies	2.50 ± 1.48	1.23 ± 0.49	9.34 ± 0.37	1.67 ± 0.15

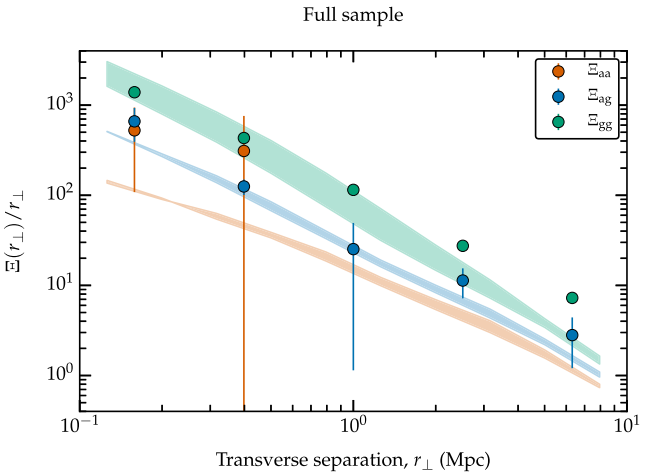

Figure 17. Same as Fig. 14, but now comparing the data to the predicted slopes and amplitudes of the correlation functions from the EAGLE simulation. The shaded regions represent the predictions over the redshift range $0.1 < z < 0.7$, and adopt the same colour scheme as the data.

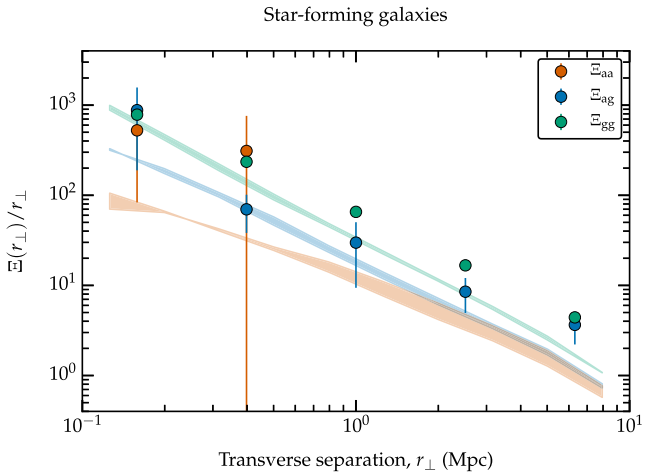

Figure 18. Same as Fig. 17, but for star-forming galaxies only.

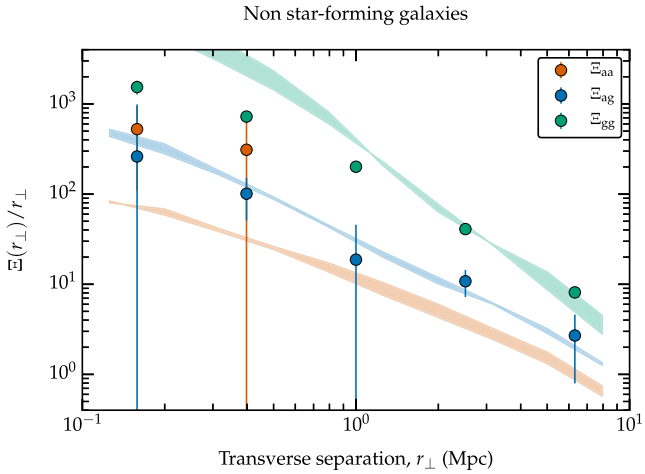

Figure 19. Same as Fig. 17, but for non-star-forming galaxies only.

non-star-forming galaxies. These slopes and amplitudes are entirely consistent with one another within the uncertainties, and this demonstrates that the likelihood of finding O VI absorbers around star-forming galaxies is similar to the likelihood of finding O VI absorbers around non-star-forming galaxies, at least over the scales probed by our data (to separations as close as ∼100 kpc).

We now compare the correlation functions projected along the LOS to predictions from the EAGLE simulation. In Fig. 17, we show the same measurements as in Fig. 14 for our full sample, with the predictions from EAGLE over the redshift range $0.1 < z < 0.7$ shown as shaded regions with the same colour scheme as the data. We see a reasonable agreement in $\Xi_{\rm ag}$ between the real data and that from the simulation given the uncertainties. The simulation also produces predictions for $\Xi_{\rm aa}$ that are consistent with the data. We also find a reasonable agreement in $\Xi_{\rm gg}$ at $<$ Mpc scales, but this agreement ceases at larger scales due to a discernible difference in the slopes of the correlation functions. As highlighted previously, it is likely that differences between the observed and simulated galaxy autocorrelation functions can be explained in part due to the bias against low-mass non-star-forming galaxies in the observational sample, owing to the difficulty in assigning redshifts to these galaxies. The relative slopes between $\Xi_{\rm gg}$ and $\Xi_{\rm ag}$ predicted by the EAGLE simulation reveal the same trend suggested by the data, whereby the slope of $\Xi_{\rm ag}$ is shallower than that of $\Xi_{\rm gg}$. It can also be seen that the slope of $\Xi_{\rm aa}$ is shallower still. This indicates that O VI absorbers and galaxies in the EAGLE simulation do not linearly trace the same underlying distribution of matter.

In Fig. 18, we show the same comparison but for star-forming galaxies only. There is again a reasonable agreement between the predicted and measured $\Xi_{\rm ag}$ within the uncertainties, albeit with a discernible difference in the slopes of the correlation functions. A good agreement is seen in Fig. 19 for the slope and amplitude of the cross-correlation function of O VI absorbers with non-star-forming galaxies. We also see that the amplitudes in $\Xi_{\rm ag}$ are in general lower than those in $\Xi_{\rm gg}$, and that this is more pronounced for non-star-forming galaxies, for which $\Xi_{\rm gg}$ is a more biased tracer of the underlying distribution of matter. Furthermore, the predicted relative slopes in $\Xi_{\rm ag}$ and $\Xi_{\rm gg}$ show the same trends suggested by the data, whereby $\Xi_{\rm ag}$ is shallower than $\Xi_{\rm gg}$ for both star-forming and non-star-forming galaxies, with this effect being more pronounced in the case of the latter. The slope in $\Xi_{\rm aa}$ is also consistently shallower than both of these. We can therefore infer in the simulation that O VI absorbers and galaxies do not linearly trace the same underlying distribution of matter, and that this is most markedly the case for non-star-forming galaxies. In addition, the likelihood of

finding O VI absorbers close to star-forming galaxies is similar to the likelihood of finding O VI absorbers close to non-star-forming galaxies over the scales probed. These inferences are all entirely consistent with those that can be drawn from the data.

Finally, it is important to note that all of our comparisons with the EAGLE simulation are subject to the caveats highlighted in Section 5.2, whereby the statistical properties of the simulated and observed absorbers differ. As a result, the simulated absorbers required a systematic +0.3 dex shift in their column densities as a matter of procedure. Nevertheless, bearing this caveat in mind, and despite the disparity in their statistical properties, the clustering and dynamical properties of absorbers with galaxies in the EAGLE simulation show encouraging qualitative agreement with the data.

8 DISCUSSION

In light of these new results on the cross-correlation functions of O VI absorbers and galaxies, we now explore some possible interpretations, and make comparisons to similar studies in the literature.

8.1 Comparison with previous results

Chen & Mulchaey (2009) performed a very similar study to the one presented here, in which they measured the two-point cross-correlation function of O VI absorbers and galaxies projected along the LOS at $z < 1$. We note that all of the galaxy data used in that study form a small subset of the galaxy data used here, whilst their absorption-line data from the Space Telescope Imaging Spectrograph (STIS) and the Far-Ultraviolet Spectroscopic Explorer (FUSE) have been updated to that available from COS. We find that our results contrast with these earlier results in two primary aspects: (i) Chen & Mulchaey (2009) find that the clustering amplitudes of O VI absorbers around star-forming galaxies are comparable to those of star-forming galaxies with themselves, whilst we find that in general they are smaller; and (ii) Chen & Mulchaey (2009) find that the clustering amplitudes of O VI absorbers around star-forming galaxies are weaker than those around non-star-forming galaxies, whilst we find that they are comparable. To explain these differences, it is important to note the sample sizes. Chen & Mulchaey (2009) used a sample of 13 O VI absorbers and 670 galaxies, which is substantially smaller than the sample assembled for this study. The quoted uncertainties on their measurements are also Poissonian, which underestimates the true uncertainties (see Section 7.1). We therefore believe that our measurements are more statistically robust, and that with more conservative estimates on the uncertainties, the results of these two studies may in fact be consistent.

Turner et al. (2014) examined the distribution of O VI, H I, and other metal ions around star-forming galaxies at $z \approx 2.4$. Their study uses a different technique to ours, in which they measure the median optical depth of O VI in spectral pixels binned in terms of their transverse and LOS separation from galaxies. Their study is optimized to investigate the association between absorbers and star-forming high-redshift galaxies on small (<2 Mpc) scales, whereas ours is optimized to investigate the association between absorbers and galaxies at $z < 1$ on predominantly large scales, spanning the range $0.1 \lesssim r_{\perp} \lesssim 10$ Mpc. Given the difference in approach between these studies, our comparison is restricted to being purely qualitative. Nevertheless, we note that the distribution of O VI around galaxies presented in Turner et al. (2014) reveals stronger anisotropies along the LOS than are evident in our study. It is important to note that these anisotropies are revealed on scales of a few hundred (proper) kpc, which are substantially smaller than the $\sim$ Mpc scales (comoving) that are considered here. It will be interesting to investigate the evolution in the distribution and dynamics of O VI around galaxies from high redshifts, around the peak in star formation activity at $z \sim 2$, to the present day, but there are no straight-forward comparisons at present.

Prochaska et al. (2011b) investigated the incidence of H I and O VI absorbers around galaxies at $z \lesssim 0.5$. Note that we make use of a subset of their galaxy redshift data in this study (Prochaska et al. 2011a). By examining the incidence rate of O VI absorbers as a function of galaxy impact parameter, they infer that the covering fraction of O VI around sub-$L^{\star}$ galaxies is nearly unity to impact parameters of 300 kpc. Comparing to the total incidence rate per absorption path length, they conclude that the majority of O VI absorbers to an equivalent width limit $W > 0.03$ Å in O VI λ1031 arise within 300 kpc of sub-$L^{\star}$ galaxies in the 'extended CGM', and that they rarely originate in the intergalactic WHIM predicted by cosmological hydrodynamical simulations (e.g. Cen & Ostriker 1999; Davé et al. 2001). The strategy of Prochaska et al. (2011b) is (i) to identify galaxies at small (<300 kpc) impact parameters from the QSO sight-lines in their sample observed with the Goddard High Resolution Spectrograph (GHRS), STIS and FUSE, then to search for O VI absorbers close to those sight-lines; and (ii) to search for galaxies close to particular O VI absorbers. Using the latter approach, they find that 5 out of the 30 O VI absorbers do not have a galaxy within 1 Mpc, although this result is sensitive to the completeness limit of the galaxy survey, which will miss low-luminosity galaxies. Our results are robust to galaxy completeness, and indicate that the clustering amplitudes of O VI absorbers with galaxies are weaker than those of galaxies with themselves. This may imply that not all O VI absorbers can be found very close to galaxies, or that covering factors of O VI around galaxies on the scales probed ($\gtrsim$100 kpc) are less than 100 per cent (or both). The conclusions of Prochaska et al. (2011b) are based on assumed covering factors of O VI around sub-$L^{\star}$ galaxies of close to 100 per cent to 300 kpc, which highlights a potential source of tension with our results. Similar studies to Prochaska et al. (2011b), e.g. Tumlinson & Fang (2005) and Stocke et al. (2013), suggest that the covering fractions are significantly below 100 per cent. We suggest that studies such as these may need to probe galaxies to fainter luminosities in order to reconcile results that may be biased by the galaxy survey completeness limits.

Tumlinson et al. (2011) performed another study of the incidence of O VI absorbers around galaxies as part of the COS-haloes survey, and found a dichotomy between the incidence rate around star-forming galaxies and non-star-forming galaxies. In particular, they find that O VI absorbers are nearly ubiquitous within 150 kpc of star-forming galaxies, but that only a small fraction of non-star-forming galaxies show O VI absorption within this distance. This result contrasts with ours, in which we find that the likelihood of finding O VI around star-forming galaxies is similar to the likelihood of finding O VI around non-star-forming galaxies. However, it is important to note that our study is optimized for scales $\gtrsim$100 kpc away from galaxies, and so if this dichotomy only exists on the smallest scales, we would not have been able to detect it. Furthermore, in terms of appropriately calculated virial radii for the COS-haloes survey galaxy sample, their QSO sight-lines intersect numerous star-forming galaxies at impact parameters greater than the virial radius, but no non-star-forming galaxies beyond this point (Shull 2014). Our approach is not subject to this particular bias, which may also help explain the differences between these two findings.

8.2 Interpretation of the results

The principal findings of this study are as follows.

(i) O VI absorbers show little velocity dispersion ($\lesssim$100 km s^{-1}) with respect to galaxies on $\sim$ Mpc scales.

(ii) The correlation length of the O VI–galaxy cross-correlation function is smaller (at the 4σ significance level for our full sample), and the slope potentially shallower, than that of the galaxy autocorrelation function in general, which indicates that O VI absorbers and galaxies may not linearly trace the same underlying distribution of matter.

(iii) The O VI–galaxy cross-correlation function split by star formation activity gives correlation lengths that are entirely consistent within their 1σ bootstrap uncertainties. Thus, on the scales probed ($\gtrsim$100 kpc), the likelihood of finding O VI absorbers around star-forming galaxies is similar to the likelihood of finding O VI absorbers around non-star-forming galaxies.

Given that the enrichment of the IGM is attributable to galaxy feedback, it is convenient to think of these results from a galaxy-centric viewpoint, as follows.

In item (ii) from the list above, the lower correlation amplitudes in the O VI–galaxy cross-correlation functions compared to the galaxy autocorrelation functions imply that either O VI absorbers are not ubiquitous near to the galaxies in our sample, or that their distribution around them is patchy, i.e. the covering factor of O VI around galaxies is substantially less than 100 per cent, as was suggested in the previous section. In reality, both of these inferences could be true. This may be a function of O VI column density and/or Doppler broadening parameter, although we have not attempted to split the O VI absorber sample in this study due to low number statistics. We also note that this conclusion may not necessarily apply to lower column density O VI absorbers and/or fainter galaxies, below the detection limits of the present survey.

Also in item (ii), indications of a shallower slope in the O VI–galaxy cross-correlation function compared to the galaxy autocorrelation function could indicate that the distribution of O VI around galaxies is in general more extended than the distribution of galaxies around themselves. From the data, we cannot rule out the possibility that O VI absorbers are primarily attached to galaxies, and that the cross-correlation amplitudes are primarily driven by clustering of the galaxies with themselves. Nevertheless, the inference of an extended distribution of O VI around galaxies is supported by predictions from the EAGLE simulation, which also indicate a shallower slope in the O VI–galaxy cross-correlation function, and an even shallower slope in the O VI autocorrelation function. The difference in the slopes and amplitudes of these correlation functions leads to the possibility that O VI absorbers and the galaxies in our sample do not linearly trace the same underlying distribution of matter, and that some O VI absorbers may be found far from galaxies that are bright enough to be included in our survey. This picture agrees well with the inferences made by Stocke et al. (2006, 2013), who find that broad O VI absorbers may trace hot gas ($T \approx 10^6$ K) that extends to large distances ($\sim$400–600 kpc) around galaxies, distributed in the cosmic web or in supercluster filaments. They argue that for a close correspondence between O VI and galaxies, low-luminosity galaxies ($L < 0.1L^\star$) must contribute to the total cross-section to match the observed O VI absorption-line frequency. Our results are entirely compatible with these inferences.

A similar scenario is seen in various cosmological hydrodynamical simulations, where O VI is distributed far from galaxies and often with a relatively flat radial profile (Cen & Chisari 2011; Smith et al. 2011; Oppenheimer et al. 2012; Shull et al. 2012; Ford et al. 2013, 2014; Hummels et al. 2013). Nevertheless, obtaining predictions that match observational findings is often met with varied success, and found to be sensitive to the subgrid physics prescriptions, particularly with reference to the physical and statistical properties of the absorbers. Comparisons made here with the EAGLE simulation largely echo these findings, but do confirm that the simulations are at least broadly capable of reproducing the observed distribution and dynamics of O VI absorbers around galaxies, as inferred from the slope and amplitude of the projected O VI–galaxy cross-correlation function, and the anisotropies in the two-dimensional O VI–galaxy cross-correlation function. These results also raise the tantalizing possibility that a fraction of O VI absorbers do in fact arise in the WHIM, outside of galaxy haloes and groups at temperatures of $10^5 < T < 10^7$ K, as predicted by the simulations. Nevertheless, a targeted approach to detecting the WHIM is still needed if we are to confirm these predictions (e.g. Tejos et al. 2016), and it is not yet clear whether the commonly assumed tracers of the WHIM (O VI, Ne VIII, broad Lyα) trace the bulk of this hot intergalactic plasma (e.g. Richter et al. 2006; Lehner et al. 2007; Danforth, Stocke & Shull 2010; Narayanan et al. 2011; Tepper-García et al. 2011, 2012; Tepper-García, Richter & Schaye 2013).

In many respects, the level of agreement on the clustering of O VI absorbers around galaxies between the EAGLE simulation and the real data is quite surprising when we consider the potential origins the O VI absorbers, some of which are expected to arise from conductive and turbulent interfaces that are not resolved by the cosmological simulations at present. Either these processes are relatively unimportant for the overall population of O VI absorbers detected by current instrumentation, or they are co-spatial with other production mechanisms (e.g. photoionization, collisional ionization, shocks), the physics of which is captured by the simulations. Given the level of agreement seen here, we also suggest that the shortcomings of our approach to extracting O VI absorbers from the EAGLE simulation, described in Section 5.2, have little overall effect in the measured clustering signal of O VI around galaxies, although it will be important in future work to verify our results with a rigorous Voigt profile fitting procedure.

Item (i) in our list of findings from the present study indicates that we have not found substantial evidence for gas outflows or inflows traced by O VI around galaxies on $\sim$ Mpc scales at low redshifts. This scenario is also consistent with that predicted by the EAGLE simulation, which implements subgrid prescriptions for effective feedback from star formation and AGN in order to match the present-day statistics of the galaxy population. Our constraints on the velocity dispersion of O VI around galaxies are consistent with a scenario in which the majority of the O VI absorbers detected with current instrumentation move with the galaxies, and that those bound to galaxy haloes are not moving with velocities sufficient to escape their local gravitational potential. This is then suggestive of a scenario in which the wider IGM not bound to individual galaxy haloes may have been enriched early in the history of the Universe (e.g. Wiersma et al. 2010; Cen & Chisari 2011; Oppenheimer et al. 2012).

Our other main finding indicates that the presence of O VI absorbers on $\gtrsim$100 kpc scales around galaxies is not strongly biased towards whether those galaxies are star forming or not. This situation is clearly echoed in the EAGLE simulation, and indicates that the instantaneous star formation activity in galaxies bears no relation to the overall distribution of metals around them. This further supports the inference that a significant proportion of the metals in the IGM have been distributed into the IGM early, and that the extent

of ongoing star formation has no discernible effect on the metal enrichment of the low-redshift IGM on $\gtrsim$100 kpc scales.

9 SUMMARY AND CONCLUSIONS

Our analysis of the two-point cross- and autocorrelation functions of O VI absorbers and galaxies at $z < 1$ resulted in the findings outlined below:

(i) O VI absorbers show very little velocity dispersion with respect to galaxies on $\sim$ Mpc scales at low redshifts. We estimate that this velocity dispersion amounts to $\lesssim$100 km s^{-1}.

(ii) The slope of the O VI–galaxy cross-correlation function is potentially shallower than that of the galaxy autocorrelation function. We therefore find that these populations may not linearly trace the same underlying distribution of matter. In particular, these results indicate that the distribution of O VI around galaxies could be more extended than the distribution of galaxies around themselves. We therefore speculate that a fraction of the O VI absorbers might trace the WHIM predicted by cosmological hydrodynamical simulations.

(iii) The clustering amplitudes of O VI absorbers around star-forming galaxies are consistent with those around non-star-forming galaxies within the uncertainties. We therefore find that the likelihood of finding O VI absorbers around star-forming galaxies is similar to the likelihood of finding O VI absorbers around non-star-forming galaxies, at least on scales $\gtrsim$100 kpc.

(iv) The amplitude of the O VI–galaxy cross-correlation is typically lower than that of the galaxy autocorrelation function by factors of a few. This indicates that O VI absorbers are either not ubiquitous to galaxies, that they are predominantly attached to fainter galaxies than those typical in our sample, or that their distribution around them is patchy on scales $\gtrsim$100 kpc (or any combination of these possibilities), at least for the column densities at which most are currently detected.

(v) We find that predictions from the EAGLE cosmological hydrodynamical simulation, subject to a correction in the predicted O VI column densities, are reasonably consistent with the observational findings outlined above. This suggests that simulations such as these may be regarded as a useful tool for understanding the distribution and dynamics of metal-enriched gas traced by O VI around galaxies on $\gtrsim$100 kpc scales.

ACKNOWLEDGEMENTS

We would first like to thank the referee, whose comments and suggestions improved this paper.

CWF would like to acknowledge the support of an STFC studentship (ST/J201013/1). NHMC thanks the Australian Research Council for *Discovery Project* grant DP130100568 which supported this work. TT acknowledges support from the Interuniversity Attraction Poles Programme initiated by the Belgian Science Policy Office ([AP P7/08 CHARM]). The research was supported in part by the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement 278594-GasAroundGalaxies.

We thank the contributors to SCIPY,[16] MATPLOTLIB[17] and the PYTHON programming language,[18] the free and open-source community and the NASA Astrophysics Data system[19] for software and services. This work also made use of ASTROPY; a community-developed core PYTHON package for Astronomy (Astropy Collaboration et al. 2013).

The work in this paper was in part based on observations made with the NASA/ESA *HST* under programmes GO 11585 and GO 12264, obtained at the STScI, 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, under programmes 070.A-9007, 086.A-0970 and 087.A-0857. 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 NASA. 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. 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, Technología e Innovación Productiva (Argentina).

GAMA is a joint European-Australasian project based around a spectroscopic campaign using the Anglo-Australian Telescope. The GAMA input catalogue is based on data taken from the Sloan Digital Sky Survey and the UKIRT Infrared Deep Sky Survey. Complementary imaging of the GAMA regions is being obtained by a number of independent survey programmes including *GALEX* MIS, VST KiDS, VISTA VIKING, *WISE*, *Herschel*-ATLAS, GMRT and ASKAP providing UV to radio coverage. GAMA is funded by the STFC (UK), the ARC (Australia), the AAO, and the participating institutions. The GAMA website is http://www.gama-survey.org/.

This work also made use of the DiRAC Data Centric system at Durham University, operated by the Institute for Computational Cosmology on behalf of the STFC DiRAC HPC Facility[20]. This equipment was funded by BIS National E-infrastructure capital grant ST/K00042X/1, STFC capital grant ST/H008519/1 and STFC DiRAC Operations grant ST/K003267/1 and Durham University. DiRAC is part of the National E-Infrastructure.

The raw data from *HST*/COS may be accessed from the MAST archive.[21] That from the European Southern Observatory may be accessed from the ESO Archive,[22] and that from the Gemini Observatory may be accessed from the Gemini Science Archive.[23] Absorption line data from Danforth et al. (2016) can be accessed from http://archive.stsci.edu/prepds/igm/.

REFERENCES

Abazajian K. N. et al., 2009, ApJS, 182, 543
Adelberger K. L., Steidel C. C., Shapley A. E., Pettini M., 2003, ApJ, 584, 45
Aguirre A., Schaye J., Theuns T., 2002, ApJ, 576, 1
Astropy Collaboration et al., 2013, A&A, 558, A33
Baldry I. K. et al., 2012, MNRAS, 421, 621

[16] http://www.scipy.org
[17] http://matplotlib.org
[18] http://www.python.org
[19] http://adswww.harvard.edu
[20] http://www.dirac.ac.uk
[21] http://archive.stsci.edu
[22] http://archive.eso.org/eso/eso_archive_main.html
[23] http://www.cadc-ccda.hia-iha.nrc-cnrc.gc.ca/en/gsa/

Baldry I. K. et al., 2014, MNRAS, 441, 2440
Baldwin J. A., Phillips M. M., Terlevich R., 1981, PASP, 93, 5
Baugh C. M., Lacey C. G., Frenk C. S., Granato G. L., Silva L., Bressan A., Benson A. J., Cole S., 2005, MNRAS, 356, 1191
Bertin E., 2006, in Gabriel C., Arviset C., Ponz D., Enrique S., eds, ASP Conf. Ser. Vol. 351, Astronomical Data Analysis Software and Systems XV. Astron. Soc. Pac., San Francisco, p. 112
Bertin E., Arnouts S., 1996, A&AS, 117, 393
Borkowski K. J., Balbus S. A., Fristrom C. C., 1990, ApJ, 355, 501
Bower R. G., Benson A. J., Malbon R., Helly J. C., Frenk C. S., Baugh C. M., Cole S., Lacey C. G., 2006, MNRAS, 370, 645
Brinchmann J., Charlot S., White S. D. M., Tremonti C., Kauffmann G., Heckman T., Brinkmann J., 2004, MNRAS, 351, 1151
Carswell R. F., Whelan J. A. J., Smith M. G., Boksenberg A., Tytler D., 1982, MNRAS, 198, 91
Cen R., Chisari N. E., 2011, ApJ, 731, 11
Cen R., Ostriker J. P., 1999, ApJ, 514, 1
Cen R., Miralda-Escudé J., Ostriker J. P., Rauch M., 1994, ApJ, 437, L9
Chen H.-W., Mulchaey J. S., 2009, ApJ, 701, 1219 (C09)
Colless M. et al., 2001, MNRAS, 328, 1039
Crain R. A. et al., 2009, MNRAS, 399, 1773
Crain R. A. et al., 2015, MNRAS, 450, 1937
Creasey P., Theuns T., Bower R. G., 2013, MNRAS, 429, 1922
Crighton N. H. M. et al., 2013, MNRAS, 433, 178
Danforth C. W., Shull J. M., 2005, ApJ, 624, 555
Danforth C. W., Shull J. M., 2008, ApJ, 679, 194
Danforth C. W., Stocke J. T., Shull J. M., 2010, ApJ, 710, 613
Danforth C. W. et al., 2014, preprint (arXiv:1402.2655)
Danforth C. W. et al., 2016, ApJ, 817, 111
Davé R., Hernquist L., Katz N., Weinberg D. H., 1999, ApJ, 511, 521
Davé R. et al., 2001, ApJ, 552, 473
Davé R., Oppenheimer B. D., Katz N., Kollmeier J. A., Weinberg D. H., 2010, MNRAS, 408, 2051
Davis M., Peebles P. J. E., 1983, ApJ, 267, 465
Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, ApJ, 292, 371
Dekel A., Birnboim Y., 2006, MNRAS, 368, 2
Dekel A., Sari R., Ceverino D., 2009, ApJ, 703, 785
Dolag K., Borgani S., Murante G., Springel V., 2009, MNRAS, 399, 497
Driver S. P. et al., 2011, MNRAS, 413, 971
Finn C. W. et al., 2014, MNRAS, 440, 3317
Ford A. B., Oppenheimer B. D., Davé R., Katz N., Kollmeier J. A., Weinberg D. H., 2013, MNRAS, 432, 89
Ford A. B., Davé R., Oppenheimer B. D., Katz N., Kollmeier J. A., Thompson R., Weinberg D. H., 2014, MNRAS, 444, 1260
Fukugita M., Peebles P. J. E., 2004, ApJ, 616, 643
Fukugita M., Hogan C. J., Peebles P. J. E., 1998, ApJ, 503, 518
Green J. C. et al., 2012, ApJ, 744, 60
Groth E. J., Peebles P. J. E., 1977, ApJ, 217, 385
Guzzo L. et al., 2014, A&A, 566, A108
Haardt F., Madau P., 2001, in Neumann D. M., Tran J. T. V., eds, Clusters of Galaxies and the High Redshift Universe Observed in X-rays. Commissariat à l'Energie Atomique, p. 64
Hernquist L., Katz N., Weinberg D. H., Miralda-Escudé J., 1996, ApJ, 457, L51
Hopkins A. M. et al., 2013, MNRAS, 430, 2047
Hummels C. B., Bryan G. L., Smith B. D., Turk M. J., 2013, MNRAS, 430, 1548
Hussain T., Muzahid S., Narayanan A., Srianand R., Wakker B. P., Charlton J. C., Pathak A., 2015, MNRAS, 446, 2444
Johnson S. D., Chen H.-W., Mulchaey J. S., 2013, MNRAS, 434, 1765 (J13)
Kaiser N., 1987, MNRAS, 227, 1
Keeney B. A., Danforth C. W., Stocke J. T., France K., 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., Green J. C., 2013, ApJ, 765, 27
Kennicutt R. C., Jr, 1998, ARA&A, 36, 189
Kereš D., Katz N., Weinberg D. H., Davé R., 2005, MNRAS, 363, 2
Kewley L. J., Dopita M. A., Sutherland R. S., Heisler C. A., Trevena J., 2001, ApJ, 556, 121
Knuth K. H., 2006, preprint (arXiv:physics/0605197)
Kwak K., Shelton R. L., 2010, ApJ, 719, 523
Lagos C. D. P., Cora S. A., Padilla N. D., 2008, MNRAS, 388, 587
Landy S. D., Szalay A. S., 1993, ApJ, 412, 64
Le Fèvre O. et al., 2004, A&A, 417, 839
Le Fèvre O. et al., 2005, A&A, 439, 845
Lehner N., Savage B. D., Richter P., Sembach K. R., Tripp T. M., Wakker B. P., 2007, ApJ, 658, 680
Lehner N. et al., 2013, ApJ, 770, 138
Li C., White S. D. M., 2009, MNRAS, 398, 2177
Liang C. J., Chen H.-W., 2014, MNRAS, 445, 2061
Liske J. et al., 2015, MNRAS, 452, 2087
McGaugh S. S., Schombert J. M., de Blok W. J. G., Zagursky M. J., 2010, ApJ, 708, L14
Meiring J. D., Tripp T. M., Werk J. K., Howk J. C., Jenkins E. B., Prochaska J. X., Lehner N., Sembach K. R., 2013, ApJ, 767, 49
Miralda-Escudé J., Cen R., Ostriker J. P., Rauch M., 1996, ApJ, 471, 582
Morris S. L., Jannuzi B. T., 2006, MNRAS, 367, 1261
Moustakas J., Kennicutt R. C., Jr, Tremonti C. A., 2006, ApJ, 642, 775
Narayanan A. et al., 2011, ApJ, 730, 15
Newman J. A. et al., 2013, ApJS, 208, 5
Oppenheimer B. D., Schaye J., 2013a, MNRAS, 434, 1043
Oppenheimer B. D., Schaye J., 2013b, MNRAS, 434, 1063
Oppenheimer B. D., Davé R., Katz N., Kollmeier J. A., Weinberg D. H., 2012, MNRAS, 420, 829
Osterbrock D. E., 1989, Astrophysics of Gaseous Nebulae and Active Galactic Nuclei. University Science Books, Mill Valley, CA
Planck Collaboration XVI, 2014, A&A, 571, A16
Prochaska J. X., Weiner B., Chen H.-W., Cooksey K. L., Mulchaey J. S., 2011a, ApJS, 193, 28 (P11)
Prochaska J. X., Weiner B., Chen H.-W., Mulchaey J., Cooksey K., 2011b, ApJ, 740, 91
Rees M. J., Ostriker J. P., 1977, MNRAS, 179, 541
Richter P., Savage B. D., Sembach K. R., Tripp T. M., 2006, A&A, 445, 827
Savage B. D., Lehner N., Narayanan A., 2011, ApJ, 743, 180
Savage B. D., Kim T.-S., Wakker B. P., Keeney B., Shull J. M., Stocke J. T., Green J. C., 2014, ApJS, 212, 8
Sawala T. et al., 2015, MNRAS, 448, 2941
Scannapieco C. et al., 2012, MNRAS, 423, 1726
Schaye J., 2001, ApJ, 559, L1
Schaye J. et al., 2010, MNRAS, 402, 1536
Schaye J. et al., 2015, MNRAS, 446, 521
Scodeggio M. et al., 2005, PASP, 117, 1284
Shull J. M., 2014, ApJ, 784, 142
Shull J. M., Smith B. D., Danforth C. W., 2012, ApJ, 759, 23
Shull J. M., Danforth C. W., Tilton E. M., 2014, ApJ, 796, 49
Simcoe R. A., Sullivan P. W., Cooksey K. L., Kao M. M., Matejek M. S., Burgasser A. J., 2012, Nature, 492, 79
Smith B. D., Hallman E. J., Shull J. M., O'Shea B. W., 2011, ApJ, 731, 6
Springel V., White S. D. M., Tormen G., Kauffmann G., 2001, MNRAS, 328, 726
Stocke J. T., Penton S. V., Danforth C. W., Shull J. M., Tumlinson J., McLin K. M., 2006, ApJ, 641, 217
Stocke J. T., Keeney B. A., Danforth C. W., Shull J. M., Froning C. S., Green J. C., Penton S. V., Savage B. D., 2013, ApJ, 763, 148
Stocke J. T. et al., 2014, ApJ, 791, 128
Storey P. J., Zeippen C. J., 2000, MNRAS, 312, 813
Stoughton C. et al., 2002, AJ, 123, 485
Suresh J., Bird S., Vogelsberger M., Genel S., Torrey P., Sijacki D., Springel V., Hernquist L., 2015, MNRAS, 448, 895
Tejos N. et al., 2014, MNRAS, 437, 2017 (T14)
Tejos N. et al., 2016, MNRAS, 455, 2662
Tepper-García T., Richter P., Schaye J., Booth C. M., Dalla Vecchia C., Theuns T., Wiersma R. P. C., 2011, MNRAS, 413, 190

Tepper-García T., Richter P., Schaye J., Booth C. M., Dalla Vecchia C., Theuns T., 2012, MNRAS, 425, 1640
Tepper-García T., Richter P., Schaye J., 2013, MNRAS, 436, 2063
Theuns T., Leonard A., Efstathiou G., Pearce F. R., Thomas P. A., 1998, MNRAS, 301, 478
Thom C., Chen H.-W., 2008, ApJS, 179, 37
Tripp T. M., Giroux M. L., Stocke J. T., Tumlinson J., Oegerle W. R., 2001, ApJ, 563, 724
Tumlinson J., Fang T., 2005, ApJ, 623, L97
Tumlinson J. et al., 2011, Science, 334, 948
Tumlinson J. et al., 2013, ApJ, 777, 59
Turner M. L., Schaye J., Steidel C. C., Rudie G. C., Strom A. L., 2014, MNRAS, 445, 794
van de Voort F., Schaye J., Booth C. M., Haas M. R., Dalla Vecchia C., 2011, MNRAS, 414, 2458
Vasiliev E. O., Ryabova M. V., Shchekinov Y. A., 2015, MNRAS, 446, 3078
Veilleux S., Cecil G., Bland-Hawthorn J., 2005, ARA&A, 43, 769
Vogelsberger M. et al., 2014, MNRAS, 444, 1518
Werk J. K., Prochaska J. X., Thom C., Tumlinson J., Tripp T. M., O'Meara J. M., Peeples M. S., 2013, ApJS, 204, 17
Werk J. K. et al., 2014, ApJ, 792, 8
White S. D. M., Frenk C. S., 1991, ApJ, 379, 52
White S. D. M., Rees M. J., 1978, MNRAS, 183, 341
Wiersma R. P. C., Schaye J., Dalla Vecchia C., Booth C. M., Theuns T., Aguirre A., 2010, MNRAS, 409, 132
Wiersma R. P. C., Schaye J., Theuns T., 2011, MNRAS, 415, 353
Young P. J., Sargent W. L. W., Boksenberg A., Carswell R. F., Whelan J. A. J., 1979, ApJ, 229, 891

This paper has been typeset from a TEX/LATEX file prepared by the author.