

# Large-scale clustering of Lyman $\alpha$ emission intensity from SDSS/BOSS

Rupert A. C. Croft,<sup>1,2★</sup> Jordi Miralda-Escudé,<sup>3,4★</sup> Zheng Zheng,<sup>5★</sup> Adam Bolton,<sup>5</sup>  
 Kyle S. Dawson,<sup>5</sup> Jeffrey B. Peterson,<sup>1</sup> Donald G. York,<sup>6</sup> Daniel Eisenstein,<sup>7</sup>  
 Jon Brinkmann,<sup>8</sup> Joel Brownstein,<sup>5</sup> Renyue Cen,<sup>9</sup> Timothée Delubac,<sup>10</sup>  
 Andreu Font-Ribera,<sup>11</sup> Jean-Christophe Hamilton,<sup>12</sup> Khee-Gan Lee,<sup>13</sup> Adam Myers,<sup>14</sup>  
 Nathalie Palanque-Delabrouille,<sup>10</sup> Isabelle Pâris,<sup>15</sup> Patrick Petitjean,<sup>16</sup>  
 Matthew M. Pieri,<sup>17</sup> Nicholas P. Ross,<sup>11</sup> Graziano Rossi,<sup>18</sup> David J. Schlegel,<sup>11</sup>  
 Donald P. Schneider,<sup>19,20</sup> Anže Slosar,<sup>21</sup> José Vazquez,<sup>21</sup> Matteo Viel,<sup>15,22</sup>  
 David H. Weinberg<sup>23</sup> and Christophe Yèche<sup>10</sup>

*Affiliations are listed at the end of the paper*

Accepted 2016 January 21. Received 2016 January 21; in original form 2015 April 9

## ABSTRACT

We present a tentative detection of the large-scale structure of Ly  $\alpha$  emission in the Universe at redshifts  $z = 2\text{--}3.5$  by measuring the cross-correlation of Ly  $\alpha$  surface brightness with quasars in Sloan Digital Sky Survey/Baryon Oscillation Spectroscopic Survey. We use a million spectra targeting luminous red galaxies at  $z < 0.8$ , after subtracting a best-fitting model galaxy spectrum from each one, as an estimate of the high-redshift Ly  $\alpha$  surface brightness. The quasar–Ly  $\alpha$  emission cross-correlation is detected on scales  $1 \sim 15 h^{-1}$  Mpc, with shape consistent with a  $\Lambda$ CDM model with  $\Omega_m = 0.30^{+0.10}_{-0.07}$ . The predicted amplitude of this cross-correlation is proportional to the product of the mean Ly  $\alpha$  surface brightness,  $\langle \mu_\alpha \rangle$ , the amplitude of mass fluctuations and the quasar and Ly  $\alpha$  emission bias factors. We infer  $\langle \mu_\alpha \rangle (b_\alpha/3) = (3.9 \pm 0.9) \times 10^{-21} \text{ erg s}^{-1} \text{ cm}^{-2} \text{ \AA}^{-1} \text{ arcsec}^{-2}$ , where  $b_\alpha$  is the Ly  $\alpha$  emission bias. If star-forming galaxies dominate this emission, we find  $\rho_{\text{SFR}} = (0.28 \pm 0.07)(3/b_\alpha) \text{ yr}^{-1} \text{ Mpc}^{-3}$ . For  $b_\alpha = 3$ , this value is  $\sim 30$  times larger than previous estimates from individually detected Ly  $\alpha$  emitters, but consistent with the total  $\rho_{\text{SFR}}$  derived from dust-corrected, continuum UV galaxy surveys, if most of the Ly  $\alpha$  photons from these galaxies avoid dust absorption and are reemitted after diffusing in large gas haloes. Heating of intergalactic gas by He II photoionization from quasar radiation or jets may alternatively explain the detected correlation, and cooling radiation from gas in galactic haloes may also contribute. We also detect redshift space anisotropy of the quasar–Ly  $\alpha$  emission cross-correlation, finding evidence at the  $3.0\sigma$  level that it is radially elongated, which may be explained by radiative-transfer effects. Our measurements represent the first application of the intensity mapping technique to optical observations.

**Key words:** cosmology: observations.

## 1 INTRODUCTION

The Ly  $\alpha$  emission line of neutral hydrogen is a strong feature that has been used to detect galaxies at a wide range of redshifts (e.g. Hu & McMahon 1996; Keel et al. 1999; Fujita et al. 2003; Cowie, Barger & Hu 2010). Another potentially useful technique is that

of intensity mapping (e.g. Chang et al. 2010; Carilli 2011; Gong et al. 2011; Peterson & Suarez 2012; Silva et al. 2013; Doré et al. 2014; Pullen, Doré & Bock 2014), which seeks to map the large-scale structure using one emission line or more (see e.g. Wyithe & Morales 2007; Visbal & Loeb 2010), without resolving individual sources (such as galaxies or gas clouds). By measuring this structure, one is sensitive to all clustered emission, without the observational biases which arise from source detection and luminosity measurement (such as detection limits, determination of backgrounds and finite aperture size). In this paper, we seek to perform the first

\* E-mail: [rcroft@cmu.edu](mailto:rcroft@cmu.edu) (RACC); [miralda@icc.ub.edu](mailto:miralda@icc.ub.edu) (JM-E); [zhengzheng@astro.utah.edu](mailto:zhengzheng@astro.utah.edu) (ZZ)

cosmological measurement of intensity mapping in the Ly  $\alpha$  line, using a large data set of spectra from the Sloan Digital Sky Survey III (SDSS-III; Eisenstein et al. 2011) Baryon Oscillation Spectroscopic Survey (BOSS; Dawson et al. 2013). We use spectra that were targeted at massive galaxies at  $z < 0.8$ . After subtracting best-fitting model galaxy spectra, we expect that any high-redshift Ly  $\alpha$  emitters that are within the fibre aperture result in a residual flux, present also in sky fibres. Even if not detectable as individual sources, we can search for large-scale structure in this emission by determining its spatial cross-correlation function with the positions of BOSS quasars, which are tracers of structure with a known bias factor (e.g. White et al. 2012) at redshifts  $z > 2$ , where the Ly  $\alpha$  emission line is in the optical part of the spectrum.

Following the early prediction by Partridge & Peebles (1967) that galaxies should be detectable at high redshift from their Ly  $\alpha$  emission line, many surveys have been designed to detect individual galaxies as sources of Ly  $\alpha$  emission. These include narrow-band imaging (e.g. Ouchi et al. 2003; Gronwall et al. 2007), serendipitous slit spectroscopy (e.g. Cassata et al. 2011) and integral field spectroscopy (e.g. van Breukelen, Jarvis & Venemans 2005; Blanc et al. 2011). These techniques have resulted in the compilation of catalogues of several hundred to a few thousand Ly  $\alpha$  emitting galaxies from redshifts  $z \sim 2.1$  to the redshifts associated with the end of reionization. These samples have been used to show that Ly  $\alpha$  emitters of line luminosity  $L_\alpha = 10^{42} \text{ erg s}^{-1}$  found at  $z = 3$  have space densities of  $\sim 10^{-3} \text{ Mpc}^{-3}$  (e.g. Gawiser et al. 2007; Cassata et al. 2015) and are therefore expected to be the progenitors of  $L_*$  galaxies at redshift  $z = 0$ . The clustering of these galaxies has been measured on scales of up to 10 Mpc by Gaiuta et al. (2010) and Gawiser et al. (2007), who find that at redshifts  $z = 2\text{--}3$  they have a bias factor with respect to the underlying matter (in CDM models) of  $b \sim 1.5\text{--}2$ . Integrating the luminosity functions of Ly  $\alpha$  emitting galaxies, assuming a power-law extrapolation for the faint-end slope, has revealed that the comoving volume emissivity of Ly  $\alpha$  photons declines significantly from  $z \sim 6$  to 2 (Gronwall et al. 2007; Ouchi et al. 2008; Cassata et al. 2011). This behaviour can be compared to the opposite evolution in redshift of galaxies measured in optically thin parts of the rest-frame UV spectrum (or using the H  $\alpha$  line; e.g. Hayes et al. 2011). This comparison has been used to infer (e.g. by Cassata et al. 2011; Hayes et al. 2011) that the escape fraction of Ly  $\alpha$  photons produced in star-forming regions has significantly decreased from  $z = 6$  to 2.

Because Ly  $\alpha$  photons have a high cross-section for scattering off neutral hydrogen, extended Ly  $\alpha$  emission is expected to be common in many environments. For example, Ly  $\alpha$  radiation from star-forming regions in galaxies should undergo hundreds or thousands of scatterings in gas in any circumgalactic medium before finally escaping or else being absorbed by dust. The existence of a general fluorescent emission from the intergalactic medium (IGM) was also hypothesized by Hogan & Weymann (1987) and Gould & Weinberg (1996). Theoretical work applying line radiative transfer on gas distribution in cosmological hydrodynamic simulations has made predictions for Ly  $\alpha$  emission around galaxies and quasars (e.g. Cantalupo et al. 2005; Laursen & Sommer-Larsen 2007; Kollmeier et al. 2010; Zheng et al. 2011a), as well as metal line emission (Bertone & Schaye 2012). These studies have resulted in predictions of extended Ly  $\alpha$  haloes around galaxies with sizes of hundreds of kpc, with a strong dependence of their properties on environment that can lead to new effects on galaxy clustering (Zheng et al. 2011a).

Observational evidence for extended emission includes the discovery and characterization of the so-called Ly  $\alpha$  ‘blobs’ (Steidel

et al. 2000). Deep spectroscopic searches for diffuse Ly  $\alpha$  emission have been completed by Rauch et al. (2008), finding faint Ly  $\alpha$  emitting galaxies. Stacking of spectra of damped Ly  $\alpha$  absorbers in quasars has also produced measurements of residual Ly  $\alpha$  emission (Rahmani et al. 2010; Noterdaeme et al. 2014). Recently, Martin et al. (2014a,b) published the first results from the Cosmic Web Imager, an integral field spectrograph designed to map low surface brightness emission, detecting Ly  $\alpha$  emission from filamentary structures around a  $z = 2.8$  quasar as long as 250–400 proper kpc. Diffuse Ly  $\alpha$  haloes around high-redshift galaxies have been found to be ubiquitous by Steidel et al. (2011) and Matsuda et al. (2012). Momose et al. (2014) have assembled several samples of up to 3600 Ly  $\alpha$  emitters from Subaru narrow-band imaging at a range of redshifts from  $z = 2.2$  to 6.6 and, after controlling for atmospheric and instrumental artefacts, they detect diffuse extended Ly  $\alpha$  haloes with exponential scalelengths of  $\sim 5\text{--}10$  kpc from  $z = 2.2$  to 5.7. The large-scale studies in our paper are an alternative, complementary observational strategy to these earlier studies, which involve deep integrations over small fields of view.

All of these sources should be clustered on large scales and should contribute to the mean Ly  $\alpha$  emission intensity in the Universe. This mean emission is detectable if it cross-correlates as expected with other tracers of large-scale structure that we can observe at the same redshift. We shall use the quasars found by BOSS at  $z > 2$  to correlate with Ly  $\alpha$  emission in this work. This clustered Ly  $\alpha$  emission is extremely faint, but as we shall demonstrate it can be detected with BOSS thanks to the enormous number of spectra that are observed. While large-scale clustering measurements cannot easily allow separation of the signal into various sources, we may expect faint Ly  $\alpha$  emitting galaxies to dominate over quasars due to their much larger number density. If this is the case, then the mean Ly  $\alpha$  emission intensity clustered with quasars can be used as a measure of the global star formation rate (e.g. Cassata et al. 2011), times the mean (luminosity weighted) bias factor of the distribution of these galaxies. Our measurement of Ly  $\alpha$  emission will therefore be useful as a probe of star formation which takes into account all sources of Ly  $\alpha$  emission, reaching to arbitrarily faint luminosities and surface brightnesses from extended haloes to faint galaxies.

The structure of this paper is as follows. Section 2 describes the data samples we use in our work, which include the galaxy and sky spectra from SDSS DR10, along with the quasar catalogue from that data release. We present our measurement of quasar–Ly  $\alpha$  emission correlations in Section 3, including the evolution with redshift and clustering parallel and transverse to the line of sight. Section 4 describes our tests involving fitting and subtraction of emission lines. In Section 5, we convert our determination of the Ly  $\alpha$  surface brightness into a star formation rate density and compare to other measurements. In Section 6, we summarize our results and in Section 7 discuss them further. There are also three appendices to the paper, A–C; in these we measure stray light contamination, determine a large-scale surface brightness correction and perform some sample tests of our results.

## 2 DATA SAMPLES

This study makes use of data from the SDSS BOSS survey Data Release 10 (DR10; Ahn et al. 2014), including quasar position and redshift data, galaxy spectra and sky fibre spectra. The SDSS camera and telescope are described in Gunn et al. (1998) and Gunn et al. (2006), respectively. Full information on the SDSS/BOSS spectrographs can be found in Smee et al. (2013). The wavelength coverage of the spectrograph is from  $\lambda = 3560$  to  $10\,400 \text{ \AA}$  the

resolving power is  $R \sim 1400$  for the range  $\lambda = 3800\text{--}4900 \text{ \AA}$ , and is kept above  $R = 1000$  for the remainder of the wavelength range. The fibres have a diameter of  $120 \mu\text{m}$ , corresponding to 2 arcsec in angle. We restrict the redshift range of data we use in our analysis to  $2.0 < z < 3.5$ , due to the spectrograph cutoff at low redshift and the limited number of observed quasars at high redshift.

## 2.1 Spectra

The 987 482 galaxy spectra in our sample are of targeted luminous red galaxies (LRGs) which are within redshifts  $z \sim 0.15$  and  $z \sim 0.7$ . The redshift range of the original targets is not important to our study, as for each spectrum we make use only of the pixels for which the Ly  $\alpha$  emission line lies within the redshift range specified above ( $2.0 < z < 3.5$ ). In observed wavelength units this is  $3647\text{--}5470 \text{ \AA}$ . We also make use of 146 065 sky fibre spectra.

The main BOSS LRG programme consists of two galaxy target samples (see Dawson et al. 2013), designated CMASS (for constant mass) and LOWZ (for low redshift). The LOWZ galaxy sample is composed of massive red galaxies spanning the redshift range  $0.15 \lesssim z \lesssim 0.4$ . The CMASS galaxy sample is composed of massive galaxies spanning the redshift range  $0.4 \lesssim z \lesssim 0.7$ . Both samples are colour-selected to provide near-uniform sampling over the combined volume. The faintest galaxies are at  $r = 19.5$  for LOWZ and  $i = 19.9$  for CMASS. Colours and magnitudes for the galaxy selection cuts are corrected for Galactic extinction using Schlegel, Finkbeiner & Davis (1998) dust maps. We do not differentiate between CMASS and LOWZ samples in our analysis.

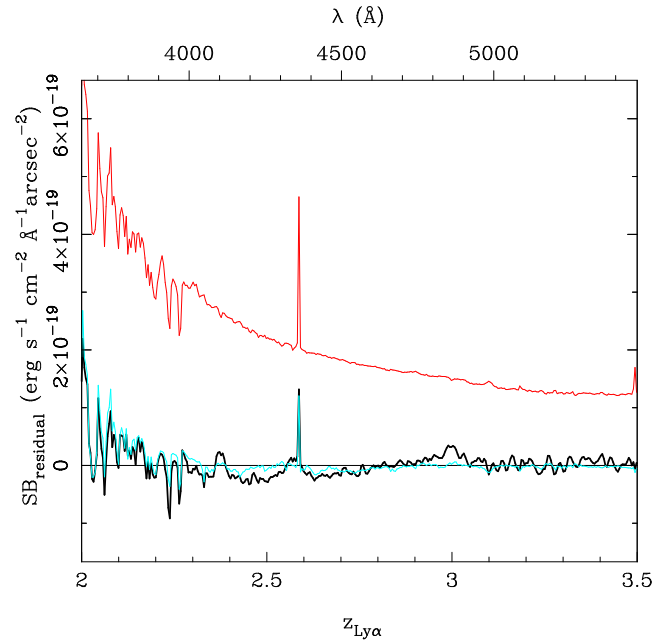
The spectroscopic measurement pipeline for BOSS is described in detail in Bolton et al. (2012). The most important data products that are used in the present analysis are: (i) wavelength-calibrated, sky-subtracted, flux-calibrated and co-added object spectra, which have been rebinned on to a uniform baseline of  $\Delta \log_{10} \lambda = 10^{-4}$  (about  $69 \text{ km s}^{-1} \text{ pixel}^{-1}$ ); (ii) statistical error-estimate vectors for each spectrum (expressed as inverse variance) incorporating contributions from photon noise, CCD read noise and sky-subtraction error; (iii) mask vectors for each spectrum.

### 2.1.1 Data preparation

For each of the LRG spectra, we subtract the best-fitting model spectrum provided by the pipeline. This template model spectrum (see Bolton et al. 2012 for details) is computed using least-squares minimization comparison of each galaxy spectrum to a full range of galaxy templates. A range of redshifts is explored, with trial redshifts spaced every pixel. At each redshift, the spectrum is fitted with an error-weighted least-squares linear combination of redshifted template eigenspectra in combination with a low-order polynomial. The polynomial terms absorb Galactic extinction, intrinsic extinction and residual spectrophotometric calibration errors (typically at the 10 per cent level) that are not fully spanned by the eigenspectra; there are three polynomial degrees of freedom for galaxies. The template basis sets are derived from rest-frame principal-component analyses of training samples of galaxies, and have four degrees of freedom (eigenspectra).

After subtraction of the best-fitting template spectrum, we compute the average residual spectrum of all galaxies. This is displayed in Fig. 1 where the horizontal axis is labelled in units of the redshift of the Ly  $\alpha$  line. Fig. 1 also presents the mean sky spectrum and the sky-subtracted sky spectrum.

We can see that the residual surface brightness per unit wavelength (hereafter shortened to ‘surface brightness’ – we use this term to refer to the quantity measured throughout the paper, which is most



**Figure 1.** Black line: the average residual surface brightness in all 987 482 LRG spectra after subtraction of the best-fitting galaxy model. The bottom horizontal axis is in units of the redshift of the Ly  $\alpha$  line and the top in units of observed wavelength. Red line (top): the average sky fibre surface brightness in all 146 065 sky fibre spectra. Cyan line: the average sky fibre surface brightness with model sky subtracted from all sky fibre spectra. In all curves, the prominent emission line at wavelength  $\lambda = 4358 \text{ \AA}$  is due to terrestrial airglow from Mercury streetlamps (the Hg G-line).

precisely the flux density per unit solid angle per unit wavelength) in the galaxy fibres is within  $\pm 10^{-19} \text{ erg s}^{-1} \text{ cm}^{-2} \text{ \AA}^{-1} \text{ arcsec}^{-2}$  for most of the redshift range. There are however significant excursions corresponding to features including the zero redshift calcium H and K lines (at  $3969$  and  $3934 \text{ \AA}$ ) and a strong Mercury G line from streetlamps (at  $4358 \text{ \AA}$ ). In our analysis we subtract the mean residual surface brightness, from all spectra before cross-correlating them as we are only interested in the fluctuations in the Ly  $\alpha$  surface brightness. In order to reduce noise, we also mask two regions corresponding to large features in the residual surface brightness,  $40$  and  $30 \text{ \AA}$  windows centred on wavelengths  $3900$  and  $4357 \text{ \AA}$ , respectively (corresponding to redshifts  $z = 2.21$  and  $2.58$ ).

Comparing the sky fibre and galaxy fibre residual spectra, we can see that there are differences at the  $\sim 10^{-19} \text{ erg s}^{-1}$  level over much of the spectra. We attribute these to galaxy surface brightness that was not subtracted perfectly by the galaxy model. In our cross-correlation technique for measuring clustering in the Ly  $\alpha$  emission, we necessarily subtract the mean surface brightness, therefore residual fluctuations seen in Fig. 1 are not problematic except for the noise they contribute.

## 2.2 Quasars

We use quasars from the SDSS/BOSS DR10 catalogue (Ahn et al. 2014). The quasar target sample included both colour-selected candidates and known quasars (Bovy et al. 2011; Kirkpatrick et al. 2011; Ross et al. 2012, 2013). The candidate quasar spectra were all visually inspected and redshift estimates computed using a principal component analysis (see Pâris et al. 2012 for the details of the procedure as applied to DR9 quasars). We select the 130 812 quasars in the DR10 data set that have redshifts in the range  $2.0 < z < 3.5$ .

Because the galaxy pixels cover this redshift range uniformly, the central redshift of our measurements is the mean redshift of these quasars,  $z = 2.55$ .

### 3 QUASAR–LY $\alpha$ EMISSION CROSS-CORRELATION

Before computing the quasar–Ly  $\alpha$  emission cross-correlation, we first split the sample of galaxy spectra into 100 subsamples of approximately equal sky area based on contiguous groupings of plates. We then convert the galaxy spectrum pixels and the quasar angular positions and redshifts into comoving Cartesian coordinates using a flat cosmological model with matter density  $\Omega_m = 0.315$ , consistent with the *Planck*, Planck Collaboration XVI et al. (2014) results (cosmological constant density  $\Omega_\Lambda = 0.685$ ). This fiducial model is used throughout the paper.

We compute the quasar–Ly  $\alpha$  emission surface brightness cross-correlation,  $\xi_{q\alpha}(r)$ , using a sum over all quasar–galaxy spectrum pixel pairs separated by  $r$  within a certain bin:

$$\xi_{q\alpha}(r) = \frac{1}{\sum_{i=1}^{N(r)} w_{ri}} \sum_{i=1}^{N(r)} w_{ri} \Delta_{\mu,ri}, \quad (1)$$

where  $N(r)$  is the number of pixels in the bin centred on quasar–pixel distance  $r$ , and  $\Delta_{\mu,ri} = \mu_{ri} - \langle \mu(z) \rangle$  is the residual surface brightness in the spectrum at pixel  $i$  for the bin  $r$ . Note here that we have a different list of pixels labelled as  $i$  for each bin in the separation  $r$  between a pixel and a quasar, which has Ly  $\alpha$  surface brightness  $\mu_{ri}$ . The residual flux at each pixel is obtained by subtracting the mean at each redshift,  $\langle \mu(z) \rangle$ . We weight each pixel by  $w_{ri} = 1/\sigma_{ri}^2$ , where  $\sigma_{ri}^2$  is the pipeline estimate of the inverse variance of the flux at each pixel. We first present our results as a function of only the modulus of the quasar–pixel separation  $r$  in comoving  $h^{-1}$  Mpc, in 20 bins logarithmically spaced between  $r = 0.5 h^{-1}$  Mpc and  $r = 150 h^{-1}$  Mpc. In Section 3.3, we will also examine redshift-space anisotropies in the correlation function  $\xi_{q\alpha}$  by considering bins in the parallel and perpendicular components of  $r$ , using the same formulation.

When evaluating equation (1), a possible significant systematic error is caused by stray light from the quasars themselves contaminating spectra of nearby galaxies. This occurs because the light from the various fibres is dispersed on to a single CCD, so that extraction of each spectrum along one dimension (Bolton et al. 2012) may include light from adjacent fibres of bright sources. We see strong evidence of this stray light from quasars in galaxy spectra when the quasar and galaxy spectra are four fibres apart or fewer, in the list of fibres as they are ordered in the CCD. The effect is discussed in detail in Appendix A. When the galaxy and quasar spectra in a pair are more than four fibres apart, we see no evidence for this contamination, and the results are statistically consistent with using only pairs of quasars and galaxies on different plates (see Appendix A). In order to safely eliminate this stray quasar light when computing the flux cross-correlation with equation (1), we therefore apply the constraint that the quasar and galaxy fibres must be at least six fibres apart.

There is also the possibility that some clustering in the plane of the sky is generated by effects (e.g. galactic obscuration) which modulate both Ly  $\alpha$  surface brightness and quasar target selection. Appendix B presents measurements of  $\xi_{q\alpha}$  for quasar–pixel pairs which are close together on the sky (i.e. in the transverse separation) but widely separated along the line of sight. This measurement enables us to quantify how much clustering could be caused by ef-

fects such as Galactic obscuration and to compute a  $\xi_{q\alpha}$  correction term to be subtracted from our fiducial clustering result. We also measure  $\xi_{q\alpha}$  for pairs which are close in the line-of-sight separation but widely separated on the sky. This latter measurement constrains how much spurious clustering is caused by large-scale variations in the line-of-sight direction, for example redshift evolution in the efficiency of galaxy subtraction, or flux calibration errors with wavelength that may be associated with sky lines. We apply the corrections to  $\xi_{q\alpha}$  from Appendix B to our analysis below and in the other sections of the paper. We discuss the small-scale anisotropy of  $\xi_{q\alpha}$  in Section 3.3 below. For now we note that application of the correction factors described above changes the amplitude and shape of our measured  $\xi_{q\alpha}$  by less than  $1\sigma$  in all cases.

We perform the pairwise computation of equation (1) for each of our 100 subsamples, and then compute the mean and standard deviation of  $\xi_{q\alpha}(r)$  using a jackknife estimator. The jackknife estimator is also used to compute the covariance matrix of  $\xi_{q\alpha}(r)$ :

$$C_{ij} = \sum_{k=1}^M [\xi_{q\alpha,k}(r_i) - \overline{\xi_{q\alpha}}(r_i)][\xi_{q\alpha,k}(r_j) - \overline{\xi_{q\alpha}}(r_j)], \quad (2)$$

where  $\xi_{q\alpha,k}(r_i)$  is the cross-correlation in bin  $i$  for jackknife sample  $k$ ,  $\overline{\xi_{q\alpha}}(r_i)$  is the cross-correlation for bin  $i$  for the full data set, and the number of jackknife samples is  $M = 100$ .

#### 3.1 Fiducial result

We show  $\xi_{q\alpha}(r)$  for our fiducial sample (which is the entire data set over the redshift range  $2.0 < z < 3.5$ ) in Fig. 2. The mean Ly  $\alpha$  redshift of the galaxy pixels in this sample is  $z = 2.71$  and of the quasars  $z = 2.55$ . Because the galaxy pixel distribution is uniform in redshift, and because quasar–pixel pairs with small separations contribute most to the clustering signal, we adopt the effective mean redshift of our fiducial measurement to be  $z = 2.55$ . The cross-correlation function is in units of the surface brightness of Ly  $\alpha$  emission, and its amplitude is directly proportional to that surface brightness. The  $\xi_{q\alpha}(r)$  points reveal that there is significant measurable large-scale structure present in the Ly  $\alpha$  emission, on scales from 1 to  $\sim 15 h^{-1}$  Mpc. Fig. 2 also displays a linear CDM fit to the cross-correlation function, which is consistent with observational results on scales from 1 to  $100 h^{-1}$  Mpc. We turn to this fit in the next subsection.

#### 3.2 Model fit

If the Ly  $\alpha$  emission clustering is due to a linearly biased version of the density field, then a model for the isotropically averaged quasar–Ly  $\alpha$  cross-correlation  $\xi_{q\alpha}(r)$  is as follows:

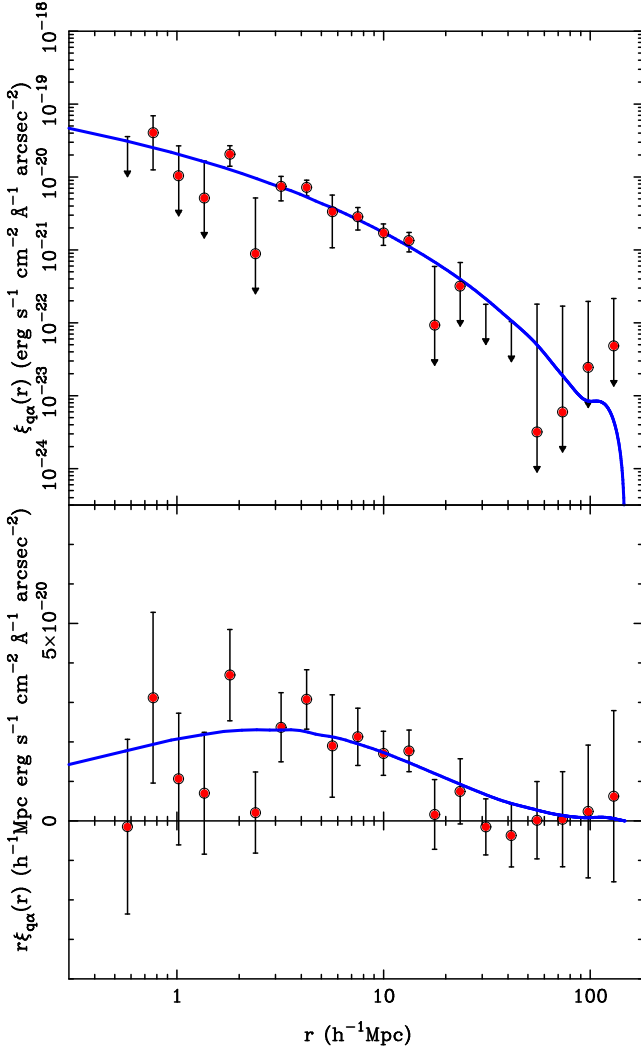
$$\xi_{q\alpha}(r) = b_q b_\alpha f_\beta \langle \mu_\alpha \rangle \xi(r), \quad (3)$$

where  $\langle \mu_\alpha \rangle$  is the mean surface brightness of Ly  $\alpha$  emission,  $b_q$  and  $b_\alpha$  are the quasar and Ly  $\alpha$  emission linear bias factors,  $\xi(r)$  is the linear  $\Lambda$ CDM mass correlation function and  $f_\beta$  is a constant enhancement to the correlation function on linear scales that is caused by peculiar velocity redshift-space distortions (Kaiser 1987).

It is important to note that  $b_\alpha$  is the bias factor for Ly  $\alpha$  surface brightness fluctuations, and is different in definition from the usually quoted bias factor of Ly  $\alpha$  emitters,  $b_{\text{LAE}}$  (e.g. as measured by Gawiser et al. 2007; Guaita et al. 2010). The bias factor  $b_{\text{LAE}}$  reflects the relation between the fluctuations  $\delta_n$  in the number density  $n$  of Ly  $\alpha$  emitters and that in the matter density  $\delta$ ,

$$\delta_n = \frac{n - \langle n \rangle}{\langle n \rangle} = b_{\text{LAE}} \delta, \quad (4)$$





**Figure 2.** The quasar–Ly  $\alpha$  emission cross-correlation function,  $\xi_{q\alpha}(r)$  (see equation 1). The points represent results for the fiducial sample that covers redshift range  $2.0 < z < 3.5$ . The error bars have been calculated using a jackknife estimator and 100 subsamples of the data. The smooth curve is a best-fitting linear CDM correlation function (see Section 3.2). The top panel shows the  $\xi_{q\alpha}(r)$  results with a log y-axis scale, and the bottom panel displays  $r\xi_{q\alpha}(r)$  on a linear scale in order to allow points which are negative to be visible.

where  $\delta = (\rho - \langle \rho \rangle) / \langle \rho \rangle$  and  $\rho$  is the matter density field. The factor  $b_\alpha$  in equation (3) relates fluctuations  $\delta_\mu$  in the Ly  $\alpha$  surface brightness  $\mu$  to matter fluctuations according to

$$\delta_\mu = \frac{\mu - \langle \mu \rangle}{\langle \mu \rangle} = b_\alpha \delta. \quad (5)$$

In the absence of radiative transfer effects (Zheng et al. 2011a; see Section 6.4), the Ly  $\alpha$  surface brightness  $\mu$  is proportional to the Ly  $\alpha$  luminosity density  $\rho_L$  of the underlying star-forming galaxy population. The fluctuations  $\delta_L$  of the latter can be characterized by the bias factor  $b_L$ ,

$$\delta_L = \frac{\rho_L - \langle \rho_L \rangle}{\langle \rho_L \rangle} = b_L \delta, \quad (6)$$

and we have  $b_\alpha = b_L$ . As  $b_L$  reflects weighting by luminosity rather than by number, it is likely to be significantly higher than  $b_{LAE}$ , because higher luminosity emitters tend to be more strongly clustered.

The radiative transfer effect modifies this relation to  $b_\alpha = b_L + \alpha_1$  with  $\alpha_1$  a positive number (see Section 6.4). Overall, we expect  $b_\alpha$  to be substantially higher than  $b_{LAE}$ . We will return to this topic in Section 5.

We use the linear CDM transfer function of Lewis, Challinor & Lasenby (2000) to compute  $\xi(r)$ . In our computations we choose to vary the shape of the correlation function by changing  $\Omega_m$ , the matter density, in the context of the  $\Lambda$ CDM model, keeping the other parameters which influence the shape (such as  $h$ , and the baryon density  $\Omega_b$ ) fixed. Parameters have been reported for the best-fitting  $\Lambda$ CDM model to the *Planck* satellite data by Planck Collaboration XVI et al. (2014). We assume *Planck* values for  $\Omega_b = 0.049$  and the spectral index  $n_s = 0.9603$  but set  $h = 0.7$ . Note that we are merely using  $\Omega_m$  to parametrize the shape of the correlation function to see if it is consistent with other observations, and are not presenting our results for  $\Omega_m$  as properly marginalized measurements of that parameter.

The other free parameter is the amplitude,  $b_q b_\alpha f_\beta \langle \mu \rangle$ . We assume in all cases that the underlying amplitude of mass fluctuations  $\sigma_8(z=0) = 0.83$ , and therefore that  $\sigma_8(z=2.55) = 0.294$ , again consistent with Planck Collaboration XVI et al. (2014).

We fit our model to the data in Fig. 2 by varying these two parameters. The  $\chi^2$  value is given by

$$\chi^2 = \sum_N \left[ \xi_{q\alpha}^{\text{obs}}(r_i) - \xi_{q\alpha}^{\text{mod}}(r_i) \right] C_{ij}^{-1} \left[ \xi_{q\alpha}^{\text{obs}}(r_j) - \xi_{q\alpha}^{\text{mod}}(r_j) \right], \quad (7)$$

where the sum is over the  $N = 20$  bins,  $\xi_{q\alpha}^{\text{obs}}(r_i)$  is the observed cross-correlation measured in bin  $i$ ,  $\xi_{q\alpha}^{\text{mod}}(r_i)$  is the model prediction for bin  $i$ , and  $C_{ij}$  is the covariance matrix computed using our 100 jackknife samples in equation (2).

The best-fitting values and  $1\sigma$  error bars are as follows:

$$b_q b_\alpha f_\beta \langle \mu \rangle = 3.33_{-0.43}^{+0.41} \times 10^{-20} \text{ erg s}^{-1} \text{ cm}^{-2} \text{ \AA}^{-1} \text{ arcsec}^{-2}, \quad (8)$$

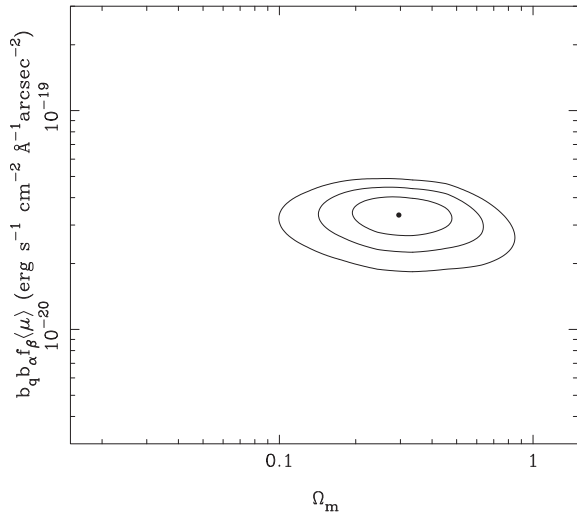
and  $\Omega_m = 0.296_{-0.071}^{+0.103}$ . The shape parameter  $\Omega_m$  is consistent with the best-fitting value from the *Planck* satellite results (Planck Collaboration XVI et al. 2014,  $\Omega_m = 0.30$ ). The  $1\sigma$ ,  $2\sigma$  and  $3\sigma$  confidence contours in these parameters considered together (i.e.  $\Delta\chi^2 = 2.3, 6.17$  and  $11.8$ ) are displayed along with the best-fitting values in Fig. 3.

Our linear fit to the isotropically averaged correlation function neglects non-linear effects and redshift measurement errors. We leave more detailed non-linear and halo modelling of the correlation function to future work.

### 3.3 Clustering transverse and parallel to line of sight

We also compute the quasar–Ly  $\alpha$  cross-correlation as a function of  $r_\parallel$  and  $r_\perp$ , the quasar–pixel pair separation along and across the line of sight, shown in Fig. 4, on a linear scale. We can see that the contours are relatively symmetric about the  $r_\parallel = 0$  axis and somewhat stretched along the  $r_\parallel$  direction. Font-Ribera et al. (2013) found a redshift offset between quasars and the Ly  $\alpha$  forest of  $\delta_z = -160 \text{ km s}^{-1}$ , due to the quasar catalogue redshifts being on average too small by this amount. This correlation resulted in the quasar–Ly  $\alpha$  forest cross-correlation being shifted upwards by this amount. The precision of the quasar–Ly  $\alpha$  emission cross-correlation in our paper is smaller, but visual inspection of Fig. 4 reveals that the position of the centroid of the cross-correlation is consistent with a small upward shift of this magnitude ( $1.6 h^{-1} \text{ Mpc}$  at this redshift).

On the scales where the cross-correlation is easy to discern ( $r \lesssim 20 h^{-1} \text{ Mpc}$ ), there is no sign of compression due to linear infall



**Figure 3.** Fit parameters for the amplitude  $b_q b_{\alpha} f_{\beta}(\mu)$  and shape  $\Omega_m$  (for fixed  $h$  and other parameters) of a linearly biased CDM model fit to the Ly  $\alpha$  cross-correlation function plotted in Fig. 2. The dot indicates the best-fitting parameters and the contours show the  $1\sigma$ ,  $2\sigma$  and  $3\sigma$  confidence contours.

(Kaiser 1987). The redshift-space distortions should be less prominent compared to the Ly  $\alpha$  forest because of the expected higher bias for the Ly  $\alpha$  emission. In reality there appears to be stretching along the line of sight (we quantify this below), which might be due to a combination of quasar redshift errors, the intrinsic velocity

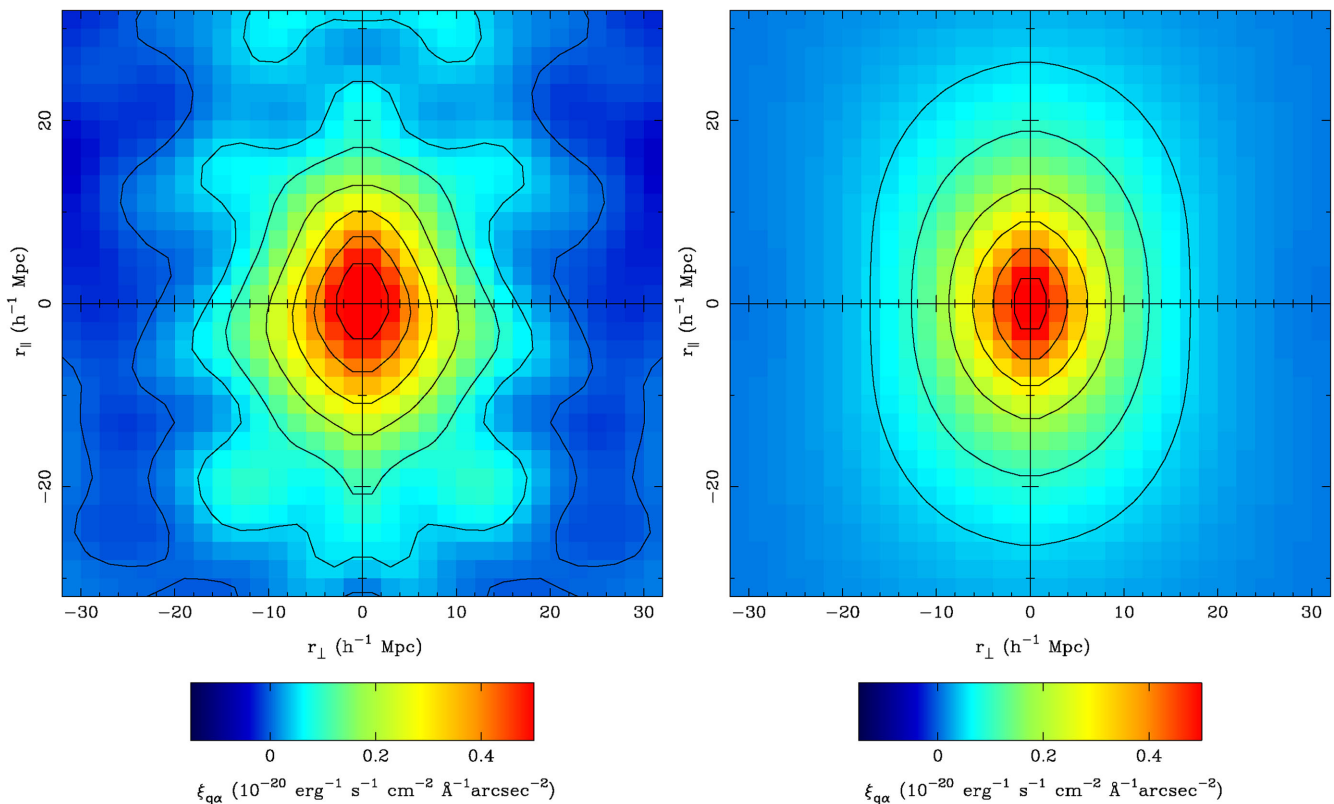
dispersion of quasars in their host haloes, or the intrinsic velocity dispersion of the sources of Ly  $\alpha$  emission.

Another source of apparent clustering anisotropy could be the radiative transfer effects predicted by Zheng et al. (2011a). It was shown by these authors, using cosmological radiation hydrodynamic simulations, that Ly  $\alpha$  radiative transfer has a strong environmental dependence which can cause the apparent spatial distribution of Ly  $\alpha$  emission to become anisotropic with respect to the line-of-sight direction. Density fluctuations along the line-of-sight direction are found to preferentially emit the Ly  $\alpha$  radiation in that direction in overdense regions, mainly because of the effect of peculiar velocity gradients on the Ly  $\alpha$  radiative transfer. This causes a suppression of the line-of-sight fluctuation, which can be modelled similarly to the Kaiser effect (also caused by the peculiar velocity gradient), even though the sign of the effect is opposite.

### 3.4 Fitting redshift-space distortions

In order to approximately quantify the level of distortion in Fig. 4 and its statistical significance, we have investigated fitting a redshift-space distortion model to the  $\xi_{q\alpha}(r_{\perp}, r_{\parallel})$  data. To compute the model for  $\xi_{q\alpha}(r_{\perp}, r_{\parallel})$ , we first assume the linear  $\Lambda$ CDM correlation function shape used in equation (3) and then use a model for peculiar velocities to distort it in redshift space. Our peculiar velocity model includes standard linear infall for large-scale flows (Kaiser 1987) and a small-scale random velocity dispersion (e.g. Davis & Peebles 1983).

The parametrization of the model for linear infall allows for stretching (outflow) as well as squashing (infall) along the line



**Figure 4.** Left-hand panel: the quasar–Ly  $\alpha$  cross-correlation  $\xi_{q\alpha}$  as a function of  $r_{\parallel}$  and  $r_{\perp}$ . The units (of Ly  $\alpha$  surface brightness) are the same as in Fig. 2. The contours are spaced at values of  $10^{-21} \text{ erg s}^{-1} \text{ cm}^{-2} \text{ \AA}^{-1} \text{ arcsec}^{-2}$ . To reduce noise in the image, the data set was smoothed with a Gaussian filter with  $\sigma = 4 h^{-1} \text{ Mpc}$  (2 cells) before plotting. Right-hand panel: the model fit to the quasar–Ly  $\alpha$  cross-correlation including redshift-space distortions (see Section 3.4).

of sight. Although gravitational processes are expected to result in infall, as mentioned above Zheng et al. (2011b) have shown that radiative transfer effects on the anisotropy of clustering can be approximately parametrized with the same model. We do this here, allowing a net linear outflow measured by the model to be interpreted as the radiative transfer effect.

The effects of coherent flows on the correlation function in linear theory were presented by Hamilton (1992). We use the formulation of Hawkins et al. (2003), with modifications to make it appropriate for the case of cross-correlation functions. This modification involves the use of the two bias factors  $b_q$  and  $b_\alpha$  from equation (3) to compute redshift-space distortion factors  $\beta_q = \Omega_m(z = 2.55)^{0.6}/b_q$  and  $\beta_\alpha = \Omega_m(z = 2.55)^{0.6}/b_\alpha$ . The linearly distorted quasar–Ly $\alpha$  emission cross-correlation function is then given by

$$\xi'_{q\alpha}(r_\perp, r_\parallel) = b_q b_\alpha \langle \mu_\alpha \rangle [\xi_0(s)P_0(\mu) + \xi_2(s)P_2(\mu) + \xi_4(s)P_4(\mu)], \quad (9)$$

where  $\mu = r_\parallel/r$ , and

$$\xi_0(s) = \left[ 1 + \frac{1}{3}(\beta_q + \beta_\alpha) + \frac{1}{5}\beta_q\beta_\alpha \right] \xi(r) \quad (10)$$

$$\xi_2(s) = \left[ \frac{2}{3}(\beta_q + \beta_\alpha) + \frac{4}{7}\beta_q\beta_\alpha \right] [\xi(r) - \bar{\xi}(r)], \quad (11)$$

$$\xi_4(s) = \frac{8}{35}\beta_q\beta_\alpha \left[ \xi(r) + \frac{5}{2}\bar{\xi}(r) - \frac{7}{2}\bar{\bar{\xi}}(r) \right], \quad (12)$$

with

$$\bar{\xi}(r) = \frac{3}{r^3} \int_0^r \xi(r')r'^2 dr', \quad (13)$$

$$\bar{\bar{\xi}}(r) = \frac{5}{r^5} \int_0^r \xi(r')r'^4 dr'. \quad (14)$$

Here  $\xi(r)$  is the linear  $\Lambda$ CDM correlation function of equation (3).

We use these relations to create a model  $\xi'_{q\alpha}(r_\perp, r_\parallel)$  which we convolve with the distribution function of random pairwise motions,  $f(v)$ , to produce the final model  $\xi_{q\alpha}(r_\perp, r_\parallel)$ :

$$\xi_{q\alpha}(r_\perp, r_\parallel) = \int_{-\infty}^{\infty} \xi'_{q\alpha} \left( r_\perp, r_\parallel - \frac{(1+z)v}{H(z=2.55)} \right) f(v) dv \quad (15)$$

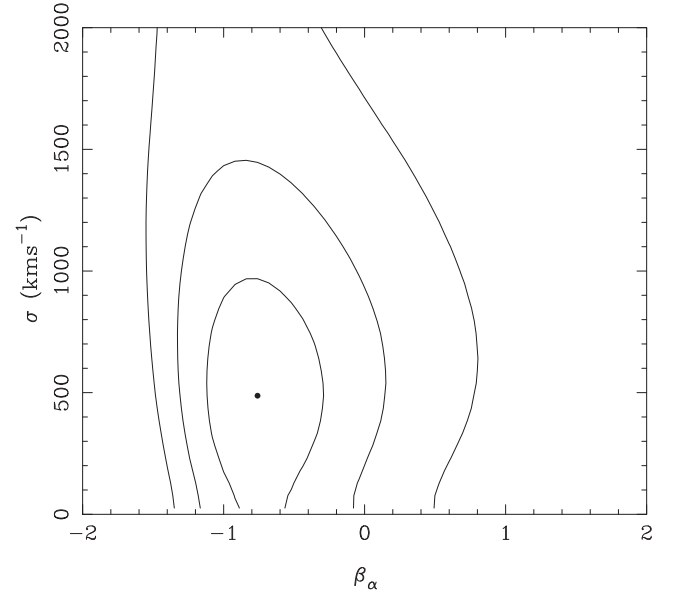
The random velocity dispersion we use is an exponential model, thus the distribution function of velocities is

$$f(v) = \frac{1}{\sigma\sqrt{2}} \exp\left(-\frac{\sqrt{2}|v|}{\sigma}\right), \quad (16)$$

where  $\sigma$  is the pairwise velocity dispersion of quasars and Ly $\alpha$  emission, which we assume to be independent of pair separation.

There is a strong degeneracy between  $\beta_q$  and  $\beta_\alpha$ , but the bias factor for BOSS quasars is reasonably well measured. The first BOSS measurement, of  $b_q = 3.6 \pm 0.6$ , was made using the quasar autocorrelation function by White et al. (2012). Font-Ribera et al. (2013), find an even more precise value of  $b_q = 3.64^{+0.13}_{-0.15}$  from the cross-correlation of BOSS quasars with the Ly $\alpha$  forest. Because of this we set  $b_q = 3.64$ , giving  $\beta_q = 0.27$ . The free parameters in our distortion model are therefore  $\beta_\alpha$ ,  $\sigma$  and an amplitude parameter  $b_q b_\alpha \langle \mu_\alpha \rangle$ .

We set the parameter governing the shape in our model,  $\Omega_m$ , equal to the *Planck* value,  $\Omega_m = 0.30$ . We compute  $\xi_{q\alpha}(r_\perp, r_\parallel)$  for our model for a grid of values of varying  $\beta_\alpha$ ,  $\sigma$  and  $b_q b_\alpha \langle \mu \rangle$ . We then compare our model to the observed  $\xi_{q\alpha}(r_\perp, r_\parallel)$  from



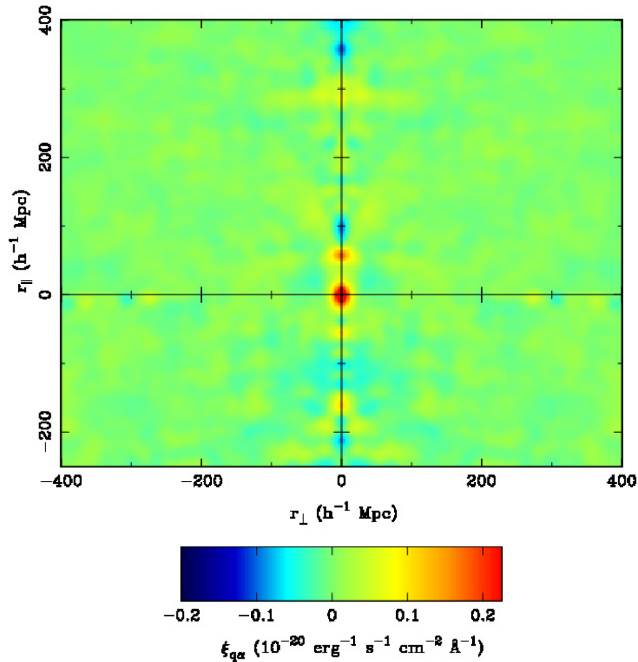
**Figure 5.**  $1\sigma$ ,  $2\sigma$  and  $3\sigma$  confidence contours for the redshift-space distortion parameters  $\beta$  and  $\sigma$  found from the fit to the quasar–Ly $\alpha$  emission cross-correlation (Section 3.4). The best-fitting values of  $\beta$  and  $\sigma$  are shown with a dot.

Fig. 4, performing a  $\chi^2$  fit to all points within  $r = 40 h^{-1}$  Mpc. We again use jackknife error bars computed from 100 subsamples, but because of difficulties with noisy matrix inversion, we do not use the off-diagonal ( $C_{i \neq j}$ ) terms when inverting the covariance matrix. We marginalize over the amplitude parameter  $b_q b_\alpha \langle \mu_\alpha \rangle$ , and show the confidence contours (for the two remaining degrees of freedom) in Fig. 5.

We have allowed the  $\beta_\alpha$  parameter to be negative in our fit not because we believe that outflow of Ly $\alpha$  emission is likely around quasars but, as mentioned above, because this allows quantifying the stretching along the line of sight seen in Fig. 4 and its significance. Our best-fitting values with  $1\sigma$  error bars are  $\beta_\alpha = -0.76 \pm 0.36$  and  $\sigma = 490 \pm 300 \text{ km s}^{-1}$ . From these values and by observing the contours in Fig. 5, we infer a detection of anisotropies in the quasar–Ly $\alpha$  emission correlation function that is opposite in sign to that expected from peculiar velocity flows due to gravitational evolution: our constraint on  $\beta_\alpha$  is  $2.1\sigma$  from  $\beta_\alpha = 0$ . We interpret this result as indicating that there are strong non-gravitational effects on the Ly $\alpha$  emission causing the elongation of the cross-correlation contours along the line of sight extending to large separations; the good fit obtained with the redshift-space distortion model with negative  $\beta$  needs to be understood in this case as a coincidence, since the model is not physically correct.

If we limit the fit to points with  $20 h^{-1}$  Mpc, we find central values for  $\beta_\alpha$  and  $\sigma$  that are consistent with those for our fiducial fit. As might be expected, the error bars are larger, however, by a factor of 40 per cent for  $\sigma$  and 65 per cent for  $\beta_\alpha$ .

If we assume a Ly $\alpha$  emission bias factor  $b_\alpha = 3$ , corresponding to highly biased star-forming galaxies, (see Section 5 for further discussion of this value), then  $\beta_\alpha = 0.32$  at redshift  $z = 2.55$ . For the expected value of  $\sigma$ , we can use as a guide the results of Font-Ribera et al. (2013) who constrained  $\sigma < 370 \text{ km s}^{-1}$  at the  $1\sigma$  confidence level from the redshift-space quasar–Ly $\alpha$  forest cross-correlation function. We can see from Fig. 5 that although the  $\sigma$  measurement is consistent with Font-Ribera et al.,  $\beta_\alpha$  and  $\sigma$  considered jointly disagree at the  $2.5\sigma$  level from our measurement.



**Figure 6.** The quasar–Ly $\alpha$  cross-correlation  $\xi_{q\alpha}$  as a function of  $r_{\parallel}$  and  $r_{\perp}$ . This figure shows the same information as Fig. 4, except over a larger range of scales (see text).

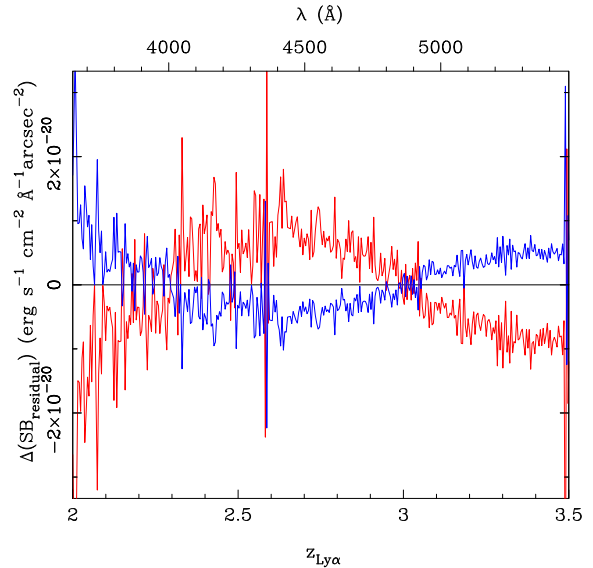
We discuss further in Section 6 how our measurements can be interpreted as a detection of clustering anisotropies due to radiative transfer effects (Zheng et al. 2011b). In this case, elongation along the line of sight is expected, which can explain the effective measurement of a negative  $\beta_{\alpha}$  parameter.

The cross-correlation contours of the best-fitting model are plotted in Fig. 4 (right-hand panel), which also reveals stretching along the line of sight. The  $\chi^2$  value for the fit is 610 measured from 400 bins, with three free parameters, a reduced  $\chi^2$  of 1.5. The fit is therefore not good, and the discrepancy arises in large part in the central region, where the model has lower surface brightness than the observations. This result may be a sign that adding a one-halo term to the correlation function would provide a better fit, and should be addressed in future work with a larger data sample.

### 3.5 Large-scale tests

The cross-correlation across and along the line of sight over the whole spectrum offers a way to test whether the detected signal is reasonable. One can search for any significant cross-correlation signal if a different wavelength other than Ly $\alpha$  is used, which would indicate either contributions from other emission lines, or that other effects are causing the cross-correlation signal measured. An equivalent approach is to extend the line-of-sight range of the cross-correlation to large distances. Fig. 6 extends the contours shown in Fig. 4 to much larger scales. The positive signal seen in Fig. 4 is the most significant feature, centred at  $r_{\parallel} = 0$ ,  $r_{\perp} = 0$ . This is a good indication that Ly $\alpha$  emission is the dominant contribution to our signal. Signal from lines at longer (shorter) wavelengths than Ly $\alpha$  would appear at positive (negative) values of  $r_{\parallel}$ .

The second most prominent feature, at  $r_{\perp} = 0$ ,  $r_{\parallel} \sim 60 h^{-1}$  Mpc, is significant at the  $\sim 1.5\sigma$  level (the pixel at the centre of the feature is  $1.5\sigma$  from the zero level, using the jackknife error bars from Section 3.4), and so is consistent with noise. Strong lines that might be an



**Figure 7.** Average residual flux for two subsamples of LRG spectra selected to be close to the centre (blue line) or edge (red line) of the spectrograph camera, after subtracting the residual flux for the whole (fiducial) sample plotted as the black line in Fig. 1. The residual flux is that left after subtraction of the best-fitting galaxy model from each spectrum.

issue, such as Lyman  $\beta$  and Carbon IV, are very far away (at  $r_{\parallel} = -490$  and  $+640 h^{-1}$  Mpc, respectively) and so are not a concern. Si III would appear at  $r_{\parallel} = -22 h^{-1}$  Mpc if it was present.

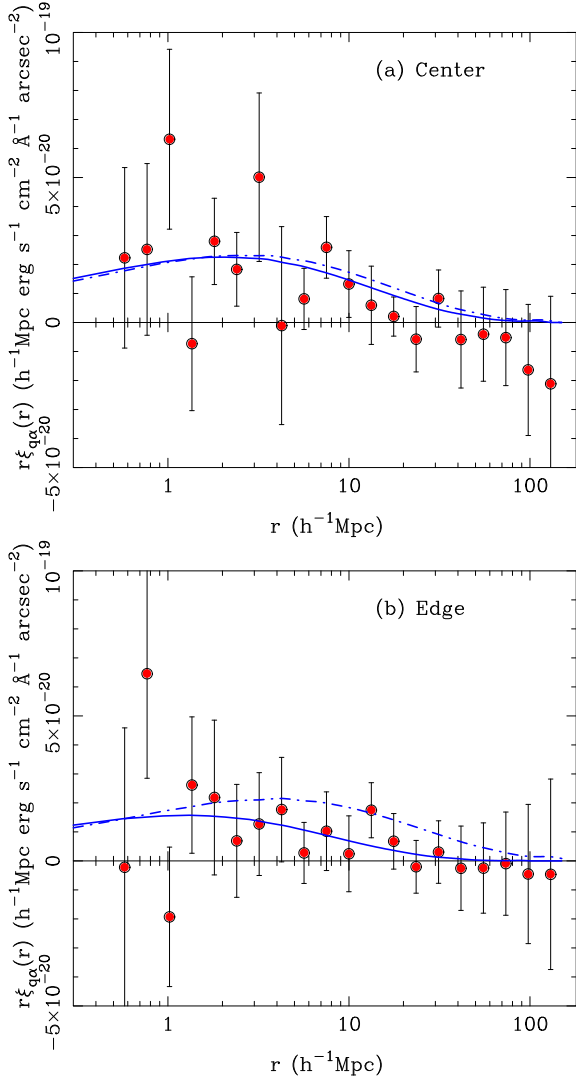
### 3.6 Flux cross-correlation and fibre position in spectrograph

As mentioned above, Appendix C contains tests of quasar–Ly $\alpha$  cross-correlation based on dividing our sample depending on the luminosity or redshift of the foreground LRGs and quasars. Other sample tests can be designed based on observational parameters, which could help identify sources of systematic error. We have found one test in particular that shows a possible contamination of our results due to the variable optical performance of the spectrograph across a plate. Due to the possible importance of this effect we report this in the main body of the paper, while other tests are described in the appendices.

The test itself involves splitting the sample into two subsets based on fibre positions in the spectrograph. Of the 1000 fibres in each plate, those labelled with numbers 100–400 and 600–900 are positioned towards the centre of the spectrograph and the rest are positioned more towards the edge. The optical quality of the instrument cameras degrades towards the edge, which can affect the sensitivity and sky background, possibly causing systematic errors through effects on the quasar selection function and the residual light in galaxy spectra.

We carry out this split, and find that the average residual flux after subtraction of the model LRG spectra is somewhat different for the two subsamples. The residual flux for the fiducial sample, plotted in Fig. 1, roughly fluctuates within  $\pm 3 \times 10^{-20} \text{ erg s}^{-1} \text{ cm}^{-2} \text{ \AA}^{-1} \text{ arcsec}^{-2}$  for most of the relevant redshift range. In Fig. 7, we plot for each subsample (edge and centre)  $\Delta(\text{SB}_{\text{residual}})$ , the residual flux for the subsample minus the residual flux for the fiducial sample. We can see that there are indeed differences: the residual flux in fibres close to the edge of the cameras is systematically higher than the fiducial sample (and the centre of camera sample) by  $\sim 10^{-20} \text{ erg s}^{-1} \text{ cm}^{-2} \text{ \AA}^{-1} \text{ arcsec}^{-2}$  over the



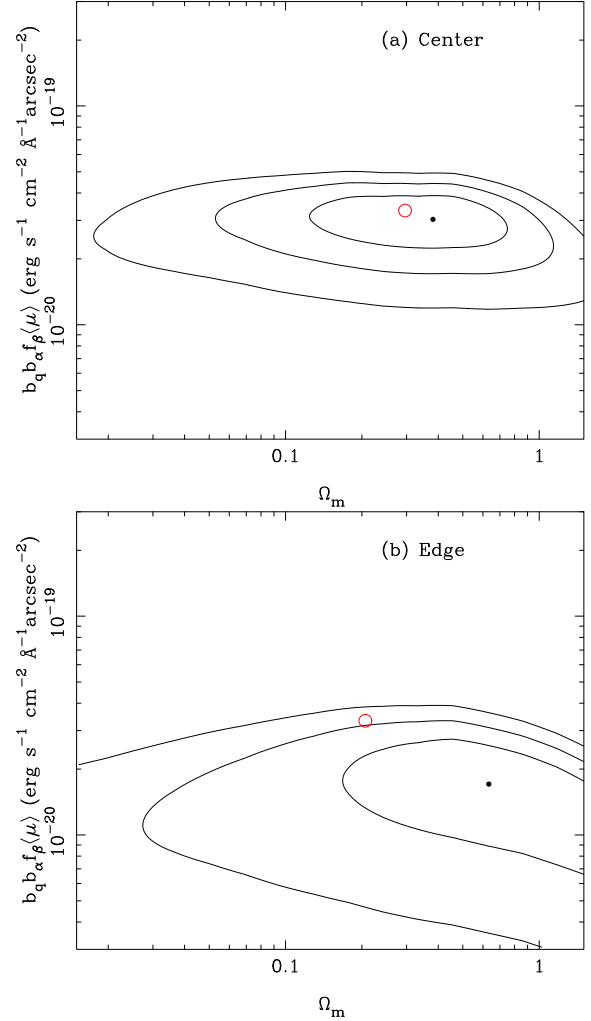


**Figure 8.** A test of fibre position in the field of view on the results for the quasar–Ly  $\alpha$  emission cross-correlation function  $\xi_{q\alpha}(r)$  (see equation 1). The results using fibres in the central region of field of view are shown in panel (a), and those on the edge (where the optical quality of the spectrograph camera degrades) are shown in panel (b). The smooth curve is the best-fitting linear CDM correlation function (see Section 3.2) and the dash–dotted line is the best-fitting CDM curve for the fiducial sample.

wavelength range  $\sim 4000\text{--}5000 \text{ \AA}$ . As this covers the important Ly  $\alpha$  redshift range  $z = 2.3\text{--}3.0$ , this could indeed have consequences for our measurements.

Using the appropriate mean residual flux for each subsample, we compute the quasar–Ly  $\alpha$  emission cross-correlation  $\xi_{q\alpha}$ . It should be noted that the subsamples are not quite the same size (there are 400 ‘edge’ fibres per plate versus 600 ‘centre’ fibres). We show the  $\xi_{q\alpha}$  results in Fig. 8, where we can see by eye that the data points appear to be approximately consistent with the fiducial results.

We fit the usual CDM model parameters. The contours are shown in Fig. 9. We find  $b_q b_{\alpha} f_{\beta}(\mu) = 3.1^{+0.5}_{-0.5} \times 10^{-20} \text{ erg s}^{-1} \text{ cm}^{-2} \text{ \AA}^{-1} \text{ arcsec}^{-2}$  and  $\Omega_m = 0.38^{+0.21}_{-0.21}$  for the central fibres, and  $b_q b_{\alpha} f_{\beta}(\mu) = 1.7^{+0.7}_{-0.8} \times 10^{-20} \text{ erg s}^{-1} \text{ cm}^{-2} \text{ \AA}^{-1} \text{ arcsec}^{-2}$  and  $\Omega_m = 0.63^{+0.85}_{-0.35}$  for the edge fibres. The central fibre results are within  $1\sigma$  of the fiducial results, but the edge fibre parameters have larger errors and are  $2\sigma$  below the fiducial results. As the edge



**Figure 9.** The effect of fibre position in the field of view on the shape and amplitude of the quasar–Ly  $\alpha$  cross-correlation. We show the fit parameters for the amplitude  $b_q b_{\alpha} f_{\beta}(\mu)$  and shape  $\Omega_m$  of a linearly biased CDM model fit to the Ly  $\alpha$  cross-correlation function plotted in Fig. 14. Panel (a) is for fibres in the central region of the field of view and panel (b) shows results for fibres on the edge (where the optical quality of the spectrograph camera degrades). The dots indicate the best-fitting parameters and the contours show the  $1\sigma$ ,  $2\sigma$  and  $3\sigma$  confidence contours on the fit parameters. The open circles show the best fit to the fiducial results (from Fig. 3).

and centre subsample sizes are different, simple Poisson statistics would predict that the error bars should be related by a factor of  $\sqrt{60/40} = 1.22$ . The error bars on the edge sample are instead larger by a factor of 1.5. We note also that the average result of the two subsamples does not equal the result of the whole, fiducial sample, which can be due to several effects, including fluctuations in the quasar redshift distribution among the two subsamples and the way the fit responds to noise in the measured cross-correlation. We have tried averaging the  $\xi_{q\alpha}$  results from the two subsamples before fitting and find fit parameters consistent at the  $1\sigma$  level with the fiducial results. The difference between the subsamples is not statistically significant, so we cannot be conclusive that it indicates the presence of a systematic error, but it is nevertheless a warning about a possible contamination of our detection.

We generally caution that inhomogeneities in the residual galaxy flux similar to those seen in Fig. 7 might cause a spurious clustering signal if they are correlated with inhomogeneities in the quasar

**Table 1.** The amplitude parameter  $b_q b_{\alpha f \beta}(\mu)$  for different redshift bins.

$\langle z \rangle$	$z_{\min}$	$z_{\max}$	$b_q b_{\alpha f \beta}(\mu)$ (using $\Omega_m = 0.30$ ) ( $10^{-20}$ erg s $^{-1}$ cm $^{-2}$ Å $^{-1}$ arcsec $^{-2}$ )
2.20	2.0	2.29	$5.0^{+1.1}_{-1.2}$
2.37	2.29	2.46	$4.0^{+0.8}_{-0.9}$
2.59	2.46	2.75	$3.4^{+0.8}_{-0.9}$
3.04	2.75	3.50	$2.0^{+0.8}_{-0.8}$

selection as a function of redshift. We believe this is not a strong effect because our signal is detected on scales much smaller than the plate size, where these effects are unlikely to occur, and because the various tests in the appendix are consistent with a cosmological origin of the signal. Moreover, as we see below, there is also no strange dependence of the cross-correlation amplitude on the quasar redshift.

### 3.7 Evolution with redshift

The redshift coverage of our data sample ( $z = 2-3.5$ ) is sufficient that we can separate it into different bins in redshift and search for evolution. We do this for four different redshift bins where each bin contains one quarter of the quasar data. The bin boundaries and mean quasar redshifts of each bin are given in Table 1.

The quasar–Ly  $\alpha$  emission cross-correlation results are shown in the four panels of Fig. 10. The global CDM model fit to the full sample averaged over all redshifts is indicated as the dash-dotted line in every panel. Although the results for the redshift bins are relatively noisy, as expected, they are broadly similar to the global result in shape and amplitude and show no clear evidence for any redshift evolution.

Fig. 11 presents the best-fitting CDM shape and amplitude parameters from Section 3.2, along with confidence contours. The fitting was again performed using jackknife error bars. Fig. 11 demonstrates that the fiducial model results all lie within the  $2\sigma$  confidence contours for the different redshift bins.

Assuming that the shape of the cross-correlation function remains fixed in comoving coordinates (as it would do if governed by linear biasing), we can search for changes in the amplitude of clustering and the mean Ly  $\alpha$  surface brightness as a function of redshift. We set  $\Omega_m = 0.30$  (the CDM shape determined by the *Planck* results; Planck Collaboration XVI et al. 2014) and then determine the best-fitting amplitude parameters,  $b_q b_{\alpha f \beta}(\mu)$ , at each redshift bin. The result, shown in Fig. 12, indicates a decreasing cross-correlation amplitude with redshift, although the errors are large and a horizontal line would not be an unreasonable fit ‘by eye’. To express this quantitatively, we have carried out a  $\chi^2$  fit to the function  $\log(b_q b_{\alpha f \beta}(\mu)) = a + bz$ , finding a slope parameter  $b = -0.40 \pm 0.20$ , meaning that the hint of redshift evolution is significant at the  $2.0\sigma$  level only. The values for the fit parameters  $b_q b_{\alpha f \beta}(\mu)$  for the different redshift bins (which were used to plot Fig. 11) are listed in Table 1.

We have also looked at redshift evolution of the CDM shape governed by the parameter  $\Omega_m$ . Within the assumption of linear biasing, the shape should remain constant with redshift. The results are examined in Fig. 13, where the results are indeed consistent with a constant  $\Omega_m$ , within the uncertainties. The *Planck* value ( $\Omega_m = 0.30$ ) is also shown and is consistent with our results.

## 4 EMISSION LINES

Given that the sources of the clustered Ly  $\alpha$  emission seen in Section 3 could be discrete objects such as star-forming galaxies, we must investigate see if individual emission lines can be detected in our spectra and whether discrete detectable lines can account for the signal. The BOSS spectra have relatively short integrations on a small (2.5 m) telescope, and cannot be expected to compete in individual detections with other surveys such as that described in Cassata et al. (2011): we can only detect the most luminous objects. However, our cross-correlation technique enables us to find the mean total surface brightness, which includes all emission line objects no matter how faint they may be. The difference between the cross-correlation signal with and without individually identified lines therefore enables us to discover what fraction of the Ly  $\alpha$  surface brightness lies below our line detection limit.

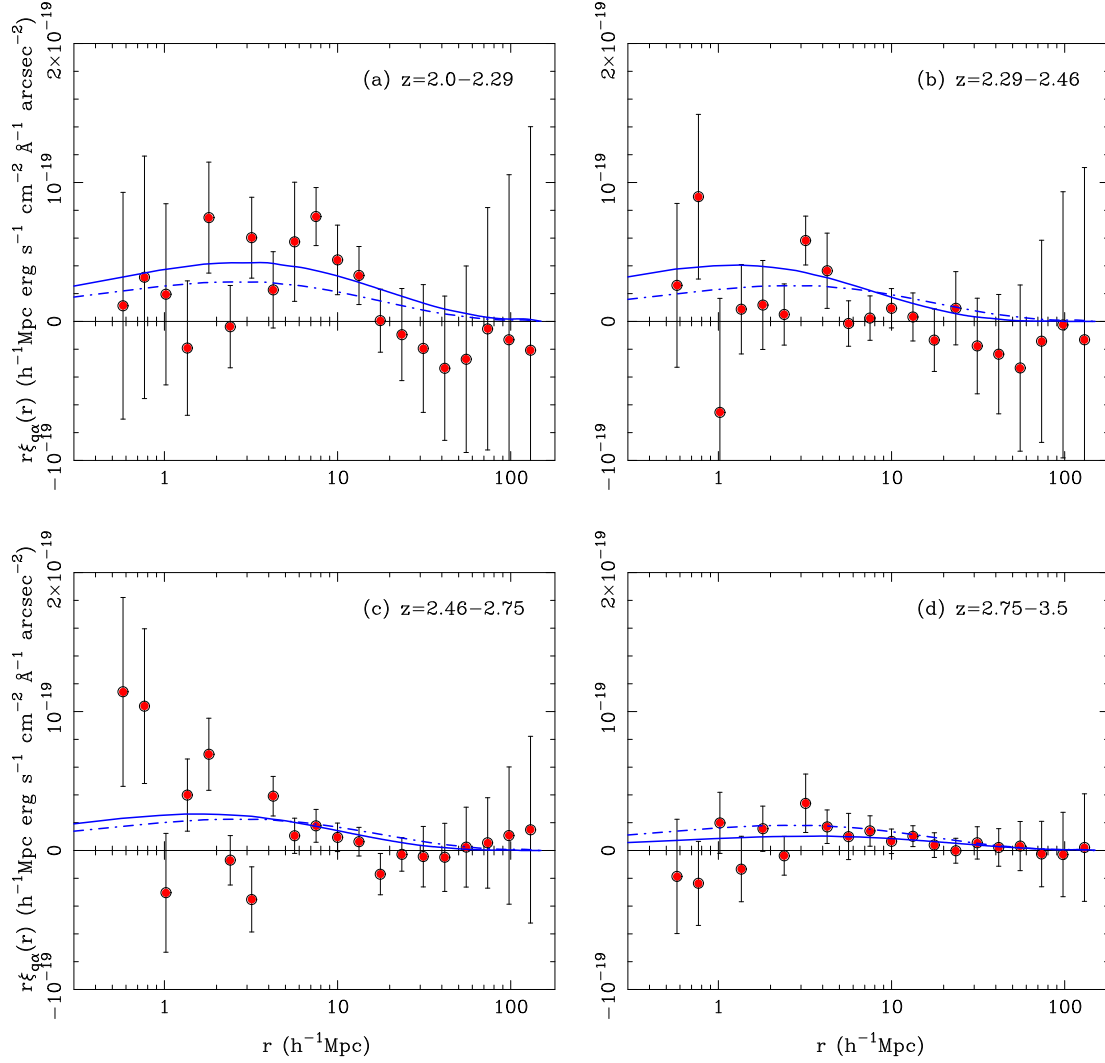
We note that our line detection procedure is less sophisticated than that in the BELLS survey (Brownstein et al. 2012), which used line detections of galaxies behind LRGs to find gravitational lenses. In particular, we are not seeking confirmed detections of objects (which requires multiple emission lines) and we do not deal with interlopers, except statistically.

### 4.1 Line fitting

For each LRG spectrum (see Section 2.1), we subtract the best-fitting galaxy spectrum model, as we do in our fiducial cross-correlation analysis. We then fit lines to this residual flux, centring our fitted line profile on the centre of each spectrum pixel, one at a time. In this first stage, each spectrum pixel is therefore the centre of a best-fitting line profile – we remove overlapping lines later. For each of a grid of values of amplitude  $A$  and line width  $\sigma$  (between 0 and 20 Å), we compute the  $\chi^2$  value of a positive Gaussian emission line profile  $G(x)$ , where the profile has the form  $G(x) = A \exp[-x^2/(2\sigma^2)]$  and  $x$  is the separation between the line centre and the pixels we include in our fit. We use pixels in our fit that are in a 40 Å region centred on the line centre, excluding masked regions as in Section 2.1.1. Once the best-fitting values of  $A$  and  $\sigma$  are found, we estimate the significance of each fitted line from the  $\chi^2$  difference between the line fit and a flat interval with zero flux. After fitting to all the pixels we pass through the list and eliminate overlapping lines, removing the lower significance line when there is an overlap.

We find 5200 lines with a nominal significance of  $5\sigma$  ( $\Delta\chi^2 > 30.1$  for the 2 degrees of freedom fitted), and  $1.6 \times 10^6$  lines with a nominal significance of  $3\sigma$  ( $\Delta\chi^2 > 11.8$ ). There are  $1.3 \times 10^9$  pixels in the search regions of the spectra. The detected lines are constrained to be at least 40 Å (37 pixels) apart, but for rare lines this should not change the random expectation. One would therefore expect to find approximately 370 and  $1.8 \times 10^6$   $5\sigma$  and  $3\sigma$  lines, respectively, from positive noise fluctuations alone.

For the  $3\sigma$  lines, the fact that we have detected fewer lines than even pure noise fluctuations predict is likely to be a sign that the fluctuations do not exactly obey a Gaussian noise model. An additional complication is that the noise estimate from the standard data pipeline which we have used has been shown to be underestimated by up to 16 per cent for the relevant wavelengths (Palanque-Delabrouille et al. 2013). The detection of more  $5\sigma$  lines than randomly expected is likely to indicate that there are false detections arising from unsubtracted features in the galaxy spectra, sky lines we have not accounted for, and other systematics. There is also the possibility of interloping [O II] emission lines from lower redshift



**Figure 10.** The quasar–Ly  $\alpha$  emission cross-correlation function  $\xi_{q\alpha}(r)$  for different redshift ranges, which correspond to one quarter of the full data set each. The solid lines indicate the best-fitting CDM model at each redshift bin and the dash–dotted line represents the best fit to the full data set (shown in Fig. 3). Both the shape and amplitude of the model fit are consistent at the  $2\sigma$  level with no evolution over the redshift range (this is shown in Figs 12 and 13).

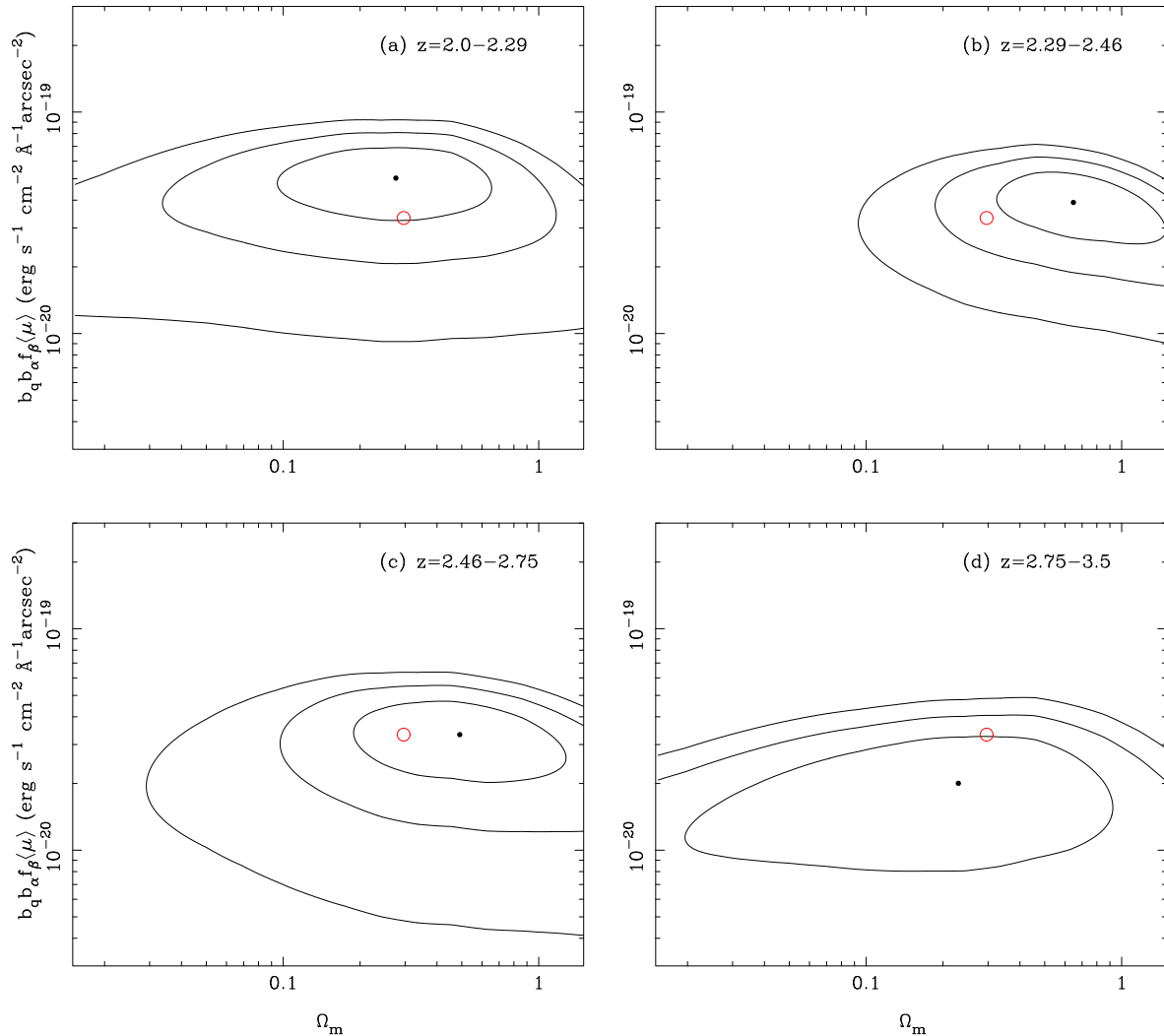
galaxies (see Noterdaeme et al. 2010; Menard et al. 2011). For our wavelength coverage of 3800–5500 Å, this interloper emission would arise from between  $z = 0.02$  and 0.48. Without a significant additional effort to remove false detections and interlopers, our data set is not useful for computing the luminosity function of Ly  $\alpha$  lines. Instead, we turn to the statistical cross-correlation to test for the fraction of these lines which are really Ly  $\alpha$  emission lines in our redshift interval.

#### 4.2 Cross-correlation

We subtract the flux in the lines detected in Section 4.1 from each LRG spectrum, and then recalculate the cross-correlation of quasars and Ly  $\alpha$  emission (equation 1). The results for  $\xi_{q\alpha}(r)$  (again computing the error bars using a jackknife estimator and 100 subsamples) are shown in Fig. 14, using our two thresholds on the significance of the removed lines,  $5\sigma$  and  $3\sigma$ . We can see that in each case, the clustering signal is still visible and the shape traces that of a CDM curve. This shows that most of the surface brightness of Ly  $\alpha$  emission is not accounted for by these lines. As expected, most of the lines are due to noise features. By subtracting these

lines, however, we are also subtracting any possible real lines, and the change in the amplitude of  $\xi_{q\alpha}(r)$  is a measure of the fraction of surface brightness that is actually contributed by strong lines.

The shape and amplitude fitting parameters ( $\Omega_m$  and  $b_q b_{\alpha f_{\beta}}(\mu)$ ) for these two cases ( $>5\sigma$  and  $>3\sigma$  lines subtracted) are shown in Fig. 15. The shape parameter  $\Omega_m$  is very similar in the two cases and almost the same as in the fiducial case. The amplitude is lower, as would be expected for subtraction of some real lines, but the fiducial result (with no line subtraction) lies well within the  $1\sigma$  error contour of both of the panels in Fig. 15, implying that the contribution to the cross-correlation from emission lines that are detected and removed is not statistically significant. Quantitatively, this can be seen by considering that we find the amplitude parameter for the  $>5\sigma$  case to be  $b_q b_{\alpha f_{\beta}}(\mu) = 3.18^{+0.39}_{-0.41} \times 10^{-20} \text{ erg s}^{-1} \text{ cm}^{-2} \text{ \AA}^{-1} \text{ arcsec}^{-2}$ , and for the  $3\sigma$  case to be  $b_q b_{\alpha f_{\beta}}(\mu) = 2.89^{+0.43}_{-0.37} \times 10^{-20} \text{ erg s}^{-1} \text{ cm}^{-2} \text{ \AA}^{-1} \text{ arcsec}^{-2}$ . The amplitude parameter is therefore  $4 \pm 12$  per cent and  $13^{+11}_{-13}$  per cent lower than the fiducial case for the  $>5\sigma$  and  $>3\sigma$  line removal cases, but both of these are consistent with zero within the errors. The analysis is therefore consistent with our line fitting having found no true Ly  $\alpha$  emission lines at all.



**Figure 11.** Likelihood contours for power spectrum parameters  $\Omega_m$  (which we are using to parametrize the CDM shape) and  $b_q b_{\alpha} f_{\beta} \langle \mu \rangle$  (the amplitude) for different redshift ranges. The filled dot shows the best-fitting values for that redshift range and the open circles the best-fitting values for the full data set (shown in Fig. 3).

In order to relate the significance levels to line luminosity, we have computed the luminosity from the surface brightness for each line (bearing in mind that our measurements are restricted to a 1 arcsec radius fibre aperture). We find that the median luminosity of the  $>5\sigma$  lines is  $L = 9.0 \times 10^{42} \text{ erg s}^{-1}$  and the  $>3\sigma$  lines have a median luminosity  $L = 1.9 \times 10^{42} \text{ erg s}^{-1}$ . We can compare these luminosities measured with some published values from Ly  $\alpha$  emitter surveys. The flux limit of the Guaita et al. (2010) data sample was  $2 \times 10^{-17} \text{ erg s}^{-1} \text{ cm}^{-2}$  (emission line flux) at  $z=2.1$ . This corresponds to a Ly  $\alpha$  luminosity of  $5 \times 10^{41} \text{ erg s}^{-1}$ . For the Gawiser et al. (2007) sample at  $z=3.1$ , the line flux limit was  $1.5 \times 10^{-17} \text{ erg s}^{-1} \text{ cm}^{-2}$ , corresponding to a line luminosity of  $1.3 \times 10^{42} \text{ erg s}^{-1}$ . Therefore, if any of our  $3\sigma$  lines were real, they would have luminosities just above those of the Gawiser et al. and Guaita et al. samples, although our small aperture implies that the intrinsic luminosity of our fitted line emitters would likely be even higher.

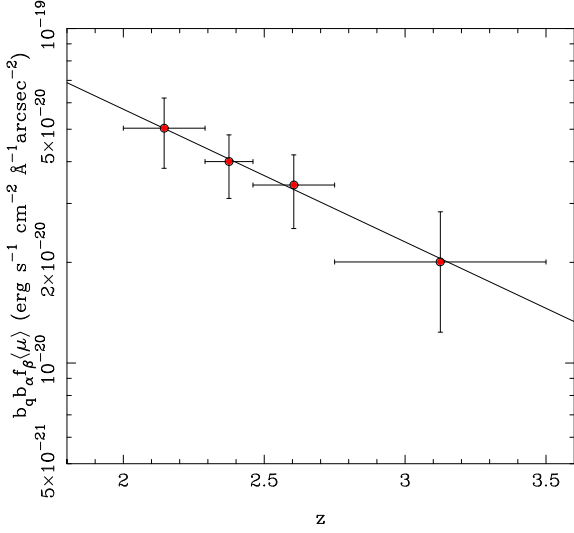
Before turning to the interpretation of the signal of Ly  $\alpha$  surface brightness around quasars, we present one further test of the cross-correlation, using the line-subtracted spectra. After subtraction of either the  $3\sigma$  or  $5\sigma$  significance lines, we have carried out a fit of the linear redshift-space distortion model of Section 3.4 to the

$\xi_{q\alpha}(r_{\perp}, r_{\parallel})$  results. We find that the  $\xi_{q\alpha}$  contours have a similar appearance to the results without line subtraction. The fit parameters with  $1\sigma$  confidence intervals are  $\beta_{\alpha} = -0.76 \pm 0.32$  and  $\sigma < 440 \text{ km s}^{-1}$  (after subtraction of  $5\sigma$  lines) and  $\beta_{\alpha} = -0.76^{+0.64}_{-0.48}$  and  $\sigma < 480 \text{ km s}^{-1}$  (after subtraction of  $3\sigma$  lines). The central values of  $\beta_{\alpha}$  are the same as for the fiducial case, and the similarity of the results can be taken as an additional sign that the detection of redshift-space elongation in Section 3.4 is reliable.

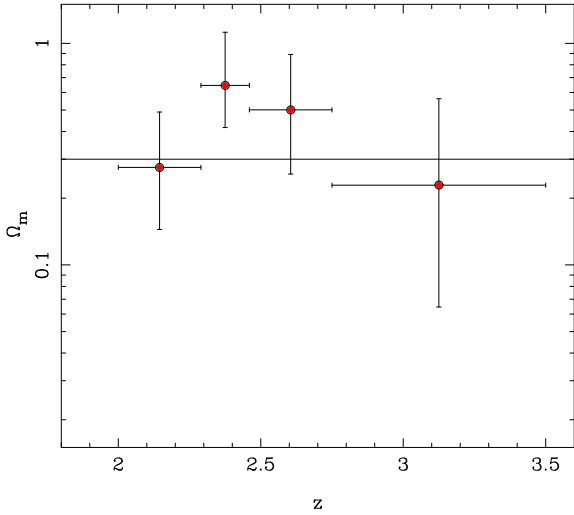
## 5 STAR FORMATION RATE DENSITY

In this section, we discuss the consequences of assuming that the signal we report is due to star-forming galaxies that trace the same large-scale structure as quasars. This assumption is not necessarily the one that can best explain what we measure: in particular, note that the cross-correlation signal arises from the small fraction of the volume of the BOSS survey that is within a comoving distance of  $\sim 15 h^{-1} \text{ Mpc}$  of a quasar (see Fig. 2). Other possibilities are discussed at the end of Section 6.





**Figure 12.** The evolution of  $b_q b_{\alpha} f_{\beta}(\mu)$  with redshift. The results from Fig. 10 are used, but here we have fixed the CDM shape to one for a  $\Lambda$ CDM model with  $\Omega_m = 0.30$  (see Section 3.2). The error bars are  $1\sigma$  and are estimated using the maximum likelihood fit to the amplitude of the CDM model. The solid line is a log-linear fit to the data (see text).

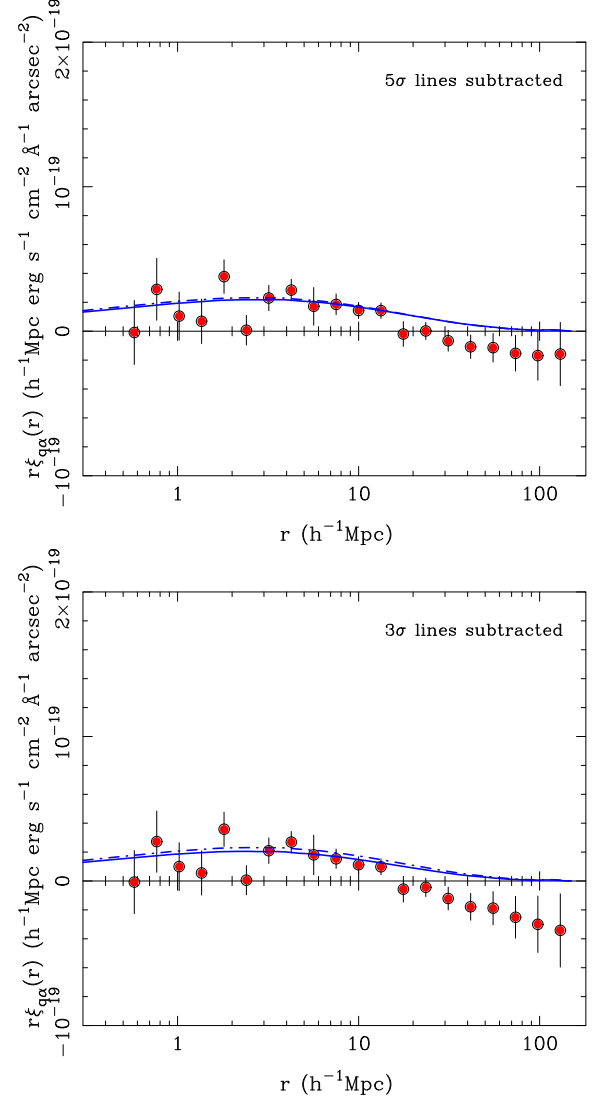


**Figure 13.** The value of  $\Omega_m$  (which we are using to parametrize the shape of the  $\Lambda$ CDM correlation function, holding other parameters fixed) versus redshift. The results from Fig. 10 were used. The solid line is the best fit to the *Planck* results (Planck Collaboration XVI et al. 2014),  $\Omega_m = 0.30$ .

Traditionally, narrow-band surveys have been used to detect Ly  $\alpha$  emitters, compute their luminosity function and integrate it to compute a mean Ly  $\alpha$  luminosity density  $\epsilon_{\alpha}$  (e.g. Gronwall et al. 2007), which can be converted into a star formation rate using a relationship such as

$$\text{SFR}/(M_{\odot} \text{ yr}^{-1}) = L_{\alpha}/(1.1 \times 10^{42} \text{ erg s}^{-1}) \quad (17)$$

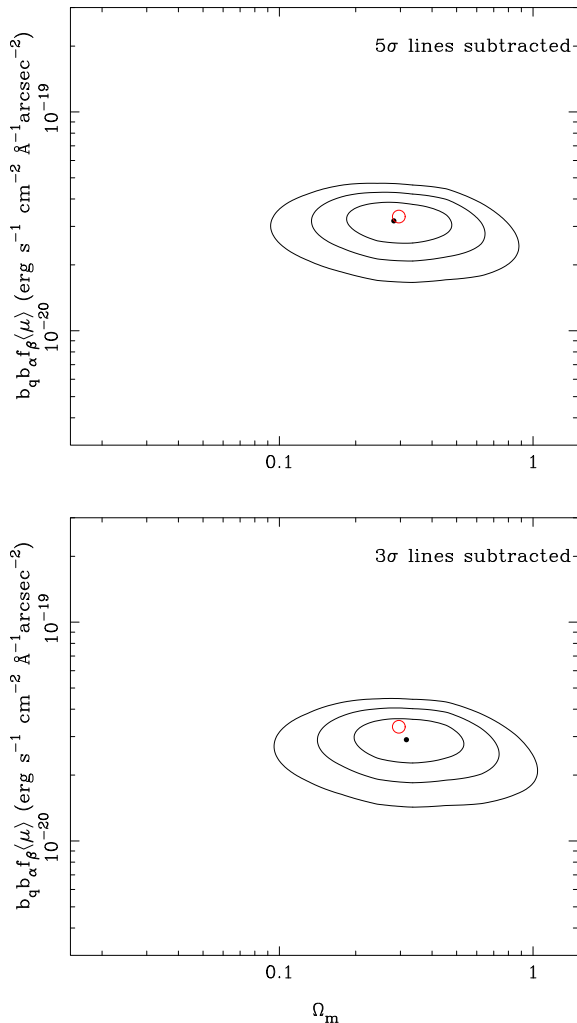
(Cassata et al. 2011), where  $L_{\alpha}$  is the Ly  $\alpha$  luminosity. The conversion factor is based on a stellar population with a Salpeter initial mass function (IMF) and with no correction for effects like dust and escape fraction, which is accurate to within a factor of a few for a range of population age, high mass cutoff of stars and metallicity (Leitherer et al. 1999).



**Figure 14.** The quasar–Ly  $\alpha$  emission cross-correlation function  $\xi_{q\alpha}(r)$  (see equation 1), as in Fig. 2, but computed after subtracting emission lines that are apparently detected in the spectra at the  $5\sigma$  significance level in panel (a), and the  $3\sigma$  level in panel (b). The smooth curve is the best-fitting linear CDM correlation function (see Section 3.2) and the dash–dotted line is the best-fitting CDM curve for the fiducial sample (i.e. before subtracting the apparent emission lines).

This method assumes that the surveys of Ly  $\alpha$  emitters are able to capture all the radiation from young stars. However, these surveys can only detect the high surface brightness portion of sources within a small angular aperture, and may be missing much of the Ly  $\alpha$  line intensity when it is scattered far out into the galaxy halo. In our case, the statistical cross-correlation technique we are using should not be affected by any threshold in Ly  $\alpha$  surface brightness. We should therefore be able to compute the total star formation rate density from our measurement. One large uncertainty is absorption due to dust, which is known to significantly affect UV continuum and line estimators of star formation.

We recall that our measurement is of the quantity  $b_q b_{\alpha} f_{\beta}(\mu)$ , so to compute the Ly  $\alpha$  surface brightness we need to have independent knowledge of  $b_q$ ,  $b_{\alpha}$  and  $f_{\beta}$ . For the quasar bias factor, we use the



**Figure 15.** The effect of removing strong lines on the shape and amplitude of the quasar–Ly $\alpha$  cross-correlation. We show the fit parameters for the amplitude  $b_q b_{\alpha} f_{\beta} \langle \mu_{\text{Ly}\alpha} \rangle$  and shape  $\Omega_m$  of a linearly biased CDM model fit to the Ly $\alpha$  cross-correlation function plotted in Fig. 14. Panel (a) is after removing  $>5\sigma$  significance lines and panel (b)  $>3\sigma$  significance lines. The dots indicate the best-fitting parameters and the contours show the  $1\sigma$ ,  $2\sigma$  and  $3\sigma$  confidence contours on the fit parameters. The open circles show the best-fitting results to the fiducial results (from Fig. 3).

value measured for BOSS quasars by Font-Ribera et al. (2013),  $b_q = 3.64^{+0.13}_{-0.15}$ . The bias factor  $b_{\alpha}$  is related to a luminosity-weighted bias factor  $b_L$ , from the definitions in equations (5)–(6) ( $b_{\alpha} = b_L$  in the absence of radiative transfer effect), and  $b_L$  is different from the number weighted bias factor of Ly $\alpha$  emitters,  $b_{\text{LAE}}$ . To understand the difference in the values of the two bias factors, we start with the following simple model for the Ly $\alpha$  emission.

If we assume that there are  $\langle N(M) \rangle$  galaxies per dark matter halo of mass  $M$  and that Ly $\alpha$  emission comes from galaxies in haloes above a mass limit  $M_{\text{min}}$ , then the spatial bias can be computed as follows (e.g. Berlind & Weinberg 2002):

$$b_{\text{LAE}} = \int_{M_{\text{min}}}^{\infty} b_h(M) \langle N(M) \rangle \frac{dn}{dM} dM \bigg/ \int_{M_{\text{min}}}^{\infty} \langle N(M) \rangle \frac{dn}{dM} dM, \quad (18)$$

where  $b_h(M)$  is the bias factor for haloes of mass  $M$  and  $dn/dM$  is the halo mass function. For the luminosity weighted bias factor, we have

$$b_L = \int_{M_{\text{min}}}^{\infty} b_h(M) L(M) \frac{dn}{dM} dM \bigg/ \int_{M_{\text{min}}}^{\infty} L(M) \frac{dn}{dM} dM, \quad (19)$$

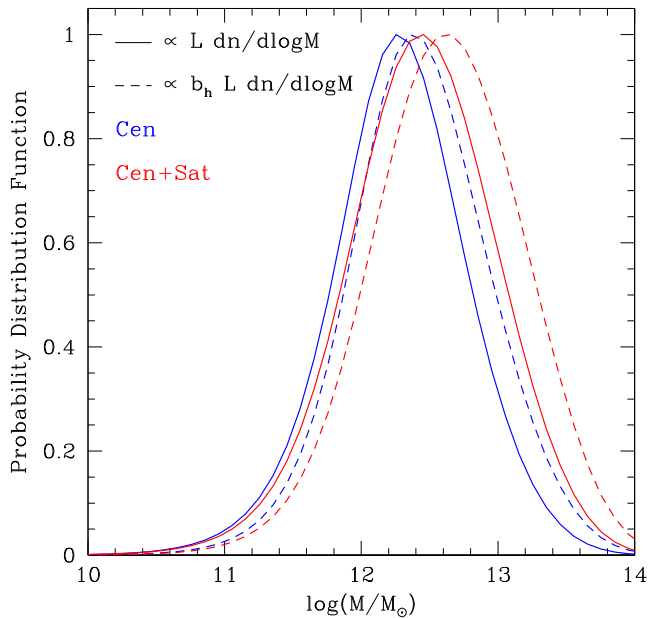
where  $L(M)$  is the average Ly $\alpha$  luminosity in haloes of mass  $M$ .

Observationally, a value of  $b_{\text{LAE}} = 1.75 \pm 0.23$  results from an average of the values measured by Gawiser et al. (2007) and Guaita et al. (2010), who find  $b_{\text{LAE}} = 1.7^{+0.3}_{-0.4}$  at  $z = 3.1$  and  $b_{\text{LAE}} = 1.8 \pm 0.3$  at  $z = 2.1$ , respectively. Using equation (18), we find that  $b_{\text{LAE}} = 1.75$  corresponds to  $M_{\text{min}} = 10^{11} M_{\odot}$ . In the calculation, we assume one galaxy per halo ( $\langle N(M) \rangle = 1$ ), which overestimates  $M_{\text{min}}$  by only a small factor (see the appendix of Zheng, Coil & Zehavi 2007). The mass estimation is consistent with, e.g. the analysis of Gawiser et al. (2007).

To estimate the luminosity-weighted bias factor  $b_L$ , we need to know the relation between luminosity and halo mass. Assuming that  $L(M) \propto M^p$ , for  $M_{\text{min}} = 10^{11} M_{\odot}$  we obtain  $b_L = 2.38, 4.40$  and  $6.84$  for the values  $p = 1, 2$  and  $3$ , respectively. The differences between  $b_{\text{LAE}}$  and  $b_L$  are therefore substantial, the latter being usually much larger. To proceed further, we need to consider the likely relation between  $L$  and  $M$ . The Ly $\alpha$  luminosity is related to the star formation rate in galaxies. Along the star-forming sequence, the star formation rate is inferred to be approximately proportional to the stellar mass (e.g. Daddi et al. 2007; Pannella et al. 2009; Lee et al. 2011). We therefore first make use of the relationship between halo mass and stellar mass found with abundance matching by Moster et al. (2010). It is in the form of a softened broken power law, and at  $z \sim 2.55$  the low-mass-end (high-mass-end) slope of the  $L$ – $M$  relation is  $\sim 2.5$  ( $\sim 0.6$ ) with a transition mass around  $10^{12} M_{\odot}$ . Using this relation in equation (19) yields  $b_L = 2.82$  for  $M_{\text{min}} = 10^{11} M_{\odot}$ . In fact, the result is insensitive to  $M_{\text{min}}$  ( $b_L = 2.81$  for  $M_{\text{min}} = 10^9 M_{\odot}$ ), given the steep  $L$ – $M$  relation (so low-mass haloes are weighted less). If we modify the high-mass-end slope to  $\sim 1$  to approximately account for the luminosity contribution from satellite galaxies, we obtain  $b_L = 3.15$ .

There are various uncertainties involved in deriving  $b_L$  with this simple model. First, the slope of the star formation rate versus the stellar mass relation for the star-forming galaxy sequence can be slightly different from unity. We find that a 10 per cent deviation from unity in the above slope leads to an  $\sim 5$  per cent change in the value of  $b_L$ . Secondly, we assume that the Ly $\alpha$  luminosity is proportional to the star formation rate. The way these two quantities track each other may vary as a function of star formation rate if, for example, the escape fraction of Ly $\alpha$  photons varies. Another possibility is that a large fraction of the Ly $\alpha$  emission comes from previously undetected sources or low surface brightness haloes around galaxies. These factors will change the  $L$ – $M$  relation and therefore the derived value of  $b_L$ . Even with the above uncertainties, it is likely that  $b_L$  is around 3.

With the above model, we can compute the contribution to the Ly $\alpha$  luminosity density from haloes of different masses, which is simply proportional to the average Ly $\alpha$  luminosity in haloes of mass  $M$  times the differential halo mass function  $dn/d\log M$ . The solid blue curve in Fig. 16 uses central galaxy Ly $\alpha$  luminosity only, which peaks around  $\log(M/M_{\odot}) = 12.25$ . Including the contribution from the satellite galaxies shifts the peak slightly to a higher mass, around  $\log(M/M_{\odot}) = 12.45$ , as shown by the solid red curve. The curve gives the probability density of a random Ly $\alpha$  photon to come from a halo of mass  $M$ . The Ly $\alpha$  emission detected through the cross-correlation technique probes the halo-bias-weighted

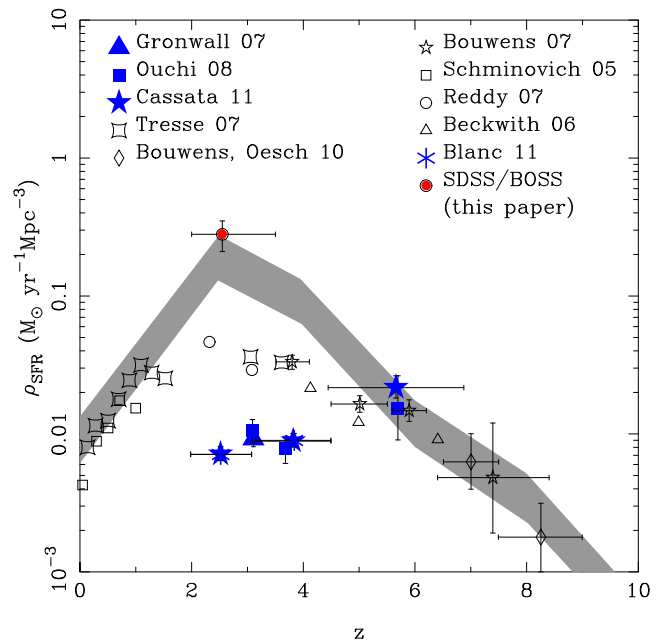


**Figure 16.** Probability distribution functions (PDFs) of Ly  $\alpha$  luminosity density (solid curves) and halo-bias-weighted Ly  $\alpha$  luminosity density (dashed curves) for our fiducial model. The PDFs are computed as proportional to the average Ly  $\alpha$  luminosity or halo-bias-weighted Ly  $\alpha$  luminosity in haloes of mass  $M$  multiplied by the differential halo mass function  $dn/d\log M$ . The blue curves use only the Ly  $\alpha$  luminosity of central galaxies, and the red curves include contributions from satellite galaxies. All curves have been normalized to unity at their respective maxima. See the text for further details concerning the model.

luminosity density distribution. The dashed curves in Fig. 16 show these probability distributions. With the satellite contribution included, the curve peaks around  $\log(M/M_\odot) = 12.6$ . Taking the full width at half-maximum (FWHM) of the curve, the fiducial model implies that the signal in the cross-correlation should mainly come from Ly  $\alpha$  emission in haloes of mass  $(1-20) \times 10^{12} M_\odot$ .

Finally, as mentioned in Section 3.2,  $b_L$  is not the same as  $b_\alpha$  once the radiative transfer effect is taken into account. A simple model shows that  $b_\alpha = b_L + \alpha_1$  with  $\alpha_1$  a positive number (see Section 6.4). The value of  $\alpha_1$  is not readily known without detailed radiative transfer modelling. Overall, we expect  $b_\alpha$  to be larger than  $\sim 3$ . We choose to parametrize derived quantities in terms of  $(3/b_\alpha)$ .

An additional uncertain factor to consider is the modification to clustering caused by redshift-space distortions. This is embodied in the  $f_\beta$  parameter. We have seen in Section 3.4 that measurements of anisotropies in clustering give a measurement of  $\beta_\alpha = \Omega_m^{0.6}(z = 2.55)/b_\alpha = -0.76 \pm 0.36$ . This negative value of  $\beta_\alpha$  is of the form expected to be caused by radiative transfer effects on clustering (Zheng et al. 2011a) and is opposite in sign to the usual Kaiser (1987) peculiar velocity redshift-space distortions. Nevertheless, this redshift-space distortion model was shown in Section 3.4 to give a reasonable fit to the data and one can use this to compute the factor  $f_\beta$  as  $f_\beta = 1 + \frac{1}{3}(\beta_q + \beta_\alpha) + \frac{1}{5}(\beta_q \beta_\alpha)$  (from equation 10). If we do this we find that the value of  $f_\beta = 0.80 \pm 0.15$ , which we take as a reasonable estimate of the reduction of the monopole term due to spreading the correlation along the line of sight. Even though gravitational evolution is not the physically correct model for interpreting our observations owing to the negative value we obtain for  $\beta_\alpha$ , a model with redshift-space distortion plus radiative transfer effect does seem to work reasonably well here (see Section 6.4).



**Figure 17.** The star formation rate density ( $\rho_{\text{SFR}}$ ) inferred from our measurement of the mean Ly  $\alpha$  surface brightness in the Universe between  $z = 2$  and  $3.5$  (see Section 5) is shown as the red point with solid line error bars, assuming that the linear bias factor for Ly  $\alpha$  emission is  $b_\alpha = 3$ , a reasonable value for the luminosity-weighted clustering of star-forming galaxies (see Section 5). The true value of  $b_\alpha$  is unknown, so this data point should be scaled by  $3/b_\alpha$ . We note this point has been calculated under the assumption that the Ly  $\alpha$  surface brightness seen is in fact largely due to star-forming galaxies, and that cooling radiation and quasar heating may also contribute as discussed in Section 6.3. Other data values plotted with open (black) symbols are from published  $\rho_{\text{SFR}}$  values which used UV estimators. The solid (blue) points show estimates of  $\rho_{\text{SFR}}$  computed from the luminosity functions of surveys for Ly  $\alpha$  emitters. The references are given in Section 5. The shaded area represents the range of dust corrected UV estimates compiled by Bouwens et al. (2010).

We use this value and propagate the errors from the bias measurements and our measurement of  $b_q b_\alpha f_\beta \langle \mu \rangle$  (for fixed shape parametrized by  $\Omega_m = 0.30$ ), to compute the mean Ly  $\alpha$  surface brightness at  $z = 2.55$ , finding

$$\langle \mu_\alpha \rangle = (3.9 \pm 0.9) \times 10^{-21} (3/b_\alpha) \text{ erg s}^{-1} \text{ cm}^{-2} \text{ \AA}^{-1} \text{ arcsec}^{-2}. \quad (20)$$

We convert this into a comoving Ly  $\alpha$  luminosity density  $\epsilon_\alpha$  using

$$\epsilon_\alpha = 4\pi \langle \mu_\alpha \rangle \frac{H(z)}{c} \lambda_\alpha (1+z)^2, \quad (21)$$

where  $c$  is the speed of light and  $\lambda_\alpha = 1216 \text{ \AA}$ . We find the value  $\epsilon_\alpha = 3.1 \times 10^{41} (3/b_\alpha) \text{ erg s}^{-1} \text{ Mpc}^{-3}$ . We then use equation (17) to convert this into a measurement of the star formation rate density

$$\rho_{\text{SFR}}(z = 2.55) = (0.28 \pm 0.07) \frac{3}{b_\alpha} M_\odot \text{ yr}^{-1} \text{ Mpc}^{-3}. \quad (22)$$

As mentioned before, the conversion depends on the assumption about the underlying stellar population. A younger population and lower metallicity would lead to a lower inferred SFR, which could be an important effect for interpreting our results. Keeping this possibility in mind, we proceed with the discussion by using the above result from the commonly adopted conversion factor.

We plot this result in Fig. 17 as the red point, for the chosen value of  $b_\alpha = 3$ . We note that the true value for the parameter  $b_\alpha$  is not

well determined, and that the  $\rho_{\text{SFR}}$  data point scales as the factor  $3/b_\alpha$ . Our discussion above suggests that  $b_\alpha$  is likely to be larger than 3, so that it is more probable for  $\rho_{\text{SFR}}$  to be scaled downwards than upwards. Fig. 17 also shows various estimates of  $\rho_{\text{SFR}}$  from Ly  $\alpha$  emitter surveys as well as UV continuum estimates of the star formation rate. We can see that our measurement is about 30 times higher (for  $b_\alpha = 3$ ) than the Ly  $\alpha$  emitter based measurements of Gronwall et al. (2007), Ouchi et al. (2008) or Cassata et al. (2011). Note that these Ly  $\alpha$  emitter based measurements result from a direct integration over the observed Ly  $\alpha$  luminosity function without corrections for any possible dust effect.

A complication which adds substantial uncertainty is dust absorption, which may affect Ly  $\alpha$  and continuum radiation differently. One estimator of the level of dust extinction in the continuum is the rest-frame UV continuum slope,  $\beta$ , which specifies how the flux density of a galaxy varies with wavelength (i.e.  $f_\lambda \propto \lambda^\beta$ ) in the UV continuum region ( $\sim 1300 - \sim 3500 \text{ \AA}$ ) of its spectrum. If an intrinsic  $\beta$ -dust extinction relationship is assumed (usually that measured empirically from  $z \sim 0$  galaxies by Meurer, Heckman & Calzetti 1999), one can use observations of  $\beta$  for high redshift galaxies to infer a dust-corrected UV luminosity and star formation rate. This has been done by several authors, including Bouwens et al. (2009, 2010). In Fig. 17, we show as a grey band the compilation of dust-corrected UV star formation densities from Bouwens et al. (2010) computed using this technique. According to Bouwens et al. (2010), the dust correction for a limiting luminosity of  $0.3 L_{z=3}^*$  is  $6.0 \pm 2.5$  at redshift  $z = 2.5$ . Support for the validity of these corrections comes from the agreement of dust-corrected UV values with  $\rho_{\text{SFR}}$  estimated from infrared observations (see the recent review by Madau & Dickinson 2014).

As for the effect of dust on the Ly  $\alpha$  radiation, Cassata et al. (2011) have speculated that it could be even stronger than proposed by Bouwens et al. (2010), as the Ly  $\alpha$  emitter inferred  $\rho_{\text{SFR}}$  is less than 20 per cent of the non-dust corrected UV continuum value (as can be seen in Fig. 17). One reason which favours this interpretation is the fact that resonantly scattered radiation has to cover a longer path length than continuum radiation before it leaves a galaxy. On the other hand, there is one well-known mechanism (Neufeld 1991) which could preferentially enhance the escape of Ly  $\alpha$  radiation: in a clumpy medium of dusty clouds, continuum (UV) photons are absorbed as soon as their path crosses an optically thick dust cloud, whereas Ly  $\alpha$  photons can bounce off the cloud surfaces and find their way through the clouds to escape, leading to a lower extinction for Ly  $\alpha$  than for continuum photons if the intercloud medium is sufficiently devoid of dust. The anisotropic escape of Ly  $\alpha$  radiation (Zheng & Wallace 2014) caused by, for example, a bipolar galactic wind, can also help make Ly  $\alpha$  photons follow the path of lower extinction optical depth to escape, while UV continuum photons are emitted isotropically and on average experience more extinction. From Fig. 17, it appears that some mechanism of this sort is needed if we are to explain our results with star-forming galaxies tracing the large-scale structure. We discuss these issues further in Section 6.

In conclusion, the rather surprising result seen in Fig. 17 is that the fiducial value of the Ly  $\alpha$  surface brightness inferred from our measurement is consistent with emission from star-forming galaxies only if the majority of Ly  $\alpha$  photons produced in stars at  $z = 2.55$  escape from their host galaxies and are detected. The dust-corrected results of Bouwens et al. (2010) imply  $\rho_{\text{SFR}} = 0.19^{+0.08}_{-0.06} M_\odot \text{ yr}^{-1} \text{ Mpc}^{-3}$  at  $z = 2.55$ , and from our measurement,  $\rho_{\text{SFR}} = (0.28 \pm 0.07)(3/b_\alpha) M_\odot \text{ yr}^{-1} \text{ Mpc}^{-3}$ . This means

that  $b_\alpha > 3.0$  is needed for our measurement not to imply detection of more Ly  $\alpha$  photons than are actually produced at more than the  $1\sigma$  level, even with no dust absorption of these Ly  $\alpha$  photons. We note that the intensity mapping technique we use in this paper will detect Ly  $\alpha$  photons which are scattered into our sightline from arbitrarily large distances from the emitting galaxy, and at arbitrarily low surface brightness. One can therefore characterize this implication of our measurement as a ‘total escape fraction’ of Ly  $\alpha$  photons from star-forming galaxies that is close to 100 per cent, whereas the ‘detected escape fraction’ measured from traditional surveys of Ly  $\alpha$  emitters, which have surface brightness limitations, is clearly much lower.

### 5.1 Low surface brightness Ly $\alpha$ haloes around galaxies

The Ly  $\alpha$  luminosity from observed Ly  $\alpha$  emitters is usually inferred from a set of pixels around a source above a certain surface brightness threshold, or measured within a fixed aperture (e.g. 2 arcsec; Ouchi et al. 2008). Our measurement, on the other hand, is not subject to any surface brightness limit, so the high total escape fraction of Ly  $\alpha$  photons that we infer may be hidden in extended haloes around galaxies of low surface brightness that are not included in Ly  $\alpha$  emitter surveys. These diffuse Ly  $\alpha$  haloes, resulting from scattering by neutral gas in the circumgalactic and intergalactic media, are predicted to exist around a star-forming galaxy based on radiative transfer modelling (Zheng et al. 2011b), and have been detected from stacking analyses of Ly  $\alpha$  emitting galaxies (Steidel et al. 2011; Matsuda et al. 2012; Momose et al. 2014).

Can the Ly  $\alpha$ -quasar cross-correlation we have detected be explained by known diffuse Ly  $\alpha$  haloes? To address this question, we use the stacked Ly  $\alpha$  surface brightness profiles and the fits to the radial profiles of diffuse Ly  $\alpha$  haloes in Momose et al. (2014) to estimate the contribution of these haloes to our detection. The luminosity inside the aperture of radius 2 arcsec roughly corresponds to the Ly  $\alpha$  luminosity from the Ly  $\alpha$  emitter survey. We find that at both  $z = 2.2$  and 3.1, the diffuse Ly  $\alpha$  emission outside of 2 arcsec is about one third of the luminosity inside the aperture. Therefore, the observed diffuse Ly  $\alpha$  haloes, regardless of their origin, may only increase the contribution from galaxies detected in Ly  $\alpha$  emitter surveys by an additional  $\sim 30$  per cent.

This implies that our hypothesized extended emission from scattering haloes that may account for our measurement would have to arise in many more galaxies than those detected in Ly  $\alpha$ -emission surveys. Haloes that are highly extended would need to be present around the majority of star-forming galaxies.

## 6 DISCUSSION

We can frame further discussion of our results in terms of the following questions.

(i) Can the observed Ly  $\alpha$  surface brightness be explained by known Ly  $\alpha$  emitters? We have seen in Section 5 that the answer is no.

(ii) Can the observed Ly  $\alpha$  surface brightness be explained by faint Ly  $\alpha$  emitters below the threshold of published surveys? Ouchi et al. (2008) have shown that changing the extrapolated luminosity function faint-end slope from  $-1.0$  to  $-2.0$  changes the total integrated Ly  $\alpha$  luminosity density they infer down to their observed luminosity limit by only 20 per cent. The answer to this question is therefore also no, unless there is an extremely large amount of star formation in undetected dwarf galaxies at  $z \simeq 2.5$ .



(iii) Can the observed Ly  $\alpha$  surface brightness be explained by extended haloes around the known Ly  $\alpha$  emitters? We have seen in Section 5.1 that this is not the case, and that extended haloes around known Ly  $\alpha$  emitters, while adding to the mean surface brightness, fall short of accounting for our results.

(iv) Can the observed Ly  $\alpha$  surface brightness be explained by star-forming galaxies, when these are estimated based on the extinction-corrected UV continuum SFR density? We have seen in Section 5 that there appears to be just enough star formation per unit volume in the Universe at  $z = 2.5$  that if most of it led to escaped Ly  $\alpha$  emission, this could explain what we are seeing. We further discuss the implications of this below in Sections 6.2, 6.5 and 7.2.

(v) What is the contribution to our measurement of sources of Ly  $\alpha$  emission beyond star-forming galaxies? We address this in Section 6.3 below, showing that other contributions are likely to be small, although cooling of gas in galactic haloes and heating associated with ionizing radiation and jets from quasars may contribute to some of the Ly  $\alpha$  emission we measure.

## 6.1 Potential systematic errors in the measurement

Our measurement of Ly  $\alpha$  intensity clustering relies on statistical cross-correlation techniques applied to a large sample of spectra with relatively low signal to noise, all of which were targeted at bright foreground galaxies that we have removed in post-processing. There is therefore ample scope for small instrumental or other effects to influence the signal we measure. Bearing this in mind, we have carried out a range of tests, detailed in Appendices A–C, to make as certain as possible that the signal is real. Most importantly, these include tests of our methods for eliminating contaminating light from neighbouring fibres, which does have a strong effect. We have also tested the effect of eliminating quasars clustered with the quasars we are using in our cross-correlation, and we have checked the dependence of the signal on the luminosity of the galaxy fibre target and on the quasar luminosity. For these latter tests, we have found no significant effect and therefore conclude that the signal is real, to the extent we have been able to ascertain.

There remains the possibility that some other effect, instrumental or otherwise, is responsible for the cross-correlation signal. In particular, we have seen in Section 3.6 that the signal differs at the  $\sim 2\sigma$  level between fibres that are at the edge or central regions of plates. We can imagine various systematics working along these lines, so that in different survey areas, the mean residual light obtained after subtracting the galaxy model (Fig. 7) is modified and at the same time the selection function of the detected quasars at different redshifts is also modified (for example, due to Galactic dust absorption). Such effects could be artificially contributing to the correlation we detect. We cannot rule this out a priori, and additional tests with the BOSS data can hardly tell us very much because our total signal is detected only at the  $8\sigma$  level. Our arguments for believing that the correlation we have detected is real are that there is nothing anomalous in the redshift dependence of the correlation, and that most of our signal is at scales substantially smaller than the plate size. Moreover, the clustering signal would have to coincidentally agree with the shape expected in  $\Lambda$ CDM and pass all the tests mentioned above. In particular, contamination by quasar light is unlikely once we have eliminated the effect of neighbouring fibres in the BOSS camera and we have tested the absence of a dependence on the quasar measured flux. However, the possibility of systematics should be borne in mind until future independent work is able to confirm our measurement.

The effect of gravitational lensing on our results should be zero, even though we use spectra of bright galaxies, because our Ly  $\alpha$  measurements are of surface brightness, which is conserved under lensing. If on the other hand we were detecting Ly  $\alpha$  emitters in the fibres and obtaining their luminosity function, this would be subject to the well-known magnification bias (e.g. Turner 1980). In our case, we are computing the cross-correlation function of the surface brightness measured from all fibres with quasars, and the magnification of Ly  $\alpha$  sources cannot change the amplitude of the cross-correlation. Dust associated with the foreground galaxies might reduce the Ly  $\alpha$  emission coming from higher redshift, but this could only further increase the inferred brightness of the Ly  $\alpha$  background.

## 6.2 Star-forming galaxies and the photon budget

We have found in Section 5 that our detected signal of cross-correlation of Ly  $\alpha$  surface brightness with quasars implies a brightness for the mean Ly  $\alpha$  photon background given by equation (20). This at the same time implies an emissivity of Ly  $\alpha$  radiation of  $\epsilon_\alpha = 3.1 \times 10^{41} (3/b_\alpha) \text{ erg s}^{-1} \text{ Mpc}^{-3}$ . This emissivity can be re-expressed in terms of the rate at which Ly  $\alpha$  photons must have been emitted for each baryon in the Universe at the mean redshift of our observation,  $z = 2.55$ . Using the comoving number density of baryons  $n_b = 2.5 \times 10^{-7} \text{ cm}^{-3}$ , and an expansion rate at  $z = 2.55$  of  $H(z = 2.55) = 261 \text{ km s}^{-1} \text{ Mpc}^{-1}$  (using the parameters  $\Omega_b h^2 = 0.0221$ ,  $H_0 = 68 \text{ km s}^{-1} \text{ Mpc}^{-1}$  and  $\Omega_m = 0.315$ , consistent with the most recent determinations from *Planck* in Planck Collaboration XVI et al. 2014), we find the following result:

$$\frac{\epsilon_\alpha}{h\nu_\alpha n_b H(z)} = 306 \frac{3 \text{ photons}}{b_\alpha \text{ baryon}}. \quad (23)$$

The first, most simple assumption we make is that these photons are mostly originating from star formation in galaxies. The Ly  $\alpha$  photons created by recombinations in the H II regions produced around massive stars can then be scattered out to gaseous haloes surrounding galaxies, from which they give rise to the background we detect in the quasar–Ly  $\alpha$  emission cross-correlation. As discussed in Section 5, this implies a very large star formation rate at  $z = 2.55$ . Equation (17) can be recast in terms of the number of Ly  $\alpha$  photons emitted for each baryon that forms stars,  $n_\alpha/n_{\text{bs}}$ :

$$\frac{n_\alpha}{n_{\text{bs}}} = \frac{m_p (1.1 \times 10^{42} \text{ erg s}^{-1} \text{ yr})}{h\nu_\alpha M_\odot} = 1800 \frac{\text{photons}}{\text{baryon}}. \quad (24)$$

Comparing to equation (23), we see that this implies that, for  $b_\alpha = 3$ , about 10 per cent of all the baryons in the universe would have to turn into stars if the star formation rate is maintained over the age of the universe at  $z = 2.55$ ,  $(2/3)H^{-1}(z = 2.55) = 2.5 \times 10^9 \text{ yr}$ . The difficulty with this very high star formation rate is twofold: estimates of the *total* fraction of baryons in the form of stars at present are near 6 per cent (Fukugita, Hogan & Peebles 1998), and as described in the previous section, the total star formation rate at the redshift of our measurement can reach this value only for the maximum estimates of dust absorption, which would imply that while the UV continuum has to be absorbed by factors of  $\sim 5$ , the Ly  $\alpha$  photons would have to emerge suffering little dust absorption.

A first possible solution to the problem of this extremely high inferred star formation rate is to modify the IMF of the stellar population that is assumed in deriving the Ly  $\alpha$  photons emitted per baryon in equation (23) from population synthesis models. If the slope of the IMF is flatter in the high-mass range of  $10\text{--}100 M_\odot$ , then stars above  $\sim 20 M_\odot$ , which dominate the production of

ionizing photons and therefore of Ly  $\alpha$  photons from H II regions, increase their abundance compared to  $\sim 10 M_{\odot}$  stars, which dominate the observed UV continuum. If the IMF stays flat down to lower masses, that can also greatly reduce the total star formation rate that is implied, as well as the stellar mass that is derived for the present universe which is measured from the old stellar population dominating the present luminosity of galaxies. If this flat IMF occurs particularly in massive galaxies with high metallicity, then the UV continuum observed at  $z = 2.55$  can be further reduced due to the suppression of blue horizontal branch stars, and the luminosity-weighted bias factor of the Ly  $\alpha$  emission can be further increased. A top-heavy IMF during the epoch when most stars were formed implies a large increase in the production of heavy elements, but this may be consistent with observations that show relatively high metallicities in the intracluster medium and in massive galaxies (Renzini & Andreon 2014).

### 6.3 Other sources of Ly $\alpha$ emission beyond star-forming galaxies

We shall discuss here six possible contributions to the quasar–Ly  $\alpha$  emission cross-correlation not arising from star-forming galaxies clustered around the quasars. (1) Scattering of the quasar–Ly  $\alpha$  broad emission line by the Ly  $\alpha$  forest. (2) Fluorescence of the ionizing radiation from the quasar. (3) Gas heating due to helium ionization from the quasar radiation and the quasar jet energy dissipation. (4) Fluorescence of the general cosmic ionizing background by the overdense IGM around quasars. (5) Scattering of the cosmic UV background by the Ly  $\alpha$  forest in the overdense IGM around quasars. (6) Ly  $\alpha$  cooling radiation from radiative dissipation of gas in haloes that are correlated with quasars.

#### 6.3.1 Scattering of the quasar–Ly $\alpha$ broad emission line

The average observed flux of our sample of BOSS quasars within the central  $\sim 2000 \text{ km s}^{-1}$  of the Ly  $\alpha$  broad emission line is close to  $f_{\alpha} \sim 10^{-16} \text{ erg s}^{-1} \text{ cm}^{-2} \text{ \AA}^{-1}$ . In general, the mean fraction of light that is found to be absorbed by the Ly  $\alpha$  forest at  $z = 2.55$  is  $1 - \bar{F} \simeq 0.2$  ( $\bar{F}$  is the mean transmitted fraction; e.g. Faucher-Giguère et al. 2008). At a characteristic impact parameter of  $\sim 10 h^{-1} \text{ Mpc}$  (inside which our cross-correlation signal is strongest), corresponding to an angular separation  $\theta \sim 500 \text{ arcsec}$ , the surface brightness of the scattered radiation should be  $f_{\alpha}(1 - \bar{F})/(\psi\pi\theta^2)$ , where  $\psi$  is a dimensionless number that depends on the geometry of the scattering gas around the quasar and the shape of the Ly  $\alpha$  emission line, and has a value  $\psi \simeq 4$  for a uniform gas density and a flat emission line profile. This yields a surface brightness  $\sim 7 \times 10^{-24} \text{ erg s}^{-1} \text{ cm}^{-2} \text{ \AA}^{-1} \text{ arcsec}^{-2}$ , more than two orders of magnitude lower than our measured excess surface brightness at an impact parameter of  $10 h^{-1} \text{ Mpc}$  from a quasar, as shown in Fig. 2.

Scattered light from the quasar broad Ly  $\alpha$  emission line is therefore a negligible contribution to our detected Ly  $\alpha$  background. This is consistent with our test in Appendix C4 showing no dependence of the cross-correlation amplitude on the quasar luminosity.

#### 6.3.2 Fluorescence of the quasar ionizing radiation

An excess of Ly  $\alpha$  emission around the quasar may also arise from fluorescence of the ionizing radiation from the quasar. Hydrogen is photoionized in intergalactic absorption systems, and part of the

energy is reemitted as Ly  $\alpha$  photons when the recombinations that maintain ionization equilibrium take place. This is similar to the ‘quasar fuzz’ effect hypothesized by Haiman & Rees (2001) to arise in a halo of a forming galaxy, but over a more extended region around the quasar. These Ly  $\alpha$  photons should be predominantly emitted from the same systems that dominate the Lyman limit absorption, with column densities  $N_{\text{HI}} \sim 10^{17} \text{ cm}^{-2}$ , which determine a mean free path  $\ell_i$  for ionizing photons, and have also a bias factor  $b_i$  that determines their cross-correlation with quasars in the linear regime.

To derive the Ly  $\alpha$  surface brightness expected around the quasar from fluorescence, we define  $f_{i0}$  to be the observed flux from a quasar per unit  $\log \lambda$  at the hydrogen Lyman limit,  $\lambda_{i0} = (1 + z)912 \text{ \AA}$ . The proper flux that would be observed by a local observer at redshift  $z$ , at a proper distance  $d$  from the quasar which is small compared to the horizon,  $d \ll c/H(z)$ , is  $f_i = f_{i0}(D_{\Lambda}/d)^2(1 + z)^4$ , at wavelength  $\lambda_i = 912 \text{ \AA}$ , where  $D_{\Lambda}$  is the cosmological angular diameter distance to redshift  $z$ . Fluorescence produces a proper Ly  $\alpha$  emissivity  $\epsilon_{\alpha} = C_{\alpha} f_i / \ell_i$ , where  $\ell_i$  is the proper mean free path of ionizing photons, and  $C_{\alpha}$  is the fraction of the energy flux  $f_i$  that is converted to Ly  $\alpha$  photons. For the typical quasar frequency spectral index  $f_{\nu} d\nu \propto \nu^{-1.5} d\nu$ ,  $C_{\alpha} \simeq 0.3$ . Applying the equation of radiative transfer, this emissivity results in an observed surface brightness over an observed wavelength range  $d\lambda$  equal to

$$dS_{\alpha} = \frac{\epsilon_{\alpha}}{4\pi(1+z)^4} \frac{c}{H(z)} \frac{d\lambda}{\lambda_{\alpha 0}}, \quad (25)$$

where  $\lambda_{\alpha 0} = (1 + z)1216 \text{ \AA}$  is the wavelength at which the Ly  $\alpha$  brightness is observed, resulting from the integration of the emissivity over the proper path length  $cH^{-1}(z) d\lambda/\lambda_{\alpha 0}$ . Note that the surface brightness is reduced by the factor  $(1 + z)^4$  owing to the cosmological redshift. The final expression is therefore given by

$$\frac{dS_{\alpha}}{d\lambda} = C_{\alpha} \frac{f_{i0}}{\lambda_{\alpha 0}} \frac{D_{\Lambda}^2}{4\pi d^2} \frac{c}{\ell_i H(z)}. \quad (26)$$

This is intuitively simple to understand: the observed quasar flux that is converted to Ly  $\alpha$ ,  $C_{\alpha} f_{i0}$ , is spread over a solid angle in the sky of angular size  $d/D_{\Lambda}$ , and over a wavelength range  $\lambda_{\alpha 0} \ell_i H(z)/c$ .

Using the mean observed BOSS quasar flux  $f_{i0}/\lambda_{\alpha 0} \simeq 2 \times 10^{-17} \text{ erg s}^{-1} \text{ cm}^{-2} \text{ \AA}^{-1}$ ,  $C_{\alpha} = 0.3$ , choosing a comoving separation  $d(1 + z) = 10 h^{-1} \text{ Mpc}$  corresponding to  $d/D_{\Lambda} \simeq 500 \text{ arcsec}$ , and a mean free path of ionizing photons at  $z = 2.55$   $\ell_i H(z)/c \simeq 0.1$  (Prochaska et al. 2014), we find  $dS_{\alpha}/d\lambda \simeq 2 \times 10^{-23} \text{ erg s}^{-1} \text{ cm}^{-2} \text{ \AA}^{-1} \text{ arcsec}^{-2}$ . This falls short by two orders of magnitude to explain the detected surface brightness of  $\sim 2 \times 10^{-21} \text{ erg s}^{-1} \text{ cm}^{-2} \text{ \AA}^{-1} \text{ arcsec}^{-2}$  at this separation (see Fig. 2).

We note that the Ly  $\alpha$  fluorescent emission from the quasar should not be of uniform intensity, but would predominantly arise from Lyman limit system absorbers in the quasar vicinity (see e.g. the simulations in Kollmeier et al. 2010). However, our measurement is sensitive only to the mean surface brightness around the quasar. The clustering of Lyman limit systems around quasars should increase this mean surface brightness by a factor  $1 + b_q b_{f\beta} \xi[d(1 + z)]$ , where  $\xi$  is the mass autocorrelation function of equation (3), and  $b_i$  is the Lyman limit system bias factor. At  $z = 2.55$  and  $d(1 + z) = 10 h^{-1} \text{ Mpc}$ , the mass correlation is  $\xi[d(1 + z)] \simeq 0.05$ , and assuming  $b_q = 3.5$  and that the Lyman limit systems do not have a bias factor larger than that of DLAs ( $b_i < 2$ ; see Font-Ribera et al. 2012), we see that this enhancement is less than a factor of 2 and cannot increase the contribution from hydrogen fluorescence to a relevant value to explain our detection. In any case, our discussion suggests that the main difficulty for detecting this component of

fluorescent emission caused by individual quasars probably lies in distinguishing it from the Ly  $\alpha$  emission scattered in gas haloes surrounding star-forming galaxies that are clustered with the quasar.

The recent detection of Ly  $\alpha$  emission from cosmic web filaments around quasars by Martin et al. (2014a,b) and Cantalupo et al. (2014) is attributed to fluorescent emission, but on smaller scales (100–460 proper kpc) than we are able to probe with our measurement. This limitation on our part is due to the small number of fibre sightlines passing close to quasars in BOSS. From our estimate described above, we conclude that the large-scale signal we see is not likely to arise from the same fluorescent emission mechanism seen in this other work.

We also point out here that the quasar fluorescence contribution is affected by the anisotropy of the quasar emission and the time-delay between our observation of the quasar and that of the fluorescent light, and this will be most apparent on larger scales. For example, for an absorber at a perpendicular separation  $d(1+z) \simeq 10 h^{-1}$  Mpc at  $z = 2.55$ , the quasar radiation that is illuminating the absorber was emitted  $\sim 10^7$  yr earlier than the light we observe directly from the quasar, and the luminosity at this earlier time would be systematically lower than the presently observed one because of the flux-limited selection of quasar samples. These effects are likely to be important in any detailed modelling of the fluorescent emission due to the quasar.

### 6.3.3 Gas heating due to helium ionization and jet dissipation near quasars

In addition to the direct fluorescence by hydrogen, hard ionizing photons from the quasar with energy above 54.4 eV will ionize He II into He III in the IGM. The IGM may be highly opaque to these photons, with a mean free path much shorter than for the hydrogen-ionizing photons. A large fraction of the energy in these hard photons may, after causing the double ionization of helium, end up as hydrogen Ly  $\alpha$  photons: the fast electrons that are produced heat the gas and result in enhanced Ly  $\alpha$  cooling emission. Similarly, if relativistic jets launched by quasars can carry mechanical luminosities that are comparable to the radiative ionizing luminosity of the quasar, the energy of the jet may eventually turn to heat of the surrounding IGM and be emitted in the subsequent cooling as Ly  $\alpha$  photons.

Whereas the direct hydrogen fluorescence is emitted over a mean free path around the quasar determined by hydrogen Lyman limit systems,  $H(z)\ell_i \sim 30\,000 \text{ km s}^{-1}$ , the energy produced by He II ionization and jet dissipation can be emitted from a much smaller region, close to the scale of  $\sim 1000 \text{ km s}^{-1}$  where our strongest cross-correlation signal is found. Thus, if the same amount of energy were produced by a quasar in He II ionizing photons and jet power as in hydrogen ionizing photons, and this energy could be converted to Ly  $\alpha$  photons with a similar efficiency, within a comoving distance of  $\sim 10 h^{-1}$  Mpc, then the Ly  $\alpha$  brightness resulting from this would be  $\sim 30$  times brighter than from the direct hydrogen fluorescence at  $r \sim 10 h^{-1}$  Mpc, and therefore close to our detected signal.

This suggests that it is in principle possible that much of the signal we detect results from heating of the IGM induced by quasars due to ionization of He II and dissipated jet power, reemitted as Ly  $\alpha$  photons when the gas cools. Jets associated with some quasars are in fact known to deposit a mechanical power close to that in the ionizing luminosity (e.g. McDonald et al. 2015), and this power could be released into the IGM when a massive cluster is not present around the quasar. The difficulty here may lie in converting much

of this energy into Ly  $\alpha$  photons, rather than X-rays from hot gas or synchrotron emission, at large distances from the quasar. The He II ionization heating mechanism may be a more promising possibility, although the absence of any correlation of our signal with quasar luminosity may be a possible difficulty that will need to be investigated with more data. Refining our rough estimates on these alternative emission mechanisms necessarily involves large modelling uncertainties and complexities, and we leave this for future work.

### 6.3.4 Fluorescence of the cosmic ionizing background

Fluorescence from the mean cosmic ionizing background can also contribute to the quasar–Ly  $\alpha$  cross-correlation we measure. In this case, if we assume the intensity of ionizing photons from distant sources is uniform around the quasar, the effect on the cross-correlation would arise only from the *overdensity* of Lyman limit absorbers near the quasar, because a uniform Ly  $\alpha$  brightness does not contribute to our detected signal. We consider a value of the photoionization rate at  $z = 2.5$  of  $\Gamma = 10^{-12} \text{ s}^{-1}$ , which corresponds to a proper intensity per unit wavelength of the ionizing background of  $i_\lambda \simeq 3 \times 10^{-20} \text{ erg s}^{-1} \text{ cm}^{-2} \text{ \AA} \text{ arcsec}^{-2}$  at the ionization edge  $\lambda = 912 \text{ \AA}$  (e.g. Faucher-Giguère et al. 2008). The intensity of the Ly  $\alpha$  background that is produced is

$$i_\alpha = \frac{3C_\alpha}{4} \frac{cH^{-1}}{\ell_i} i_\lambda. \quad (27)$$

As before,  $C_\alpha$  is the mean fraction of energy converted to Ly  $\alpha$  photons, and  $\ell_i$  the mean free path of ionizing photons; the additional factor 3/4 is due to the use of intensities per unit of wavelength. Using  $C_\alpha = 0.3$  and  $\ell_i = 0.1 cH^{-1}$  as above, we obtain  $i_\alpha \simeq 6 \times 10^{-20} \text{ erg s}^{-1} \text{ cm}^{-2} \text{ \AA}^{-1} \text{ arcsec}^{-2}$  at  $z = 2.55$ . When observed at the present time, the intensity of this Ly  $\alpha$  background is reduced to

$$i_{\alpha 0} = i_\alpha / (1+z)^5 \simeq 10^{-22} \text{ erg s}^{-1} \text{ cm}^{-2} \text{ \AA}^{-1} \text{ arcsec}^{-2}. \quad (28)$$

This mean intensity now needs to be multiplied by the cross-correlation function of Lyman limit absorbers and quasars, to obtain the contribution to our measured cross-correlation. As in subsection 6.3.2, we estimate  $\xi(r = 10 h^{-1} \text{ Mpc}) \simeq 0.05$  at  $z = 2.55$ , and  $b_q b_f \xi(r = 10 h^{-1} \text{ Mpc}) \simeq 0.5$ , implying a contribution to the quasar–Ly  $\alpha$  cross-correlation of  $\sim 5 \times 10^{-23} \text{ erg s}^{-1} \text{ cm}^{-2} \text{ \AA}^{-1} \text{ arcsec}^{-2}$ , a factor of  $\sim 30$  below our measured value from Fig. 2.

We note that this estimate is consistent with a cruder calculation which uses the ionizing background observationally inferred by Fontanot et al. (2014) from the comoving space density of quasars and star-forming galaxies. After exploring the likely parameter space of limiting magnitudes and escape fractions, Fontanot et al. (2014) find that a central value for the ionizing background comoving emissivity is about  $3 \times 10^{50} \text{ photons s}^{-1} \text{ Mpc}^{-3}$  at  $z \sim 2.55$ . If the emissivity of the ionizing photons that are converted to Ly  $\alpha$  photons is at a similar level, with a conversion efficiency of 2/3, the fluorescent Ly  $\alpha$  emissivity is then  $2 \times 10^{50} \text{ photons s}^{-1} \text{ Mpc}^{-3}$ . This is  $3.2 \times 10^{39} \text{ erg s}^{-1} \text{ Mpc}^{-3}$ ,  $\sim 100$  times lower than the value inferred from our results.

We therefore conclude that hydrogen fluorescence from the cosmic ionizing background cannot be responsible for the large Ly  $\alpha$  brightness that we detect to be correlated with quasars.

### 6.3.5 Scattered Ly $\alpha$ photons from the radiation background

Just as scattering of the quasar Ly $\alpha$  broad emission line by the Ly $\alpha$  forest can in principle produce some contribution to the Ly $\alpha$  light around quasars, one may think that the general ultraviolet continuum background from distant galaxies and quasars could also give rise to an excess of Ly $\alpha$  photons near quasars because of the overdensity of the Ly $\alpha$  forest that scatters this background radiation. However, this effect actually cancels out for our observation. High-redshift galaxies that are behind the quasar should show the Ly $\alpha$  forest absorption, reducing their ultraviolet flux, and this Ly $\alpha$  absorption will be enhanced because of the overdensity surrounding the quasar. The background sources (which are of course too faint to be detected individually, but nevertheless contribute to our total background) are therefore fainter near quasars compared to any random fields. This is exactly compensated by the scattered Ly $\alpha$  background from these same sources. For this reason, this contribution to the Ly $\alpha$  background could only be detected if the individual sources behind the quasar were individually detected and subtracted out, before evaluating the Ly $\alpha$  background intensity.

In addition to Ly $\alpha$  photons scattered by the Ly $\alpha$  forest, the background would also have a contribution from background photons reaching the Ly $\gamma$  wavelength when passing near the quasar and being downscattered to Ly $\alpha$ . This contribution is also very small, and is also nearly cancelled unless the population of ultraviolet sources creating the background evolves very fast over the redshift interval corresponding to the ratio of the Ly $\gamma$  to the Ly $\alpha$  wavelength. (note that the Ly $\beta$  forest does not contribute to this diffuse emission by intergalactic gas because Ly $\beta$  absorptions can only end in a 2-photon emission from the 2s atomic state of hydrogen.)

### 6.3.6 Cooling radiation

Cooling radiation from gas in galactic haloes can produce Ly $\alpha$  emission. A rough estimate of the cooling radiation can be made if we assume that cooling and star formation reach a steady state on average in a galaxy (i.e. 1 M $_{\odot}$  of gas cools per year to feed 1 M $_{\odot}$  yr $^{-1}$  star formation). For gas initially at virial temperature  $T$  dissipating the energy through cooling (as suggested by cosmological simulations; e.g. Fardal et al. 2001; Faucher-Giguère et al. 2010; Goerd et al. 2010), we can estimate the corresponding cooling luminosity as follows.

Let us assume that a fraction  $f_c$  of all baryons in the Universe fall into haloes of virial temperature  $T$ , and they dissipate all their energy by emitting Ly $\alpha$  photons. Initially the baryons fall into the halo and are shock heated to temperature  $T$  at a radius  $r_v \simeq \sigma t/6$  in the halo, where  $t$  is the age of the Universe, and in the end they have to reach a radius  $r_g$  by dissipating their energy through Ly $\alpha$  emission, where  $r_g$  is the half-radius of the galaxy.

The circular velocity  $v_c = \sqrt{2}\sigma$  is assumed to be flat, independent of radius. Then, the potential difference from  $r_v$  to  $r_g$  is  $\phi = v_c^2 \log(r_v/r_c)$ . The energy to be dissipated per baryon is therefore  $m_p \phi$ . The relation between  $v_c^2$  and the temperature  $T$  is  $\sigma^2 = kT/\mu$ , where  $\mu = 0.6 m_p$  for the fully ionized mixture of hydrogen and helium from big bang nucleosynthesis. So, the total energy dissipated per baryon in the universe is

$$\epsilon_b = f_c m_p \phi = \frac{2f_c}{0.6} kT \log(r_v/r_c), \quad (29)$$

and the number of Ly $\alpha$  photons emitted per baryon is

$$\epsilon_\alpha = \epsilon_b / (10.2 \text{ eV}) = 3.3 f_c \frac{T}{1.2 \times 10^5 \text{ K}} \log(r_v/r_c). \quad (30)$$

Using  $T = 3 \times 10^6$  K and  $r_v/r_c = 20$  we get  $\epsilon_\alpha = 240 f_c$ . With plausible values of  $f_c \sim 0.2$ , this could amount to nearly 15 per cent of the emission we observed, and more if the bias of this Ly $\alpha$  emission is high.

Cooling radiation is therefore another alternative source of Ly $\alpha$  emission that comes close to explaining our results, being plausibly less than an order of magnitude below our measurements. Cooling radiation could contribute in a substantial way to the observed Ly $\alpha$  emission.

## 6.4 Radiative transfer effect

The above estimates show that the Ly $\alpha$  emission relevant to our clustering measurements may arise from a combination of star-forming galaxies, double helium ionization or other direct heating of intergalactic gas by quasars, or radiative cooling of accreting gas, which would be spatially correlated with quasars. Regardless of the origin, as long as Ly $\alpha$  photons are scattered by neutral hydrogen, we expect them to be affected by a radiative transfer effect. The quasar–Ly $\alpha$  cross-correlation function we measure suggests that this effect is detected, shown as elongated contours along the line of sight on scales as large as  $\sim 20 h^{-1}$  Mpc. Our fits favour that this radial elongation does not solely originate from the dispersion of the relative peculiar velocity between quasars and galaxies (including redshift errors), but is consistent with the predicted main radiative effect on the clustering of Ly $\alpha$  emitters in Zheng et al. (2011a).

The radiative transfer effect predicted in Zheng et al. (2011a) is a result of the anisotropic emission of scattered Ly $\alpha$  photons from the anisotropic distribution of neutral gas density and velocity (hence anisotropic distribution of Ly $\alpha$  scattering optical depth). In particular, Ly $\alpha$  photons preferentially escape along the direction for which the neutral gas has the largest peculiar velocity gradient, implying that large-scale density fluctuations that vary along the line of sight are suppressed in Ly $\alpha$  surface brightness, whereas those that vary perpendicular to the line of sight can be enhanced. This dependence on the line-of-sight peculiar velocity gradient is analogous to that in the Kaiser effect but of opposite sign, leading to the elongated correlation contours. For all the possible origins of Ly $\alpha$  emission discussed above, this radiative transfer effect should be at work as long as Ly $\alpha$  photons interact with neutral gas on large scales.

Our measurement can be used to constrain the parameters relevant to the radiative transfer effect. Following the simple model presented in Zheng et al. (2011a), the real-space overdensity  $\delta_\alpha$  (the Ly $\alpha$  surface brightness fluctuations here) can be related to the matter linear overdensity field  $\delta$  and peculiar velocity field as

$$1 + \delta_\alpha \propto (1 + b_L \delta) \left[ 1 + \alpha_1 \delta + \alpha_2 \frac{1}{Ha} \frac{\partial v_z}{\partial z} \right], \quad (31)$$

where  $b_L$  is the Ly $\alpha$  luminosity weighted bias of the underlying galaxy population and  $v_z$  the line-of-sight peculiar velocity. The coefficient  $\alpha_1$  represents a combined effect of the dependence of the Ly $\alpha$  radiative transfer (i.e. the Ly $\alpha$  escape emission) on the density and the transverse peculiar velocity gradient, and  $\alpha_2$  specifies the impact of the line-of-sight peculiar velocity gradient (see Zheng et al. 2011a, in particular their appendix A). Both coefficients are expected to be positive. In redshift space, we also need to add the Kaiser effect, and in terms of the Fourier component of the Ly $\alpha$  surface brightness fluctuations, we have

$$\delta_{\alpha,k}^s = (b_L + \alpha_1) [1 + \beta_\alpha \mu^2] \delta_k. \quad (32)$$



The  $\beta_\alpha$  parameter (as constrained in Section 3.4) is

$$\beta_\alpha = (1 - \alpha_2) \frac{\Omega_m(z = 2.55)^{0.6}}{b_L + \alpha_1}. \quad (33)$$

The factor ‘1’ comes from the Kaiser effect. The bias factor  $b_\alpha$  in equation (3) is  $b_L + \alpha_1$ . With  $\beta_\alpha = -0.76 \pm 0.36$ , we have  $\alpha_2 = 1 + (2.35 \pm 1.11)(b_\alpha/3)$  (N.B. this is the  $C_v$  coefficient defined in Wyithe & Dijkstra 2011). That is, the radiative transfer effect (indicated by a positive  $\alpha_2$ ) shows up or has been detected at a level of  $\sim 3.0\sigma$  for the fiducial value of  $b_\alpha = 3$ .

We note that the investigation of the anisotropic clustering effect induced by radiative transfer in Zheng et al. (2011a) is based on the  $z = 5.7$  output from a cosmological simulation. The strength of the effect depends on the coupling between the Ly  $\alpha$  photons and the IGM, which should vary with redshift. At  $z = 2-3$  relevant for this paper, as pointed out in Zheng et al. (2011a), the lower matter density from the cosmic expansion and the higher ionizing UV background lead to a lower neutral hydrogen density in the IGM, reducing the coupling. However, the coupling does not need a high neutral hydrogen column density (e.g. at  $T \sim 10^4$  K,  $\sim 10^{15} \text{cm}^{-2}$  is already optically thick for line-centre Ly  $\alpha$  photons). Furthermore, the decrease in the Hubble expansion rate tends to enhance the coupling. With a simple model of galaxies, the result from an analytic model in Wyithe & Dijkstra (2011) shows that even at  $z = 2-3$  the clustering effect induced by radiative transfer can still be significant. The picture can change depending on the strength of galactic winds, as the gas kinematics inside galaxies shapes the Ly  $\alpha$  line profile before Ly  $\alpha$  photons escape the galaxies, affecting the coupling to the IGM. Given the required large dynamical range and the uncertainty in modelling galaxy formation, a definite answer lacks for the strength of the anisotropic clustering effect at  $z = 2-3$  from studies with both simulations and analytic models (e.g. Wyithe & Dijkstra 2011; Zheng et al. 2011a; Behrens & Niemeyer 2013).

Overall, the quasar–Ly  $\alpha$  cross-correlation provides tentative evidence for the clustering effect caused by Ly  $\alpha$  radiative transfer (Zheng et al. 2011a). A better measurement with a larger data set and a more detailed modelling will help understand this effect to separate it from non-linear peculiar velocities and use it to constrain the neutral gas distribution.

### 6.5 Escape fraction and detected fraction

If the Ly  $\alpha$  emission correlated with quasars is produced by star-forming galaxies, our fiducial value of the star formation rate density of equation (22) is consistent with the whole dust corrected star formation rate in Bouwens et al. (2010). At face value, this indicates that dust has little effect in reducing the Ly  $\alpha$  emission, i.e. almost 100 per cent of the Ly  $\alpha$  photons produced from star formation escape. Clumpy dust clouds (Neufeld 1991) and anisotropic Ly  $\alpha$  escape (Zheng & Wallace 2014) may be possible mechanisms for explaining the high escape fraction of Ly  $\alpha$  photons, together with the much lower one for continuum UV photons.

We emphasize that the escape fraction is not the same as the fraction that is detected in Ly  $\alpha$  emitter surveys. The latter, the detected fraction, comes only from the central, high surface brightness part of the Ly  $\alpha$  emission from galaxies, and not from the extended Ly  $\alpha$  haloes of low surface brightness (Steidel et al. 2011; Zheng et al. 2011b; Matsuda et al. 2012; Momose et al. 2014). The detected fraction for Ly  $\alpha$  emission at  $z = 2-3$ , inferred from comparing the Ly  $\alpha$  luminosity density and H  $\alpha$  or H  $\beta$  luminosity density, is about 5 per cent (e.g. Hayes et al. 2011; Ciardullo et al. 2014). This fraction is consistent with the ratio between the star forma-

tion rate density inferred from Ly  $\alpha$  emitters and that from quasar–Ly  $\alpha$  cross-correlation in this work (see the blue and red points in Fig. 17).

## 7 SUMMARY AND CONCLUSIONS

### 7.1 Summary

We have carried out a cross-correlation analysis of residual light in SDSS/BOSS galaxy spectra and SDSS/BOSS quasars at redshifts from  $z = 2.0$  to  $3.5$ . We have used the Ly  $\alpha$  emission line (which is at wavelengths  $\lambda = 3647-5471$  Å at these redshifts) to trace structure in the cross-correlation. Our main findings are as follows.

(1) We measure large-scale structure in the quasar–Ly  $\alpha$  emission cross-correlation at a mean redshift  $z = 2.55$  at the  $8\sigma$  level. The cross-correlation function shape is consistent with the linear  $\Lambda$ CDM correlation function.

(2) Looking at the clustering as a function of separation across and along the line of sight we see evidence at the  $3.0\sigma$  level for distortions of clustering of the form expected to be caused by radiative transfer effects.

(3) We detect clustering independently in four subsamples at different redshifts, finding that the shape of the cross-correlation function is consistent with the fiducial sample. The amplitude of the cross-correlation increases by a factor of 3 between  $z = 3.5$  and  $2.0$ , although this detection of evolution is marginal, being consistent with no evolution at the  $2.0\sigma$  level.

(4) We find that the varying optical quality of the spectrograph camera across the field of view influences how well the quasar–Ly  $\alpha$  emission cross-correlation can be measured. Using only with the lower quality edge fibres results in a cross-correlation amplitude  $2\sigma$  below the full sample.

(5) Although the spectra are too shallow to allow making a good catalogue of emission lines, we are able to weakly constrain the contribution of emission lines to our signal statistically by fitting lines, subtracting them and remeasuring the quasar–Ly  $\alpha$  cross-correlation function. We find that lines with luminosities (measured in our 1 arcsec radius aperture) of  $L_{\text{Ly}\alpha} > 8 \times 10^{41} \text{erg s}^{-1}$  may contribute  $13^{+11}_{-13}$  per cent of the quasar–Ly  $\alpha$  cross-correlation amplitude at the relevant redshifts, but this contribution is still consistent with zero.

(6) In one of our sample tests, we measure the cross-correlation to be independent of quasar luminosity. This is evidence that the large-scale clustering of Ly  $\alpha$  surface brightness we measure arises mostly from Ly  $\alpha$  emission associated with star formation, and not from any systematic error associated with the quasar light.

(7) We estimate the plausible contribution to the quasar–Ly  $\alpha$  surface brightness we measure from a variety of physical processes alternative to star-forming galaxies that follow the same large-scale structure as quasars, such as fluorescence, scattering of the quasar emission, or the ionizing background. We find that cooling from intergalactic gas that is heated by quasars, either through double ionization of helium or by jet mechanical power, may contribute significantly to the detected Ly  $\alpha$  emission, although this is not favoured by the lack of correlation of the Ly  $\alpha$  brightness with quasar luminosity. Cooling radiation from gas accreting into galactic haloes may also contribute at about the 15 per cent level.

(8) Using measurements of clustering from SDSS/BOSS quasars, we convert the measured amplitude of the cross-correlation function  $b_q b_{\alpha} f_{\beta}(\mu)$  to the product of mean Ly  $\alpha$  sky brightness at  $z = 2.55$  and linear bias factor of Ly  $\alpha$  emission, finding

$\langle \mu_\alpha \rangle (b_\alpha/3) = (3.9 \pm 0.9) \times 10^{-21} \text{ erg s}^{-1} \text{ cm}^{-2} \text{ \AA}^{-1} \text{ arcsec}^{-2}$ . Assuming that this Ly  $\alpha$  surface brightness is due to star formation [see points (6) and (7) above], we convert our measured value to the mean star formation rate density in the Universe at redshift  $z = 2.55$ , finding  $\rho_{\text{SFR}} = (0.28 \pm 0.07)(3/b_\alpha) M_\odot \text{ yr}^{-1} \text{ Mpc}^{-3}$ . This is consistent with dust-corrected UV continuum based estimates of star formation, but more than an order of magnitude higher than previous estimates of the Star Formation Rate Density from surveys of individual Ly  $\alpha$  emitters.

## 7.2 Implications

A possible conclusion to draw from our work is that a high intensity Ly  $\alpha$  background at  $z \simeq 2.5$  is produced by Ly  $\alpha$  emitting galaxies that are clustered with quasars. If our measurement is confirmed and this conclusion is correct, the consequences for our understanding of galaxy formation and evolution are dramatic. The Ly  $\alpha$  emission directly observed in Ly  $\alpha$  emitting galaxies so far at these redshifts contributes only  $0.01 M_\odot \text{ yr}^{-1} \text{ Mpc}^{-3}$  to the mean star formation rate density. Extended haloes that have been seen around these Ly  $\alpha$  emitting galaxies detected in surveys can only contribute an additional  $\sim 30$  per cent to this value. We have detected  $21\text{--}35$  ( $\pm 1\sigma$  range) times more Ly  $\alpha$  photons than the Ly  $\alpha$  emitter surveys (with the uncertainty due to the factor  $(3/b_\alpha)$ ), and we have argued that most of these photons also arise from star formation. This amount of star formation represents most of the massive stars that are estimated to have formed in the universe and to have generated the present heavy elements.

To contribute to our measurement, these Ly  $\alpha$  photons cannot be absorbed by dust before escaping the galaxies. The question that remains then is how this Ly  $\alpha$  radiation could have been missed by previous observations. Putting forward and testing a detailed scenario is beyond the scope of this paper, but we speculate that all star-forming galaxies at these redshifts, even if they do not have any detectable Ly  $\alpha$  emission line in their central parts, may be surrounded by low surface brightness haloes that nevertheless have a high total integrated Ly  $\alpha$  luminosity. These low surface brightness haloes could allow the bulk of Ly  $\alpha$  photons to be below the detectable levels in narrow-band Ly  $\alpha$  emitter surveys. As much of the star formation in the Universe at redshifts  $z = 2\text{--}4$  occurs in massive galaxies, the implication is that a large fraction of the Ly  $\alpha$  emission we detect is from giant low surface brightness haloes around massive, bright, star-forming galaxies, which absorb most of their continuum photons to reradiate the energy in the infrared, and yet let their Ly  $\alpha$  photons escape.

An alternative interpretation of our measurement is that much of the Ly  $\alpha$  emission we have detected is physically associated with quasars, through the processes we have discussed of helium double ionization and dissipated energy from jets. We note that because the statistically significant signal we have measured is comprised within  $\sim 15 h^{-1} \text{ Mpc}$  of quasars, this is actually probing a very small fraction of the Universe. This scenario can therefore be best tested by measurements of other cross-correlations which use a set of probes with higher space density (such as Ly  $\alpha$  emitters or the Ly  $\alpha$  forest), or by measurement of the Ly  $\alpha$  emission autocorrelation.

## 7.3 The future

We note that the sky area covered by the million SDSS/BOSS (for our purposes randomly placed) fibres in the current study is  $\sim 3 \times 10^6 \text{ arcsec}^2$ . This is approximately  $1/200\,000$  of the entire sky area,

showing that Ly  $\alpha$  intensity mapping holds an enormous promise as a probe of structure in the Universe. In addition to the quasar–Ly  $\alpha$  emission cross-correlation employed in this paper, one can imagine carrying out Ly  $\alpha$  forest–Ly  $\alpha$  emission cross-correlations and Ly  $\alpha$  emission autocorrelation measurements. Correlations of Ly  $\beta$  absorption and Ly  $\alpha$  emission, and vice versa would avoid common mode systematics in the fluxing of spectrographs and may reduce the possibility of related error. As the Ly  $\alpha$  signal is distributed on large scales, a way to efficiently carry out intensity mapping surveys (for example for baryonic acoustic oscillation experiments) may be to use integral field spectroscopy with relatively low angular ( $\sim 10 \text{ arcsec}$ ) resolution on a wide-field ( $\sim \text{few deg}$ ) telescope. If bright point sources could be masked such an instrument could in principle capture a data set containing five orders of magnitude more information on large-scale clustering in the Universe than our present study.

## ACKNOWLEDGEMENTS

RACC was supported by NSF Awards AST-1009781, AST-1109730, OCI-0749212, the Moore Foundation and by the Leverhulme Trust's award of a Visiting Professorship at the University of Oxford. RACC would like to acknowledge the hospitality of the Astrophysics Subdepartment in Oxford where the initial stage of this work was carried out. RACC would also like to thank Matt McQuinn and Simon White for useful discussions. JM is partially supported by Spanish grant AYA2012-33938. ZZ is partially supported by NSF grant AST-1208891 and NASA grant NNX14AC89G.

Funding for SDSS-III has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation and the US Department of Energy Office of Science. The SDSS-III website is <http://www.sdss3.org/>. SDSS-III is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS-III Collaboration including the University of Arizona, the Brazilian Participation Group, Brookhaven National Laboratory, University of Cambridge, Carnegie Mellon University, University of Florida, the French Participation Group, the German Participation Group, Harvard University, the Instituto de Astrofísica de Canarias, the Michigan State/Notre Dame/JINA Participation Group, Johns Hopkins University, Lawrence Berkeley National Laboratory, Max Planck Institute for Astrophysics, Max Planck Institute for Extraterrestrial Physics, New Mexico State University, New York University, Ohio State University, Pennsylvania State University, University of Portsmouth, Princeton University, the Spanish Participation Group, University of Tokyo, University of Utah, Vanderbilt University, University of Virginia University of Washington and Yale University.

## REFERENCES

- Ahn C. et al., 2014, *ApJS*, 211, 17  
 Behrens C., Niemeyer J., 2013, *A&A*, 556, A5  
 Beifiori A. et al., 2014, *ApJ*, 789, 92  
 Berlind A. A., Weinberg D. H., 2002, *ApJ*, 575, 587  
 Bertone S., Schaye J., 2012, *MNRAS*, 419, 780  
 Blanc G. A. et al., 2011, *ApJ*, 736, 31  
 Bolton A. et al., 2012, *AJ*, 144, 144  
 Bouwens R. J. et al., 2009, *ApJ*, 705, 936  
 Bouwens R. J. et al., 2010, *ApJ*, 709, L133  
 Bovy J. et al., 2011, *ApJ*, 729, 141  
 Brownstein J. R. et al., 2012, *ApJ*, 744, 41  
 Busca N. G. et al., 2013, *A&A*, A96, 1  
 Cantalupo S., Porciani C., Lilly S. J., Miniati F., 2005, *ApJ*, 628, 61

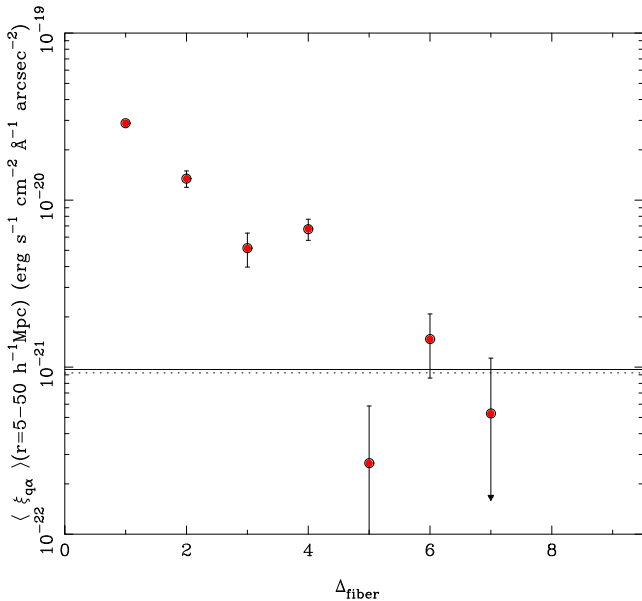
- Cantalupo S., Arrighi-Battaia F., Prochaska J. X., Hennawi J. F., Madau P., 2014, *Nature*, 506, 63
- Carilli C. L., 2011, *ApJ*, 730, L30
- Cassata P. et al., 2011, *A&A*, 525, A143
- Cassata P. et al., 2015, *A&A*, 573, A24
- Chang T.-C., Pen U.-L., Bandura K., Peterson J. B., 2010, *Nature*, 466, 463
- Ciardullo R. et al., 2014, *ApJ*, 796, 64
- Cowie L. L., Barger A. J., Hu E. M., 2010, *ApJ*, 711, 928
- Daddi E. et al., 2007, *ApJ*, 670, 156
- Davis M., Peebles P. J. E., 1983, *ApJ*, 267, 465
- Dawson K. S. et al., 2013, *AJ*, 145, 10
- Eisenstein D. J. et al., 2011, *AJ*, 142, 72
- Doré O. et al., 2014, preprint ([arXiv:1412.4872](https://arxiv.org/abs/1412.4872))
- Fardal M. A., Katz N., Gardner J. P., Hernquist L., Weinberg D. H., Davé M., 2001, *ApJ*, 562, 605
- Faucher-Giguère C.-A., Lidz A., Hernquist L., Zaldarriaga M., 2008, *ApJ*, 688, 85
- Faucher-Giguère C.-A., Kereš D., Dijkstra M., Hernquist L., Zaldarriaga M., 2010, *ApJ*, 725, 633
- Font-Ribera A. et al., 2012, *J. Cosmol. Astropart. Phys.*, 11, 059
- Font-Ribera A. et al., 2013, *J. Cosmol. Astropart. Phys.*, 05, 018
- Fontanot F., Cristiani S., Pfrommer C., Cupani G., Vanzella E., 2014, *MNRAS*, 438, 2097
- Fukugita M., Hogan C. J., Peebles P. J. E., 1998, *ApJ*, 503, 518
- Fujita S. S. et al., 2003, *AJ*, 125, 13
- Gauita L. et al., 2010, *ApJ*, 714, 255
- Gawiser E. et al., 2007, *ApJ*, 671, 278
- Goerdt T., Dekel A., Sternberg A., Ceverino D., Teyssier R., Primack J. R., 2010, *MNRAS*, 407, 613
- Gong Y., Cooray A., Silva M., Santos M., Lubin P., Rawlings S., 2011, *ApJ*, 728, L46
- Gould A., Weinberg D. H., 1996, *ApJ*, 468, 462
- Gronwall C. et al., 2007, *ApJ*, 667, 79
- Guaita L. et al., 2010, *ApJ*, 714, 255
- Gunn J. E. et al., 1998, *AJ*, 116, 3040
- Gunn J. E. et al., 2006, *AJ*, 131, 2332
- Haiman Z., Rees M. J., 2001, *ApJ*, 556, 87
- Hamilton A. J. S., 1992, *ApJ*, 385, L5
- Hawkins E. et al., 2003, *MNRAS*, 346, 78
- Hayes M., Schaerer D., Östlin G., Mas-Hesse J. M., Atek H., Kunth D., 2011, *ApJ*, 730, 13
- Hogan C. J., Weymann R. J., 1987, *MNRAS*, 225, 1
- Hu E. M., McMahon R. G., 1996, *Nature*, 6588, 231
- Kaiser N., 1987, *MNRAS*, 227, 1
- Keel W. C., Cohen S. H., Windhorst R. A., Waddington I., 1999, *AJ*, 118, 2547
- Kirkpatrick J. A., Schlegel D. J., Ross N. P., Meyers A. D., Hennawi J. F., Sheldon E. S., Schneider D. P., Weaver B. A., 2011, *ApJ*, 743, 125
- Kollmeier J. A., Zheng Z., Davé R., Gould A., Katz N., Miralda-Escudé J., Weinberg D. H., 2010, *ApJ*, 708, 1048
- Laursen P., Sommer-Larsen J., 2007, *ApJ*, 675, L69
- Lee K.-S. et al., 2011, *ApJ*, 733, 99
- Leitherer C. et al., 1999, *ApJS*, 123, 3
- Lewis A., Challinor A., Lasenby A., 2000, *ApJ*, 538, 473
- Madau P., Dickinson M., 2014, *ARA&A*, 52, 415
- Martin D. C., Chang D., Matuszewski M., Morrissey P., Rahman S., Moore A., Steidel C. C., 2014a, *ApJ*, 786, 106
- Martin D. C., Chang D., Matuszewski M., Morrissey P., Rahman S., Moore A., Steidel C. C., Matsuda Y., 2014b, *ApJ*, 786, 107
- McDonald M. et al., 2015, *ApJ*, 811, 111
- Matsuda Y. et al., 2012, *MNRAS*, 425, 878
- Menard B., Wild V., Nestor D., Quider A., Zibetti S., Rao S., Turnshek D., 2011, *MNRAS*, 417, 801
- Meurer G. R., Heckman T. M., Calzetti D., 1999, *ApJ*, 521, 64
- Momose R. et al., 2014, *MNRAS*, 442, 110
- Moster B. P., Somerville R. S., Maulbetsch C., van den Bosch F. C., Maccio A. V., Naab T., Oser L., 2010, *ApJ*, 710, 903
- Neufeld D. A., 1991, *ApJ*, L85
- Noterdaeme P., Srianand R., Mohan V., 2010, *MNRAS*, 403, 906
- Noterdaeme P., Petitjean P., Pâris I., Cai Z., Finley H., Ge J., Pieri M. M., York D. G., 2014, *A&A*, 566, A24
- Ouchi M. et al., 2003, *ApJ*, 582, 60
- Ouchi M. et al., 2008, *ApJS*, 176, 301
- Palanque-Delabrouille N. et al., 2013, *A&A*, 559, A85
- Pannella M. et al., 2009, *ApJ*, 698, L116
- Pâris I. et al., 2012, *A&A*, 548, A66
- Partridge R. B., Peebles P. J. E., 1967, *ApJ*, 147, 868
- Peterson J. B., Suarez E., 2012, *Proc. Moriond Cosmology*, preprint ([arXiv:1206.0143](https://arxiv.org/abs/1206.0143))
- Planck Collaboration XVI et al., 2014, *A&A*, 571, 16
- Prochaska J. X., Madau P., O'Meara J. M., Fumagalli M., 2014, *MNRAS*, 438, 476
- Pullen A. R., Doré O., Bock J., 2014, *ApJ*, 786, 111
- Rahmani H., Srianand R., Noterdaeme P., Petitjean P., 2010, *MNRAS*, 409, L59
- Rauch M. et al., 2008, *ApJ*, 681, 856
- Renzini A., Andreon S., 2014, *MNRAS*, 444, 3581
- Ross N. P. et al., 2012, *ApJS*, 199, 3
- Ross N. P. et al., 2013, *ApJ*, 773, 14
- Schlegel D. J., Finkbeiner D. P., Davis M., 1998, *ApJ*, 500, 525
- Silva M., Santos M., Gong Y., Cooray A., 2013, *ApJ*, 763, 132
- Smee S. A. et al., 2013, *AJ*, 126, 32
- Steidel C. C., Adelberger K. L., Shapley A. E., Pettini M., Dickinson M., Giavalisco M., 2000, *ApJ*, 532, 170
- Steidel C. C., Bogosavljević M., Shapley A. E., Kollmeier J. A., Reddy N. A., Erb D. K., Pettini M., 2011, *ApJ*, 736, 160
- Turner E. L., 1980, *ApJ*, 242, L135
- van Breukelen C., Jarvis M. J., Venemans B. P., 2005, *MNRAS*, 359, 895
- Visbal E., Loeb A., 2010, *J. Cosmol. Astropart. Phys.*, 11, 016
- White M. et al., 2012, *MNRAS*, 424, 933
- Wyithe J. S. B., Dijkstra M., 2011, *MNRAS*, 415, 3929
- Wyithe J. S. B., Morales M. F., 2007, *MNRAS*, 379, 1647
- Zheng Z., Wallace J., 2014, *ApJ*, 794, 116
- Zheng Z., Coil A. L., Zehavi I., 2007, *ApJ*, 667, 760
- Zheng Z., Cen R., Trac H., Miralda-Escudé J., 2011a, *ApJ*, 726, 38
- Zheng Z., Cen R., Weinberg D. H., Trac H., Miralda-Escudé J., 2011b, *ApJ*, 739, 62

## APPENDIX A: STRAY LIGHT CONTAMINATION

As mentioned in Section 3, a potentially very important systematic is the contamination of the galaxy spectra by quasar light. This is particularly relevant for our cross-correlation measurement because we are searching for light from sources that are clustered with quasars, and light from the quasar itself could mimic this if not treated carefully. The issue arises because the 1000 fibres that originate from the plate in the telescope focal plane are fed into (each red and blue) spectrograph and dispersed on to a single CCD. 1000 spectra are therefore extracted from a  $4k \times 4k$  chip and some light from neighbouring fibres can end up in CCD columns that are centred on other fibres.

### A1 Measured quasar–Lyman $\alpha$ surface brightness correlation

We can test for this contaminating light by use of two facts about the observational setup. The fibres are arranged by fibre number (from 1 to 1000) on the CCD, so that for each spectrum we can measure light from a certain number of fibres away. Quasars and galaxies that we use in our cross-correlation will also often be measured on different plates, so that there is no possibility for this contamination. We can therefore also test our result by restricting the calculation of



**Figure A1.** A test of stray light from quasars in nearby fibres contaminating the galaxy spectra. We show the quasar–Ly $\alpha$  emission surface brightness cross-correlation averaged over quasar–pixel pair separations of  $r = 5 - 50 h^{-1}$  Mpc plotted against  $\Delta_{\text{fiber}}$ , the difference in fibre number between the quasar and the galaxy spectra. A value of  $\Delta_{\text{fiber}} = 1$  means that the quasar and galaxy were in adjacent fibres and their spectra were dispersed next to each other in the CCD. The error bars are jackknife errors computed using 100 subsamples of the data in each case. The horizontal solid line is the quasar–Ly $\alpha$  emission surface brightness cross-correlation averaged over quasar–pixel pair separations of  $r = 5 - 50 h^{-1}$  Mpc for our fiducial computation (see Section 3.1), which uses information from all quasar–pixel pairs 6 fibres apart or greater. The dotted line (which lies close to the solid line) is the equivalent result but only using quasar–pixel pairs which are on different plates.

the quasar–Ly $\alpha$  cross-correlation to pairs of quasars and galaxies lying on different plates.

In the first test, we have computed the quasar–Ly $\alpha$  cross-correlation of equation (1) restricting ourselves to only quasar–galaxy spectrum pairs separated by specific values of  $\Delta_{\text{fiber}}$ , the difference in fibre number of the two spectra. Spectra with  $\Delta_{\text{fiber}} = 1$ , are adjacent on the CCD, for example, and have the highest potential for cross-contamination of light. We reduce the quasar–Ly $\alpha$  cross-correlation for each value of  $\Delta_{\text{fiber}}$  by averaging the  $\xi_{q\alpha}(r)$  results over a range from  $r = 5$  to  $50 h^{-1}$  Mpc. Our conclusions about stray light contamination are not dependent on the range picked.

The results are shown in Fig. A1, where we show  $\xi_{q\alpha}(r = 5-50 h^{-1} \text{Mpc})$  against  $\Delta_{\text{fiber}}$ . In the plot, the value of  $\xi_{q\alpha}(r = 5-50 h^{-1} \text{Mpc})$  for our fiducial sample (Section 3.1), which uses all pairs with  $\Delta_{\text{fiber}} \geq 6$  is shown as a horizontal line. We can see that for  $\Delta_{\text{fiber}} = 1$ , the mean surface brightness inferred from the cross-correlation is over 20 times the fiducial value. It remains significantly higher for all  $\Delta_{\text{fiber}} \leq 4$ . This is an indication that even when quasar and galaxy spectra are separated by three other spectra that light from the quasar is able to leak into the galaxy spectrum and contaminate it. Of course the particular region of the spectrum we are looking at (close to the Ly $\alpha$  emission line at the redshift of the quasar) is the one in which the quasar is extremely bright, and one would not expect it to contaminate other parts of galaxy spectrum as much. Nevertheless, for our project, this is the important region of the spectrum, and we therefore must apply a cut on  $\Delta_{\text{fiber}}$ . Based on Fig. A1, we have chosen that  $\Delta_{\text{fiber}} \geq 6$  in our analysis.

A related test is to compute  $\xi_{q\alpha}(r = 5-50 h^{-1} \text{Mpc})$  for quasars and galaxies that are on different plates (and therefore not dispersed at the same time on to the CCD). We show the results as a dashed line in Fig. A1, which is indistinguishable from the results from our fiducial analysis. This is good evidence that our cut  $\Delta_{\text{fiber}} \geq 6$  is sufficient to eliminate stray quasar light.

Yet another related test is to eliminate all galaxy fibres which have a quasar within a certain number of fibres. This is different from eliminating quasar–pixel pairs on a case-by-case basis because it will also eliminate any potential contamination which could come from quasars being clustered in space with other quasars. In the fiducial case, the contaminating quasars will have been eliminated directly, but those which are clustered with them could still contaminate the neighbouring galaxy spectra. One good reason to believe that this is not occurring to a detectable degree is that any contamination of this type would be a convolution of the redshift-space autocorrelation function of quasars and the contaminating surface brightness from the quasar Ly $\alpha$  line, which is highly asymmetric and extremely elongated along the line of sight (see below). This does not appear to describe the measured signal (e.g. Fig. 4).

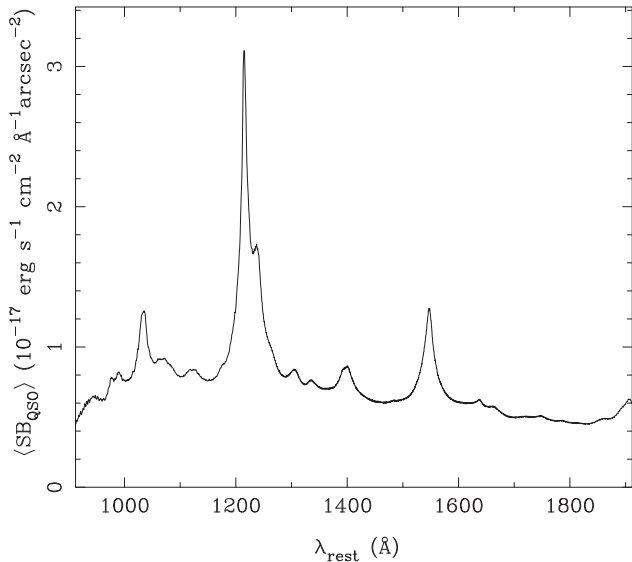
Nevertheless, we have carried out the test, which eliminates 50 per cent of the galaxy fibres from the data sample. We find an amplitude  $b_q b_{\alpha f\beta}(\mu) = 1.98^{+0.66}_{-0.65} \times 10^{-20} \text{ erg s}^{-1} \text{ cm}^{-2} \text{ \AA}^{-1} \text{ arcsec}^{-2}$ , and shape  $\Omega_m = 0.69^{+0.71}_{-0.36}$ . These are consistent at about the  $2\sigma$  with the measurement from the fiducial sample. We note that the detection level of the signal is only about  $3\sigma$  overall, compared to  $\sim 9\sigma$  for the fiducial measurement. This can be explained by the fact that eliminating all galaxies which have a quasar within 5 fibres will disproportionately affect the number of close quasar–pixel pairs. By directly counting, we find the number of quasar–pixel pairs with separations below  $40 h^{-1}$  Mpc has fallen by a factor of 3.5 rather than the factor of 2 expected if pairs were drawn uniformly from all fibre separations.

We have also carried out another, similar test, which eliminates a smaller fraction of the data set, but which should have the same effect. For this test, we remove from our list of galaxy pixels all pixels which have more than 1 quasar within  $r = 50 h^{-1}$  Mpc. In this way, we eliminate all potential cross-contamination from quasars clustered with the target quasars in the cross-correlation, but at a much reduced cost (doing this only eliminates 0.3 per cent of the galaxy spectrum pixels). After fitting the cross-correlation of this sample, we find an amplitude  $b_q b_{\alpha f\beta}(\mu) = 3.04^{+0.37}_{-0.45} \times 10^{-20} \text{ erg s}^{-1} \text{ cm}^{-2} \text{ \AA}^{-1} \text{ arcsec}^{-2}$ , and shape  $\Omega_m = 0.33^{+0.13}_{-0.10}$ . This is consistent with the fiducial result within the error bars. It is slightly lower (by less than  $1\sigma$ ), which should not be surprising, as we are presumably preferentially removing some pixels in overdense regions. Overall these two test results are a good sign that significant cross-contamination from quasars clustered with the target quasar is not occurring.

## A2 Modelling the stray light contamination

If the excess surface brightness seen above in galaxy spectra which are close on the CCD to quasar spectra is indeed due to cross-talk from quasar light, we would expect the contamination to have a quasar-like spectrum. To examine this, we first make a stacked spectrum of the DR10 quasar sample, by averaging all the spectra together in the quasar rest frame with unit weight. This is shown in Fig. A2. We can see from that figure that the quasar Ly $\alpha$  line is extremely broad, with an FWHM of  $\sim 50 \text{ \AA}$  in the rest frame. There is also the noticeable emission feature due to NV on the red wing of the line.





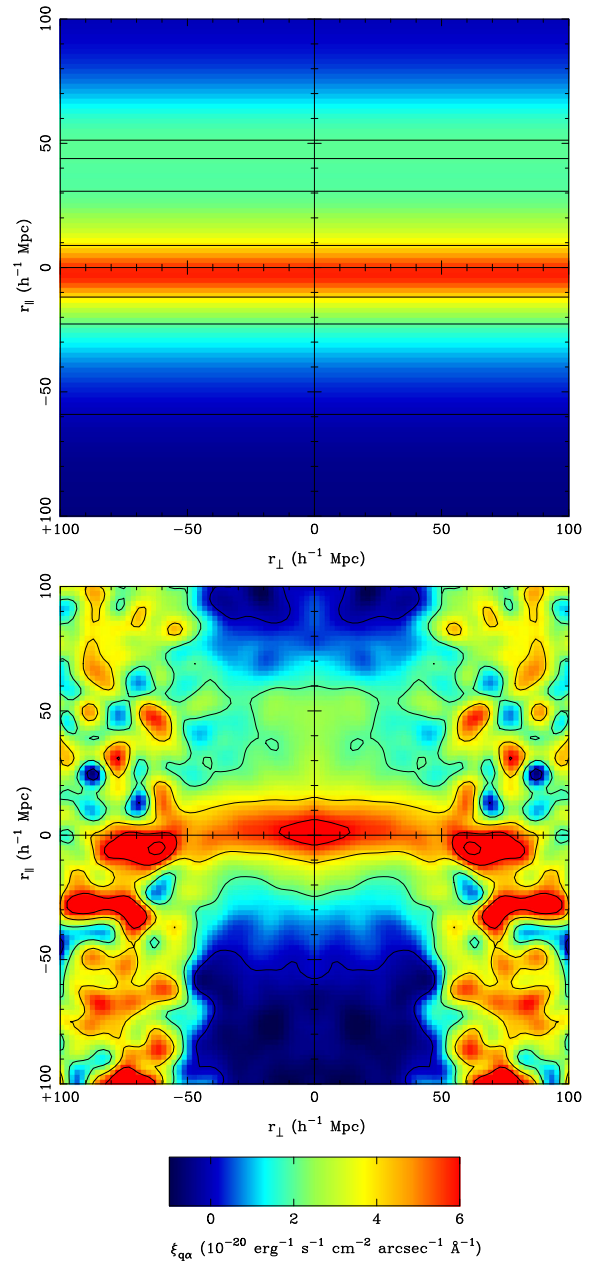
**Figure A2.** The mean surface brightness from all DR10 quasar spectra as a function of wavelength. We shift all spectra to their rest wavelength and stack them with equal weight to make the curve.

We next take this stacked spectrum, and convert the wavelength units into comoving  $h^{-1}$  Mpc at  $z = 2.55$ , the mean redshift of our measurements. If the quasar light is truly a contaminant, we expect the strength of the contamination to depend on the difference between fibre numbers ( $\Delta_{\text{fibre}}$ , as defined above) and not on the actual physical separation between quasar and pixel across the line of sight ( $r_{\perp}$ ). Plotting the quasar–Ly $\alpha$  cross-correlation for particular  $\Delta_{\text{fibre}}$  values as a function of  $r_{\perp}$  is therefore a useful check that light contamination is occurring as we believe. The prediction for the contamination, taken directly from the quasar spectrum is shown in the top panel of Fig. A3. In that figure we have scaled the overall amplitude by a factor of  $2 \times 10^{-3}$  (see below). We can see that the large width of the Ly $\alpha$  line, and its asymmetry due to the NV feature are both apparent.

The data with the strongest contamination should be from galaxy pixel–quasar pairs which are 1 fibre apart ( $\Delta_{\text{fibre}}=1$ ). We therefore plot the quasar–Ly $\alpha$  correlation for just those pixel–quasar pairs in the bottom panel of Fig. A3. It is immediately apparent that there is a stripe across the middle of the plot corresponding to the contaminating quasar Ly $\alpha$  emission line. The contamination is clearly asymmetric, extending further to positive values of  $r_{\parallel}$  than negative, and overall is visually consistent with the prediction based on the stacked quasar spectrum (top panel of Fig. A3). Because the quasar–pixel pairs that are exactly 1 fibre apart only comprise a very small fraction of the data, we expect there to be lots of noise in the plot, particularly for large values of  $r_{\perp}$ . This latter is because close pairs of quasars and galaxies on the sky are more likely to be close in  $\Delta_{\text{fibre}}$ . There is nevertheless enough range that we can examine by eye whether the contamination depends on  $r_{\perp}$ , with the answer appearing to be no. We examine this more quantitatively below.

When comparing the top and bottom panels of Fig. A3, we note that the amplitude of the stacked quasar spectrum is scaled by a factor of  $2 \times 10^{-3}$ , which is the level required to match the observations in a  $\chi^2$  fit (see below). This means that 0.2 per cent of the quasar light is scattered into neighbouring spectra (i.e. spectra with  $\Delta_{\text{fibre}} = 1$ ).

If there is a signal that is actually coming from quasar–Ly $\alpha$  emission correlations in the Universe, one would expect that to be

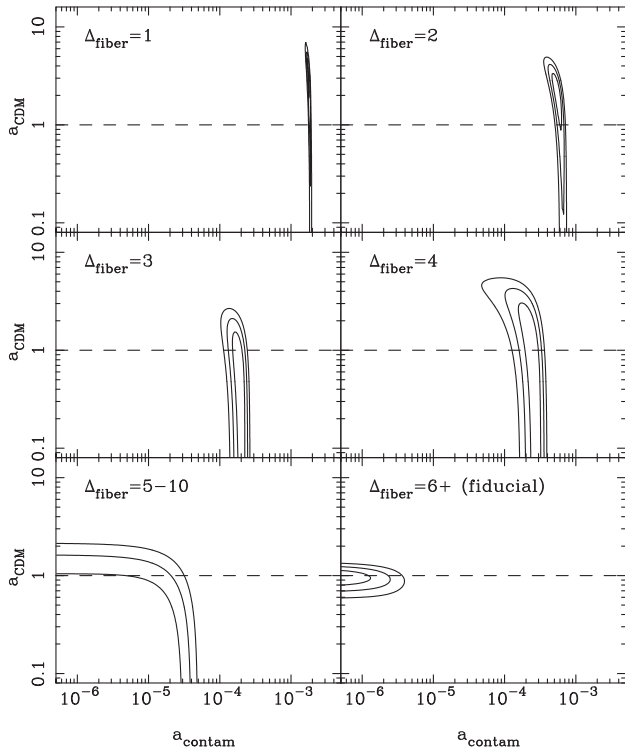


**Figure A3.** Top panel: the stacked quasar surface brightness from Fig. A2 multiplied by  $2.0 \times 10^{-3}$ . The stack is centred on the redshift of the Ly $\alpha$  line at redshift  $z = 2.55$  and the fiducial cosmology has been used to convert  $\text{\AA}$  to comoving  $h^{-1}$  Mpc. The stack is plotted as a function of  $r_{\parallel}$  and is independent of  $r_{\perp}$ . Bottom panel: The measured quasar–Ly $\alpha$  emission correlation for quasar–galaxy pixel pairs separated by 1 fibre ( $\Delta_{\text{fibre}} = 1$ ) as a function of  $r_{\parallel}$  and  $r_{\perp}$ . In both panels, a Gaussian filter with  $\sigma = 4 h^{-1}$  Mpc was used to smooth the image (as in Fig. 4).

superimposed on top of any contamination from stray light. We have therefore carried out a simple joint fit of  $\xi_{\text{q\alpha}}$  with a sum of the fiducial theoretical model from Fig. 4 (right-hand panel) scaled by an amplitude factor  $a_{\text{CDM}}$  and the stacked quasar contamination scaled by a factor  $a_{\text{contam}}$ . The fit is therefore as follows:

$$\xi_{\text{fit}}(r_{\perp}, r_{\parallel}) = a_{\text{CDM}} \xi_{\text{q\alpha, CDM}}(r_{\perp}, r_{\parallel}) + a_{\text{contam}} \xi_{\text{stack}}(r_{\perp}, r_{\parallel}), \quad (\text{A1})$$

where  $\xi_{\text{stack}}(r_{\perp}, r_{\parallel})$  (multiplied by  $2 \times 10^{-3}$ ) is shown in Fig. A2 (top panel). We separate the observational  $\xi_{\text{q\alpha}}$  results by  $\Delta_{\text{fibre}}$  and



**Figure A4.** A joint fit of quasar scattered light contamination and cosmological model for  $\xi_{q\alpha}(r)$  for different values of  $\Delta_{\text{fibre}}$ . We show  $1\sigma$ ,  $2\sigma$  and  $3\sigma$  confidence limits on the joint parameters  $a_{\text{contam}}$  and  $a_{\text{CDM}}$  from equation (A1).

then carrying out a  $\chi^2$  fit of the form given by equation (A1) to determine the best amplitude parameters  $a_{\text{contam}}$  and  $a_{\text{CDM}}$  as well as the confidence limits on those parameters. We carry out the fit for quasar–pixel pair separations  $r \leq 50 h^{-1}$  Mpc, although our conclusions are insensitive to this value.

The results are plotted in Fig. A4. For the  $\Delta_{\text{fibre}} = 1$  data set, we find a total  $\chi^2$  for the best fit of 1526 for 988 data points, which shows that the light contamination which dominates the fit is fairly well modelled by the quasar stack. It is perhaps not surprising the fit to the contamination is not perfect, as the contamination model is extremely simple. We have plotted the  $1\sigma$ ,  $2\sigma$  and  $3\sigma$  confidence contours on the parameters  $a_{\text{CDM}}$  and  $a_{\text{contam}}$  in Fig. A4. The central value of  $a_{\text{contam}}$  is  $1.75 \times 10^{-3}$ . The  $1\sigma$  confidence interval on  $a_{\text{CDM}}$  is consistent with unity, and also with zero. The results from  $\Delta_{\text{fibre}} = 1$  are therefore consistent with the sum of significant contamination from quasar light and a quasar–Ly  $\alpha$  correlation at the level predicted by the CDM model fit to the fiducial data set.

We have carried out the same fitting to data sets with different restricted values of  $\Delta_{\text{fibre}}$  in the other panels of Fig. A4. We can see that for cases with  $\Delta_{\text{fibre}} \leq 4$  there is a significant detection of a contamination contribution. The value of  $a_{\text{contam}}$  is lower for  $\Delta_{\text{fibre}} = 2-4$  than for  $\Delta_{\text{fibre}} = 1$  at the factor of  $10-4$  level. In all cases, the measurements are also consistent with an additional quasar–Ly  $\alpha$  correlation at the level of  $a_{\text{CDM}} = 1$ . For  $\Delta_{\text{fibre}} = 5-10$  and  $\Delta_{\text{fibre}} \geq 6$ , the contamination is not detectable.

The results of the fitting shown in Fig. A4 are a good sign that the signal and contamination are behaving in a particular way which can be accounted for by excising pairs of pixels from close fibres. It is useful however to examine the results for individual data points in more detail. Of particular interest is the measured quasar–Ly  $\alpha$  surface brightness as a function of  $r_{\perp}$ . As mentioned above, we

expect contamination at the CCD level for a given fibre separation to be independent of  $r_{\perp}$ . We have taken the measured  $\xi_{q\alpha}(r_{\perp}, r_{\parallel})$  values for different data sets limited by  $\Delta_{\text{fibre}}$  and averaged them between  $r_{\parallel} = \pm 50 h^{-1}$  Mpc. We plot this average,  $\langle \xi_{q\alpha} \rangle$  as a function of  $r_{\perp}$  in Fig. A5. In each panel, we also plot a dashed horizontal line showing a level of contamination that is independent of  $r_{\perp}$ , a dotted line showing the level of the signal from quasar–Ly  $\alpha$  clustering (we plot  $\langle \xi_{q\alpha} \rangle(r_{\perp})$  computed from the model shown in Fig. 4, right-hand panel), and a solid line which is the sum of the two.

We can see that in the  $\Delta_{\text{fibre}} = 1$  panel there is the strong contamination signal which is consistent with being independent of  $r_{\perp}$ . The small relative contribution of the actual signal makes little difference in this panel. In the  $\Delta_{\text{fibre}} = 2$  and  $\Delta_{\text{fibre}} = 3$  panels the level of contamination is lower (note that the y-axes are different in the different panels), but it still dominates over the expected signal, with again no sign of a dependence on  $r_{\perp}$ . For the fiducial panel (bottom right), we can see that the measured  $\langle \xi_{q\alpha} \rangle(r_{\perp})$  does have significant dependence on  $r_{\perp}$ , and this has a similar form and amplitude to the best-fitting model to the data in Section 3.2. There is no sign of a contaminating (independent of  $r_{\perp}$ ) component to this measurement.

Having carried out these tests on contamination from the Ly  $\alpha$  emission line from nearby quasars, it is pretty clear that the signal we are seeing in our fiducial data set cannot be caused by straightforward scattered light coming from nearby fibres. There remains the possibility however that there is an additional contaminating component from some other mechanism that has a dependence on  $r_{\perp}$ . This is difficult to imagine, but one can construct tests for this based on other spectral features than the Ly  $\alpha$  line. In the next two subsections we do this, first centred on the quasar C IV emission line and then using stars and the H  $\alpha$  line.

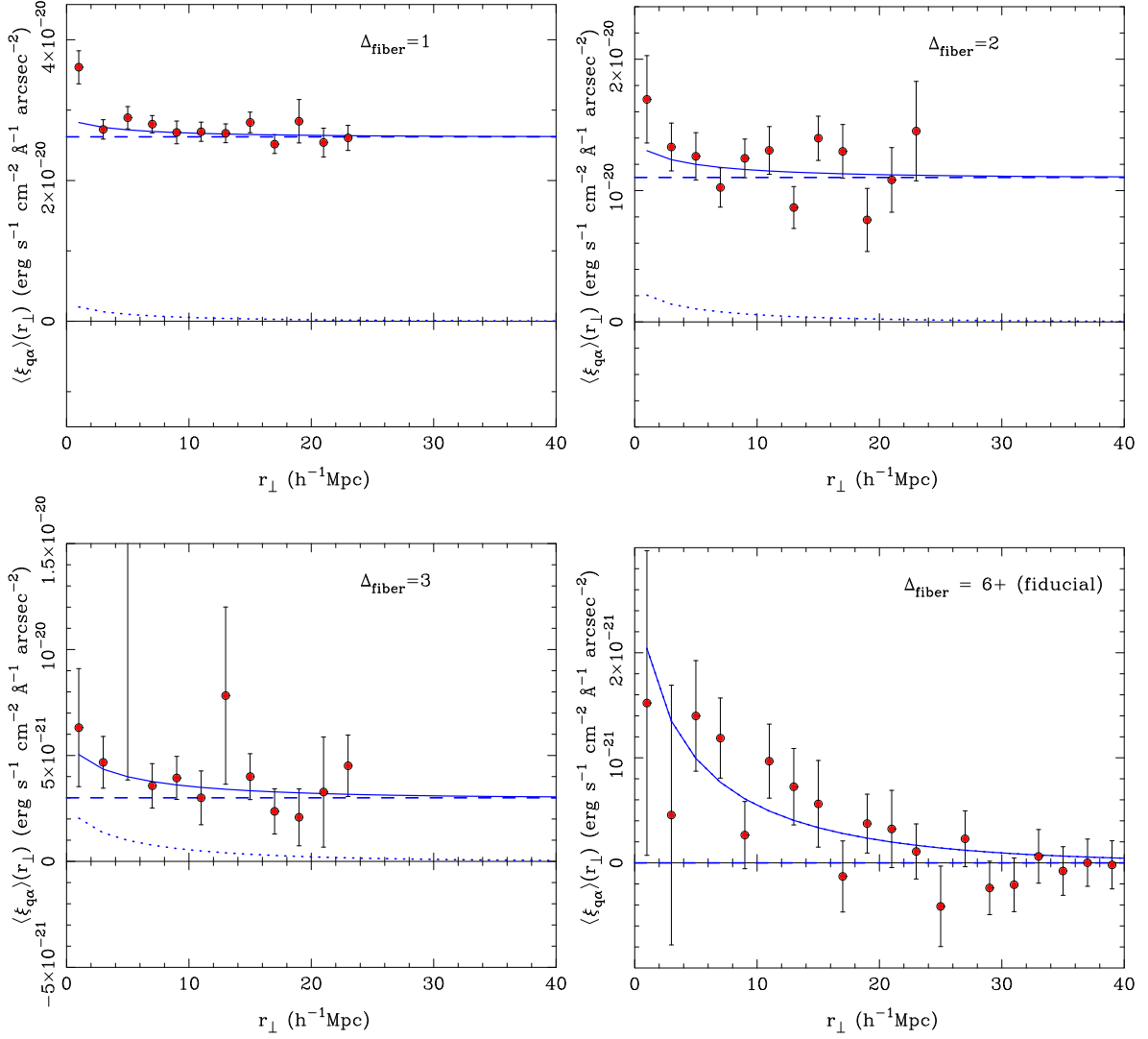
### A3 Tests with C IV

As a test of scattered light contamination, we carry out a different correlation between galaxy spectrum pixels and quasar positions. The only difference from our standard analysis (Section 3) is that we compute the cross-correlation at the wavelength of C IV at the redshift of the quasar. We use a rest wavelength of  $1550 \text{ \AA}$  for C IV (close to the centre of C IV doublet). If our interpretation of the contamination from nearby quasars from the previous subsection is correct, we would expect to see strong contamination from the C IV line for low  $\Delta_{\text{fibre}}$  and then no signal  $\xi_{q\text{CIV}}$  for large  $\Delta_{\text{fibre}}$  (because there should be no strong intergalactic C IV line emission).

As with Fig. A5, we plot  $\langle \xi_{q\alpha} \rangle(r_{\perp})$ , in Fig. A6. We show results for  $\Delta_{\text{fibre}} = 1$  and for  $\Delta_{\text{fibre}} \geq 6$ . In the case of  $\Delta_{\text{fibre}} = 1$  there is strong contamination from the quasar, which does not depend on  $r_{\perp}$ . The amplitude of the contamination is approximately three times smaller than that from Ly  $\alpha$ , which is the ratio one would expect from the relative strengths of the two lines in Fig. A2. For the  $\Delta_{\text{fibre}} \geq 6$  we see that there is no evidence for contamination, and, although the error bars are relatively large, there is no sign of any  $\xi_{q\text{CIV}}$  signal. We have plotted the fiducial  $\xi_{q\alpha}$  model on both panels with a dotted line. That the contamination behaves the same way as the Ly  $\alpha$  results in the previous subsection but that there is no  $\xi_{q\text{CIV}}$  signal is further evidence for the reality of the  $\xi_{q\alpha}$  measurement.

### A4 Tests with stars

A further test of stray light contamination which we can carry out is to cross-correlate the positions of stars with the galaxy pixels. We use all stars in the SDSS DR10 catalogue (Ahn et al. 2014), which is



**Figure A5.** The measured  $\xi_{q\alpha}$  from BOSS data averaged between  $r_{\parallel} = \pm 50 h^{-1}$  Mpc and plotted as function of  $r_{\perp}$ . We show results for different values of  $\Delta_{\text{fibre}}$  as indicated in each panel. The points with error bars show the observational results. The prediction of the theoretical model from equation (11) is shown as a dotted line. A dashed line indicates scattered light contamination of the sort discussed in Section 3.2 which is independent of  $r_{\perp}$  and the solid line is the sum of the two components.

171 612 objects in total. In this case we choose the  $H\alpha$  line, with rest wavelength  $6562.8 \text{ \AA}$  and cross-correlate star positions with galaxy pixels centred on this wavelength. We measure the resulting star- $H\alpha$  cross-correlation,  $\langle \xi_{\text{SH}\alpha} \rangle(r_{\perp})$  is for two values of  $\Delta_{\text{fibre}}$ , 1 and  $\geq 6$ . In this test, all the stars are obviously at the same rest wavelength, which effectively means that in this measurement we are stacking galaxy spectra together. This test is therefore likely to be more sensitive to the CIV test to residual artefacts in photocalibration of the spectrometer (see e.g. fig. 4 of Busca et al. 2013). Nevertheless it is useful to see whether any contamination has an  $r_{\perp}$  dependence.

We plot the results for  $\langle \xi_{\text{SH}\alpha} \rangle(r_{\perp})$  in Fig. A7, where we can see that there is again a sign of stellar light contamination when  $\Delta_{\text{fibre}} = 1$  and that this contamination appears to be independent of  $r_{\perp}$ . For  $\Delta_{\text{fibre}} \geq 6$ , the measurement is consistent with zero, again a sign that the cut in  $\Delta_{\text{fibre}}$  removes scattered light contamination.

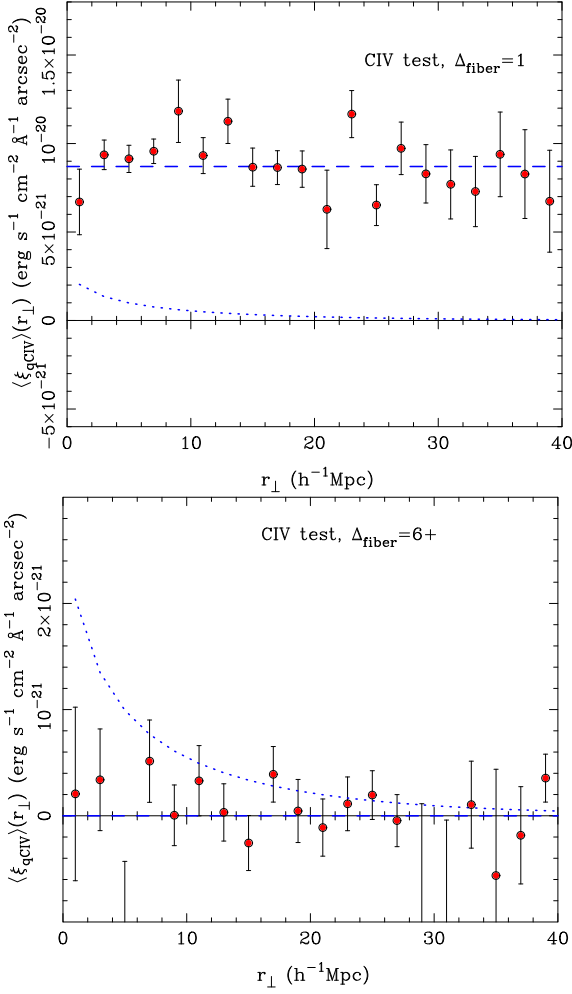
The results from the tests in Appendix A therefore suggest that although light contamination from quasars can be readily seen in the data, it can be eliminated to a high degree by excluding close pairs of fibres. When this is done, the resulting measured quasar- $\text{Ly}\alpha$

surface brightness correlation seems to be real. If instead it is produced by some as yet unknown systematic this effect would have to reproduce the  $r_{\perp}$  and  $r_{\parallel}$  dependence expected from a cosmological signal, be measurable only in the  $\text{Ly}\alpha$  wavelength range and not around other emission lines, and be independent of the position of the spectra on the detector.

Having searched for and as far as we can tell eliminated the effects of strong light contamination, we now turn to other potential systematic effects in the next two appendices.

## APPENDIX B: LARGE-SCALE SURFACE BRIGHTNESS CORRECTION

A potential systematic error on our measurement is any obscuration (such as that produced by Galactic dust) that may affect both quasar selection and  $\text{Ly}\alpha$  surface brightness, thus producing a spurious cross-correlation between the two on the plane of the sky. One can correct for such a cross-correlation by searching for an  $\xi_{q\alpha}(r_{\perp}, r_{\parallel})$

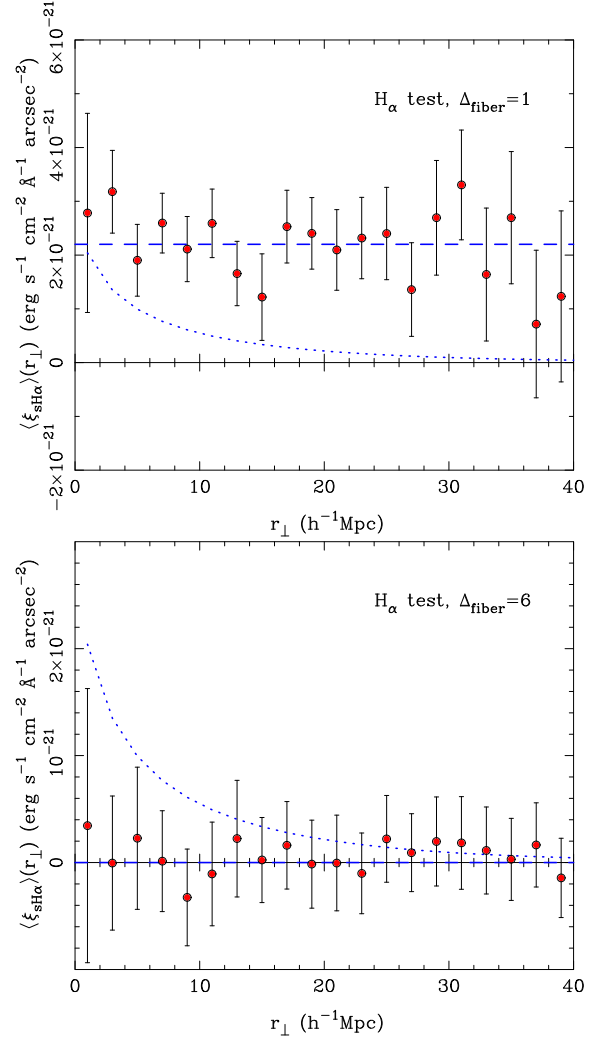


**Figure A6.** A null test using Carbon IV–quasar cross-correlations. Here we show  $\xi_{qCIV}$  measured from BOSS data averaged between  $r_{\parallel} = \pm 50 h^{-1}$  Mpc and plotted as function of  $r_{\perp}$ . We show results for two different values of  $\Delta_{\text{fibre}}$  as indicated in each panel. The points with error bars show the observational results. To aid comparison with previous plots, the prediction of the theoretical model from equation (11) for the quasar–Ly $\alpha$  cross-correlation is shown as a dotted line. A dashed line indicates scattered light contamination of the sort discussed in Section 3.2 which is independent of  $r_{\perp}$ .

signal which is non-zero for large values of  $r$ , where no physical (3D) correlation would be expected. This correlation which will be a function of  $r_{\perp}$  only (in the parallel line-of-sight approximation) can then be subtracted from the  $\xi_{q\alpha}$  signal. We have done this, computing the surface brightness correction  $\mu_{\alpha}$  given by

$$\mu_{\alpha}(r_{\perp}) = \frac{1}{2(x_a - x_b)} \left[ \int_{-x_a}^{-x_b} \xi_{q\alpha}(r_{\parallel}, r_{\perp}) dr_{\parallel} + \int_{x_b}^{x_a} \xi_{q\alpha}(r_{\parallel}, r_{\perp}) dr_{\parallel} \right], \quad (\text{B1})$$

where  $x_a = 400 h^{-1}$  Mpc and  $x_b = 80 h^{-1}$  Mpc and  $\xi_{q\alpha}(r_{\perp}, r_{\parallel})$  is the quasar–Ly $\alpha$  emission cross-correlation. The value of  $\mu_{\alpha}(r_{\perp})$  is shown in the bottom panel of Fig. B1. We can see that there is no coherent structure, as  $\mu_{\alpha}(r_{\perp})$  becomes both positive and negative as  $r_{\perp}$  varies. We have also plotted the best-fitting linear CDM model (as a function of  $r$ ) for the quasar–Ly $\alpha$  cross-correlation (from



**Figure A7.** A null test using H $\alpha$ –star cross-correlations. Here we show  $\xi_{sH\alpha}$  measured from BOSS data averaged between  $r_{\parallel} = \pm 50 h^{-1}$  Mpc and plotted as function of  $r_{\perp}$ . We show results for two different values of  $\Delta_{\text{fibre}}$  as indicated in each panel. The points with error bars show the observational results. To aid comparison with previous plots, the prediction of the theoretical model from equation (11) for the quasar–Ly $\alpha$  cross-correlation is shown as a dotted line. A dashed line indicates scattered light contamination of the sort discussed in Section 3.2 which is independent of  $r_{\perp}$ .

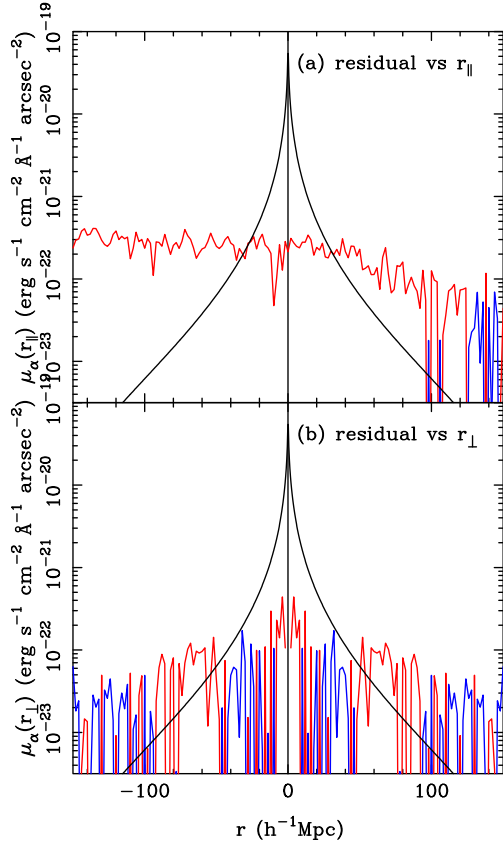
Section 3.2), and we can see that this dominates over the  $\mu_{\alpha}(r_{\perp})$  correction on small scales  $r_{\perp} < 40 h^{-1}$  Mpc, as we would hope.

There is also another analogous correction, but one that applies in the orthogonal direction. Any systematic effect which affects the Ly $\alpha$  emission in the line-of-sight direction (including redshift evolution in the Ly $\alpha$  surface brightness), and which is correlated with evolution in the quasar population with redshift (or at least evolution in the efficiency of quasar selection with redshift) could produce a spurious quasar–Ly $\alpha$  emission cross-correlation. We compute how  $\xi_{q\alpha}$  varies as a function of  $r_{\parallel}$  for large  $r_{\perp}$  values where there should be minimal physical clustering. In this case the surface brightness correction  $\mu_{\alpha}$  is given by

$$\mu_{\alpha}(r_{\parallel}) = \frac{2}{(x_a^2 - x_b^2)} \int_{x_b}^{x_a} \xi_{q\alpha}(r_{\perp}, r_{\parallel}) r_{\perp} dr_{\perp}, \quad (\text{B2})$$

where we use the same values of  $x_a$  and  $x_b$  as in equation (B1). The results for  $\mu_{\alpha}(r_{\parallel})$  are shown in the top panel of Fig. B1. We can see





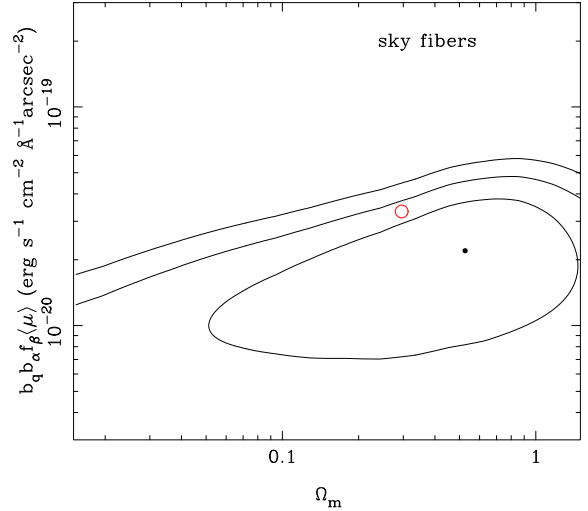
**Figure B1.** Large-scale residual flux (equations B1 and B2). Negative values are shown in blue and positive values in red.

that in this case there is a significant trend in the surface brightness correction with  $\mu_\alpha(r_\parallel)$  gradually decreasing from positive values of  $\sim 4 \times 10^{-22} \text{ erg s}^{-1} \text{ cm}^{-2} \text{ \AA}^{-1} \text{ arcsec}^{-2}$  for  $r_\parallel = -150 h^{-1} \text{ Mpc}$  to  $\sim 10^{-23} \text{ erg s}^{-1} \text{ cm}^{-2} \text{ \AA}^{-1} \text{ arcsec}^{-2}$  for  $r_\parallel = +150 h^{-1} \text{ Mpc}$ . As with the previous ( $r_\perp$ ) correction, the CDM model dominates on scales  $r_\perp < 40 h^{-1} \text{ Mpc}$ , where most of the clustering signal is located.

In our analyses in the main body of the paper, we have applied these  $\mu_\alpha(r_\parallel)$  and  $\mu_\alpha(r_\perp)$  corrections to the computation of the quasar–Ly  $\alpha$  emission cross-correlation  $\xi_{q\alpha}$ . We have done this on a quasar–pixel pair basis, computing  $r_\parallel$  and  $r_\perp$  for each pair, and then subtracting the appropriate value of  $\mu_\alpha$ . In all cases, the effect of the correction is small, as we would expect given the much large amplitude of the clustering signal compared to the surface brightness corrections that can be seen in Fig. B1. For example, in our fiducial case, without applying the large-scale surface brightness corrections, we find a best-fitting values of the shape and amplitude parameters of Section 3.2 of  $b_q b_{\alpha f_\beta} \langle \mu \rangle = 3.5 \text{ erg s}^{-1} \text{ cm}^{-2} \text{ \AA}^{-1} \text{ arcsec}^{-2}$  and  $\Omega_m = 0.26$ , which represent differences of  $0.5\sigma$  and  $0.5\sigma$  in the parameters, respectively.

## APPENDIX C: SAMPLE TESTS

In this appendix, we report on three consistency tests of our cross-correlation results. One test uses sky fibres to search for Ly  $\alpha$  emission instead of galaxy fibres, and the second looks at how the luminosity of the originally targeted galaxy affects the quasar–Ly  $\alpha$  cross-correlation. The third test investigates how quasar luminosity affects the cross-correlation.



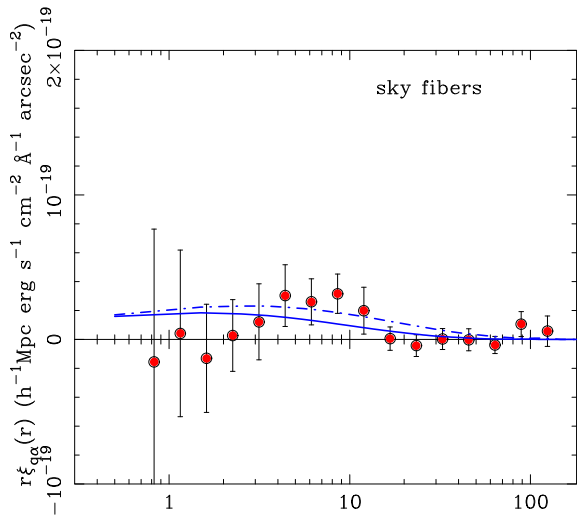
**Figure B2.** Fit parameters for the CDM model fit (as in Fig. 3) to the quasar–Ly  $\alpha$  cross-correlation function (equation 3) using sky fibres. The point shows the best-fitting values of the amplitude ( $b_q b_{\alpha f_\beta} \langle \mu \rangle$ ) and shape ( $\Omega_m$ ) and the contours show the  $1\sigma$ ,  $2\sigma$  and  $3\sigma$  confidence intervals. The open circle shows the best-fitting values of the fit parameters for the fiducial sample.

## C1 Sky fibres

Ideally one would like to be able to use fibres positioned on random areas of the sky to carry out intensity mapping, without needing to worry about subtracting a foreground galaxy. This approach is being carried out for example by HETDEX (Blanc et al. 2011). In our case, however, there are a number of such random fibres which were obtained for each plate, to use in sky subtraction. The number of fibres available for use is 146 065, approximately 15 per cent as many as there are galaxy fibres in our fiducial data set. We have taken these sky fibres and carried out the quasar–Ly  $\alpha$  cross-correlation of equation (1). The calculation was the same as in the fiducial case, including subtraction of a background from fibre (Fig. 1).

We show the results for  $\xi_{q\alpha}(r)$  for the sky fibres in Fig. C1. We can see that the data points appear to be consistent with the trend delineated by the best fit to the fiducial computation (also shown, as a dash-dotted line), although the measurement is much noisier. We fit the same CDM linear correlation function as was carried out in Section 3.2 and show the best-fitting parameters and confidence contours in Fig. B2. We find  $b_q b_{\alpha f_\beta} \langle \mu \rangle = 2.2_{-1.0}^{+1.0} \times 10^{-20} \text{ erg s}^{-1} \text{ cm}^{-2} \text{ \AA}^{-1} \text{ arcsec}^{-2}$  and  $\Omega_m = 0.52_{-0.36}^{+0.48}$ . Here we can see that the confidence contours are very wide, and that the clustering signal is present at only the  $2\sigma$  confidence level. The results are also consistent with the results from the fiducial sample at the  $1.5\sigma$  level. At present therefore, this shows that the use of galaxy and sky fibres in computing the cross-correlation is approximately equivalent, although the uncertainty is obviously large.

In the future, it would arguably be best to use randomly distributed fibres to carry out the cross-correlation, to avoid any potential selection biases. In the current sample case these biases are below the level of detectability, but one can imagine two types of bias, related to the fact that galaxy fibres are selected in regions of above average sky brightness and sky fibres selected to be in regions of below average sky brightness. Both effects could in principle bias the measurement of Ly  $\alpha$  surface brightness in opposite directions.



**Figure C1.** The quasar–Ly $\alpha$  cross-correlation function  $\xi_{q\alpha}(r)$  (equation 1) computed using sky fibres instead of galaxy-subtracted galaxy fibre. The solid line shows a fit to the CDM model and the dash-dotted line shows the fit from our fiducial sample (Fig. 3), which used galaxy-subtracted galaxy fibres.

We now carry out a further test to constrain this effect, by splitting the spectra into two halves based on target LRG luminosity.

### C2 Sample split by galaxy luminosity

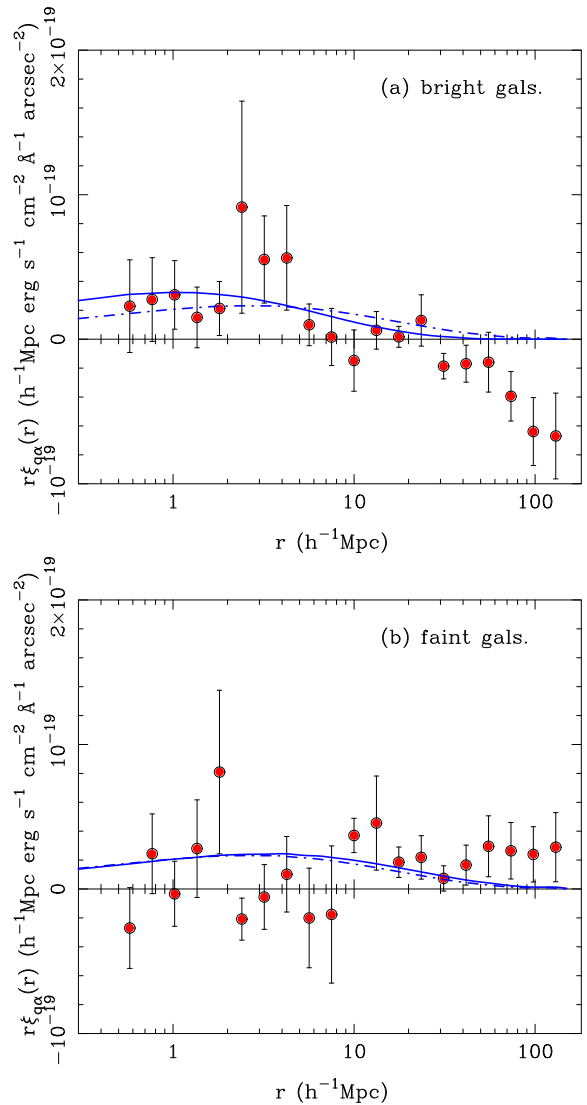
We divide the galaxy spectrum sample into two halves, based on the measured SDSS  $r$ -band luminosity (no  $k$ -correction was applied). One half consists of galaxies above the median luminosity of the whole data set ( $5.2 \times 10^{40}$  erg s $^{-1}$ ), and the other half those below it. The median  $r$ -band luminosities of the halves are  $6.9 \times 10^{40}$  and  $3.8 \times 10^{40}$  erg s $^{-1}$ .

We note that the LRG galaxy surface brightnesses measured from BOSS spectra are of the order of  $10^{-17}$  erg s $^{-1}$  cm $^{-2}$   $\text{\AA}^{-1}$  arcsec $^{-2}$ , approximately 2–3 orders of magnitude brighter than the mean Ly $\alpha$  surface brightness  $\langle\mu_{\alpha}\rangle$  that we have measured in Section 5. We therefore expect that it is unlikely that regions of excess background Ly $\alpha$  surface brightness could have caused certain LRGs to be preferentially selected and therefore bias our measurements.

We measure the quasar–Ly $\alpha$  cross-correlation  $\xi_{q\alpha}(r)$  for the two samples of spectra and show the results in Fig. C2. Both subsamples show evidence of clustering that is consistent with the CDM model shape and the fiducial amplitude. This can be seen in a quantitative manner from Fig. C3, where we show the fit parameters. We find  $b_q b_{\alpha} f_{\beta} \langle\mu\rangle = 3.1^{+0.6}_{-0.7} \times 10^{-20}$  erg s $^{-1}$  cm $^{-2}$   $\text{\AA}^{-1}$  arcsec $^{-2}$ , for the bright galaxy spectra and  $b_q b_{\alpha} f_{\beta} \langle\mu\rangle = (3.8 \pm 0.5) \times 10^{-20}$  erg s $^{-1}$  cm $^{-2}$   $\text{\AA}^{-1}$  arcsec $^{-2}$ , for the faint galaxy spectra. Both samples are consistent with the amplitude of the fiducial result at the  $1\sigma$  level (the fiducial sample fractional error bar is bar  $^{+12 \text{ per cent}}_{-11 \text{ per cent}}$  for fixed  $\Omega_m = 0.30$ ). This indicates that any bias present in the mean surface brightness  $\langle\mu_{\alpha}\rangle$  due to fibre target selection is likely to be at the  $\sim 10$  per cent level or less.

### C3 Sample split by galaxy redshift

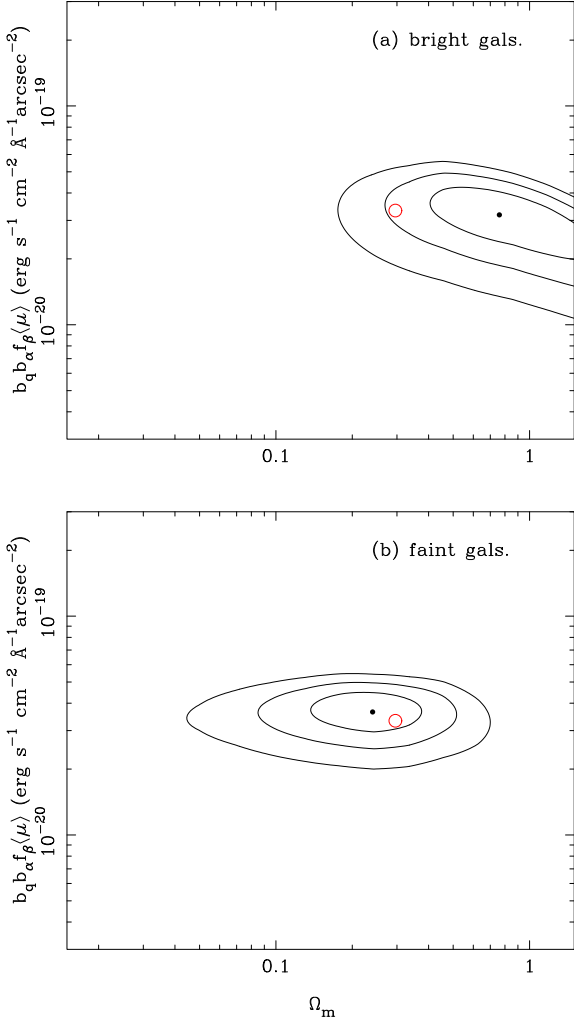
The effective radii of most BOSS LRGs are between 5–10 kpc (Beifiori et al. 2014), which is 1–2 arcsec for  $z = 0.4$  and 0.7–1.4 arcsec for  $z = 0.6$ . This means that for a given galaxy the



**Figure C2.** Quasar–Ly $\alpha$  cross-correlation  $\xi_{q\alpha}(r)$  for two subsamples of spectra originally targeted at LRGs with (a) luminosity above the median value and (b) below the median value. The solid curve shows the best-fitting CDM model and the dash-dotted line the CDM fit to the fiducial sample (all LRGs).

fraction of the LRG’s light which enters into the 1 arcsec radius fibre aperture will depend on its redshift. Therefore another useful test of the possible effect of LRG light on the Ly $\alpha$  surface brightness measurement is one based on the galaxy redshift. We divide the LRG galaxy spectrum sample into two halves. One half consists of galaxies above the median redshift of the whole data set ( $z=0.508$ ), and the other half those below it. The median galaxy redshifts of the halves are  $z = 0.579$  and  $0.403$ .

We find that both subsamples show evidence of clustering that is consistent with the CDM model shape. For reasons of space, we do not include figures. Doing the CDM model fitting, we find fit parameters  $b_q b_{\alpha} f_{\beta} \langle\mu\rangle = (3.2 \pm 0.5) \times 10^{-20}$  erg s $^{-1}$  cm $^{-2}$   $\text{\AA}^{-1}$  arcsec $^{-2}$ , for the high-redshift LRG spectra and  $b_q b_{\alpha} f_{\beta} \langle\mu\rangle = (2.5 \pm 0.7) \times 10^{-20}$  erg s $^{-1}$  cm $^{-2}$   $\text{\AA}^{-1}$  arcsec $^{-2}$ , for the low-redshift spectra. Both samples are consistent with the amplitude of the fiducial result at approximately the  $1\sigma$  level.

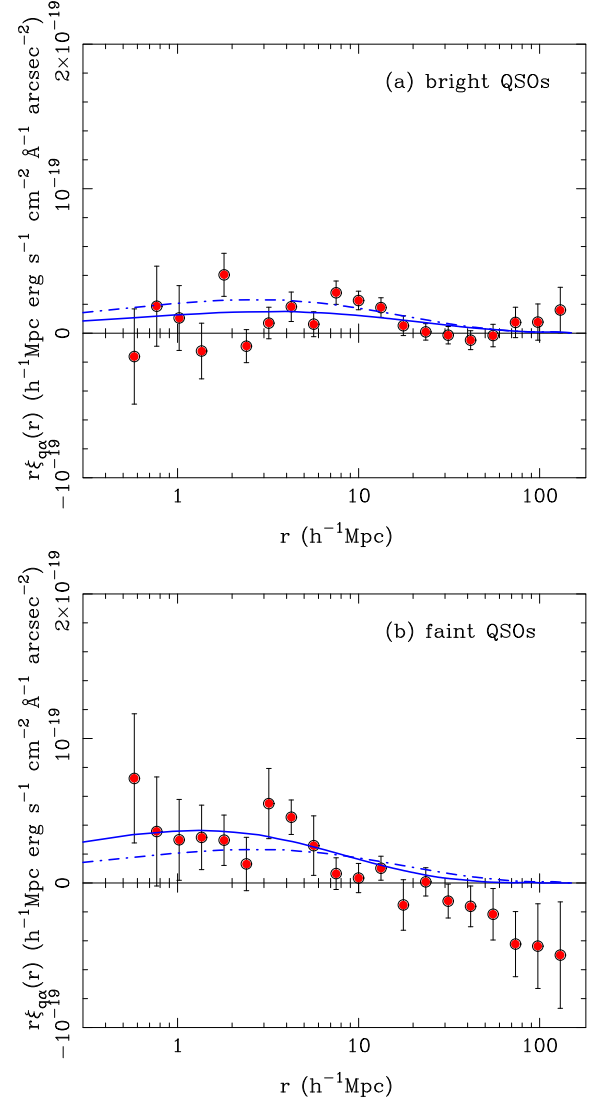


**Figure C3.** CDM amplitude and shape parameters fitted to the quasar–Ly  $\alpha$  cross-correlation  $\xi_{q\alpha}(r)$  for two subsamples of spectra originally targeted at LRGs with (a) luminosity above the median value and (b) below the median value. The points shows the best-fitting values of the amplitude ( $b_q b_{\alpha f_{\beta}}(\mu)$ ) and shape ( $\Omega_m$ ) and the contours show the  $1\sigma$ ,  $2\sigma$  and  $3\sigma$  confidence intervals. The open circles show the best-fitting values of the fit parameters for the fiducial sample.

#### C4 Sample split by quasar luminosity

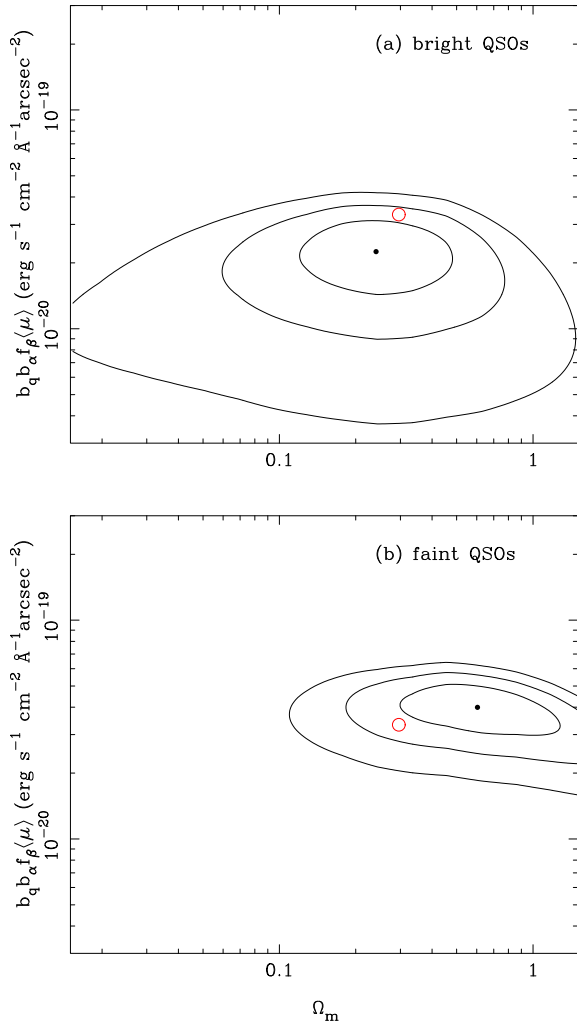
A potential source of Ly  $\alpha$  emission clustered with quasars is recombination radiation from dense IGM systems illuminated by the quasars themselves (e.g. Kollmeier et al. 2010). We have made an estimate of the amplitude of this signal in Section 6 and find it likely that it is much smaller than the signal from star-forming galaxies clustered with the quasar. One way of testing this directly is by splitting the sample into high- and low-luminosity quasars and then measuring the  $\xi_{q\alpha}(r)$  signal for each. We note that this test can also function as a diagnostic for stray light contamination from quasars, as any stray light should be correlated with quasar luminosity.

In this section we do this, measuring  $\xi_{q\alpha}(r)$  the subsample of quasars with SDSS  $r$ -band luminosity above the median, and with the subsample with luminosities below the median. The median luminosity of the bright subsample is 3.45 times the median luminosity of the faint sample, and the median redshifts of the quasars in each are  $z = 2.66$  and  $2.44$ , respectively.



**Figure C4.** Quasar–Ly  $\alpha$  cross-correlation  $\xi_{q\alpha}(r)$  for two subsamples of quasar data with (a) luminosity above the median value and (b) below the median value. The solid curve shows the best-fitting CDM model and the dash–dotted line the CDM fit to the fiducial sample (all quasars).

We can see in Figs C4 that the  $\xi_{q\alpha}(r)$  results for two subsamples do not look very different. In particular, if the quasar luminosities were causing significant fluorescent emission from nearby IGM material one might expect the brighter subsample to exhibit a steeper  $\xi_{q\alpha}(r)$  on small scales, which is not the case. From Fig. C5, we can see that the fit parameters for the bright subsample are within  $1\sigma$  of the fiducial result and the faint subsample within  $2\sigma$ . The amplitude of parameter ( $b_q b_{\alpha f_{\beta}}(\mu)$ ) (for *Planck*  $\Omega_m = 0.30$ ) is actually slightly lower for the bright subsample, being  $(2.3 \pm 0.6) \times 10^{-20} \text{ erg s}^{-1} \text{ cm}^{-2} \text{ \AA}^{-1} \text{ arcsec}^{-2}$ , whereas for the faint subsample it is  $3.9_{-0.6}^{+0.7} \times 10^{-20} \text{ erg s}^{-1} \text{ cm}^{-2} \text{ \AA}^{-1} \text{ arcsec}^{-2}$ . Given that the quasar subsamples are different in luminosities by a factor of 3.45, one would expect there to be a significant difference the  $\xi_{q\alpha}(r)$  results in the opposite direction if quasar properties were significantly affecting the large-scale Ly  $\alpha$  intensity around them. As mentioned above, this lack of dependence on quasar luminosity is also strong evidence that stray light from quasars scattered in the



**Figure C5.** CDM amplitude and shape parameters fitted to the quasar–Ly  $\alpha$  cross-correlation  $\xi_{q\alpha}(r)$  for two subsamples of quasars with (a) luminosity above the median value and (b) below the median value. The points show the best-fitting values of the amplitude ( $b_q b_{\alpha} f_{\beta}(\mu)$ ) and shape ( $\Omega_m$ ) and the contours show the  $1\sigma$ ,  $2\sigma$  and  $3\sigma$  confidence intervals. The open circles show the best-fitting values of the fit parameters for the fiducial sample.

spectrograph is not contaminating our measured cross-correlation signal.

<sup>1</sup>*McWilliams Center for Cosmology, Department of Physics, Carnegie Mellon University, Pittsburgh, PA 15213, USA*

- <sup>2</sup>*Astrophysics, University of Oxford, Keble Road, Oxford OX1 3RH, UK*  
<sup>3</sup>*Institució Catalana de Recerca i Estudis Avançats, E-08001 Barcelona, Catalonia, Spain*  
<sup>4</sup>*Institut de Ciències del Cosmos, Universitat de Barcelona/IEEC, E-08028 Barcelona, Catalonia, Spain*  
<sup>5</sup>*Department of Physics and Astronomy, University of Utah, 115 S 1400 E, Salt Lake City, UT 84112, USA*  
<sup>6</sup>*Department of Astronomy and Astrophysics, The University of Chicago, 5640 South Ellis Avenue, Chicago, IL 60615, USA*  
<sup>7</sup>*Department of Astronomy, Harvard University, 60 Garden St, Cambridge MA 02138, USA*  
<sup>8</sup>*Apache Point Observatory, PO Box 59, Sunspot, NM 88349, USA*  
<sup>9</sup>*Department of Astrophysical Sciences, Princeton University, Princeton, NJ 08544, USA*  
<sup>10</sup>*CEA, Centre de Saclay, IRFU, F-91191 Gif-sur-Yvette, France*  
<sup>11</sup>*Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA*  
<sup>12</sup>*APC, Université Paris Diderot-Paris 7, CNRS/IN2P3, CEA, Observatoire de Paris, 10, rue A. Domon & L. Duquet, F-75013 Paris, France*  
<sup>13</sup>*Max-Planck-Institut für Astronomie, Königstuhl 17, D-69117 Heidelberg, Germany*  
<sup>14</sup>*Department of Physics and Astronomy, University of Wyoming, Laramie, WY 82071, USA*  
<sup>15</sup>*INAF, Osservatorio Astronomico di Trieste, Via G. B. Tiepolo 11, I-34131 Trieste, Italy*  
<sup>16</sup>*Université Paris 6 et CNRS, Institut d’Astrophysique de Paris, 98bis Blvd. Arago, F-75014 Paris, France*  
<sup>17</sup>*A\*MIDEX, Aix Marseille Université, CNRS, LAM (Laboratoire d’Astrophysique de Marseille) UMR 7326, F-13388 Marseille, France*  
<sup>18</sup>*Department of Astronomy and Space Science, Sejong University, Seoul 143-747, Korea*  
<sup>19</sup>*Department of Astronomy and Astrophysics, The Pennsylvania State University, University Park, PA 16802, USA*  
<sup>20</sup>*Institute for Gravitation and the Cosmos, The Pennsylvania State University, University Park, PA 16802, USA*  
<sup>21</sup>*Bldg 510 Brookhaven National Laboratory, Upton, NY 11973, USA*  
<sup>22</sup>*INFN/National Institute for Nuclear Physics, Via Valerio 2, I-34127 Trieste, Italy*  
<sup>23</sup>*Department of Astronomy, Ohio State University, 140 West 18th Avenue, Columbus, OH 43210, USA*

This paper has been typeset from a  $\text{\TeX}/\text{\LaTeX}$  file prepared by the author.