Modelling Lyman α forest cross-correlations with LyMAS

Cassandra Lochhaas,[1]* David H. Weinberg,[1]* Sébastien Peirani,[2] Yohan Dubois,[2] Stéphane Colombi,[2] Jérémy Blaizot,[3] Andreu Font-Ribera,[4,5] Christophe Pichon[2,6] and Julien Devriendt[2]

[1]*Department of Astronomy and CCAPP, The Ohio State University, 140 West 18th Avenue, Columbus, OH 43210, USA*
[2]*Institut d'Astrophysique de Paris, Sorbonne Universitiès, UPMC Univ Paris 6 et CNRS, UMR 7095, 98 bis Bd Arago, F-75014 Paris, France*
[3]*Université de Lyon, F-69003 Lyon, France*
[4]*Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA*
[5]*Kavli IPMU (WPI), UTIAS, The University of Tokyo, Kashiwa, Chiba 277-8583, Japan*
[6]*Institute of Astronomy, Cambridge University, Madingley Road, Cambridge CB3 0HA, UK*

ABSTRACT

We use the Lyα Mass Association Scheme (LyMAS) to predict cross-correlations at $z = 2.5$ between dark matter haloes and transmitted flux in the Lyα forest, and compare to cross-correlations measured for quasars and damped Lyα systems (DLAs) from the Baryon Oscillation Spectroscopic Survey (BOSS) by Font-Ribera et al. We calibrate LyMAS using Horizon-AGN hydrodynamical cosmological simulations of a $(100 \ h^{-1} \ \mathrm{Mpc})^3$ comoving volume. We apply this calibration to a $(1 \ h^{-1} \ \mathrm{Gpc})^3$ simulation realized with 2048^3 dark matter particles. In the $100 \ h^{-1}$ Mpc box, LyMAS reproduces the halo-flux correlations computed from the full hydrodynamic gas distribution very well. In the $1 \ h^{-1}$ Gpc box, the amplitude of the large-scale cross-correlation tracks the halo bias b_h as expected. We provide empirical fitting functions that describe our numerical results. In the transverse separation bins used for the BOSS analyses, LyMAS cross-correlation predictions follow linear theory accurately down to small scales. Fitting the BOSS measurements requires inclusion of random velocity errors; we find best-fitting rms velocity errors of 399 and 252 km s^{-1} for quasars and DLAs, respectively. We infer bias-weighted mean halo masses of $M_\mathrm{h}/10^{12} \ h^{-1} \ \mathrm{M}_\odot = 2.19^{+0.16}_{-0.15}$ and $0.69^{+0.16}_{-0.14}$ for the host haloes of quasars and DLAs, with $\sim$0.2 dex systematic uncertainty associated with redshift evolution, intergalactic medium parameters, and selection of data fitting range.

Key words: large-scale structure of Universe.

1 INTRODUCTION

The enormous high-redshift quasar sample observed by the Baryon Oscillation Spectroscopic Survey (BOSS; Dawson et al. 2013) of SDSS-III (Eisenstein et al. 2011, SDSS = Sloan Digital Sky Survey) has enabled the first empirical studies of the large-scale three-dimensional auto-correlations of transmitted flux in the Lyα forest (e.g. Slosar et al. 2011) and of the cross-correlations between the Lyα forest and other tracers of structure such as damped Lyα systems (DLAs; see Font-Ribera et al. 2012, hereafter FR12) and quasars (Font-Ribera et al. 2013, hereafter FR13). Auto-correlation analyses allow high-precision measurements of the angular diameter distance and Hubble parameter via baryon acoustic oscillations (BAO; see Busca et al. 2013; Slosar et al. 2013; Delubac et al. 2015). Cross-correlation analyses provide a complementary route

to BAO measurements (Font-Ribera et al. 2014), and they provide novel constraints on the properties of dark matter (DM) haloes that host DLAs and quasars. On large scales, these correlations can be described by linear theory with an effective bias factor b_F and redshift-space distortion parameter β_F for the forest that depend on the underlying cosmology and the physical parameters of the intergalactic medium (IGM; see, e.g., McDonald 2003; Seljak 2012; Arinyo-i-Prats et al. 2015). Fully exploiting these measurements, however, requires a theoretical description of Lyα forest clustering that extends to non-linear scales. In this paper, we use the Lyα Mass Association Scheme (LyMAS; Peirani et al. 2014, hereafter P14) to predict the cross-correlation between DM haloes and the Lyα forest, and we use these predictions to model the measurements of DLA- and quasar-forest cross-correlations by FR12 and FR13.

Ideally, one would model Lyα forest correlations using hydrodynamic simulations that resolve the $\sim$100 kpc Jeans scale of the low-density IGM (see Cen et al. 1994; Zhang, Anninos & Norman 1995; Hernquist et al. 1996; Miralda-Escudé et al. 1996 for pioneering studies and Lidz et al. 2010; Peeples et al. 2010; Borde et al. 2014; Arinyo-i-Prats et al. 2015 for some recent examples).

* E-mail: lochhaas@astronomy.ohio-state.edu (CL); dhw@astronomy.ohio-state.edu (DHW)

However, ∼Gpc³ simulation volumes are needed to avoid artificial box size effects and exploit the statistical precision achievable with BOSS, and calculations with the required combination of volume and resolution are not currently feasible. One approach to this challenge (e.g. Slosar et al. 2009) uses variants of the fluctuating Gunn–Peterson approximation (Gunn & Peterson 1965; Croft et al. 1998; Weinberg, Katz & Hernquist 1998), a deterministic mapping between DM density and Lyα flux. LyMAS instead uses high-resolution hydrodynamic simulations to calibrate the conditional probability distribution $P(F_s|\delta_s)$ for the transmitted flux smoothed (in one dimension) on the scale of BOSS spectral resolution given the redshift-space DM density contrast δ_s smoothed (in three dimensions) on a similar scale. This conditional distribution is applied to the smoothed DM density field of a larger volume, lower resolution *N*-body simulation to determine the Lyα forest flux along skewers through the matter distribution. Additional steps are used to impose pixel-to-pixel coherence of spectra along a given line of sight and to ensure that the 1d power spectrum and unconditional probability distribution function (PDF) of flux matches that of the calibrating hydrodynamic simulations (see P14).

P14 calibrated $P(F_s|\delta_s)$ using the Horizon-MareNostrum simulation, a hydrodynamic simulation of a $(50~h^{-1}~\mathrm{Mpc})^3$ comoving volume with *Wilkinson Microwave Anisotropy Probe* 1 (*WMAP*1) cosmological parameters (Spergel et al. 2003, $h \equiv H_0/100~\mathrm{km~s^{-1}~Mpc^{-1}}$). They applied this calibration to a 1024^3 simulation of a $(300~h^{-1}~\mathrm{Mpc})^3$ cube with a Gaussian DM smoothing scale $r_s = 0.3~h^{-1}$ Mpc and to a 1024^3 simulation of a $(1~h^{-1}~\mathrm{Gpc})^3$ cube with $r_s = 1.0~h^{-1}$ Mpc. In this work, we improve the calibration using Horizon-AGN and Horizon-noAGN, two simulations of a $(100~h^{-1}~\mathrm{Mpc})^3$ volume adopting *WMAP*7 parameters (Komatsu et al. 2011), one with and one without feedback from active galactic nuclei (AGN), as described by Dubois et al. (2014). For large volumes, we use one 2048^3 simulation of a $(1~h^{-1}~\mathrm{Gpc})^3$ cube, with DM smoothing $r_s = 0.5~h^{-1}$ Mpc, and several 1024^3 simulations of the same volume with $r_s = 1.0~h^{-1}$ Mpc. These 'upgrades' improve the accuracy and precision of our predictions and give us a better handle on statistical uncertainties and the potential impact of AGN feedback.

FR12 measured the cross-correlation $\xi(\pi, \sigma)$ between DLAs and the Lyα forest, as a function of line-of-sight separation π in bins of transverse separation σ, using ∼60 000 quasar spectra from the Data Release 9 (DR9; Ahn et al. 2012) BOSS quasar sample (Paris et al. 2012; Ross et al. 2012). They modelled these measurements using linear theory over transverse scales $\sigma = 1$–$60~h^{-1}$ Mpc (comoving), excluding scales $r = (\sigma^2 + \pi^2)^{1/2} < 5~h^{-1}$ Mpc in all bins. Adopting Lyα forest linear bias parameters $b_F = -0.168$, $\beta_F = 1.0$ at effective redshift $z = 2.3$, they inferred a linear bias for DLAs of $b_{DLA} = 2.17 \pm 0.2$ at $z = 2.3$, corresponding to a characteristic halo mass $M_h = 4.5 \times 10^{11}~h^{-1}~\mathrm{M_\odot}$ (which yields the same bias factor at this redshift). FR13 applied similar techniques to the cross-correlation of DR9 quasars with the Lyα forest, extending measurements to $\sigma = 80~h^{-1}$ Mpc and excluding scales $r < 15~h^{-1}$ Mpc from their linear fits in all bins. They inferred $b_Q = 3.64^{+0.13}_{-0.15}$ for the average bias factor of DR9 quasars, with a corresponding halo mass $M_h = 3.26 \times 10^{12}~h^{-1}~\mathrm{M_\odot}$. Similar analyses using the BOSS DR12 (Alam et al. 2015) Lyα forest, DLA, and quasar samples are underway.

In Section 2, we give an explanation of how LyMAS is calibrated on a small-volume hydrodynamic simulation and then applied to large-scale DM simulations, and we briefly examine the influence of AGN feedback on the Lyα forest. In Section 3, we present our results for the dependence of the cross-correlation on various properties and provide comparisons of our calculations to linear theory

predictions for the correlation. In Section 3.1, we analyse the accuracy of LyMAS for predicting cross-correlations by examining how different smoothing lengths, box sizes, and particle numbers affect the correlation. We fit a simple, three-parameter model to the cross-correlation in Section 3.2 to determine the dependence of the strength, width, and offset of the correlation on halo mass. We compare our results to linear theory predictions for the Lyα forest auto-correlation and halo-Lyα cross-correlation in Section 3.3. We compare our calculations for halo-forest cross-correlations to measured cross-correlations of DLAs and quasars with the Lyα forest in Section 4. We present our conclusions in Section 5.

2 SIMULATIONS

A detailed account of how LyMAS operates is given by P14, but we summarize the main points and present some changes since the original paper here. LyMAS uses PDFs for the Lyα forest flux that are conditioned on the underlying DM density to predict the forest flux in a large-volume DM simulation. These conditional PDFs are determined from a full hydrodynamics and *N*-body simulation using the RAMSES code (Teyssier 2002) in a periodic box with side length of $100~h^{-1}$ Mpc ($50~h^{-1}$ Mpc in P14), adopting *WMAP*7 cosmological parameters (*WMAP*1 in P14): $\Omega_m = 0.272$, $\Omega_\Lambda = 0.7284$, $\Omega_b = 0.045$, $h = 0.704$, $\sigma_8 = 0.81$, $n_s = 0.967$. The conditional PDFs are then applied to large-scale *N*-body simulations created using GADGET2 (Springel 2005), from which pseudo-spectra are extracted.
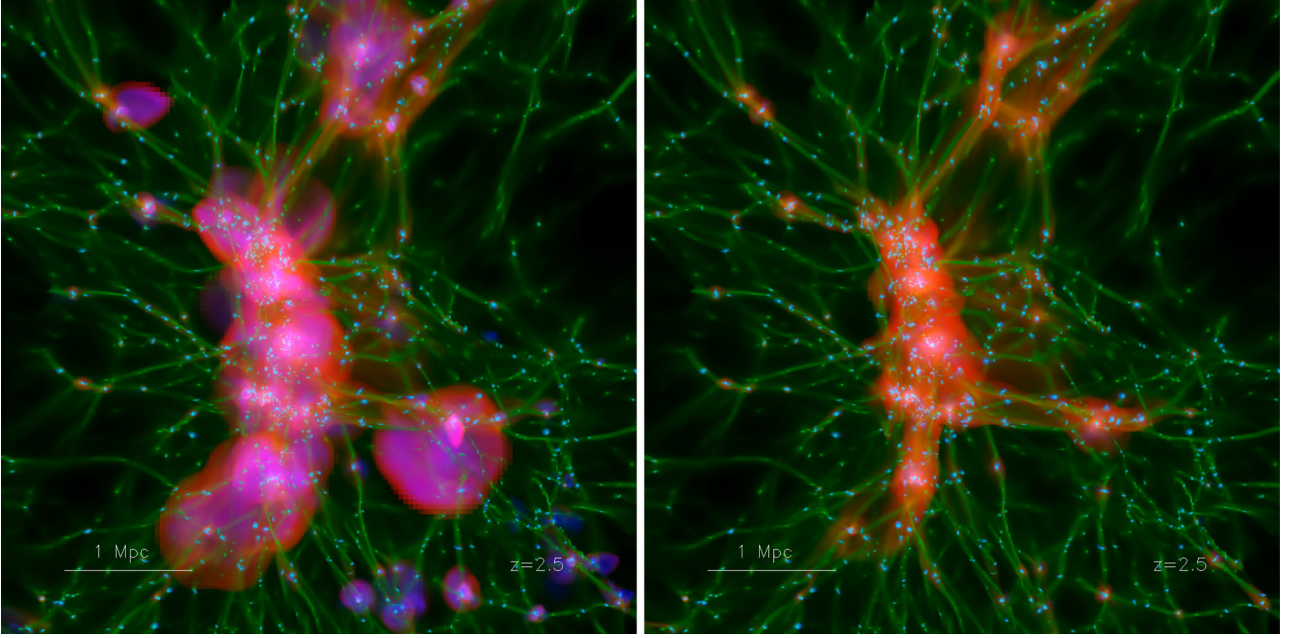
The smaller hydrodynamics simulation includes physical processes such as metal-dependent cooling, photoionization and heating from a UV background, supernova feedback, and metal enrichment. To calibrate the conditional PDFs, about one million lines of sight are extracted from this simulation, and the optical depth of Lyα absorption is calculated based on the neutral hydrogen density along the line of sight. The spectra are smoothed with a 1d Gaussian of dispersion $0.696~h^{-1}$ Mpc, equivalent to BOSS spectral resolution at $z \approx 2.5$. DM skewers that correspond to the spectra are also extracted from this simulation, and the DM is smoothed three-dimensionally with several smoothing scales r_s. The optical depth along each spectrum is converted to Lyα forest flux by $F = e^{-\tau}$, and PDFs for the values of the flux conditional on DM overdensity can thus be determined.

LyMAS operates on the assumption that the flux in one pixel is not correlated with the flux in a separate pixel *except* through correlations in the underlying DM. The flux in pixel 1 is drawn from the conditional distribution $P(F_1|\delta_1)$ without reference to the value F_2 drawn from $P(F_2|\delta_2)$ in pixel 2, but F_1 and F_2 are correlated because of the correlation between δ_1 and δ_2. This *ansatz*, that the relation between flux and matter density is 'locally stochastic', can be expressed by the equation

$$P(F_1, F_2|\delta_1, \delta_2) = P(F_1|\delta_1)P(F_2|\delta_2). \tag{1}$$

Along each individual line of sight, coherence among neighbouring pixels is imposed by generating a 'percentile field' that governs draws from $P(F_s|\delta_s)$, rather than drawing each pixel value independently. The correct 1d flux power spectrum is imposed, on average, by multiplying line-of-sight Fourier modes by numerically calibrated correction factors. Except for the change in the calibrating simulations themselves, our procedures here are the same as those presented by P14.

We apply LyMAS to DM simulations with side lengths $300~h^{-1}$ Mpc and $1~h^{-1}$ Gpc, the same sizes studied in P14, but with cosmological parameters matched to those of the hydro simulations. Our primary simulations have 1024^3 particles for $300~h^{-1}$ Mpc and

Figure 1. The same slice through a hydrodynamics simulation that includes AGN feedback (left) and one that does not (right). Blue colours trace regions of high gas metallicity, red colours trace regions of high gas temperature, and green colours trace regions of high gas density.

2048^3 particles for $1\ h^{-1}$ Gpc, and we analyse them with DM smoothing lengths of 0.3 and $0.5\ h^{-1}$ Mpc, respectively, equal to the initial mean interparticle separation. We also carried out a 1024^3 simulation of the $1\ h^{-1}$ Gpc box with the same initial conditions and two 1024^3 simulations of $1\ h^{-1}$ Gpc boxes with independent initial fluctuations, all analysed with a $1.0\ h^{-1}$ Mpc DM smoothing length. We apply all three smoothing lengths, $r_s = 0.3, 0.5, 1.0\ h^{-1}$ Mpc to the $100\ h^{-1}$ Mpc hydro simulation to calibrate conditional PDFs and to test the accuracy of LyMAS.

The hydrodynamical cosmological simulation Horizon-AGN is described in detail by Dubois et al. (2014). The initial grid for the gas distribution is 1024^3, with a comoving grid cell of $0.25\ h^{-1}$ Mpc, and adaptive mesh refinement is used up to a maximum resolution (for hydrodynamics and gravity) of 1 proper kpc. The DM distribution is represented by 1024^3 particles, with effective gravitational softening set by the cloud-in-cell interpolation of the particle mass distribution on to the adaptive grid. Fig. 1 shows a slice through a simulation that includes AGN feedback (Horizon-AGN, left) and the corresponding slice through a simulation without AGN feedback (Horizon-noAGN, right). While the gas density field (green) is similar in both simulations, regions of high temperature (red) and high metallicity (blue) are more extended in the AGN feedback run. AGN feedback has a considerable impact on the mass function and morphologies of the simulated galaxy population (Dubois et al. 2013), but the impact on conditional flux PDFs is small, and (as we show below) the impact on halo-flux correlations is negligible. When extracting Lyα forest spectra from the simulated gas distribution, the UV background intensity is chosen to give a mean transmitted Lyα forest flux $\bar{F} = \langle e^{-\tau} \rangle = 0.795$, matching the metal-corrected $z = 2.5$ value measured from high-resolution spectra by Faucher-Giguère et al. (2008).

We identify DM haloes using a friends-of-friends algorithm (Davis et al. 1985) with a linking length of 0.15. Then we compute the centre of mass of each object. Once we have pseudo-spectra determined by LyMAS and the positions and masses of DM haloes in the *N*-body simulations, we calculate the cross-correlation by ran-domly selecting a large number of pseudo-spectra and DM haloes from each volume. We bin the DM haloes by mass, and for each halo we bin the pseudo-spectra around it in transverse and line-of-sight separations σ and π. We calculate the cross-correlation ξ as

$$\xi(\pi, \sigma) = \frac{\langle F(\pi, \sigma) \rangle}{\bar{F}} - 1, \tag{2}$$

where $F(\pi, \sigma)$ is the transmitted flux fraction ($F = 1$ for no absorption by the Lyα forest, $F = 0$ for complete Lyα absorption) at line-of-sight separation π and transverse separation σ from the centre of mass of the halo, where $\bar{F}$ is the mean Lyα forest transmitted flux in the entire box. The average of the flux (indicated by $\langle \rangle$) is performed over all pixels that fall into the halo mass, π, and σ bin. Equation (2) is equivalent to the cross-correlation between the halo number density field and the flux fluctuation field

$$\delta_F = \frac{F - \bar{F}}{\bar{F}}. \tag{3}$$

We also calculate the auto-correlation as

$$\xi(r, \mu) = \langle \delta_{F1} \delta_{F2} \rangle, \tag{4}$$

where $\delta_F = (F - \bar{F})/\bar{F}$, r denotes the comoving separation between the pixels where F_1 and F_2 are measured, and $\mu = \pi/r$ denotes the cosine of the angle from the line of sight between the two pixels. All calculations are done in redshift space, incorporating the simulated peculiar velocities of the gas, DM particles, and haloes.

3 LYMAS PREDICTIONS FOR HALO-FOREST CROSS-CORRELATIONS

In this section, we assess the accuracy of LyMAS as a tool for predicting halo-forest cross-correlations by comparing results for a variety of simulations. After demonstrating the expected accuracy of our largest volume simulation, we fit a model to the numerical cross-correlation results and analyse its dependence on halo mass.
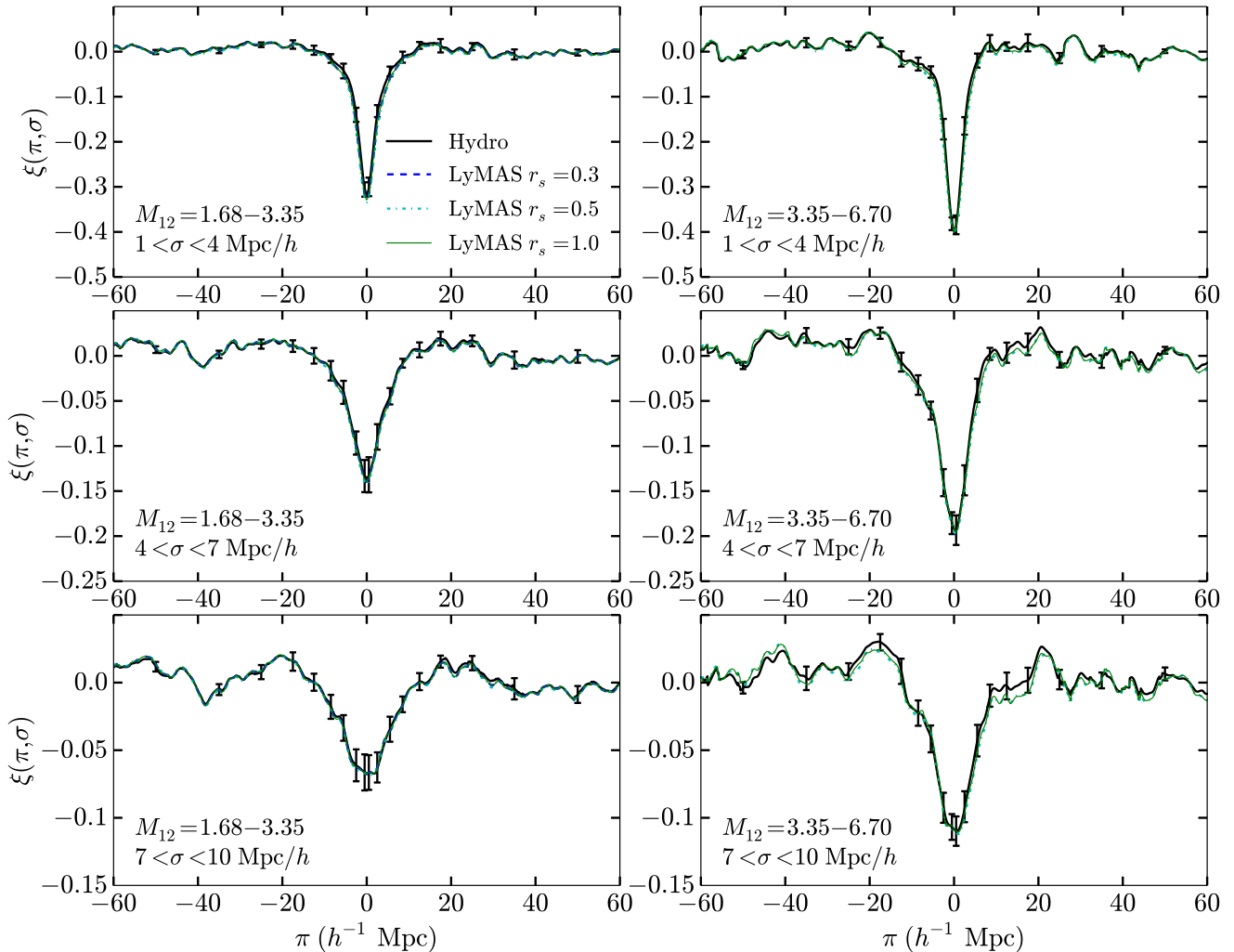
Figure 2. The cross-correlation between DM haloes and Lyα forest flux calculated from true gas spectra (black solid) and from LyMAS applied to the matter distribution with 0.3 h^{-1} Mpc 3d DM smoothing (blue dashed), 0.5 h^{-1} Mpc smoothing (cyan dot–dashed), or 1.0 Mpc h^{-1} Mpc smoothing (green solid) in the $(100\,h^{-1}$ Mpc$)^3$ simulation. Rows show transverse separation bins $\sigma = 1$–$4, 4$–7, and 7–$10\,h^{-1}$ Mpc, and columns show DM halo mass bins $M_{12} = 1.68$–3.35 and 3.35–6.70. Similar agreement holds in other mass and separation bins. Error bars are computed from the standard deviation of the mean among 16 subvolumes.

We also compare the LyMAS predictions to linear theory to infer where non-linear effects have a significant impact on auto- and cross-correlations.

3.1 Accuracy of LyMAS

Fig. 2 compares the halo-forest cross-correlation $\xi(\pi, \sigma)$ computed from the gas distribution of the noAGN hydrodynamic simulation (see Section 2) to the cross-correlations predicted by applying LyMAS to the smoothed DM density fields of the same simulation with $r_s = 0.3, 0.5$, or $1.0\,h^{-1}$ Mpc. Each panel shows the correlation as a function of line-of-sight separation π between the centre of mass of each halo and pixels in the spectra. The three rows show three different halo-pixel transverse separation bins, $\sigma = 1$–4, 4–7, and 7–$10\,h^{-1}$ Mpc, and the two columns show two different mass bins for the haloes, $M_{12} \equiv M_h/10^{12}\,h^{-1}\,M_\odot = 1.68$–$3.35$ and 3.35–6.70. As expected, the strength of the cross-correlation decreases with increasing transverse or line-of-sight separation (σ and π, respectively) and increases with increasing halo mass M_{12}. The four curves are nearly indistinguishable in every panel, and we find similar results for other transverse separation and halo mass bins. We

conclude that halo-forest cross-correlations predicted by LyMAS at BOSS spectral resolution should be virtually identical to those that would be predicted by a full high-resolution hydrodynamic simulation of the same volume, for DM smoothing lengths up to $r_s = 1.0\,h^{-1}$ Mpc.

For representative statistical error bars, we divide the $(100\,h^{-1}$ Mpc$)^3$ box into 16 subvolumes, with transverse size $25\,h^{-1}$ Mpc $\times\ 25\,h^{-1}$ Mpc and line-of-sight size 100 h^{-1} Mpc. We calculate the cross-correlation, $\xi(\pi, \sigma)$, in each of the 16 subvolumes and compute the error on the mean by first finding the mean values in bins of π (using the same bins as for σ), then computing the error in each (π, σ) bin as the standard deviation of the mean among the 16 subvolumes in the bin. We include halo-pixel pairs that cross the subvolume divisions and assign these pairs to the subvolume that contains the centre of mass of the DM halo. We always use the global mean flux $\bar{F}$ when calculating $\xi(\pi, \sigma)$, instead of the mean flux within each subvolume. This procedure for calculating errors may underestimate the statistical error we would find from the dispersion among multiple independent simulations, or from multiple $(100\,h^{-1}$ Mpc$)^3$ volumes in a larger simulation, because it underestimates the impact of coherent structure induced
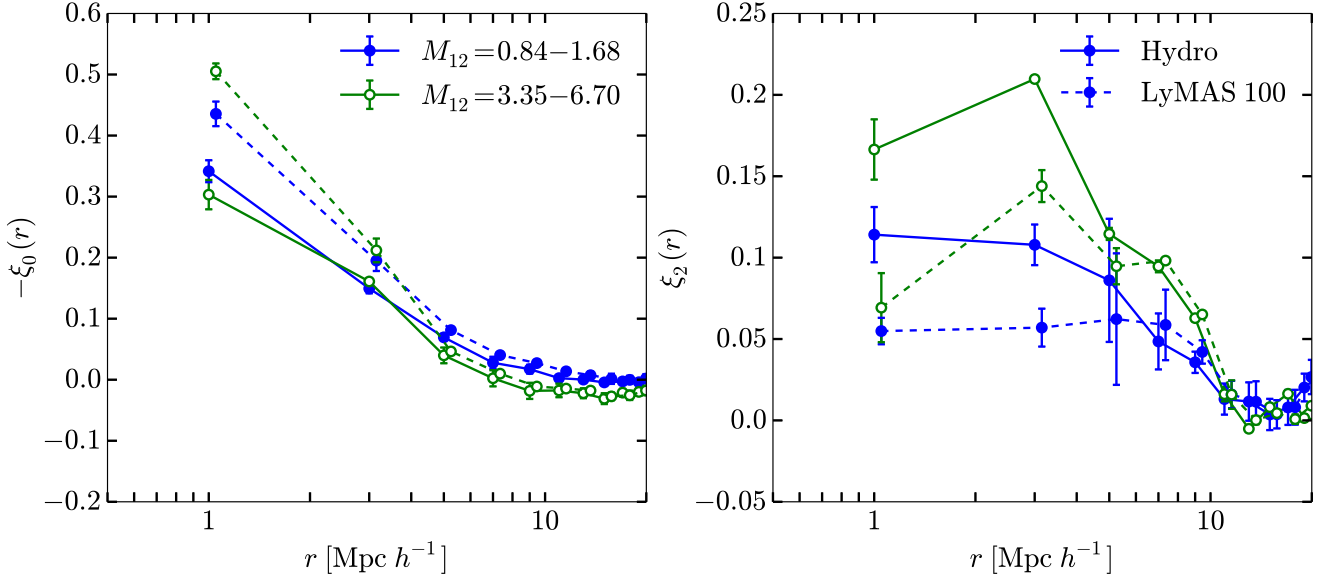
Figure 3. The monopole (left) and quadrupole (right) of the halo-flux cross-correlation in the hydro box (solid) and for the LyMAS predictions applied to the $(100\ h^{-1}\ \mathrm{Mpc})^3$ DM simulation with $r_s = 0.3\ h^{-1}$ Mpc smoothing (dashed, horizontally offset for clarity). Halo mass bins $M_{12} = 0.84-1.68$ (filled circles) and 3.35–6.70 (open circles) are shown. Error bars are computed from the standard deviation of the mean among 16 subvolumes.
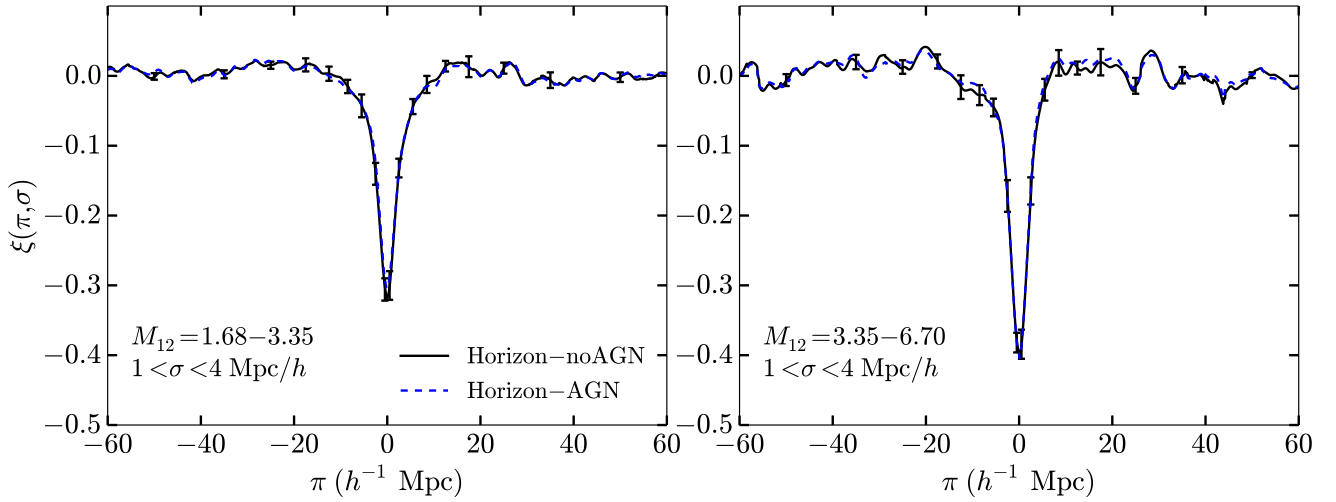

Figure 4. The cross-correlation between DM haloes and Lyα forest flux calculated in the $(100\ h^{-1}\ \mathrm{Mpc})^3$ hydrodynamic simulations using the true gas spectra either including (blue dashed) or not including (black solid) AGN feedback. Only the $\sigma = 1$–$4\ h^{-1}$ Mpc separation bin and $M_{12} = 1.68$–3.35 and 3.35–6.70 mass bins are shown, but the near-perfect agreement holds in other mass and separation bins. Error bars are computed from the standard deviation of the mean among 16 subvolumes.

by large-scale Fourier modes. The multiple curves in Fig. 2 agree to much better than our estimated statistical errors in any case because they are all derived from the same underlying full-cube matter distribution.

To further analyse how well the LyMAS predictions of the halo-flux cross-correlation agree with those from the hydrodynamic simulation, Fig. 3 shows the negative of the monopole (left-hand panel) and the quadrupole (right-hand panel) of the cross-correlation in these simulations as functions of 3d separation r for two halo mass bins (see Section 3.3 for a description of how the monopole and quadrupole are calculated). Small-scale deviations between LyMAS and full hydro are more evident in this representation than in the $\xi(\pi,\ \sigma)$ representation, which averages over a range of scales. LyMAS weakly overpredicts the strength of the monopole at $r \leq 3\ h^{-1}$ Mpc and underpredicts the strength of the quadrupole

at $r \leq 6\ h^{-1}$ Mpc, with a larger difference in the higher halo mass bin. On larger scales, the monopole and quadrupole of the LyMAS cross-correlation match the hydro correlation to well within the error bars.

As AGN feedback plays an important role in the temperature structure in high-density regions (see Fig. 1), we examine its impact on the halo-forest cross-correlation. Fig. 4 compares the correlation computed from the true gas spectra in the two hydro simulations, with and without AGN feedback (as described in Section 2), both in a $(100\ h^{-1}\ \mathrm{Mpc})^3$ volume. Only the $\sigma = 1$–$4\ h^{-1}$ Mpc transverse separation bin and the $M_{12} = 1.68$–3.35 and 3.35–6.70 mass bins are shown, but results are similar in other σ and M_{12} bins. There is virtually no difference in the cross-correlation when AGN feedback is included as compared to when it is not. Previous studies have shown that AGN feedback can have a small (few per cent) impact
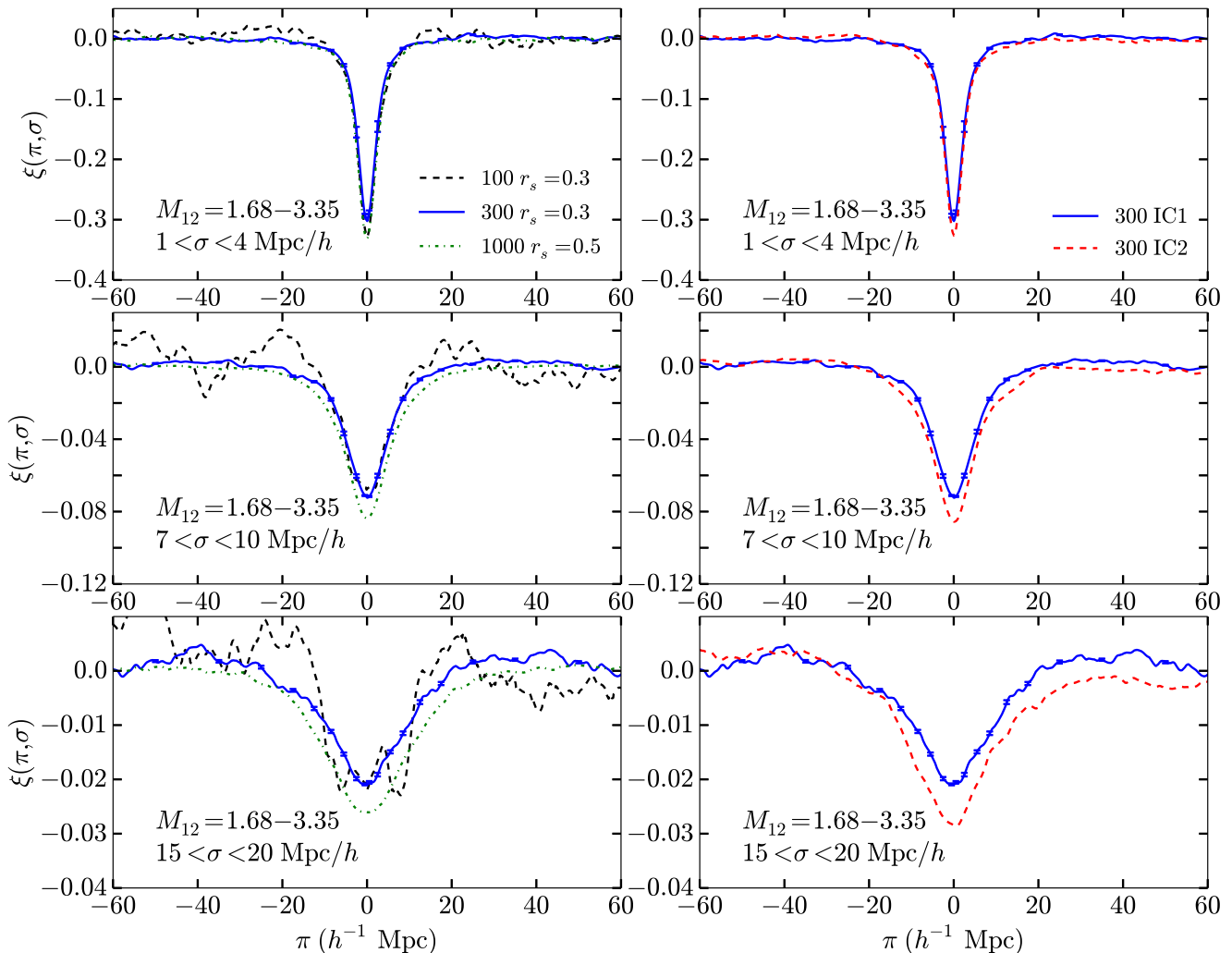
Figure 5. Left column shows the halo-forest cross-correlation calculated by applying LyMAS to $(100\ h^{-1}\ \mathrm{Mpc})^3$ (black dashed), $(300\ h^{-1}\ \mathrm{Mpc})^3$ (blue solid), and $(1\ h^{-1}\ \mathrm{Gpc})^3$ (green dot–dashed) DM volumes, in transverse separation bins $\sigma = 1$–$4, 7$–$10,$ and 15–$20\ h^{-1}$ Mpc for haloes with $M_{12} = 1.68$–3.35. Right column shows the cross-correlation from applying LyMAS to two different DM density distributions in $(300\ h^{-1}\ \mathrm{Mpc})^3$ volumes, from simulations with the same cosmological parameters but different initial conditions, in the same halo mass bin and transverse separation bins as the left column.

on the unconditional PDF and 1d power spectrum of Lyα transmitted flux (Viel, Schaye & Booth 2013), and we find similar effects in our simulations. However, while AGN feedback in these simulations has a substantial effect on the properties of massive galaxies (Dubois et al. 2013), the impact on halo-forest cross-correlations is negligible at the level of the statistical errors in our simulation predictions and the observational data. We find a similarly negligible effect if we use the conditional PDFs from either the AGN or noAGN hydro simulation as our basis for applying LyMAS to larger volume N-body simulations. In the remainder of this paper, we use the noAGN simulation as the source of these conditional PDFs, but none of our conclusions would change if we used the AGN simulation instead.

We next examine the effect on the cross-correlation of using larger volume simulations. The left column of Fig. 5 shows the correlation predicted by applying LyMAS to simulation volumes of $(100\ h^{-1}\ \mathrm{Mpc})^3$, $(300\ h^{-1}\ \mathrm{Mpc})^3$, and $(1\ h^{-1}\ \mathrm{Gpc})^3$ with 3d DM smoothing lengths of $r_s = 0.3, 0.3,$ and $0.5\ h^{-1}$ Mpc, respectively. Only the $M_{12} = 1.68$–3.35 halo mass bin is shown. Based on Fig. 2, we are confident that differences in the calculated correlation are differences in the simulation volumes rather than smoothing scale

effects. In each of the three transverse separation bins shown in the figure ($\sigma = 1$–$4, 7$–$10,$ and 15–$20\ h^{-1}$ Mpc), there are significant differences in the strength of the correlation, most easily seen at $\xi(\pi = 0, \sigma)$. This difference is quite large in the larger transverse separation bins, where the separation is a significant fraction of the smaller simulation volumes. While these differences could be partly due to random statistical fluctuations between the boxes, the weaker correlation at large σ in the smaller boxes is likely a systematic effect of the absence of Fourier modes larger than the box size.

The right column of Fig. 5 compares the correlation calculated by applying LyMAS to two different $(300\ h^{-1}\ \mathrm{Mpc})^3$ DM distributions (solid and dashed) that differ only in the statistical fluctuations of the initial conditions of the DM density. There is a significant difference in the strength of the cross-correlation between the two distributions, which means that $(300\ h^{-1}\ \mathrm{Mpc})^3$ is not a large enough volume to overcome substantial statistical fluctuations in the density distribution. For this reason, we use a $(1\ h^{-1}\ \mathrm{Gpc})^3$ box for all following analysis of the cross-correlation. Fig. 5 also makes the cautionary point that differences between simulation volumes can sometimes be much larger than that estimated by our subvolume method, though we show below that the
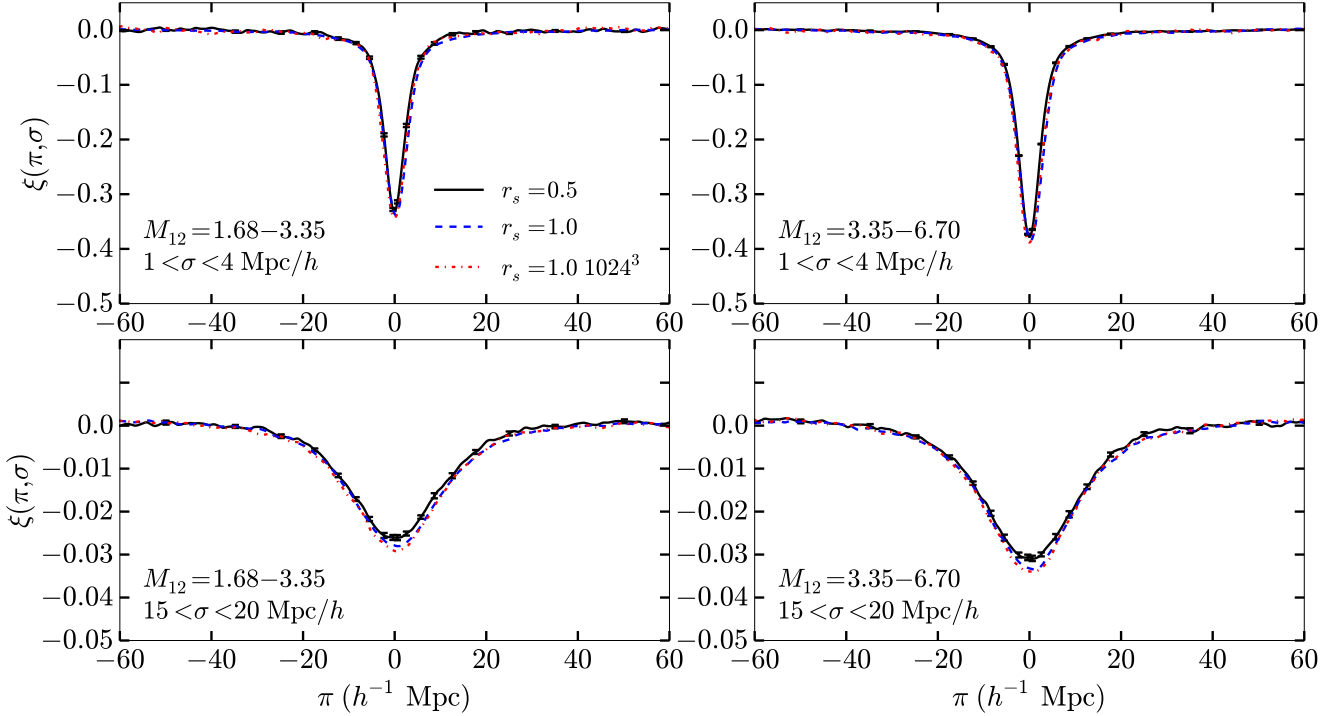
Figure 6. The cross-correlation between DM haloes and Lyα forest flux from applying LyMAS to the $(1\ h^{-1}\ \mathrm{Gpc})^3$ DM simulation with 2048^3 particles for 0.5 (black solid) or $1.0\ h^{-1}$ Mpc (blue dashed) DM smoothing scale, or for a $(1\ h^{-1}\ \mathrm{Gpc})^3$ DM simulation with 1024^3 particles and $1.0\ h^{-1}$ Mpc smoothing scale (red dot–dashed). Left and right columns show mass bins $M_{12} = 1.68$–3.35 and 3.35–6.70; top and bottom rows show transverse separation bins $\sigma = 1$–4 and 15–20 h^{-1} Mpc.

subvolume error estimate appears reasonable for the $(1\ h^{-1}\ \mathrm{Gpc})^3$ box at $\sigma > 7\ h^{-1}$ Mpc.

As a further examination of DM smoothing length and resolution effects, Fig. 6 shows the correlations in the $(1\ h^{-1}\ \mathrm{Gpc})^3$ volume using two different smoothing scales and two choices of *N*-body mass resolution. Dashed and dot–dashed curves show LyMAS predictions for our primary, 2048^3-particle simulation and a simulation with the same initial conditions but 1024^3 particles, both for DM smoothing length $r_\mathrm{s} = 1.0\ h^{-1}$ Mpc. Results are indistinguishable, indicating that a (Gaussian) DM smoothing length equal to the mean interparticle separation yields converged results for the cross-correlation. Solid curves show results for the 2048^3 simulation with a smoothing length $r_\mathrm{s} = 0.5\ h^{-1}$ Mpc, and in contrast to Fig. 2 we now see a slight difference between $r_\mathrm{s} = 0.5$ and $1.0\ h^{-1}$ Mpc in the larger ($\sigma = 15$–20 h^{-1} Mpc) separation bin. P14 found that $r_\mathrm{s} = 0.3\ h^{-1}$ Mpc yielded noticeably more accurate auto-correlation results than $r_\mathrm{s} = 1.0\ h^{-1}$ Mpc, though Fig. 2 and similar tests we have carried out for auto-correlations suggest that P14's results may have been affected by the limited $(50\ h^{-1}\ \mathrm{Mpc})^3$ volume of their hydrodynamic simulation. In what follows, we use the 2048^3 simulation with $r_\mathrm{s} = 0.5\ h^{-1}$ Mpc for our main predictions but use 1024^3 simulations with $r_\mathrm{s} = 1.0\ h^{-1}$ Mpc to estimate statistical errors.

3.2 LyMAS predictions and trends with halo mass

After experimenting with different functional forms, we have found that we can achieve a visually good fit to the predicted cross-correlations in each σ bin for all of the mass ranges we have examined using a three-parameter Lorentzian function

$$\xi(\pi, \sigma) = -\alpha \frac{\gamma^2}{\pi^2 + \gamma^2} + \Delta, \tag{5}$$

where α is the correlation strength at $\pi = 0$ and γ describes the width of the Lorentzian in each σ bin. The cross-correlation can be non-zero out to large separations, and fits are significantly more accurate if we include a large-scale vertical offset Δ as an additional parameter. The purpose of these fits is partly to allow us to report our full set of numerical results in a compact form and partly to allow us to fit observational data with a continuous model rather than interpolate across the discrete mass bins used in our computation.

We fit the Lorentzian model of equation (5) to the cross-correlation predicted by LyMAS in each of the transverse separation σ and halo mass M_{12} bins. Thick solid lines and points with error bars in Fig. 7 show the peak correlation strength, $|\xi(\pi = 0, \sigma)|$, as a function of the average halo mass within each bin, as determined by the best-fitting Lorentz profile. Errors on these points are the dispersion in the best-fitting Lorentz profile among the same 16 subvolumes described in Section 3.1, found simply by fitting the same form to $\xi(\pi, \sigma)$ in each subvolume. [For each of these 16 fits, the errors on values of $\xi(\pi, \sigma)$ were the error on the mean of all pixels in the (π, σ) bin.] As expected, the correlation strength increases with decreasing σ and with increasing halo mass. We fit a power law of the form

$$\xi(\pi = 0, \sigma) = A \left(\frac{M_\mathrm{h}}{4 \times 10^{12}\ h^{-1}\ \mathrm{M}_\odot} \right)^m \tag{6}$$

in each σ bin; these fits are shown as the thin solid lines, with best-fitting parameters and errors shown in Table A1. The powers m of halo mass are nearly identical for all transverse separation bins except the smallest, meaning that the mass dependence of the correlation strength is nearly the same at all transverse separations. The best-fitting scaling factor generally decreases with increasing
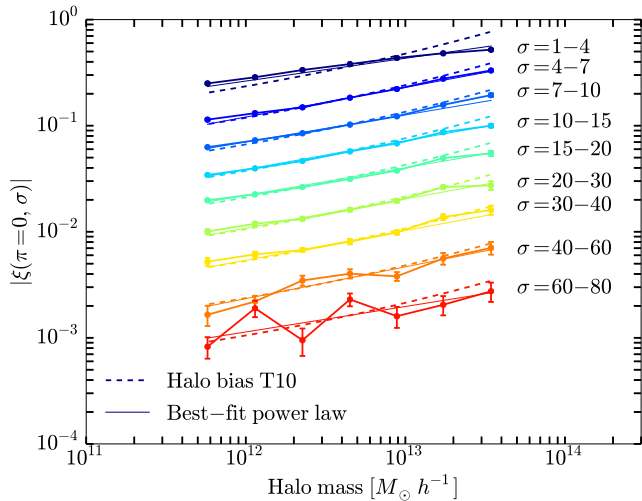
Figure 7. The peak correlation $|\xi(\pi = 0, \sigma)|$ of the best-fitting Lorentzian profiles as a function of average halo mass within each bin (thick solid lines and points with errors) for each transverse separation bin (marked next to each line on the right, increasing top to bottom). The dashed lines show the dependence of the T10 halo bias on halo mass, scaled to a height appropriate for each transverse separation bin. The thin solid lines show a best-fitting power law to each line.

Figure 8. The best-fitting full width at half-minimum 2γ of the Lorentzian profiles as functions of the centres of each transverse separation bin (solid lines). Different coloured lines show different halo mass ranges, as listed in the plot legend. The black dashed line shows the best-fitting linear function to the width for all halo mass bins.

transverse separation. Errors on the parameters of the fit are the dispersion of the best-fitting parameter in the 16 subvolumes.
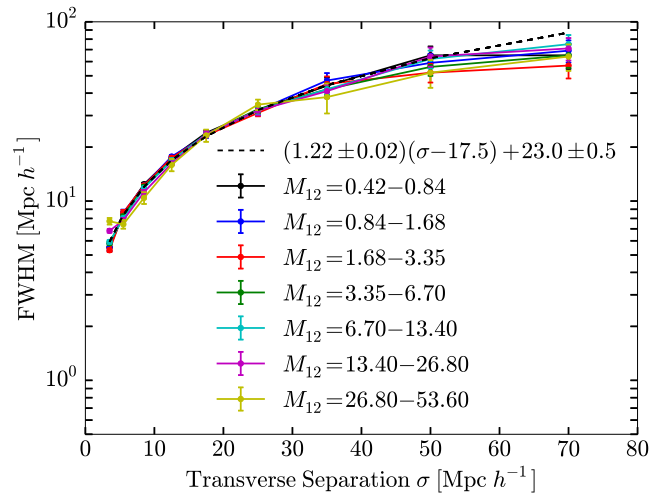
Dashed lines in Fig. 7 show the Tinker et al. (2010, hereafter T10) halo-bias formula as a function of halo mass at $z = 2.5$, scaled vertically to match the point at $\sim 4 \times 10^{12} \, h^{-1} \, \mathrm{M}_\odot$ for each transverse separation bin. The predicted slope of the halo bias with mass matches the measured slope of $\xi(\pi = 0, \sigma)$ with mass very well, except in the smallest σ bin, where our measured slope is lower than the halo-bias prediction. The T10 bias as a function of halo mass is not a perfect power law over the mass range plotted here, and our highest mass points tend to fall below this relation. This disagreement could be a consequence of our finite box size, as even a $(1 \, h^{-1} \, \mathrm{Gpc})^3$ box contains relatively few of these extreme haloes at this redshift (300 haloes in the largest mass bin, compared to ~ 1.58 million in the smallest). As a representation of our results, therefore, one can either take the power-law fits reported in Table A1 or the T10 halo-bias mass dependence normalized to the reported fit amplitude at $M_h = 4 \times 10^{12} \, h^{-1} \, \mathrm{M}_\odot$ (corresponding to $\nu = 2.75$ in the T10 formula for our adopted cosmology and $z = 2.5$). Linear theory predicts that $\xi(\pi = 0, \sigma)$ is nearly proportional to b_h, but not exactly so because of the reduced redshift-space distortion for higher halo bias (see equation 11 below).

Fig. 8 plots the dependence of the best-fitting width of the Lorentz profiles on transverse separation and halo mass. The full width at half-minimum, 2γ, of the Lorentz profile has virtually no dependence on halo mass, so we fit a single linear function to all bins simultaneously around a pivot point at $\sigma = 17.5$, obtaining

$$\mathrm{FWHM}(\sigma) = 2\gamma = (1.22 \pm 0.02)(\sigma - 17.5) + (23.0 \pm 0.5). \quad (7)$$

The relatively small errors on the slope and intercept show that this fitting function works quite well for the relative width of the correlation in each transverse separation bin. Again, errors are the error on the mean of the best parameters fitted to correlations in each of the 16 subvolumes.

We have so far considered correlations in narrow bins of halo mass, but quasars and DLAs likely occupy a wide range of host

haloes. We now examine the effect of a distribution of halo masses on the predicted correlation. We choose haloes within each of the lowest four halo mass bins, $M_{12} = 0.42$–0.84, 0.84–1.68, 1.68–3.35, and 3.35–6.70. We use three different mass distributions: one in which the number of haloes in each bin reflects the mass distribution of all haloes in the volume ('all haloes', the numbers are 158 068, 61 548, 21 150, and 6363, respectively), one in which the number of haloes in each bin is the same ('flat', the number is 6000), and one in which there are the same number of haloes in the first and last bins and no haloes in the middle two bins ('bimodal', the numbers are 6000, 0, 0, and 6000). Fig. 9 shows the halo-flux correlation for each of these distributions. Again, errors are from the subvolume method, as described in Section 3.1.

Also plotted in each panel of Fig. 9 is a Lorentz profile prediction for the correlation, with amplitude based on the best-fitting power laws in Fig. 7 (dashed lines). Within any σ bin, the cross-correlation amplitude is approximately $\propto M_h^{0.272}$. Therefore, we expect, at least for linear bias, that the clustering of the population will equal that of haloes with the effective mass

$$M_{h,\mathrm{eff}} = \left[\frac{1}{N_{\mathrm{haloes}}} \sum_{i=1}^{N_{\mathrm{haloes}}} M_{h,i}^{0.272} \right]^{1/0.272}, \quad (8)$$

where N_{haloes} is the total number of haloes used in the correlation, $M_{h,i}$ is the mass of each halo, and 0.272 is the weighted average of the best-fitting powers in the power-law dependence of the correlation on halo mass for all but the smallest transverse separation bin, as given in Table A1. Equation (8) is equivalent to the condition

$$b(M_{h,\mathrm{eff}}) = \frac{1}{N_{\mathrm{haloes}}} \sum_{i=1}^{N_{\mathrm{haloes}}} b(M_{h,i}) \quad (9)$$

for a halo-bias relation $b \propto M_h^{0.272}$. In Fig. 9, the amplitude of $|\xi(\pi = 0, \sigma)|$ is computed from this value of $M_{h,\mathrm{eff}}$ and the power-law fits in Table A1, and the width of the Lorentzian is determined from the linear fit of equation (7). The prediction using our best-fitting parameters accurately matches the actual halo-flux correlation calculated using LyMAS in all cases. In linear theory, the cross-correlation strength should depend only on the number-weighted mean bias factor of haloes. This expectation holds for the
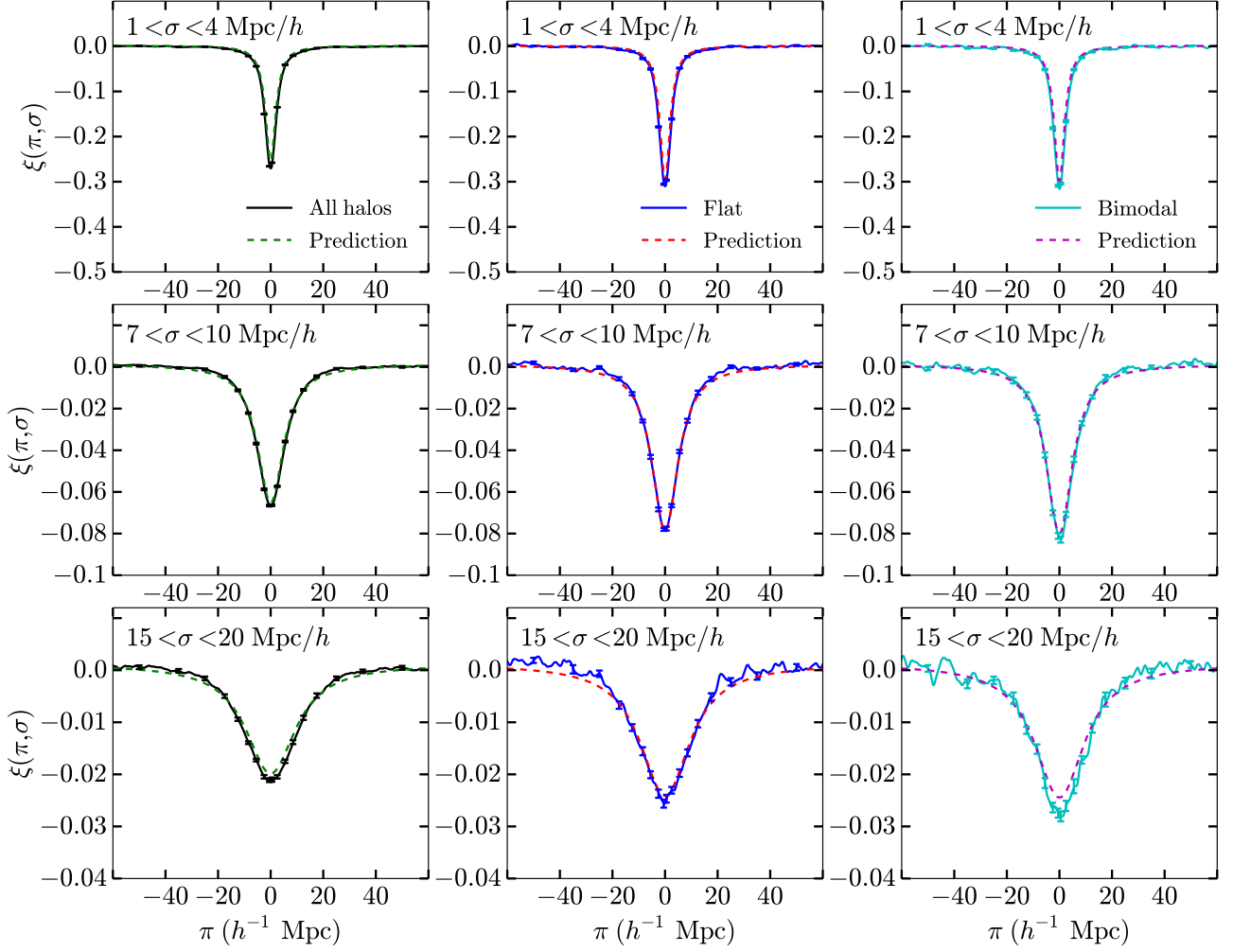
Figure 9. The cross-correlation for three transverse separation bins (rows, increasing top to bottom) and three halo mass distributions (columns) plotted as solid lines, and a Lorentz profile with amplitude, width, and offset based on the dependence on halo mass and transverse separation determined by equations (6) and (7) and Tables A1 and A2 plotted as dashed lines. Left column shows the cross-correlation from a distribution that reflects the mass distribution of all haloes in the simulation, middle column shows a flat distribution of halo masses, and right column shows a bimodal distribution with equal numbers of high- and low-mass haloes, and none between.

LyMAS predictions even down to small scales, indicating that bias-weighted cross-correlations can robustly constrain the mean halo mass but cannot discriminate among distributions with the same mean mass.

3.3 Comparison to linear theory

At sufficiently large scales, we expect the Lyα flux auto-correlation and the halo-flux cross-correlation to be adequately described by linear perturbation theory. In redshift space, the linear theory 3d flux power spectrum is

$$P^F(\boldsymbol{k}, \mu) = b_F^2 (1 + \beta_F \mu^2)^2 P_m(\boldsymbol{k}), \qquad (10)$$

where $P_m(\boldsymbol{k})$ is the linear theory matter power spectrum, μ is the cosine of the angle between the wavevector $\boldsymbol{k}$ and the line of sight, and b_F and β_F are the bias factor and redshift-space distortion parameter of the flux (Kaiser 1987; McDonald 2003). For haloes or galaxies, one expects $\beta = f(z)/b \approx [\Omega_m(z)]^{0.55}/b$, where $f(z)$ is the fluctuation growth rate. However, for flux this relation does not hold because of the non-linear relation between flux and optical depth, and β must be predicted separately (for more discussion, see

Seljak 2012; Arinyo-i-Prats et al. 2015). For the halo-flux cross-power spectrum, the linear theory prediction is

$$P^{Fh}(\boldsymbol{k}, \mu) = b_F b_h (1 + \beta_F \mu^2)(1 + \beta_h \mu^2) P_m(\boldsymbol{k}), \qquad (11)$$

where the halo bias b_h can be calibrated from numerical simulations (T10) and $\beta_h = f(z)/b_h$. Correlation functions can be computed from power spectra using the appropriate Fourier transform relations (e.g. equations 4.5–4.12 of Slosar et al. 2011). The matter power spectrum used for our linear theory calculations has the same *WMAP*7 parameters adopted for the hydro and *N*-body simulations (see Section 2). We measure $\xi(r, \mu)$ in bins of $\Delta\mu = 0.05$ and compute the monopole and quadrupole $\xi_0(r)$ and $\xi_2(r)$ from the sum

$$\xi_l(r) = (2l + 1) \sum_{\mu=0}^{\mu=1} \xi(r, \mu) P_l(\mu) \Delta\mu, \qquad (12)$$

where $P_0(\mu) = 1$, $P_2(\mu) = \frac{3}{2}\mu^2 - 1$ are the Legendre polynomials of degree l.

Fig. 10 shows the monopole and quadrupole of the flux auto-correlation in the $(1\, h^{-1}\, \mathrm{Gpc})^3$ LyMAS simulation as a function of separation. Errors are the error in the mean among 16 subvolumes
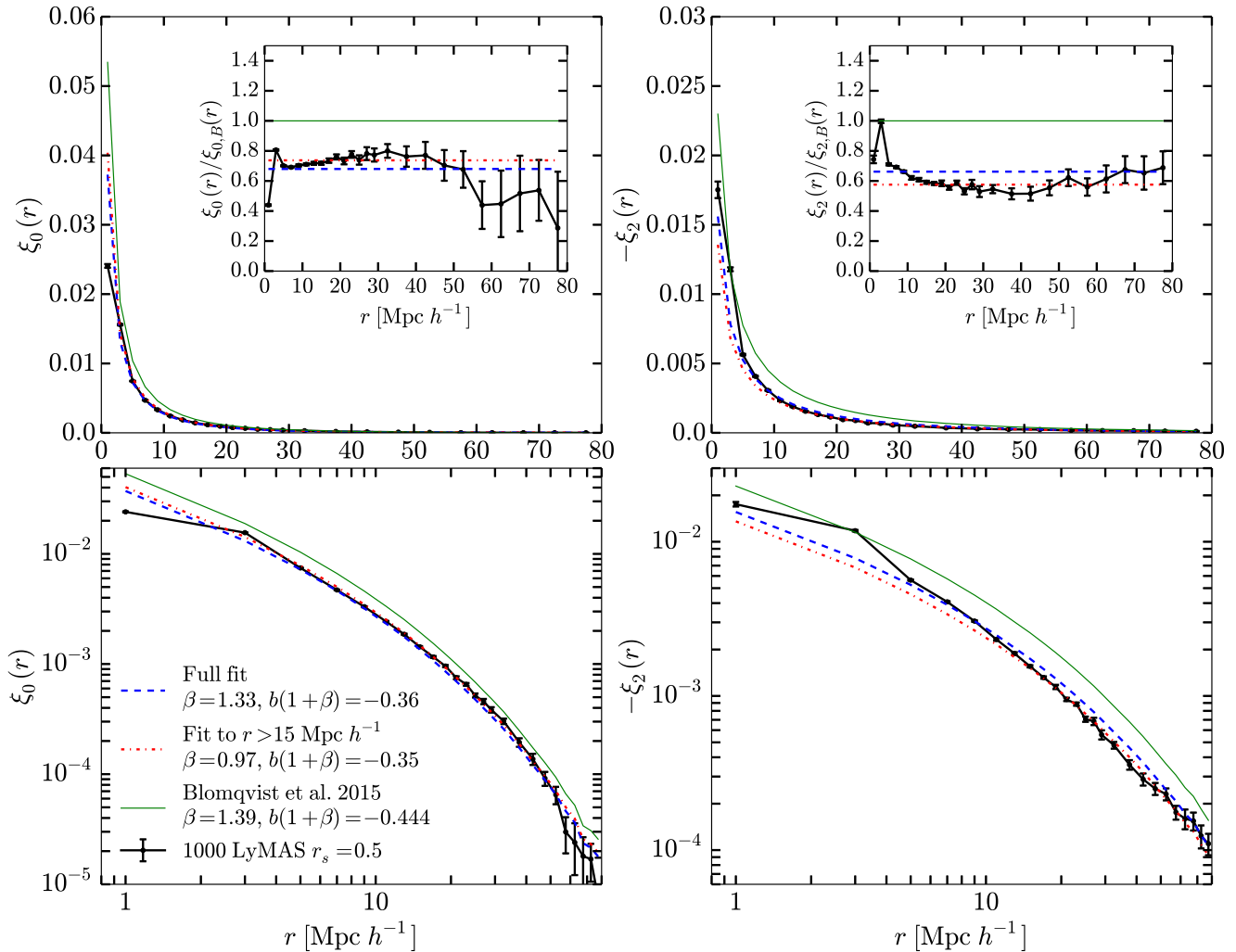
Figure 10. The monopole (left) and quadrupole (right) of the auto-correlation of Lyα forest flux as determined from LyMAS in the $(1\ h^{-1}\ \mathrm{Gpc})^3$ volume (solid black) compared to the linear theory model with parameters b and β fitted to the full range of data (blue dashed) and fitted to only $r > 15\ h^{-1}$ Mpc (red dot–dashed). Green solid curves show linear theory with values of b and β from the BOSS analysis of Blomqvist et al. (2015), scaled to $z = 2.5$. Insets show ratio to linear theory with the Blomqvist et al. (2015) values.

as described in Section 3.1. The top row shows the correlation in linear space, and the bottom row is the same correlation in log–log space. The dashed and dot–dashed lines in each panel show linear theory predictions with b_F and β_F determined by fitting the full range of data or only $r > 15\ h^{-1}$ Mpc, respectively.

The χ^2 for the large separations is 43.23 for 34 d.o.f. but the fit to the full data range is extremely poor, with $\chi^2 = 4593$ for 48 d.o.f. Visually, one can see that the fit to $r > 15\ h^{-1}$ Mpc gives a good description of the LyMAS monopole and quadrupole at these scales, but the LyMAS quadrupole rises above this fit at smaller r. At intermediate $r (\approx 10–50\ h^{-1}$ Mpc), the linear theory model fit to the full range underpredicts $\xi_0(r)$ and overpredicts $|\xi_2(r)|$. For $r > 15\ h^{-1}$ Mpc, our best-fitting parameters are $\beta = 0.970 \pm 0.016$, $b = -0.178$. There is significant degeneracy between these parameters, and Lyα forest analyses frequently quote the combination $b(1 + \beta)$, which is better determined than b or β alone. Our $r > 15\ h^{-1}$ Mpc fit yields $b(1 + \beta) = -0.3525 \pm 0.0011$.

Analyses of the BOSS Lyα forest have reported a variety of values for β and $b(1 + \beta)$ (e.g. Slosar et al. 2011; Busca et al. 2013; Delubac et al. 2015), with significant dependence on details of the data analysis. The most precise, and likely most robust, de-

termination is from Blomqvist et al. (2015), who implement a new method for removing biases imprinted by the continuum determination procedure. They find $\beta = 1.39^{+0.11}_{-0.10}$ and $b(1 + \beta) = -0.374 \pm 0.007$ at $z = 2.3$. They do not fit for redshift dependence, so here we scale the value $b(1 + \beta)$ by $[(1 + 2.5)/(1 + 2.3)]^{2.9} = 1.19$ using the redshift evolution reported by Slosar et al. (2011). We have also multiplied the Blomqvist et al. (2015) predictions by a factor $0.79/0.81$ to account for the slightly lower σ_8 value that they assumed. With these parameters, the Blomqvist et al. (2015) fit lies above the LyMAS prediction, by factors of approximately 1.25 and 1.6 for the monopole and quadrupole, respectively. If we did not apply the redshift scaling to $b(1 + \beta)$, these factors would be approximately 0.95 and 1.25.

We leave investigation of this discrepancy in the auto-correlation function to future work. There are a number of possible contributions, and it is not clear whether one dominates or a variety of small effects combine. One is incorrect cosmological parameters in the Horizon-noAGN simulation, though the *WMAP*7 parameters used here are only moderately different (most notably in Ω_m) from the Planck 2015 values (Planck Collaboration XIII 2015). A second contribution could be inaccurate IGM properties: our

adopted value of $\bar{F} = 0.795$ (based on Faucher-Giguère et al. 2008) could be incorrect, or the UV background history used in Horizon-noAGN could lead to an incorrect temperature–density relation in the diffuse IGM, perhaps because it does not include extra heating by helium reionization. A related possibility, and one that might be more capable of producing substantial differences in b and β, is that spatial variations in the UV background or the IGM temperature–density relation induce additional large-scale clustering of the Lyα forest flux that is not accounted for in our simulations (see recent discussions by Gontcho A Gontcho, Miralda-Escudé & Busca 2014; Pontzen 2014; Greig, Bolton & Wyithe 2015). Another possible explanation is that LyMAS does not reproduce full hydro simulation results for the flux auto-correlation with sufficient accuracy. Although we find excellent agreement between hydro and LyMAS calculations of the halo-flux cross-correlation in the $(100\ h^{-1}\ \mathrm{Mpc})^3$ simulations (Fig. 2), the auto-correlation monopole and quadrupole differ at the 10–20 per cent level at $r \sim 10\ h^{-1}$ Mpc, while differences at larger scales are harder to assess because the correlation function is artificially suppressed by the finite box size (Zhu, private communication). The mismatch in Fig. 10 is the cause for some caution about the mass or bias values that we infer from fitting BOSS cross-correlation measurements in Section 4. However, we also note that different methods for inferring β and b from BOSS have yielded a wide range of values because of the challenges of correcting for continuum determination biases, metal-line contamination, and DLA absorption, and the Blomqvist et al. (2015) values may not be the last word. There is also uncertainty in scaling from $z = 2.3$ to our simulation redshift of $z = 2.5$.

Recently, Arinyo-i-Prats et al. (2015), extending the work of McDonald (2003), have attempted to compute b and β theoretically using 3d flux power spectra measured from a number of hydro simulations with comoving box sizes ranging from 40 to 120 h^{-1} Mpc. Because of the limited box size, these authors fit non-linear models to the numerical $P_{\mathrm{F}}(\boldsymbol{k}, \mu)$, and the extrapolated large-scale values of b and β depend on the assumed non-linear models as well as on simulation and cosmological parameters. Their estimates of β at $z \approx 2.5$ are typically 1.2–1.4, clearly higher than the value we find here from LyMAS. The values of b are harder to compare because they are more sensitive to cosmological parameters and the adopted form of the non-linear model. LyMAS and the model-based extrapolation method of Arinyo-i-Prats et al. (2015) represent two different strategies for deriving large-scale Lyα forest clustering from limited volume hydro simulations, and over time we hope that they will converge to yield consistent predictions given the same IGM and cosmological parameters.

Returning to the cross-correlation that is the focus of this paper, Fig. 11 compares the halo-flux $\xi(\pi, \sigma)$ from LyMAS to the linear theory model for two different halo mass ranges. In this case, the linear theory is that for the cross-correlation between Lyα forest and haloes, with redshift distortion parameter $\beta_{\mathrm{F}} = 0.97$ and Lyα bias $b_{\mathrm{F}} = -0.178$, equivalent to our best-fitting parameters to the $r > 15\ h^{-1}$ Mpc fit to the LyMAS auto-correlation. For the $M_{12} = 0.42$–0.84 halo mass range, the T10 bias is $b_{\mathrm{h}} = 2.56$, and the T10 halo bias for the larger $M_{12} = 1.68$–3.35 bin is $b_{\mathrm{h}} = 3.71$. With these bias values, the LyMAS $\xi(\pi, \sigma)$ visually matches linear theory closely even down to $\sigma = 1$–4 h^{-1} Mpc and $\pi \approx 0$. However, there are statistically significant residuals for small π at $\sigma = 1$–4 and 7–10 h^{-1} Mpc, with peak differences about 10 per cent of $\xi(\pi = 0, \sigma)$.

Fig. 12 examines the comparison between LyMAS and linear theory in terms of the monopole and quadrupole of the halo-flux cross-correlation. We use the same β_{F} and b_{F} values as in Fig. 11 and the halo-bias factors from the T10 formula evaluated at the logarithmic centre of each mass bin. The monopole is well described by linear theory at $r > 3\ h^{-1}$ Mpc, with some deviation for the highest mass bin $M_{12} = 26.8$–53.6. However, the quadrupole rises above linear theory at $r = 10$–20 h^{-1} Mpc, and then flattens below it at $r < 5\ h^{-1}$ Mpc. Thus, the level of agreement between LyMAS and linear theory in Fig. 11 is to some degree a consequence of the particular quantity plotted. Fig. 13 plots $|\xi(\mu)|$ for comoving separation $r = 5\ h^{-1}$ Mpc. The two lowest mass bins show reasonably good accord with linear theory at all μ, with LyMAS rising slightly above for separations perpendicular to the line of sight ($\mu \approx 0$) and falling slightly below for separations along the line of sight ($\mu \approx 1$). However, for the two highest mass bins, LyMAS yields substantially weaker correlations at $\mu > 0.5$ ($M_{12} = 6.7$–13.4) or $\mu > 0.3$ ($M_{12} = 26.8$–53.6). The overall normalization of the linear theory curves, and to a lesser degree the μ dependence, depends on the b_{F} and β_{F} values inferred from the LyMAS flux auto-correlation, and the volume of the $(100\ h^{-1}\ \mathrm{Mpc})^3$ hydro simulation limits our ability to test the accuracy of LyMAS in predicting these quantities.

In summary, for halo masses in the range $M_{12} \approx 0.5$–3 found by FR12 and FR13, LyMAS predictions for $\xi(\pi, \sigma)$ in bins of σ follow linear theory even down to small scales. However, other representations show significant deviations between LyMAS and linear theory, especially at higher halo masses and high values of μ.

4 COMPARISON TO OBSERVED CROSS-CORRELATIONS

Now that we have examined how the cross-correlation depends on DM halo mass and separation from the halo (in both line-of-sight and transverse directions), we can compare the halo-forest correlation to other cross-correlations with the forest, such as quasars and DLAs. This will allow us to determine a characteristic mass for the host haloes of these objects.

The measured cross-correlations of the Lyα forest with DLAs and with quasars (FR12, FR13) are plotted in Fig. 14 as the red circles and green triangles, respectively. These correlations were measured using $\sim$60 000 quasar spectra in DR9 of BOSS; measurements from the final, DR12 BOSS sample are in progress. A surprising feature of the data points is that the DLA and quasar cross-correlations are similar in amplitude, even though FR12 and FR13 derive quite different bias factors ($b_{\mathrm{DLA}} = 2.17 \pm 0.2$, $b_{\mathrm{Q}} = 3.64 \pm 0.13$) for these two populations. However, the FR13 b_{Q} fit is driven by the large $r = \sqrt{\pi^2 + \sigma^2}$ bins, and the $r < 15\ h^{-1}$ Mpc bins are discarded, while smaller scale measurements carry much of the weight in the FR12 b_{DLA} fits. We have calculated the correlation function in LyMAS in the same transverse separation bins as the measured correlation to obtain the closest comparison. Lorentz profiles with amplitude, width, and offset predicted by our LyMAS calculation for haloes of mass $M_{12} = 0.5$ (black solid), 2.0 (blue dashed), and 8.0 (cyan dot–dashed) are determined by the equations (6) and (7) and Tables A1 and A2.

Because BOSS Lyα forest spectra are calibrated to reproduce the mean Lyα absorption as a function of redshift, there is a slight bias in cross-correlation measurements, since the positive correlation over the full Lyα forest is calibrated out. FR12 and FR13 apply a mean transmission correction (MTC) that removes this bias. The derivation of the MTC is laid out in appendix A of FR12. We apply this correction to our predicted cross-correlation whenever we compare to the measured correlation of quasars or DLAs. Generally, the correction shifts correlations upwards so that the cross-correlation
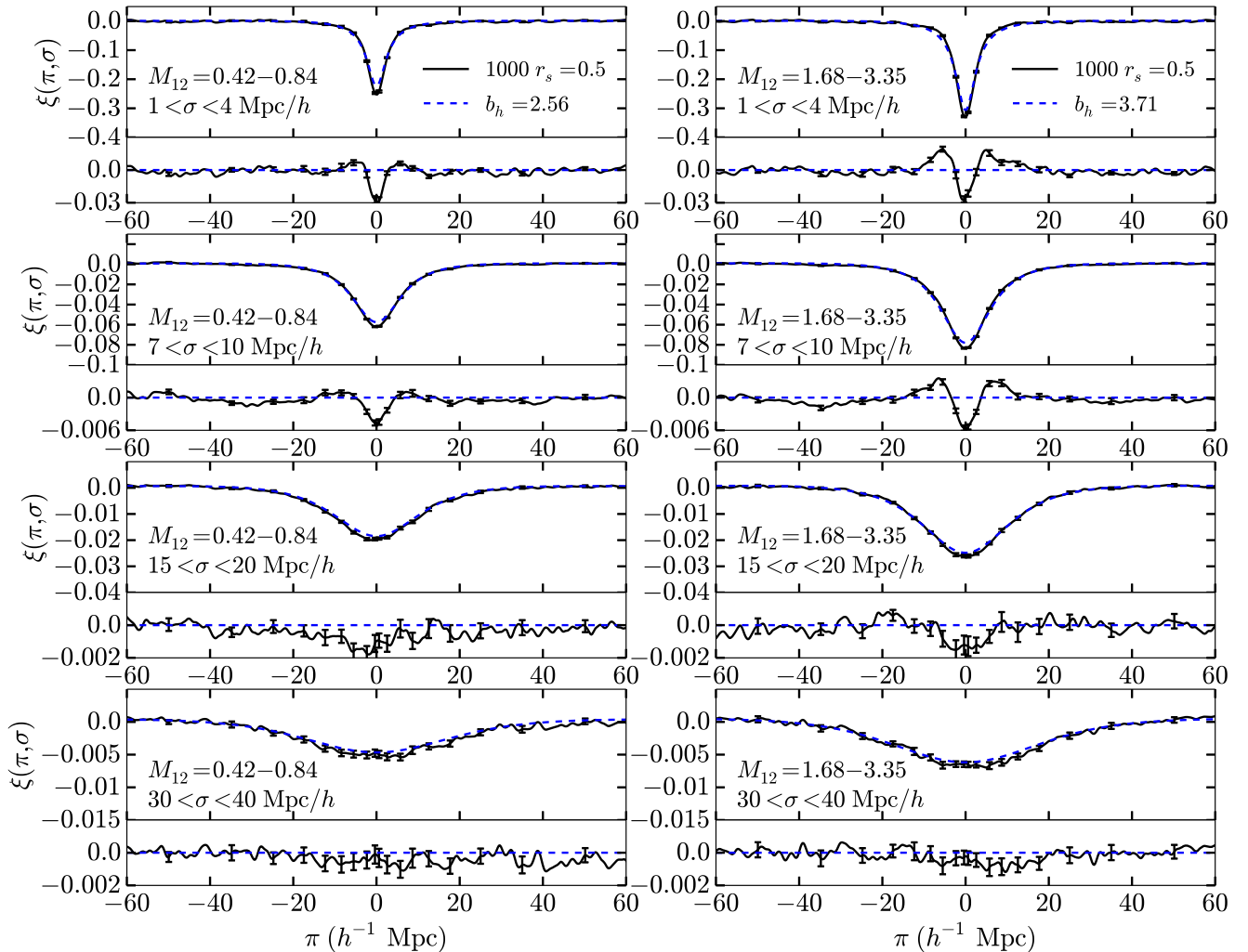
Figure 11. The cross-correlation for four transverse separation bins, $\sigma = 1$–$4, 7$–$10, 15$–$20, 30$–40 (rows) and two halo mass bins $M_{12} = 0.42$–$0.84, 1.68$–3.35 (columns) in the (1 h^{-1} Gpc)^3 LyMAS box (black solid) compared to the cross-correlation predicted by the linear theory for redshift distortion $\beta = 0.97$, Lyα forest bias $b_{\mathrm{Ly\alpha}} = -0.178$, and the T10 value for halo bias in these mass bins, $b_{\mathrm{h}} = 2.56$ or 3.71 (blue dashed).

at large $|\pi|$ becomes positive instead of near zero. It is a small effect, but important for fitting the non-zero values of the measured cross-correlation.

In the $\sigma \approx 10$–$40\,h^{-1}$ Mpc transverse separation bins, it appears as though the cross-correlation for haloes of average mass $M_{12} = 0.5$ or 2.0 fits the measurement of the cross-correlation for DLAs or quasars fairly well. However, the correlation for haloes of the same mass in the $\sigma = 1$–$4\,h^{-1}$ Mpc separation bin does not fit the data. We expect that if there is a characteristic mass distribution for haloes that host DLAs or quasars, then the correlation predicted for that mass distribution should match the observations for all separation bins. The drastic overestimation of the strength of the correlation in the smaller transverse separation bins indicates a missing component in our model of the halo-flux cross-correlation.

Fig. 15 summarizes this data–model comparison by fitting our Lorentz profile form to the DLA and quasar data in each σ bin. We show the measured amplitude as a function of σ by circles and triangles for DLAs and quasars, respectively, while solid curves show the amplitude of Lorentz profiles fitted to the simulation cross-correlations for three different halo mass ranges. As in Fig. 14, it is evident that the DLAs and quasars exhibit similar overall clustering

strength but that any halo mass that matches the observed clustering strength at $\sigma = 20$–$50\,h^{-1}$ Mpc overpredicts it at $\sigma = 1$–$10\,h^{-1}$ Mpc. If our predictions were based on linear theory, one might conclude that this trend reflects the breakdown of linear theory predictions, but LyMAS is a fully non-linear method and we have shown (Fig. 2) that it reproduces hydrodynamic simulation results essentially perfectly even at $\sigma = 1$–$4\,h^{-1}$ Mpc. The discrepancy could reflect some form of feedback or IGM physics not incorporated into our hydro simulations, but the complete insensitivity of the halo-flux predictions to the presence or absence of AGN feedback (Fig. 4) makes such an explanation unlikely.

One key difference between the simulations and the observations is that we know the redshift-space position of simulated haloes exactly while the redshifts of quasars and DLAs must be estimated observationally from emission and absorption lines. Internal motions and radiative transfer effects lead to differences of many hundreds of km s^{-1} among quasar redshifts estimated from different lines (Espey et al. 1989; Richards et al. 2002; Hewett & Wild 2010). From the asymmetry of $\xi(\pi, \sigma)$, FR13 concluded that the quasar redshifts reported by the BOSS pipeline had a systematic offset of -157 km s^{-1}, which is small but not completely negligible
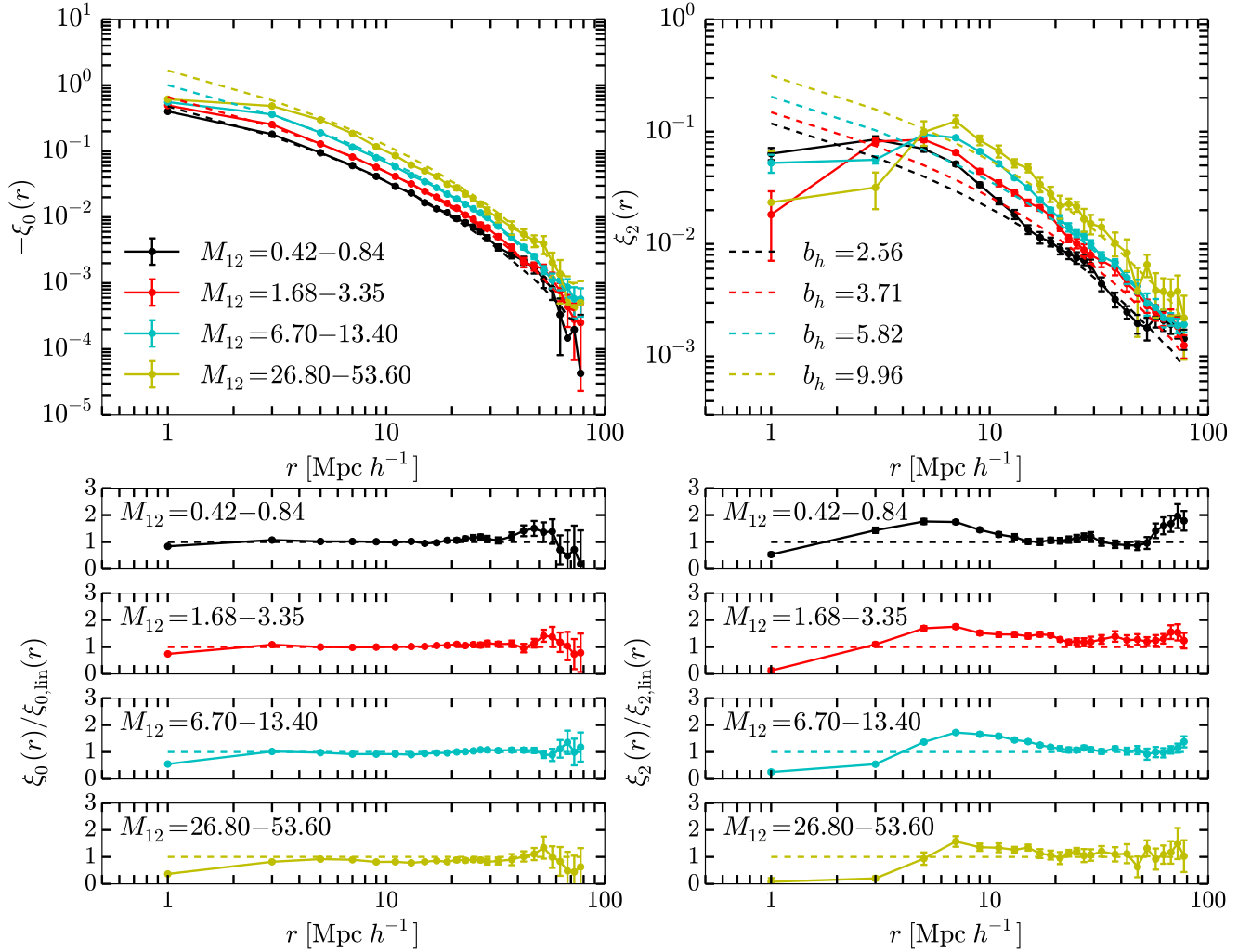
Figure 12. The monopole (left) and quadrupole (right) of the cross-correlation between halo centres and the Lyα forest. LyMAS results for four mass bins, $M_{12} = 0.42$–0.84, 1.68–3.35, 6.70–13.40, and 26.80–53.60, are shown as the solid lines connecting points with errors (mass increases bottom to top). Dashed lines show the linear theory calculation for the b_F and β_F determined by fitting the LyMAS flux auto-correlation and halo biases b_h that correspond to the log-centre of each mass bin shown using the T10 $b_h(M)$ relation, with halo redshift distortion $\beta_h = 1/b_h$ (bias increases bottom to top). Bottom panels show ratio of LyMAS cross-correlation to the linear theory calculation.

compared to the width of $\xi(\pi, \sigma)$ in bins of small σ. More importantly, random redshift errors with a velocity distribution $f(\pi)$ will convolve the predicted $\xi(\pi, \sigma)$ in the π direction, increasing the width but decreasing the depth of $\xi(\pi, \sigma)$.[1] This convolution has a significant impact when σ is comparable to the rms redshift error but becomes negligible at large σ, so it has the correct qualitative form to reconcile our predictions with the BOSS measurements.

Fig. 16 presents a fit to the FR13 data in which we add their inferred mean π offset of 157 km s^{-1} to the model predictions and add the rms velocity error as a free parameter along with halo mass. We use the full covariance matrix that FR13 estimated analytically by treating the forest as a Gaussian random field. There are significant off-diagonal correlations of errors at different π in

a given σ bin, but correlations across σ bins are generally weak. In our fit, we adopt the Lorentz function form of equation (5) convolved with a Gaussian distribution of redshift errors, using equation (7) for the Lorentzian width, the power-law form of equation (6) to predict amplitude as a function of halo mass (with parameters listed in Table A1), and the best-fitting offset Δ as a function of halo mass (determined by interpolating the offsets given in Table A2). When inferring errors on the halo mass M_{12} and rms velocity error, we incorporate only the observational measurement uncertainties, not the uncertainties in the parameters describing our fits to the LyMAS simulation, which would have a small relative effect.

It is evident from Fig. 16 that including the rms redshift error as a parameter dramatically improves the fit to the FR13 measurements, eliminating the systematic mismatch between large- and small-σ bins seen in Figs 14 and 15. The best-fitting rms error is 399 ± 21 km s^{-1}, corresponding to $\pi = 3.8 \, h^{-1}$ Mpc for our *WMAP*7 cosmological parameters. This rms error appears plausible relative to comparisons of different redshift estimators as described in the BOSS DR9 quasar catalogue paper of Paris et al. (2012). The global

---

[1] This statement holds true even though our σ bins have a finite width, a point we have tested explicitly by adding random offsets to halo positions and remeasuring $\xi(\pi, \sigma)$. Note that an offset of π comoving h^{-1} Mpc corresponds to a velocity offset $v = \pi \times (1 + z)^{-1} \times (100 \, \text{km s}^{-1}) \times H(z)/H_0$ at redshift z. For our cosmology at $z = 2.5$, $v = 101\pi$ km s^{-1}.
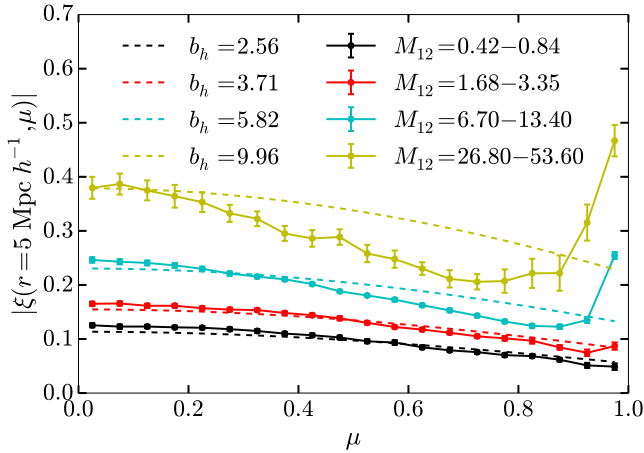
Figure 13. The cross-correlation between halo centres and the Lyα forest from LyMAS at a separation of $r = 5\ h^{-1}$ Mpc as a function of the cosine of the angle from the line of sight, μ, for the halo mass bins $M_{12} = 0.42$–0.84, 1.68–3.35, 6.70–13.40, and 26.80–53.60 (mass increases bottom to top). Solid lines show the correlation predicted by LyMAS and dashed lines show the linear theory prediction (bias increases bottom to top).
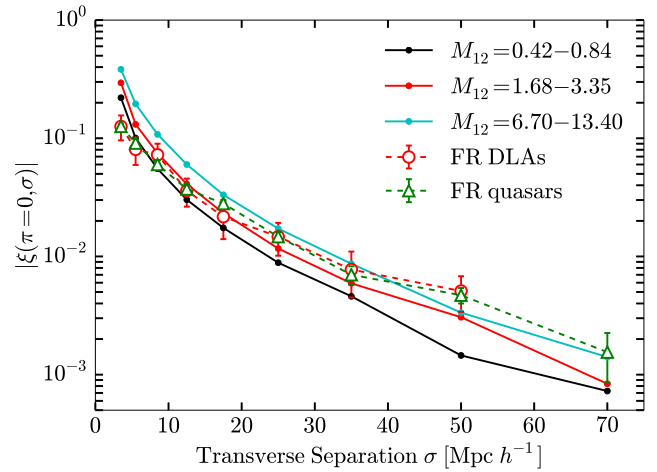
Figure 15. The strength of the cross-correlation as measured by the best-fitting amplitude of the Lorentz profile fit to LyMAS for each of the different halo mass bins (solid lines; mass increases bottom to top) and transverse separation bins. The red dashed and green dashed lines with open points show the amplitude of the measured correlation for DLAs and quasars, respectively, found by fitting the Lorentz profile to the data.
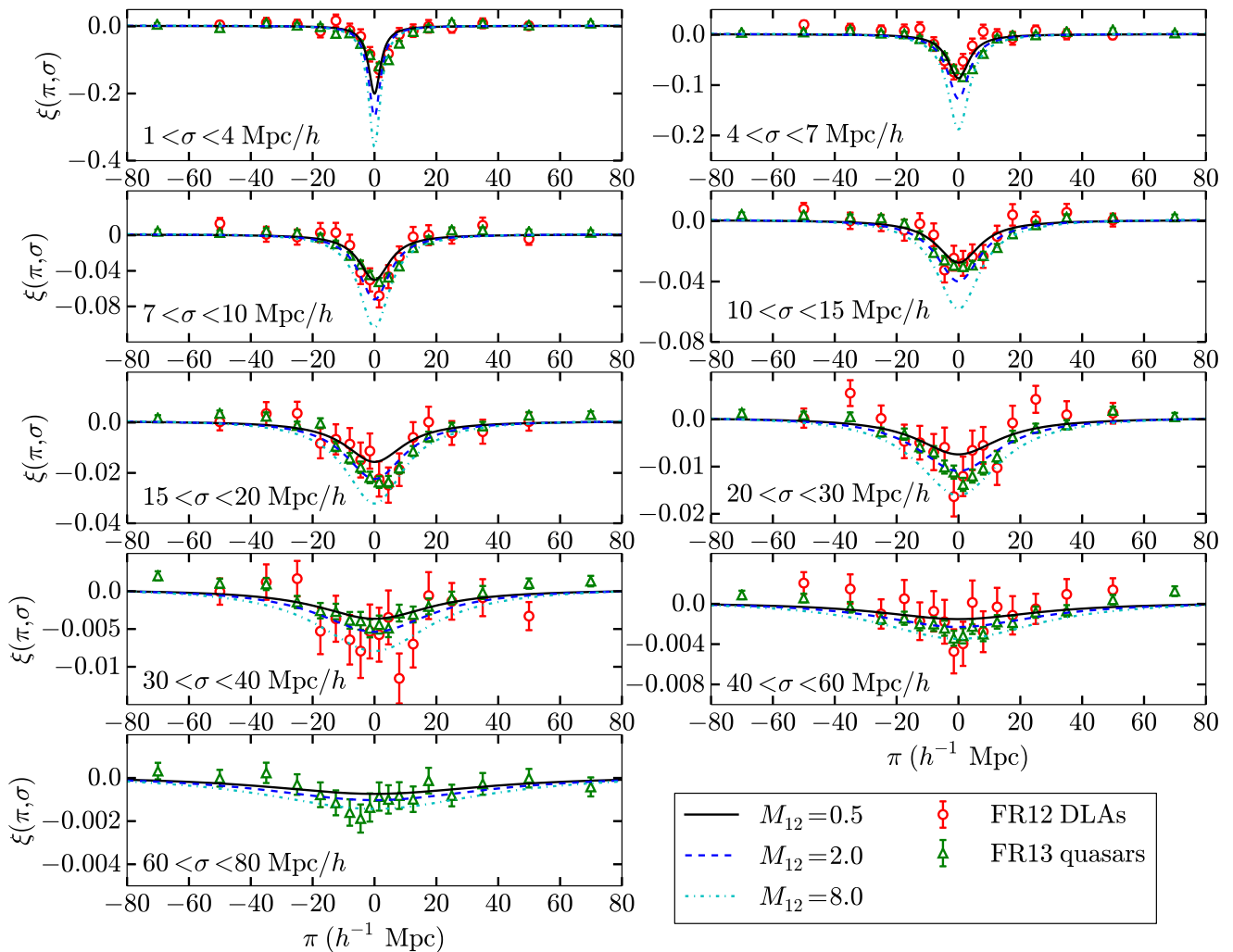
Figure 14. The cross-correlation as predicted by LyMAS (described by Lorentz profile fits, and adding the mean transmission correction) for halo masses of $M_{12} = 0.5$ (black solid), $M_{12} = 2.0$ (blue dashed), and $M_{12} = 8.0$ (cyan dot–dashed). Red circles and green triangles are the cross-correlation measured by FR12 and FR13 for DLAs and quasars, respectively. Panels show different transverse separation bins as indicated, increasing top to bottom, left to right.
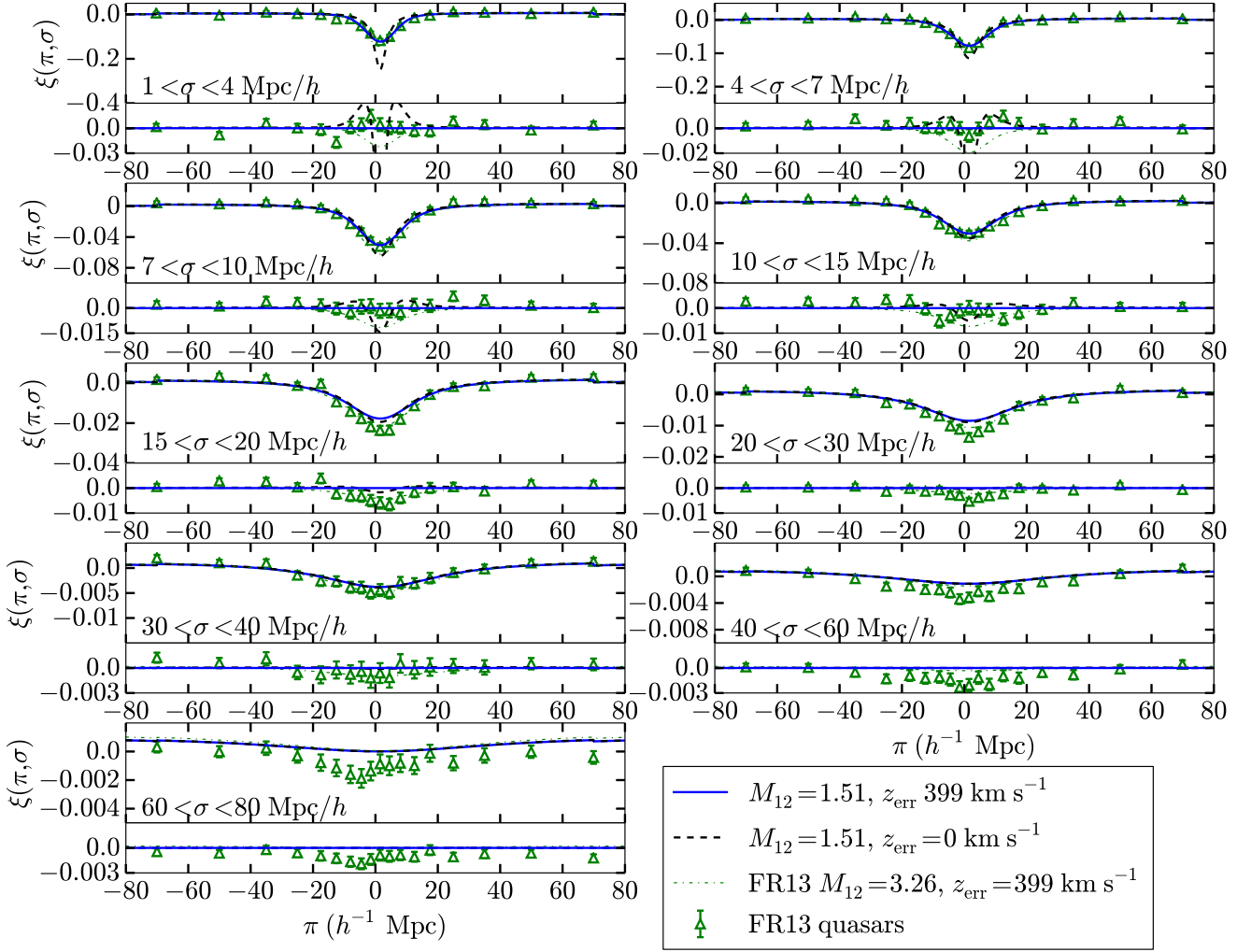
Figure 16. The Lorentz profile that describes the cross-correlation for a halo mass of $M_{12} = 1.51$ (black dashed) and this same correlation when a redshift measurement error of 399 km s^{-1} (blue solid) is included in the line-of-sight position of the DM haloes. This is the error that brings the strength of the correlation closest to that of the data. Green triangles are the cross-correlation between quasars and the Lyα forest from FR13, and green dot–dashed line is the correlation implied by FR13's best-fitting bias. Panels show different transverse separation bins as labelled. Lower panels show residuals from the best-fitting model with $z_{\rm err} = 399$ km s^{-1}.

best-fitting halo mass is $M_{12} = 1.51^{+0.11}_{-0.10}$. As discussed in Section 3.2, this should be interpreted as a bias-weighted mean host halo mass of quasars, i.e. it is the mass $\bar{M}$ for which

$$b(\bar{M}) = \frac{\int_0^\infty \mathrm{d}M\ b(M)p(M)\,\mathrm{d}n/\mathrm{d}M}{\int_0^\infty \mathrm{d}M\ p(M)\,\mathrm{d}n/\mathrm{d}M}, \qquad (13)$$

where $p(M)$ is the probability that a halo of mass M hosts a BOSS quasar at any given time and $\mathrm{d}n/\mathrm{d}M$ is the halo mass function. With the T10 halo-bias relation at $z = 2.5$, our halo mass range corresponds to $b(\bar{M}) = 3.18 \pm 0.06$. Evaluated at the central redshift of the observational measurement, $z = 2.3$, the same halo mass corresponds to $b(\bar{M}) = 2.92 \pm 0.06$.

One can see from Fig. 16 that the LyMAS predictions including rms velocity errors are a good but not perfect fit to the FR13 measurements, given the extremely small statistical errors of the latter. The overall χ^2 is 290 for 162 data points, with two fitting parameters. Fig. A2 also shows residuals between the LyMAS numerical results and our Lorentzian fits to those results, but those residuals are small compared to the differences between the model predictions and the data. We have checked that adding these residuals

(scaled in amplitude by the halo bias) to our convolved Lorentzian model – in effect, using the direct simulation results in place of the Lorentzian description of them – does not noticeably change the χ^2 of the fit or the best-fitting values of the model parameters.

Fig. 17 examines the χ^2 contributions in greater detail, with each panel showing the χ^2 from data points in a single σ bin as a function of M_{12}, for our best-fitting rms velocity error of 399 km s^{-1} and for velocity errors of 100, 200, or 600 km s^{-1}. With a small velocity error, the $\sigma = 1$–$4\ h^{-1}$ Mpc bin has a high χ^2 for any M_{12}, and, as seen previously in Fig. 14, the value of M_{12} that would match in this bin is strongly discordant with the values that fit larger σ bins. Conversely, a large velocity error makes $\xi(\pi, \sigma)$ too broad in the innermost bins, worsening χ^2. Even with the best-fitting rms error, the innermost bin favours a lower M_{12} than other bins. Apart from this bin, the σ bins with the largest χ^2 contributions are 15–20, 20–30, and 40–60 h^{-1} Mpc, where one can see visually in Fig. 16 that our best-fitting model underpredicts the measured correlation. In other bins, the global best-fitting $M_{12} = 1.51$ yields close to the minimum χ^2 contribution, and the χ^2 is not very far from the value 18 ± 4.2 expected for 18 data points.
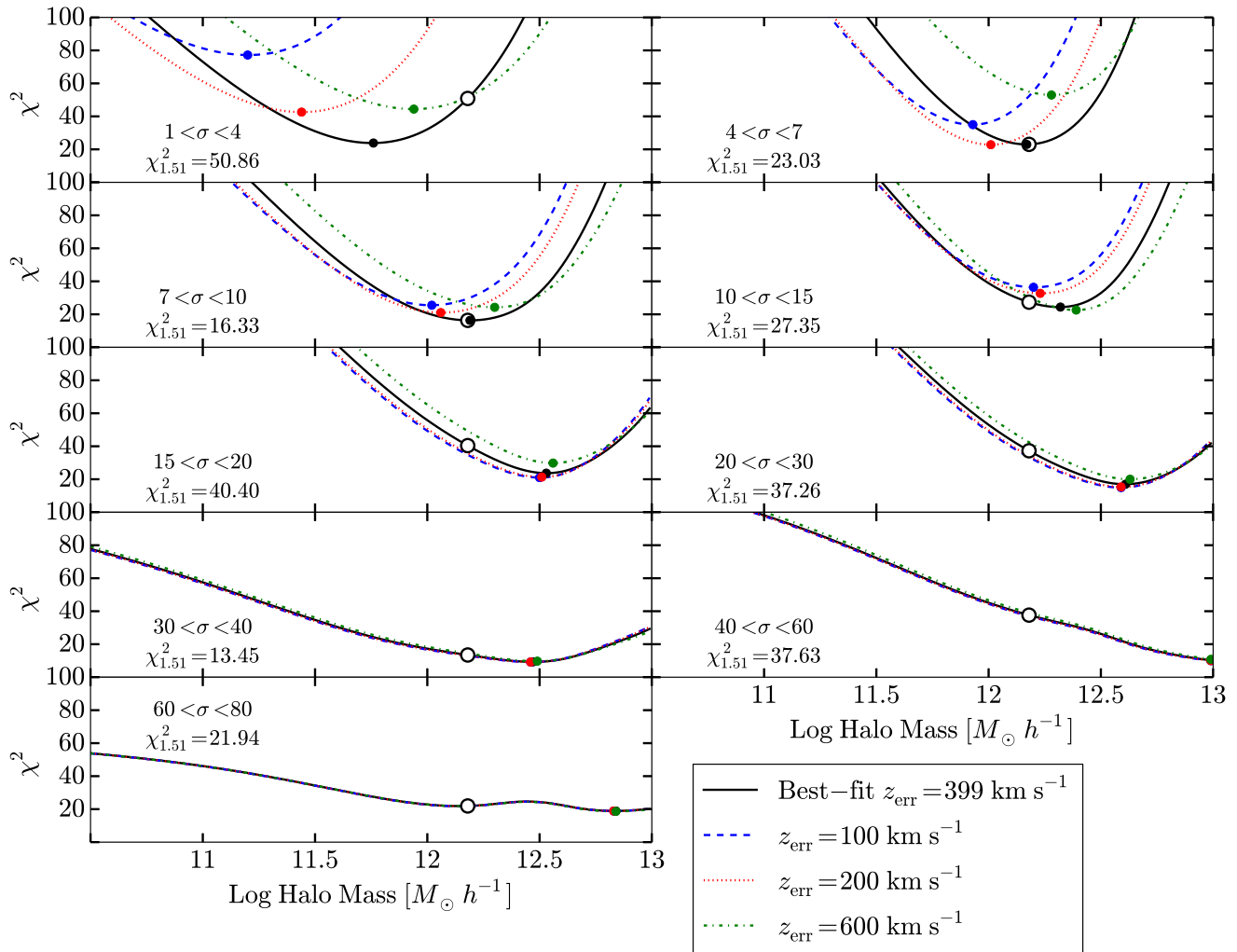
Figure 17. χ^2 as a function of halo mass in each σ bin of the FR13 quasar-flux cross-correlation measurement. Black curves adopt our global best-fitting value of 399 km s^{-1} for the rms velocity error, while dashed, dotted, and dot–dashed curves impose rms velocity errors of 100, 200, and 600 km s^{-1}, respectively. Filled circles mark the minimum of each curve, i.e. the halo mass preferred by the data in this σ bin alone for the adopted velocity error. Open black circles show the global best-fitting $M_{12} = 1.51$, and the contribution to χ^2 from each σ bin for this best-fitting model is listed. There are 18 $\xi(\pi, \sigma)$ data points in each σ bin.

The slight mismatch between our simulation output at $z = 2.5$ and the central redshift $z = 2.3$ of the FR13 observational analysis introduces some ambiguity in comparing our inferred halo bias and halo mass to the results of FR13's linear theory fit. (Ongoing work with the much larger BOSS DR12 sample will allow tailored comparisons in fairly narrow redshift bins.) Our value of $b(\bar{M}) = 3.18 \pm 0.06$ is formally inconsistent with the $b_Q = 3.64^{+0.13}_{-0.15}$ found by FR13, and our inferred halo mass is correspondingly lower. Using the T10 halo formula at $z = 2.3$, the FR13 best-fitting bias corresponds to $M_{12} = 3.26$. Green curves in Fig. 16 show LyMAS predictions for this halo mass at $z = 2.5$, with our best-fitting rms velocity error. While this comparison is inexact because of the redshift mismatch, it appears likely that most of the difference between our inferred bias and the somewhat higher value of FR13 arises because FR13 fit to points with $r > 15 \, h^{-1}$ Mpc, thus matching the $\sigma = 15$–$60 \, h^{-1}$ Mpc bins well and not penalizing the overly strong correlations predicted at small π in the smaller σ bins. The poor global χ^2 of our fit indicates that one should be cautious about our formal error bar. Furthermore, it is clear from Fig. 17 that higher M_{12} is disfavoured mainly by the innermost σ bin

and is therefore sensitive to our assumption of Gaussian velocity errors. If we keep the best-fitting velocity error but drop the innermost bin from the mass determination, we get $M_{12} = 2.19^{+0.16}_{-0.15}$, with corresponding $b(\bar{M}) = 3.63^{+0.08}_{-0.07}$ at $z = 2.5$.

The mild tension between the halo mass and bias factor inferred here and in FR13 is an indication of the current level of systematic uncertainty associated with model fitting and data selection. The FR13 value has the virtue of using a model that has an acceptable χ^2 over the range fitted ($r > 15 \, h^{-1}$ Mpc) and matching the central redshift of the data. It relies on the value of $b_F(1 + \beta_F)$ inferred from fitting the observed flux auto-correlation, and this fitting has statistical and systematic uncertainties of its own. LyMAS directly predicts the halo-flux cross-correlation and thus circumvents this step, but the mismatch between the LyMAS flux auto-correlation and the Blomqvist et al. (2015) linear theory parameters (Fig. 10) could indicate that the IGM structure in the hydrodynamic simulations on which LyMAS is calibrated is incorrect. The excellent match between LyMAS and the full hydrodynamic simulation predictions of the cross-correlation indicates that this is not an issue with LyMAS itself, but it could be a consequence of the physics and
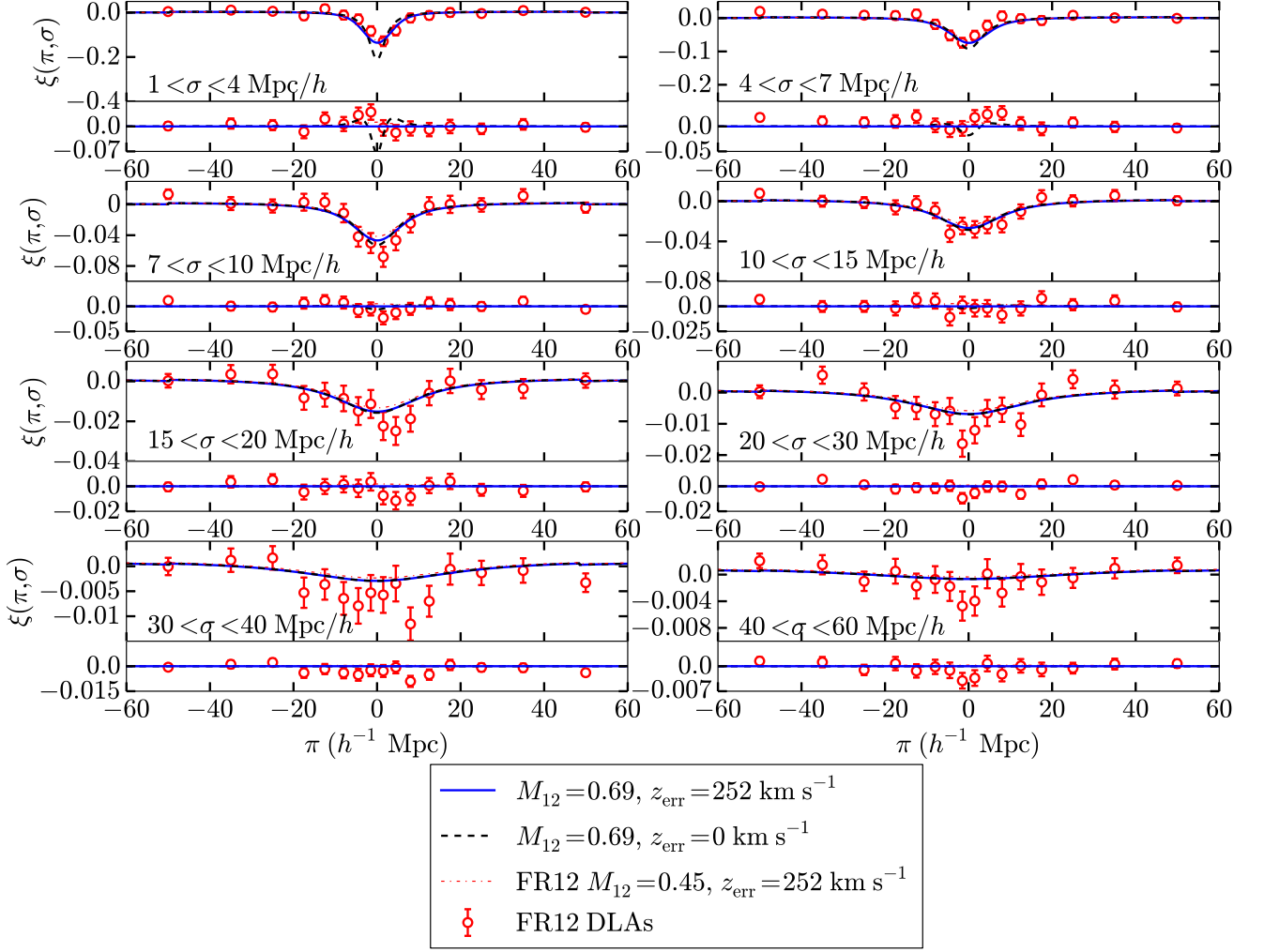
Figure 18. The Lorentz profile that describes the cross-correlation for a halo mass of $M_{12} = 0.69$ (black dashed) and this same correlation when a redshift measurement error of 252 km s^{-1} (blue solid) is included in the line-of-sight position of the DM haloes. This is the error that brings the strength of the correlation closest to that of the data. Red circles are the cross-correlation between DLAs and the Lyα forest from FR12, and red thin line is the correlation implied by FR12's best-fitting bias. Panels show different transverse separation bins as labelled. Lower panels show residuals from the best-fitting model with $z_{\text{err}} = 252$ km s^{-1}.

parameters assumed in the hydrodynamic simulations. The principal strength of the LyMAS mass determination is its use of a fully non-linear model that describes $\xi(\pi, \sigma)$ down to small scales. A hydrodynamic simulation on its own cannot make this prediction accurately because of limited volume, but the combination of hydrodynamics and LyMAS can.

Fig. 18 presents our fit to the FR12 DLA measurements including velocity errors, in the same format as Fig. 16. Here we assume no mean velocity shift, as there is nothing expected to produce a systematic offset in DLA redshifts. Our global best fit yields an rms velocity error of 252^{+63}_{-53} km s^{-1} and an effective host halo mass $M_{12} = 0.69^{+0.16}_{-0.14}$, with corresponding $b(\bar{M}) = 2.66 \pm 0.14$ at $z = 2.5$. The inclusion of velocity errors makes an important difference in the $\sigma = 1$–$4\ h^{-1}$ Mpc bin and little difference in other bins. Our model of Gaussian velocity errors with a single rms for DLAs is questionable, as there is likely a strong dependence of the error on the detection of associated metal lines, but it appears to yield a reasonable fit to the data none the less. The χ^2 of the global fit is

123 for 128 data points, with two fitting parameters. At $z = 2.3$, our best-fitting halo mass corresponds to a bias $b(\bar{M}) = 2.39^{+0.13}_{-0.11}$. Our inferred bias and halo mass are higher than those of the FR12 linear theory fit, which yields $b_{\text{DLA}} = 2.17 \pm 0.2$ and corresponding $M_{12} = 0.45$, but the two determinations are consistent at the $\sim 1\sigma$ level. The FR12 fit did not include velocity errors, and it excluded points with $r < 5\ h^{-1}$ Mpc. Red curves in Fig. 18 show the prediction for $M_{12} = 0.45$ and our best-fitting velocity error, and one can see that the differences from our best-fitting model prediction are small compared to the observational error bars.

Analysis of the full BOSS DR12 data set will allow more detailed and more precise measurements for the cross-correlations of both quasars and DLAs, allowing narrower redshift bins for evolution measurements, dividing samples into subsets of luminosity, equivalent width, or metal-line strength, and achieving good statistical constraints at large separations. Modelling these measurements with both linear theory and non-linear predictions from LyMAS should yield numerous insights into the quasar and DLA populations, and

comparisons of these model fits will either remove or sharpen the mild tensions found here for the quasar cross-correlations. With the DR12 data, it may be possible to infer the rms velocity of DLAs within their host haloes by matching the small-scale correlation for the subset of DLAs that have metal-line absorption and thus small redshift measurement errors. This measurement would be an interesting new constraint on the physics of DLA absorption.

5 CONCLUSIONS

Cross-correlation with the densely mapped Lyα forest opens a new window on the properties of quasars, DLAs, and potentially other populations of high-redshift objects. At large scales, one expects these correlations to be described by linear perturbation theory and linear bias, and precise measurements of quasar and DLA bias can be translated (for a specified cosmological model) to precise constraints on the masses of their host DM haloes (FR12; FR13). Interpreting small-scale cross-correlations requires a fully non-linear hydrodynamic model of the Lyα forest. However, cosmological hydrodynamic simulations with the required resolution cannot model large enough volumes to provide accurate predictions of 3d cross-correlations.

LyMAS provides a way to bootstrap hydrodynamic simulation results on to large-volume *N*-body simulations and thus predict halo-flux cross-correlations over the full range of observable scales. When applied to the DM distribution of the Horizon-AGN or Horizon-noAGN simulations, LyMAS reproduces the cross-correlations $\xi(\pi, \sigma)$ predicted using the full hydrodynamic Lyα forest spectra almost exactly (Fig. 2). The LyMAS predictions are insensitive to the DM mass resolution or smoothing scale over the range investigated here (Fig. 6). Furthermore, the predicted $\xi(\pi, \sigma)$ is almost completely insensitive to AGN feedback as implemented in Horizon-AGN (Fig. 4), implying that this measure of cosmological structure is insensitive to feedback assumptions even at scales $\sigma = 1–4\ h^{-1}$ Mpc.

We have applied LyMAS to a 2048^3-particle *N*-body simulation of a $1\ h^{-1}$ Gpc box, with the same *WMAP*7 cosmological parameters used for the Horizon simulations, to predict $\xi(\pi, \sigma)$ for halo masses in the range $M_{12} \approx 0.5–50$ at $z = 2.5$. Our numerical results (Fig. A1) can be described by the empirically chosen Lorentzian form of equation (5), which can be used along with equations (6) and (7) and the parameters in Tables A1 and A2 to reproduce our results for any of the FR12/FR13 σ bins and any halo mass in this range. The difference between the Lorentzian fit and the numerical results is significant at the level of our simulation's small statistical errors, but it is small compared to current observational errors. Except for the $\sigma = 1–4\ h^{-1}$ Mpc bin, the trend of cross-correlation amplitude with halo mass follows the standard large-scale $b(M_h)$ relation from T10 (Fig. 7). Furthermore, even on small scales, $\xi(\pi, \sigma)$ for a halo population with a range of masses depends only on the number-weighted halo bias (equation 9). Modelling quasar or DLA observations down to these scales therefore improves constraints on the characteristic halo mass but does not provide information about the width or shape of the halo mass distribution.

We have compared the LyMAS results to linear theory predictions using bias parameters b_F and β_F for the Lyα forest determined by fitting the LyMAS flux auto-correlation, and T10 values for the halo bias. We find surprisingly good agreement in $\xi(\pi, \sigma)$ even at $\sigma = 1–4\ h^{-1}$ Mpc, with maximum deviations of $\sim$10 per cent near $\pi = 0\ h^{-1}$ Mpc (Fig. 11). However, in a monopole–quadrupole representation, linear theory and LyMAS differ significantly in the

quadrupole at $r < 15\ h^{-1}$ Mpc, while agreement for the monopole is good down to $r = 3\ h^{-1}$ Mpc (Fig. 12). In a $\xi(r, \mu)$ representation (Fig. 13), deviations are μ-dependent, and they are larger for higher halo masses.

When comparing to the BOSS measurements of FR12 and FR13, we find that any choice of halo mass that fits $\xi(\pi, \sigma)$ at $\sigma > 7\ h^{-1}$ Mpc overpredicts the amplitude of the cross-correlation at smaller scales (Fig. 14). The most likely explanation for this discrepancy is scatter between halo velocities and measured redshifts in BOSS, which can arise from a combination of redshift measurement errors and random peculiar velocities of quasars or DLAs within their host haloes; the velocities of the haloes themselves are already accounted for in the LyMAS predictions. We fit the quasar (FR13) and DLA (FR12) measurements with a model that convolves the LyMAS predictions with a Gaussian distribution of redshift offsets of rms amplitude z_{err}, which removes the tension between large-scale and small-scale predictions. For quasars, we also incorporate the -157 km s^{-1} mean redshift offset inferred by FR13 from the asymmetry of $\xi(\pi, \sigma)$.

For quasars (Fig. 16) we find $z_{err} = 399 \pm 21$ km s^{-1}, a magnitude that is compatible with other estimates of quasar redshift errors from BOSS spectra (Paris et al. 2012). Our best-fitting halo mass is $M_{12} = 1.51^{+0.11}_{-0.10}$, which corresponds to a halo bias $b(\bar{M}) = 3.18 \pm 0.06$ at $z = 2.5$ for our *WMAP*7 cosmology (using the T10 halo-bias relation). For DLAs (Fig. 18), we find $z_{err} = 252^{+63}_{-53}$ km s^{-1}, a value that appears high for pure measurement errors but may reflect a quadrature sum of measurement errors and halo internal velocities. Our best-fitting halo mass is $M_{12} = 0.69^{+0.16}_{-0.14}$, with corresponding $b(\bar{M}) = 2.66 \pm 0.14$ at $z = 2.5$.

Relative to linear theory modelling, the LyMAS results have the advantage of fitting a fully non-linear model to the full range of measurements, and (as a necessary consequence) fitting for redshift scatter as well as effective halo mass. However, there are several caveats to our numbers. First, the slight mismatch between our simulation's $z = 2.5$ redshift output and the $z = 2.3$ central redshift of the BOSS measurements introduces ambiguity; our best-fitting halo masses correspond to bias factors that are about 10 per cent lower at $z = 2.3$ than at $z = 2.5$ (see Section 4), and it is not clear which is the more appropriate characterization of our inferred halo bias. Secondly, the flux auto-correlation from the $1\ h^{-1}$ Gpc LyMAS box is significantly below that from the linear theory fit of Blomqvist et al. (2015) to the BOSS auto-correlation measurements at $z = 2.3$ (Fig. 10), and our best-fitting values of β and $b_F(1 + \beta)$ do not agree with the values extrapolated from smaller volume hydro simulations by Arinyo-i-Prats et al. (2015). This difference could reflect a combination of incorrect thermal structure of the diffuse IGM in our calibrating hydro simulations (e.g. if the real Universe has a different history of helium reionization), inaccurate prediction of the large-scale flux auto-correlation by LyMAS, uncertainties on scaling results between $z = 2.3$ and 2.5, and errors in the Blomqvist et al. (2015) measurements, which themselves require substantial corrections for the impact of continuum determination in the observed spectra.

Finally, for quasars, our fit has a statistically unacceptable χ^2/d.o.f. $= 290/160$. The largest contribution to χ^2 comes from our innermost σ bin, where predictions are most sensitive to our assumption of Gaussian velocity errors. If we eliminate this bin from our mass fit (but retain the best-fitting z_{err}), we find $M_{12} = 2.19^{+0.16}_{-0.15}$, with corresponding $b(\bar{M}) = 3.63^{+0.08}_{-0.07}$ at $z = 2.5$. We regard this as our most reliable estimate of halo mass and halo bias for BOSS quasars because of its lower sensitivity to our

$z_{\rm err}$ model. The χ^2/d.o.f. improves to 197/142, but it is still high, reflecting a tension between lower masses favoured by the $\sigma = 4$–$15\ h^{-1}$ Mpc data and the higher masses favoured by $\sigma = 15$–$60\ h^{-1}$ Mpc data (Fig. 17). If instead we eliminate all data points with $r < 15\ h^{-1}$ Mpc but retain our best-fitting $z_{\rm err} = 399$ km s^{-1}, we obtain a best-fitting halo mass of $M_{12} = 2.69^{+0.33}_{-0.35}$, corresponding to $b(\bar{M}) = 3.85^{+0.14}_{-0.16}$ at $z = 2.5$. The χ^2/d.o.f. for this fit remains high at 221/132. For DLAs, the measurement errors are larger, and our fit (with $M_{12} = 0.69^{+0.16}_{-0.14}$ and $b(\bar{M}) = 2.39^{+0.13}_{-0.11}$) has an acceptable χ^2/d.o.f. =123/126.

Our DLA halo bias is consistent with that inferred by FR12 using linear theory modelling at separations $r > 5\ h^{-1}$ Mpc. Our quasar halo bias is lower than that inferred by FR13; an exact comparison is complicated by the slight difference in model redshift, but the best-fitting FR13 parameters correspond to a predicted cross-correlation that is stronger by $\approx$24 per cent (see Fig. 16). The main source of difference appears to be the selection of fitting range, as FR13's elimination of $r < 15\ h^{-1}$ Mpc separations removes the points that are favouring lower bias in our fit. The surprisingly good agreement between LyMAS and linear theory predictions of $\xi(\pi, \sigma)$ suggests that smaller separations can be retained even in a linear theory fit, provided one accounts for quasar redshift errors.

Analyses of the final BOSS data set will allow cross-correlation measurements in bins of redshift and tracer properties (quasar luminosity, DLA metallicity) while retaining useful statistical precision on large scales. Comparison of observations, linear theory fits, and LyMAS modelling across these bins will allow tests for internal consistency (e.g. expected redshift scaling) and investigations of the relation between tracer properties and halo mass. The importance of $z_{\rm err}$ in our fits highlights a new opportunity to constrain the velocity distribution of quasars or DLAs within their host haloes. For quasars, one can imagine measuring the mean offset and dispersion of redshifts inferred from different emission lines as a diagnostic of gas motions and ionization states in the immediate surroundings of the central black hole. Ionizing radiation from quasars should also affect cross-correlations on small scales (the 'transverse proximity effect'), and it may be possible to separate this impact from redshift dispersion effects via the expected dependence on quasar luminosity. For DLAs, the comparison of rms velocities to halo circular velocities may be a useful diagnostic for the kinematics of the metal-line absorption, perhaps distinguishing rotating discs from more chaotic gas distributions. The first studies of Lyα forest cross-correlations have already yielded the most precise measurements of quasar and DLA clustering strength, and this tool holds considerable promise as a probe of the physics of high-redshift systems.

ACKNOWLEDGEMENTS

REFERENCES

Ahn C. P. et al., 2012, ApJS, 203, 21
Alam S. et al., 2015, ApJS, 219, 12
Arinyo-i-Prats A., Miralda-Escudé J., Viel M., Cen R., 2015, J. Cosmol. Astropart. Phys., 12, 017
Blomqvist M. et al., 2015, J. Cosmol. Astropart. Phys., 11, 034
Borde A., Palanque-Delabrouille N., Rossi G., Viel M., Bolton J. S., Yèche C., LeGoff J.-M., Rich J., 2014, J. Cosmol. Astropart. Phys., 7, 005
Busca N. G. et al., 2013, A&A, 552, A96
Cen R., Miralda-Escudé J., Ostriker J. P., Rauch M., 1994, ApJ, 437, L9
Croft R. A. C., Weinberg D. H., Katz N., Hernquist L., 1998, ApJ, 495, 44
Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, ApJ, 292, 371
Dawson K. S. et al., 2013, AJ, 145, 10
Delubac T. et al., 2015, A&A, 574, A59
Dubois Y., Gavazzi R., Peirani S., Silk J., 2013, MNRAS, 433, 3297
Dubois Y. et al., 2014, MNRAS, 444, 1453
Eisenstein D. J. et al., 2011, AJ, 142, 72
Espey B. R., Carswell R. F., Bailey J. A., Smith M. G., Ward M. J., 1989, ApJ, 342, 666
Faucher-Giguère C.-A., Prochaska J. X., Lidz A., Hernquist L., Zaldarriaga M., 2008, ApJ, 681, 831
Font-Ribera A. et al., 2012, J. Cosmol. Astropart. Phys., 11, 59 (FR12)
Font-Ribera A. et al., 2013, J. Cosmol. Astropart. Phys., 5, 18 (FR13)
Font-Ribera A. et al., 2014, J. Cosmol. Astropart. Phys., 5, 027
Gontcho A Gontcho S., Miralda-Escudé J., Busca N. G., 2014, MNRAS, 442, 187
Greig B., Bolton J. S., Wyithe J. S. B., 2015, MNRAS, 447, 2503
Gunn J. E., Peterson B. A., 1965, ApJ, 142, 1633
Hernquist L., Katz N., Weinberg D. H., Miralda-Escudé J., 1996, ApJ, 457, L51
Hewett P. C., Wild V., 2010, MNRAS, 405, 2302
Kaiser N., 1987, MNRAS, 227, 1
Komatsu E. et al., 2011, ApJS, 192, 18
Lidz A. et al., 2010, ApJ, 718, 199
McDonald P., 2003, ApJ, 585, 34
Miralda-Escudé J., Cen R., Ostriker J. P., Rauch M., 1996, ApJ, 471, 582
Paris I. et al., 2012, A&A, 548, A66
Peeples M. S., Weinberg D. H., Davé R., Fardal M. A., Katz N., 2010, MNRAS, 404, 1281
Peirani S., Weinberg D. H., Colombi S., Blaizot J., Dubois Y., Pichon C., 2014, ApJ, 784, 11 (P14)
Planck Collaboration XIII, 2015, preprint (arXiv:1502.01589)
Pontzen A., 2014, Phys. Rev. D, 89, 083010
Richards G. T., Vanden Berk D. E., Reichard T. A., Hall P. B., Schneider D. P., SubbaRao M., Thakar A. R., York D. G., 2002, AJ, 124, 1
Ross N. P. et al., 2012, ApJS, 199, 3
Seljak U., 2012, J. Cosmol. Astropart. Phys., 3, 004
Slosar A., Ho S., White M., Louis T., 2009, J. Cosmol. Astropart. Phys., 10, 019
Slosar A. et al., 2011, J. Cosmol. Astropart. Phys., 9, 001
Slosar A. et al., 2013, J. Cosmol. Astropart. Phys., 4, 026
Spergel D. N. et al., 2003, ApJS, 148, 175
Springel V., 2005, MNRAS, 364, 1105
Teyssier R., 2002, A&A, 385, 337
Tinker J. L., Robertson B. E., Kravtsov A. V., Klypin A., Warren M. S., Yepes G., Gottlöber S., 2010, ApJ, 724, 878 (T10)
Viel M., Schaye J., Booth C. M., 2013, MNRAS, 429, 1734
Weinberg D. H., Katz N., Hernquist L., 1998, in Woodward C. E., Shull J. M., Thronson H. A., Jr, eds, ASP Conf. Ser. Vol. 148, Origins. Astron. Soc. Pac., San Francisco, p. 21
Zhang Y., Anninos P., Norman M. L., 1995, BAAS, 27, 1412

APPENDIX A: CROSS-CORRELATION FITTING FUNCTION

Fig. A1 shows the LyMAS predictions for the cross-correlation of the $M_{12} = 0.84$–1.68 haloes from the 2048^3 simulation of the $(1\ h^{-1}\ {\rm Gpc})^3$ volume with $r_{\rm s} = 0.5\ h^{-1}$ Mpc, for transverse separation bins that range from $\sigma = 1$–$4\ h^{-1}$ Mpc up to $\sigma = 60$–$80\ h^{-1}$ Mpc. The correlation at $\pi = 0$ weakens steadily with increasing σ as expected, from $|\xi(\pi = 0, \sigma)| \approx 0.3$ in the $\sigma = 1$–$4\ h^{-1}$ Mpc bin to 0.04 at $\sigma = 10$–$15\ h^{-1}$ Mpc to 0.002 at $\sigma = 60$–$80\ h^{-1}$ Mpc. Results for other halo mass bins are qualitatively
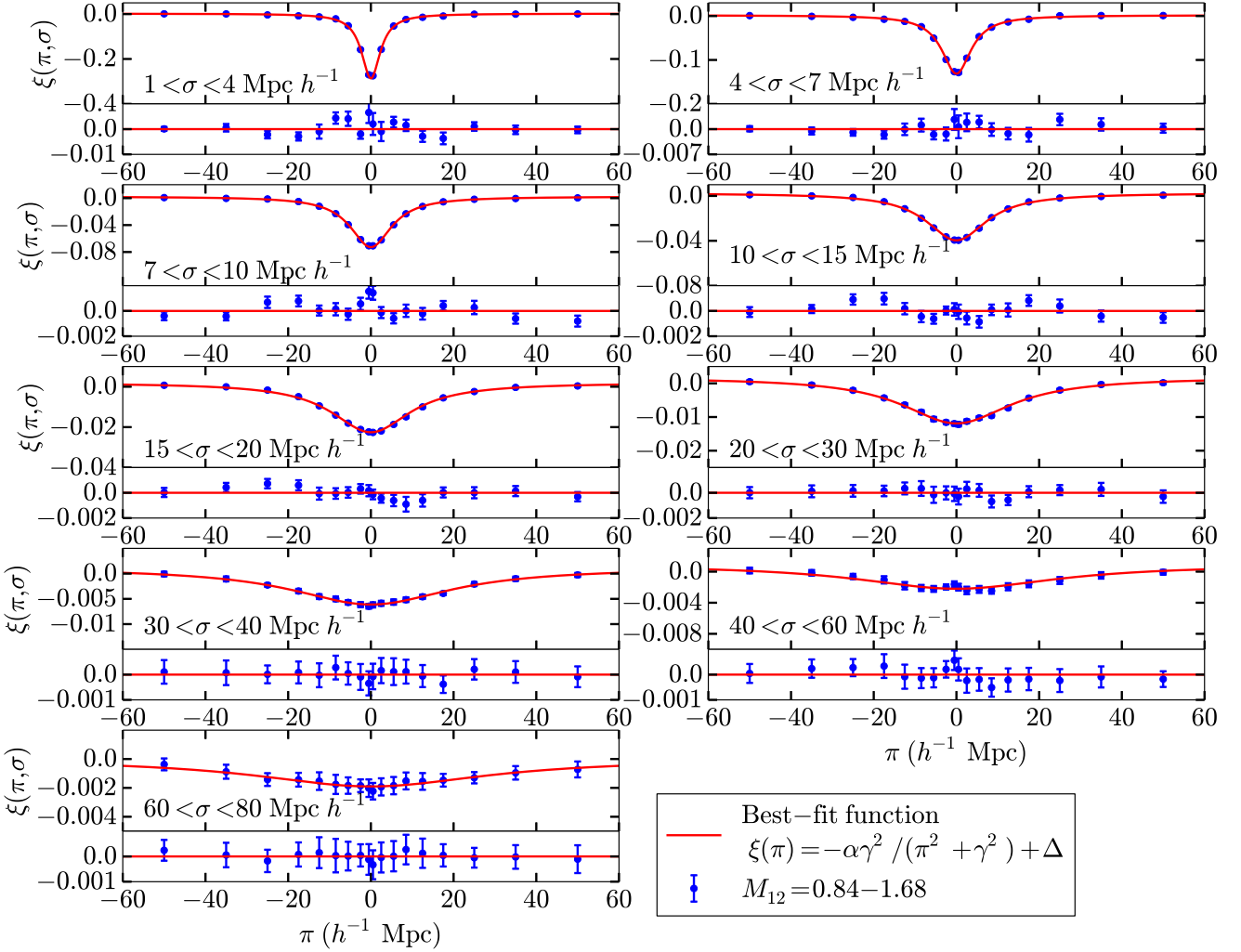
Figure A1. The cross-correlation between DM haloes and the Lyα forest flux from applying LyMAS to the 2048^3 ($1\ h^{-1}$ Gpc)^3 DM simulation with $r_s = 0.5\ h^{-1}$ Mpc for haloes in the range $M_{12} = 0.84$–1.68 (blue points, with error bars estimated from 16 subvolumes). Red curves show the best-fitting Lorentzian fitting function $\xi(\pi) = -\alpha\gamma^2/(\pi^2 + \gamma^2) + \Delta$. Panels show different transverse separation bins, increasing left to right and top to bottom, and residuals are plotted beneath each panel.

similar, with the expected trend of stronger correlations for higher mass haloes. Error bars on these points are estimated by the subvolume method. Curves in Fig. A1 show the least-squares fit of equation (5) in each σ bin, where we have used the subvolume error bars (treated as diagonal) to fit α, γ, and Δ. Formally, the fits are often not acceptable, but this may be a consequence of ignoring error covariances and/or underestimated errors from our subvolume method.

To test if the subvolume method is an appropriate measure of cosmic variance on the scale of the full volume, we compare three simulations with different initial conditions for the DM density field, each ($1\ h^{-1}$ Gpc)^3 in size, with 1024^3 particles and a smoothing scale of $r_s = 1.0\ h^{-1}$ Mpc. Fig. A2 shows the residual of the measured cross-correlations ($\xi_{\text{meas}} - \xi_{\text{fit}}$) in each simulation from a Lorentz profile fit to the simulation labelled 'IC 3', which has the same DM density field initial conditions as our standard 2048^3 simulation. We plot error bars on each simulation curve computed by applying the subvolume method to that simulation. In the smallest

transverse separation bins, the difference among the three curves is significantly larger than the subvolume error bars, but for $\sigma = 7$–10 h^{-1} Mpc and larger, the run-to-run variations are comparable in size to the error bars. This means the subvolume method underestimates the errors expected from cosmic variance in the smallest separation bins, but for larger separations the subvolume errors are reasonable. Because we do not have many $1\ h^{-1}$ Gpc simulations available, we use the subvolume errors for our analysis.

Tables A1 and A2 can be used together with equations (5)–(7) to accurately reproduce the $z = 2.5$ LyMAS predictions of $\xi(\pi, \sigma)$ for any value of M_h, with typical residuals comparable to those shown in Fig. A1. They thus provide a compact quantitative summary of our numerical results. Table A2 gives the mean value of Δ in each halo mass bin, averaged over the transverse separation bins. There does not appear to be any significant trend in Δ with transverse separation, and only a weak trend with halo mass. Nevertheless, we include it in our model, using the average values recorded in the table.
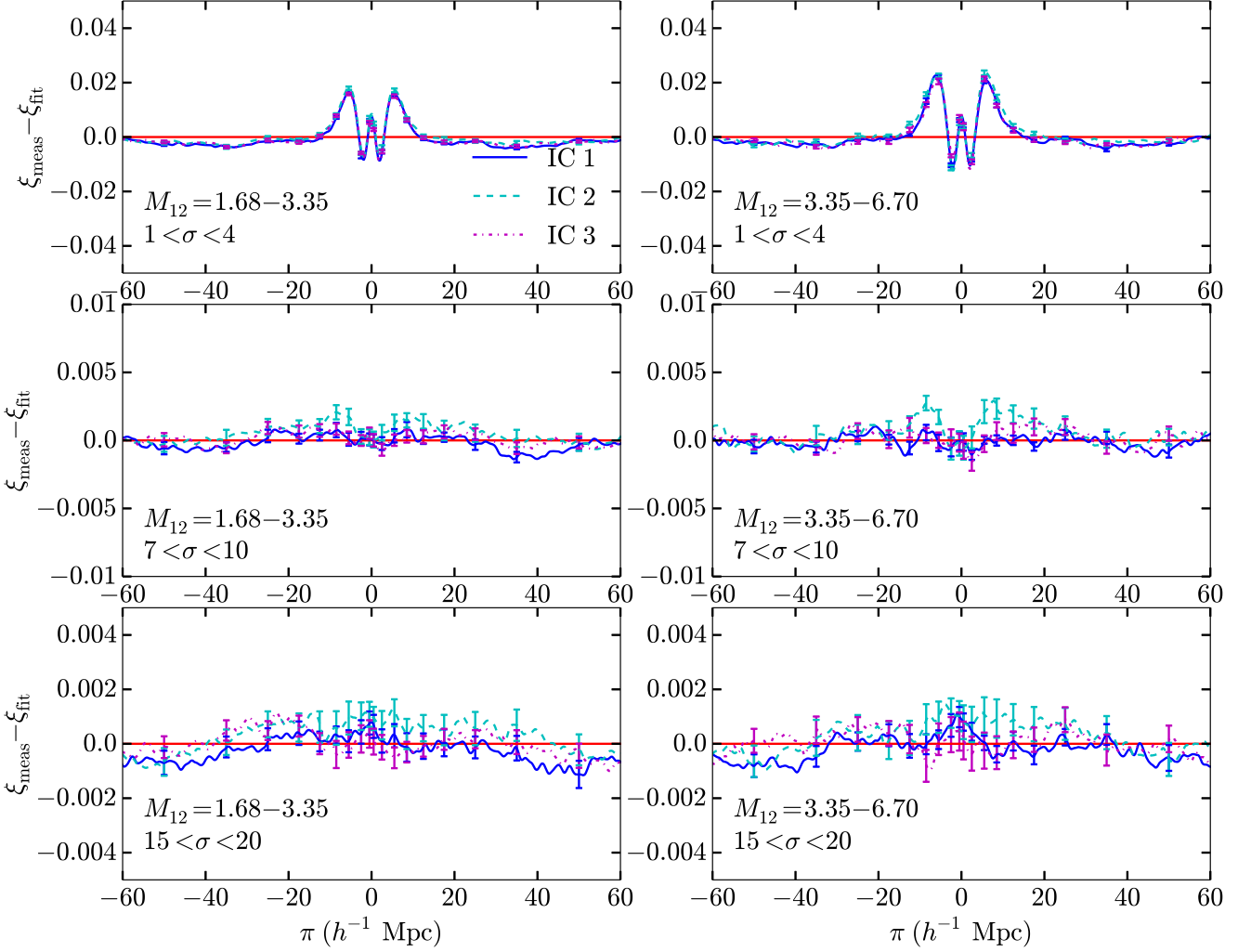
Figure A2. The residuals of the halo-flux cross-correlation from fitting a Lorentz profile to the simulation marked 'IC 3'. The three curves show residuals from three 1024^3-particle $(1\,h^{-1}\,\text{Gpc})^3$ DM simulations with $r_s = 1.0\,h^{-1}$ Mpc and different sets of initial conditions for the DM density field (blue solid, cyan dashed, and magenta dot–dashed lines). Rows show different transverse separation bins $1 < \sigma < 4$, $7 < \sigma < 10$, and $15 < \sigma < 20\,h^{-1}$ Mpc, and columns show different halo mass bins $M_{12} = 1.68$–3.35 and $M_{12} = 3.35$–6.70. Error bars are computed by applying the subvolume method to each simulation individually.

Table A1. Best-fitting parameters for fits to the trend in halo-flux cross-correlation strength with halo mass M_h and transverse separation σ bin, $|\xi(\pi = 0, \sigma)| = A(M_h/4 \times 10^{12}\,h^{-1}\,\text{M}_\odot)^m$. Errors on parameters are from the subvolume method.

σ bin (h^{-1} Mpc)	$\log A$	m
1–4	-0.45 ± 0.003	0.215 ± 0.005
4–7	-0.75 ± 0.003	0.280 ± 0.007
7–10	-1.00 ± 0.004	0.255 ± 0.013
10–15	-1.25 ± 0.006	0.270 ± 0.012
15–20	-1.50 ± 0.008	0.260 ± 0.015
20–30	-1.80 ± 0.011	0.275 ± 0.025
30–40	-2.10 ± 0.020	0.280 ± 0.025
40–60	-2.45 ± 0.026	0.300 ± 0.038
60–80	-2.80 ± 0.036	0.240 ± 0.035

Table A2. The Δ value reported here is the mean across σ bins in the corresponding halo mass bin, since we do not detect any significant trend in Δ with σ bin. Errors on parameters are from the subvolume method.

M_{12} bin	Δ
0.42–0.84	0.0012 ± 0.0003
0.84–1.68	0.0011 ± 0.0003
1.68–3.35	0.0014 ± 0.0003
3.35–6.70	0.0014 ± 0.0003
6.70–13.40	0.0015 ± 0.0003
13.40–26.80	0.0019 ± 0.004
26.80–53.60	0.0019 ± 0.0004

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