

On the possibility of baryon acoustic oscillation measurements at redshift $z > 7.6$ with the Roman space telescope

Siddharth Satpathy^{1,2}★, Zhaozhou An,^{1,2} Rupert A. C. Croft,^{1,2,3,4}★ Tiziana Di Matteo,^{1,2,3,4}★
Ananth Tenneti^{1,2}★, Yu Feng,⁵ Katrin Heitmann⁶ and Graziano Rossi⁷

¹Department of Physics, Carnegie Mellon University, 5000 Forbes Ave., Pittsburgh, PA 15213, USA

²The McWilliams Center for Cosmology, Carnegie Mellon University, 5000 Forbes Ave., Pittsburgh, PA 15213, USA

³School of Physics, The University of Melbourne, VIC 3010, Australia

⁴ARC Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), Australia

⁵Berkeley Center for Cosmological Physics, Department of Physics, University of California Berkeley, Berkeley, CA 94720, USA

⁶HEP and MCS Divisions, Argonne National Laboratory, Lemont, IL 60439, USA

⁷Department of Physics and Astronomy, Sejong University, Seoul 143-747, Korea

Accepted 2020 August 28. Received 2020 August 18; in original form 2019 December 28

ABSTRACT

The Nancy Grace Roman Space Telescope (RST), with its field of view and high sensitivity will make surveys of cosmological large-scale structure possible at high redshifts. We investigate the possibility of detecting baryon acoustic oscillations (BAO) at redshifts $z > 7.6$ for use as a standard ruler. We use data from the hydrodynamic simulation BLUETIDES in conjunction with the gigaparsec-scale Outer Rim simulation and a model for patchy reionization to create mock RST High Latitude Survey grism data for Lyman α emission line selected galaxies at redshifts $z = 7.4$ to $z = 10$, covering 2280 deg^2 . We measure the monopoles of galaxies in the mock catalogues and fit the BAO features. We find that for a line flux of $L = 7 \times 10^{-17} \text{ erg s}^{-1} \text{ cm}^{-2}$, the 5σ detection limit for the current design, the BAO feature is partially detectable (measured in three out of four survey quadrants analysed independently). The resulting root mean square error on the angular diameter distance to $z = 7.7$ is 7.9 per cent. If we improve the detection sensitivity by a factor of two (i.e. $L = 3.5 \times 10^{-17} \text{ erg s}^{-1} \text{ cm}^{-2}$), the distance error reduces to 1.4 per cent. We caution that many more factors are yet to be modelled, including dust obscuration, the damping wing due to the intergalactic medium, and low redshift interlopers. If these issues do not strongly affect the results, or different observational techniques (such as use of multiple lines) can mitigate them, RST, or similar instruments may be able to constrain the angular diameter distance to the high redshift universe.

Key words: galaxies: statistics – dark energy – large-scale structure of Universe.

1 INTRODUCTION

Baryon acoustic oscillations (BAO) have become very useful as a standard ruler to constrain cosmological models (see e.g. the review by Weinberg et al. 2013). Due to the distinctive nature of their signature, the BAO scale constitutes one of the most effective observables in cosmology. The BAO features can be detected as peaks in two point correlation functions or wiggles in power spectra. These signatures of BAO can be used to measure the Hubble constant and the angular diameter distance as function of redshift. As such, it is used as a robust and unique probe to investigate the nature of dark energy.

The sound horizon at the time of recombination leaves its imprint on both the cosmic microwave background (CMB) and 3D distribution of matter. Using simulations, in this paper we explore whether it may be possible to measure the BAO scale at redshifts $z > 7.6$ using galaxies detected by the RST satellite (formerly the *Wide Field*

*Infrared Survey Telescope, WFIRST*¹). These measurements would represent the first BAO detection between $z = 2.3$ and $z = 1100$.

The first observations made of BAO (Cole et al. 2005; Eisenstein et al. 2005) at low redshifts ($z \sim 0\text{--}0.4$) used luminous red galaxies (LRGs) as tracers of structure, and these have been supplemented by even larger samples at higher redshifts (e.g. Zhai et al. 2017; Bautista et al. 2018), up to $z = 0.72$. The density field at redshifts $z \sim 2\text{--}3.5$ is accessible using the Lyman α forest of absorption in quasar spectra, and its measured clustering has led to the determination of the BAO scale at $z \sim 2.3\text{--}2.4$ (Busca et al. 2013; Slosar et al. 2013; du Mas des Bourboux et al. 2017; de Sainte Agathe et al. 2019). The gap between low redshift galaxy determinations and the forest was recently filled by a determination using quasar clustering at $z = 1.52$ (Ata et al. 2018). Emission line galaxies at $z = 0.85$ (Raichoor et al. 2017) and $z = 0.9$ (Comparat et al. 2016) from the Extended Baryon Oscillation Spectroscopic Survey (eBOSS) will also soon be adding constraints in this crucial redshift range, around the time that the universe started to accelerate. Radio

* E-mail: siddharthsatpathy.ss@gmail.com (SS); rcroft@cmu.edu (RACC); tizianad@andrew.cmu.edu (TDM)

¹<https://wfirst.gsfc.nasa.gov/>

observations of 21 cm emission will increase the volume covered (Vanderlinde & Chime Collaboration 2014). At higher redshifts, galaxies have been discovered as far away as $z = 11.1$ (Oesch et al. 2016), in small *Hubble Space Telescope* (*HST*) fields at present. The RST satellite¹ will have a much wider field of view than *HST*, and the planned RST grism survey holds the promise of measuring the large-scale structure of the universe at early times with galaxy survey techniques which have been hitherto been confined to low redshifts.

Among the reasons for looking at the BAO scale in a different redshift regime are the possibility of early dark energy (Hill & Baxter 2018), or models where dark matter decays into radiation and affects the expansion rate (Wang et al. 2014). There is also current tension between locally determined H_0 measurements (Riess et al. 2018) and those determined from the CMB in the context of cold dark matter (CDM) (Planck Collaboration VI 2020a). Using multiple redshifts to determine the consistency (or not) of BAO measurements should be useful in finding a solution to this controversy.

Cosmological surveys designed to measure the BAO feature must cover large volumes, as the comoving length scale involved is 150 megaparsecs, and cosmic variance must be suppressed. The number density of sampling objects must be large enough to overcome shot noise. When there is domination by shot-noise, one can still obtain BAO measurements and the error will just scale as $1/\sqrt{N}$, where N is the number of objects. For example, eBOSS quasars are shot-noise dominated (Zarrouk et al. 2018). The shot-noise dominated regime carries the implication that one can return to the same volume and sample it further to obtain more precise results. Conversely, when one is cosmic variance limited, one gains more from sampling a new volume instead of sampling the same volume.

The current number density of quasars at $z \sim 3$ in the largest survey (final SDSS-III BOSS DR09; Eftekharzadeh et al. 2015) is $1.057 \pm 0.007 (h^{-1} \text{Mpc})^3$, which is too low. Galaxies at similar redshifts and higher exist in sufficient numbers, but their relatively low luminosity has made gathering samples that cover sufficient volume impossible with current technology. RST and *Euclid* will enable the universe at $z = 6$ to be studied with a more representative sample of galaxies than have been possible at present (small current fields of view lead to rarer brighter galaxies being missed). So far, galaxies have been found up to redshifts $z = 12$ with dropout techniques (Ellis et al. 2013). The largest samples with photometric redshifts at $z \geq 6$ are Chaikin, Tyulneva & Kaurov (2018), Yan et al. (2011), Bouwens et al. (2011), Laporte et al. (2017), Eyles et al. (2005), and Mobasher et al. (2005) which contain 72, 20, 1, 1, 2, and 1 galaxies, respectively. In a few cases spectral confirmation has been obtained by identifying the Lyman α line (refs for $z > 7.5$ galaxies; Finkelstein et al. 2013; Song et al. 2016). Between $z = 6$ and 7, the Candels survey (Pentericci et al. 2018) has found 260 spectroscopically confirmed galaxies. Narrow band imaging surveys searching directly for the Lyman α line offer a route to large samples in principle. An early example of such a search includes Tilvi et al. (2010) at $z = 7.7$ which was not able to confirm any objects. More recently, the LAGER survey (Hu et al. 2017) has followed up candidates with further spectroscopy and found a 66 per cent success rate at $z \sim 7.0$.

There are number of obstacles to the finding and use of $z > 7.6$ galaxies as a BAO probe. The most important is the actual visibility of lines that would enable spectroscopic identification. The Lyman α line is subject to multiple scatterings and absorption by dust within galaxies. It is also absorbed by neutral hydrogen content of the intergalactic medium (IGM), and Lyman α emitting galaxies at the highest redshifts have proved difficult to find. One the other hand,

at $z \sim 6.5$, Bagley et al. (2017) have found that Lyman α emitters are detectable and may lie mainly in ionized bubbles, sufficiently large for the Lyman α photons to redshift out of resonance before encountering the neutral IGM gas. At lower redshifts (e.g. $z \sim 2-3$), Lyman α emission appears to have an escape fraction of $\sim 5-10$ per cent (Gronwall et al. 2007; Francis et al. 2012) but at higher redshifts $z > 6$ it appears to climb to near 100 per cent (Cassata et al. 2011; Sobacchi & Mesinger 2015).

Interlopers are another important issue. For example there will be galaxies at redshifts $1.1 < z < 1.9$ for which Hydrogen Balmer lines will fall in the spectral window for Lyman α at redshifts ~ 8 . Intermediate redshift ($1.7 < z < 2.8$) galaxies will also contribute [O III] 5007 emission. How well RST grism spectra could be used to identify galaxies at the highest redshifts and remove interlopers is an open question.

In this paper, we carry out the simplest analysis possible, using a combination of a large dark matter simulation (Outer Rim; Li et al. 2019) and a hydrodynamic simulation (BLUETIDES; Wilkins et al. 2016a, b) to see whether BAO could be detected at redshifts $z \sim 8$ with RST's spectral coverage. We make many simplifying assumptions, each of which are likely to strongly affect the outcome, but can be tested with future work. We assume that all Lyman α emission can be seen. In other words, we suppose that the Lyman α emitting galaxies lie in ionized bubbles (Kakiichi et al. 2016; Yajima, Sugimura & Hasegawa 2018). We ignore galaxy-scale extinction of Lyman α emission. That is, we assume that the raw star formation in our hydrodynamic simulation of galaxy formation, BLUETIDES (Wilkins et al. 2016a, b) can be directly translated into Lyman α luminosity. We also do not directly address interlopers and do not do a complete analysis of spectral overlap in the grism spectra. We restrict ourselves to seeing whether mock observations of an RST grism survey under these conditions can detect the BAO feature. Our analysis therefore acts as a first step – if we find that it is possible, then these restrictions should be relaxed and tested to see whether it is worth pursuing the idea further.

The plan for the paper is as follows. In Section 2 we describe the different cosmological data sets we use. In Section 3 we show how these were used to construct mock catalogues for the RST high latitude survey. We describe details results obtained from mock catalogues in Section 4. We present comprehensive discussion and analysis of our results in Section 5.

2 DATA SETS

2.1 The BLUETIDES simulation

BLUETIDES is a hydrodynamic simulation which was designed to cover the high redshift Universe (Feng et al. 2015, 2016). The model uses the Smoothed Particle Hydrodynamics code MP-GADGET to evolve a cubical volume with 2×7040^3 particles. These particles are used to model dark matter, gas, stars, and supermassive black holes. This simulation evolved a $(400/h \approx 577)^3 \text{ cMpc}^3$ cube to redshift $z = 7.3$. The resolution, volume, and number of galaxy haloes in BLUETIDES are well matched to depth and area of current observational surveys targeting the high-redshift Universe. Additional details of the BLUETIDES simulation can be found in Feng et al. (2015, 2016); Wilkins et al. (2016a, b, 2017).

A friends-of-friends (FoF) algorithm was used to identify galaxy haloes. At the high redshifts relevant here these haloes are sufficiently accurate representations of the galaxies that would be found in the simulations if observational determinations were used. This was seen directly in Feng et al. (2016) by comparing FoF haloes with galaxy

Table 1. Simulation cosmological parameters based on assumed Λ CDM cosmological model.

Name	BLUETIDES	Outer rim
Ω_Λ	0.7186	0.7352
Ω_{Matter}	0.2814	0.2648
Ω_{Baryon}	0.0464	0.04479
h	0.697	0.71
σ_8	0.820	0.8
n_s	0.971	0.963

haloes extracted from BLUETIDES mock images and identified with the Source Extractor algorithm (Bertin & Arnouts 1996).

Feng et al. (2016) used results from the BLUETIDES simulation to find good correspondence between the predicted galaxy luminosity function and observations in the observable range $-18 \leq M_{\text{UV}} \leq -22.5$ with some dust extinction required to match the abundance of brighter objects. Additionally, Feng et al. (2016) also find good agreement between the star formation rate density in BLUETIDES and current observations in the redshift range $8 \leq z \leq 10$.

Wilkins et al. (2017) use galaxy haloes from BLUETIDES simulation to show that intrinsic and observed luminosity functions at $z \in \{8, 9, 10\}$ exhibit the fast expected build up of the galaxy population at high-redshifts. The work presented in Wilkins et al. (2017) reveals that the intrinsic luminosity function is broadly similar to the observed luminosity function at faint luminosities ($M > -20$). At brighter luminosities, the luminosity function is shown to have a stronger steepening which indicates the increasing strength of dust attenuation. In their work, Wilkins et al. (2017) also demonstrate that a Schechter function provides a good overall fit to the shape of the dust attenuated far-UV luminosity function for $z \in \{8, 9, 10\}$.

In Waters et al. (2016), it was shown that the number density of galaxies that will be observable by the RST high latitude survey should be high enough that BAO measurement at redshift $z \sim 8$ could be possible. The BLUETIDES volume being too small to make direct mock HLS observations and measurements of the BAO, it was necessary to follow up with different techniques, and this is the role of this paper.

The BLUETIDES simulation is used in this paper in two roles: First, to make some example RST grism fields to check the degree of spectral overlap between high redshift galaxies and others at similar redshifts. Secondly, we use the BLUETIDES galaxies to populate the dark matter haloes in a much larger dark matter only simulation. For the dark matter only simulation, we use the outer rim catalogue. Details of the Outer Rim simulation are described in 2.2.

2.2 The outer rim simulation

The outer rim simulation (Heitmann et al. 2019; Li et al. 2019) is a large-scale dark matter only simulation which we use to measure large-scale clustering and BAO. It was run using the Hardware/Hybrid Accelerated Cosmology Code (HACC; Habib et al. 2013, 2016), on Mira, a BG/Q system at the Argonne Leadership Computing Facility. The cosmology used is a Λ CDM model close to the best-fitting model from WMAP-7 (Komatsu et al. 2011), with parameters given in Table 1. The comoving box size of the simulation is $3000 h^{-1}$ Mpc, yielding a volume 400 times that of BLUETIDES. 1.07 trillion particles were evolved, leading to a particle mass of $m_p = 1.85 \times 10^9 h^{-1} M_\odot$. The model was run to redshift $z = 0$, but in this paper we use only outputs at $z = 7$ and above.

Haloes are identified with a friends-of-friends (Davis et al. 1985) halo finder with a linking length of $b = 0.168$.

2.3 Galaxy Lyman α flux

At the redshifts $z \sim 8$ which concern us here, we will simulate the Lyman α line luminosity of galaxies, which will lie within the RST grism spectral range (see section 3.1 below). An example of Lyman α line detection from a galaxy at these redshifts is given in the work of Lehnert et al. (2010). The redshift of the galaxy UDFy-381355395 is $z = 8.6$, and Lehnert et al. (2010) measured a Lyman α line flux of $(6.1 \pm 1.0) \times 10^{-18} \text{ erg s}^{-1} \text{ cm}^{-2}$, corresponding to a luminosity of $5.5 \pm 1.0 \pm 1.8 \times 10^{42} \text{ erg s}^{-1}$.

In BLUETIDES, the star formation rate (SFR) for each galaxy is available to us, and we translate this quantity directly into Lyman α luminosity using the following scaling equation (Cassata et al. 2011).

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

In equation (1), $L_{\text{Ly}\alpha}$ denotes the Lyman α luminosity. This conversion factor is based on a stellar population with a Salpeter initial mass function (IMF) and is accurate to within a factor of a few for a range of population ages, high mass cutoffs of stars, and metallicities (Leitherer et al. 1999). Most relevant in our case is that there is no correction for effects like dust, escape fraction, and intergalactic or circumgalactic absorption. This means that using the raw star formation rate measurements from the simulation represents a best case for the detectability of galaxies and this should be borne in mind when interpreting the results. We will return to this in Section 5. Due to the paucity of observational data at these redshifts, the magnitude of these obscuring effects is not well known, but they are likely to be significant. In the example of the galaxy mentioned in Lehnert et al. (2010), the continuum SFR was estimated to be $2\text{--}4 M_\odot \text{ yr}^{-1}$. At the same time, Lehnert et al. (2010) evaluated the SFR from the Lyman α line to be $0.3\text{--}2.1 M_\odot \text{ yr}^{-1}$, which leads to Lyman α overall escape fraction from 8 per cent – 100 per cent.

The BLUETIDES simulation volume is not sufficiently large to make mock RST survey catalogues that can be used to measure large-scale structure. We use the halo catalogues from the Outer Rim simulation to do this, painting SFRs from BLUETIDES galaxies on to correspondingly massive Outer Rim haloes. Fig. 1 shows the relationship between mass and SFR for BLUETIDES at redshift $z = 8$. We can see that there is an approximately linear relationship between $\log(M)$ and $\log(\text{SFR})$. Our SFR assignment scheme is as follows: for each Outer Rim halo, we randomly pick a BLUETIDES halo with total mass within $5 \times 10^{10} h^{-1} M_\odot$, and assign the SFR for that halo to the Outer Rim object. In this way, we account for some aspects of the scatter between SFR and halo mass, although not any spatially correlated structure. For large mass galaxies, we increase the range to make sure there are enough sample haloes in BLUETIDES.

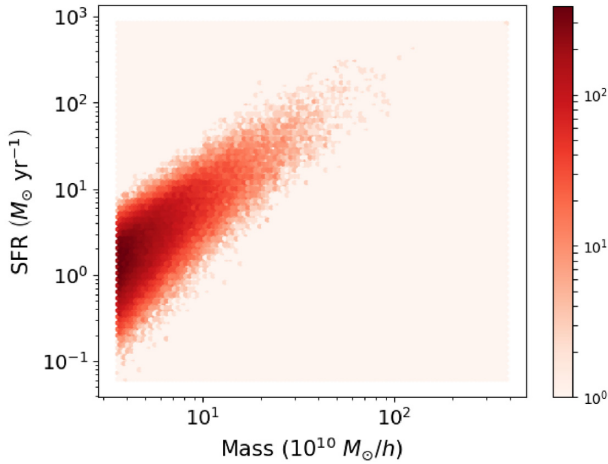
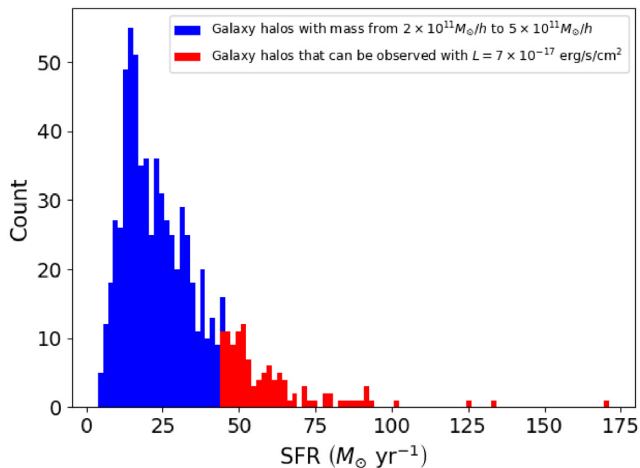
In one of the RST design documents, the 7σ line flux sensitivity for the high latitude survey is $10^{-16} \text{ erg s}^{-1} \text{ cm}^{-2}$.² We discuss RST and the flux limits on our mock surveys further in Section 3.1.1.

Fig. 2 shows the SFR distribution of BLUETIDES galaxies at $z = 8$ with mass from $2 \times 10^{11} h^{-1} M_\odot$ to $5 \times 10^{11} h^{-1} M_\odot$. We can see that when the line sensitivity is $7 \times 10^{-17} \text{ erg s}^{-1} \text{ cm}^{-2}$, then the minimum detectable luminosity is $5 \times 10^{43} \text{ erg s}^{-1}$ at $z = 8$, which

²https://wfirst.gsfc.nasa.gov/science/sdt_public/WFIRST-AFTA_SDT_Report_150310_Final.pdf

Table 2. This table shows the numbers of haloes present in, and the mean redshifts of the three quarter-mock catalogues (L7, L3.5, L1).

Line detection sensitivity	Mock 1	Mock 2	Mock 3	Mock 4
Number of haloes (before patchy reionization)				
7 erg s ⁻¹ cm ⁻²	15 484	15 344	15 280	15 144
3.5 erg s ⁻¹ cm ⁻²	37 091	37 091	37 154	36 896
1 erg s ⁻¹ cm ⁻²	194 684	194 522	195 157	194 107
Number of haloes (with patchy reionization)				
7 erg s ⁻¹ cm ⁻²	13 999	14 040	13 989	13 910
3.5 erg s ⁻¹ cm ⁻²	32 124	32 238	32 465	32 232
1 erg s ⁻¹ cm ⁻²	147 463	148 150	148 994	147 529
Mean redshifts (before patchy reionization)				
7 erg s ⁻¹ cm ⁻²	7.6474	7.6498	7.6530	7.6519
3.5 erg s ⁻¹ cm ⁻²	7.6895	7.6912	7.6962	7.6953
1 erg s ⁻¹ cm ⁻²	7.7852	7.7863	7.7884	7.7867
Mean redshifts (with patchy reionization)				
7 erg s ⁻¹ cm ⁻²	7.6645	7.6650	7.6657	7.6711
3.5 erg s ⁻¹ cm ⁻²	7.7156	7.7168	7.7197	7.7228
1 erg s ⁻¹ cm ⁻²	7.8276	7.8280	7.8294	7.8297

**Figure 1.** The variation of star formation rate with mass for BLUETIDES galaxies at $z = 8$.**Figure 2.** A histogram of star formation rates for galaxies with masses from $2 \times 10^{11} M_{\odot} h^{-1}$ to $5 \times 10^{11} M_{\odot} h^{-1}$ in redshift 8. The red bins represent galaxies that can be observed with line detection sensitivity, $L = 7 \times 10^{-17} \text{ erg s}^{-1} \text{ cm}^{-2}$.

corresponds to $40 M_{\odot} \text{ yr}^{-1}$. Only the most highly star-forming galaxies are therefore likely to be included in any BAO survey.

3 METHODOLOGY

3.1 Mock grism images

Working with grism spectroscopy is complex (McCarthy et al. 1999; Fossati et al. 2017; Colbert et al. 2018), with relatively low resolution possible, mixing of spatial and spectral structure in background and stray light, and overlap of sources. We leave full simulation of mock RST grism surveys (including foregrounds and interlopers) to other work (Gehrels et al. 2015; Colbert et al. 2018; Malhotra 2019). We will however make some extremely simple grism images using BLUETIDES simulation data, in order to assess the level of spectral overlap from galaxies at similar redshifts.

3.1.1 The RST High latitude grism survey

The RST high latitude survey (HLS) is planned to have a nominal sky coverage of 2400 deg^2 , which will greatly increase the number of galaxies available at redshifts $z \geq 8$. This survey will be conducted by RST's wide field instrument (WFI) and is expected to be a dual imaging and grism spectroscopy survey. The imaging survey will obtain direct imaging through four filters (Y106, J129, H158, and F184) and grism spectroscopy will be done at four roll angles. The filters Y106, J129, H158, and F184 are expected to enable imaging in wavelength ranges $0.927\text{--}1.192$, $1.131\text{--}1.454$, $1.380\text{--}1.774$, and $1.683\text{--}2.000 \mu\text{m}$, respectively. The imaging survey will therefore cover the wavelength range ($0.927\text{--}2 \mu\text{m}$). The grism coverage is the most relevant for our work, and the planned accessible range of wavelengths specified in the *WFIRST*-AFTA SDT report 2015 (Spergel et al. 2015) is $1\text{--}1.93 \mu\text{m}$. With these specifications, Lyman α emission at $7.22 < z < z = 14.88$ will be visible.

3.1.2 Image construction

We take the $z = 8$ snapshot of BLUETIDES and project the positions of galaxies into two dimensions using Euclidean geometry. The simulation box comoving side length of $400 h^{-1} \text{ Mpc}$ corresponds

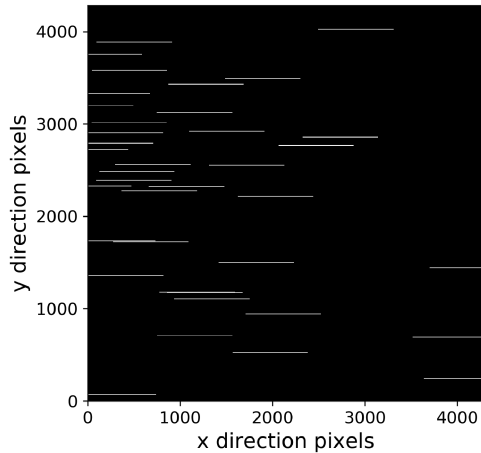


Figure 3. A simulation of a grism image with galaxies at redshift $z = 8$ from the BLUETIDES simulation. Each horizontal line denotes the spectrum of a galaxy. The brightness is proportional to the flux in the pixels.

to the comoving radial distance between $z = 7.5$ and $z = 9.5$. As the number density of galaxies decreases rapidly with redshift, our estimate of the overlap between galaxies in our crude image simulation is likely to be conservative. At each galaxy position, we plot a dispersed galaxy spectrum using the relationship between angular position and wavelength appropriate for RST ($1\text{--}2\ \mu\text{m}$). We include the effect of the redshift distortions on mock grism image by shifting the wavelengths of the spectrum of each galaxy according to the peculiar velocity along the line of sight. The galaxy spectrum data corresponds to models described in fig. 2(d) of Wilkins et al. (2013). Fig. 2(d) of Wilkins et al. (2013) illustrates the spectral energy distribution of a stellar population which has formed stars at a constant rate for 100 Myr for $z = 8.0$. For our present purposes, identical spectra for each galaxy are sufficient and we do this for ease of computation. In future work with more realistic simulations, the spectra for each individual galaxy can be determined using population synthesis models which are available and could be used (Wilkins et al. 2020).

The angular extent of the spectra perpendicular to the dispersion direction is determined by the approximate angular size of each individual galaxy. For the angular size part, we refer to Feng et al. (2016) to get M_{UV} from SFR. After this we use relationship of galaxy half-light radii versus M_{UV} shown in fig. 4 of Feng et al. (2015) to get galaxy half-life radius from M_{UV} . Using this information, we find that galaxies of mass $10^{11}\ h^{-1}\ M_{\odot}$ have an angular size of 0.11 arcsec in the grism image.

Fig. 3 shows one of the simulated grism images. The brightness shown in the picture is proportional to the flux in the pixels.

3.1.3 Spectral overlap

It is common that there will be overlap between galaxies in the grism image because some galaxies are close to each other and the direction is close to the direction between two galaxies. This will make the process of extracting galaxy spectra harder and lead to a decrease in the number of observed galaxies with reliable position information. It is therefore necessary to consider the overlap in the grism image. Usually we can mitigate this effect by taking grism images with different dispersion direction. RST will take grism images at four angles: 0° , 15° , 180° , and 195° . To estimate the upper bound of overlap, we will only consider a grism image

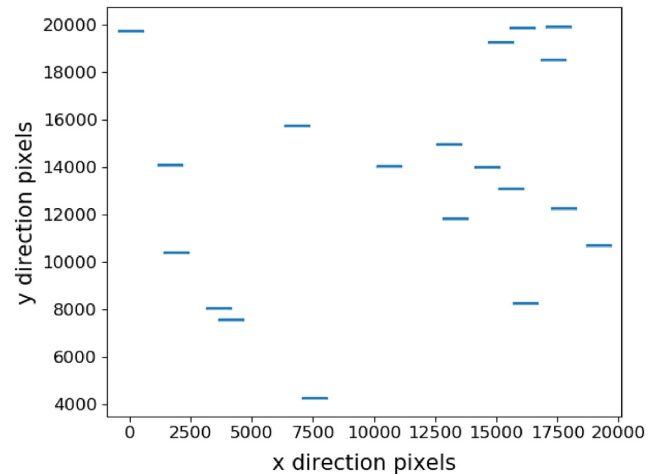


Figure 4. A grism image for a model for $L = 7 \times 10^{-17}\ \text{erg s}^{-1}\ \text{cm}^{-2}$. Each blue block represents a grism spectrum for a single galaxy. We use this plot to estimate overlap.

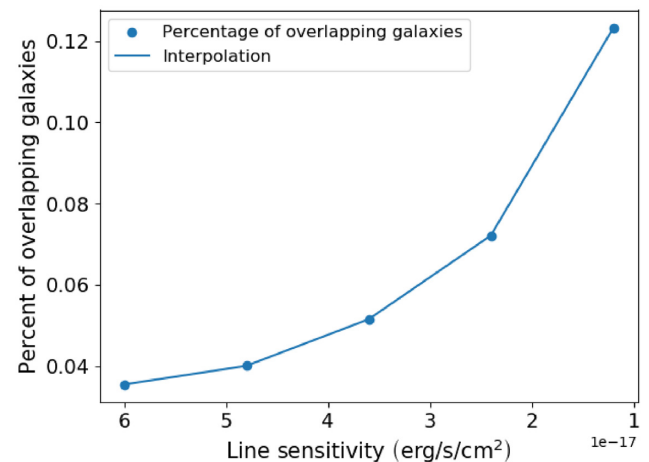


Figure 5. The percentage of overlapping galaxies as a function of line sensitivity. We notice a monotonically increasing trend in the variation of the per cent of overlapping galaxies with line sensitivity. The solid line denotes a 1D interpolation between data points.

with dispersion in one direction. Fig. 4 shows the simplified model that we use to estimate overlap. We use rectangular blocks that have the same length as grism spectrum to replace the galaxies, and find the blocks that overlap with each other and mark the galaxies as overlapping galaxies. We calculate the percentage of overlapping galaxies. Fig. 5 shows the relationship between the percentage of overlapping galaxies and line sensitivity. We can see the percentage increase when line sensitivity decreases, with the percentage around 7 per cent at $L = 3.5 \times 10^{-17}\ \text{erg s}^{-1}\ \text{cm}^{-2}$ (we will see later that this is around the level that accurate BAO fitting is possible). We calculate the upper bound of overlap by neglecting multiple dispersion directions and regarding partial overlap as full overlap, so the real overlap should be less than this.

We note of course that here we are only treating overlap from galaxies in the same high redshift range that we are interested in. These will be greatly outnumbered by a much higher density of foreground galaxies from below $z = 7.4$ in an actual survey. We leave

the problem of overlap from these galaxies and also non-overlapping interlopers (galaxies with other spectral lines in the same wavelength range as Lyman α at high redshifts) to future work.

3.2 Mock survey data

The data we use from the Outer Rim simulation covers snapshots from $z = 7.4$ to $z = 10$. Our mock RST HLS data, will cover four $3000 \times 3000 \times 700 h^{-3} \text{Mpc}^3$ cuboids drawn from the $3000^3 h^{-3} \text{Mpc}^3$ simulation volume. This corresponds to a sky area of $4 \times 570 = 2280 \text{deg}^2$, approximately equal to the total sky coverage of the RST HLS. The redshift coverage in the line of sight dimension is $z = 7.4$ to $z = 10$, with the survey volume covering $4 \times 6.3 = 25.2 \text{comoving} (\text{Gpc } h^{-1})^3$. We use the flat sky approximation for simplicity, and the conversion from distance to angle above refers to the edge of the volume, at $z = 10$. Because the whole of the Outer Rim simulation is necessary to produce a single mock RST HLS, we will analyse the four separate cuboids drawn from Outer Rim separately, as quarter mocks. The scatter between the results from the different quarter mocks will give us a crude estimate of cosmic variance.

In order to include redshift evolution in the mock surveys, we choose a cuboid with fixed position in different snapshots, and calculate the comoving radial distance of each redshift. We then extract the corresponding slice in the cuboid from different snapshots according to the relevant comoving radial distance. Stacking all the slices in order of redshift yields the mock survey. The redshift evolution is discrete rather than a true light cone – we use redshifts at $z = 7.4, 7.7, 7.9, 8.4, 8.6, 8.8, 8.0, 9.2, 9.4, 9.6, 9.8, \text{ and } 10.0$. We include the effect of redshift distortions in the mock surveys using the peculiar velocities of the simulated galaxies.

As the Outer Rim simulation covers a $3000 \times 3000 \times 3000 h^{-3} \text{Mpc}^3$ cube, we are able to create four mock cuboids from the Outer Rim data, using different parts of the volume. Each of these covers $\sim 1/4$ of the RST HLS area, so that there are therefore four approximately independent quarter-mock surveys to use for baryon acoustic peak fitting.

As stated above, we assign BLUE TIDES galaxy luminosities to Outer Rim haloes according to their mass. We make several sets of mock surveys corresponding to different galaxy detection thresholds (different line sensitivities):

(1) L7 quarter mocks. The 2015 *WFIRST* technical report (Casertano et al. 2015) estimates that with current design parameters, the RST HLS grism survey will have a 7σ H α line detection sensitivity of $6 - 12 \times 10^{-17} \text{erg s}^{-1} \text{cm}^{-2}$ for sources with effective radius $< 0.3 \text{arcsecs}$. This is for H α spectral line ($\lambda_{\text{rest}} = 6563 \text{\AA}$) in the redshift range $z = 1.05$ to $z = 1.88$. In Spergel et al. (2015), a value of $1 \times 10^{-17} \text{erg s}^{-1} \text{cm}^{-2}$ at 7σ is quoted, which is in the same range. In our analysis, we use this value, but assume a 5σ detection limit, which results in a lower limit to the detectable line flux of $7 \times 10^{-17} \text{erg s}^{-1} \text{cm}^{-2}$. As the RST survey strategies are as yet unfinalized, we assume that there could be some flexibility, both in the selection of the fiducial sensitivity, and also possibilities of deeper integrations over part of the survey area.

We assume that this sensitivity can be applied to the Lyman α line over the same observed range of wavelengths. The line sensitivity for resolved sources is approximately a factor of two worse, but we expect the galaxies to be unresolved. For example, Feng et al. (2015) find half-light radii of $0.5 \sim 1 \text{kpc}$ for the most massive galaxies in BLUE TIDES at redshift $z = 8$, which corresponds to $0.1\text{--}0.2 \text{arcsec}$. For our first mocks, those closest to the RST design sensitivity, the

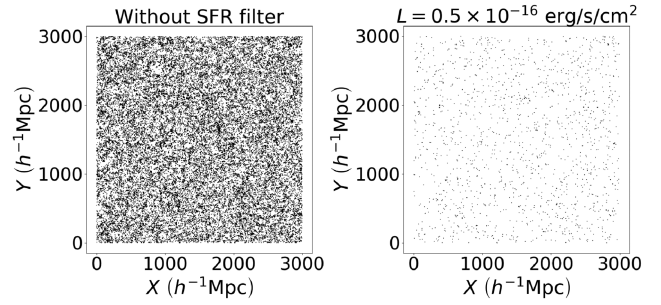


Figure 6. The figure on the left is cross-section of original data. The figure on the right is the cross-section for same position in simulated data with $L = 6 \times 10^{-17} \text{erg s}^{-1} \text{cm}^{-2}$.

mimimum Lyman α line flux is $7 \times 10^{-17} \text{erg s}^{-1} \text{cm}^{-2}$. We name these the L7 quarter mocks in what follows.

(2) L3.5 quarter mocks. Unfortunately, even with the above parameter choices, which are in themselves relatively optimistic regarding the capabilities of RST, we shall see below that the possibility of detecting BAO is marginal. We therefore make another set of quarter mocks with a 50 per cent less restrictive line detection sensitivity limit, $3.5 \times 10^{-17} \text{erg s}^{-1} \text{cm}^{-2}$. This sensitivity limit could in principle be reached by increasing the integration time and lowering the sky area covered by the HLS. We name these the L3.5 quarter mocks.

(3) L1 quarter mocks. The final set of mocks we use for analysis have a line detection sensitivity of $1 \times 10^{-17} \text{erg s}^{-1} \text{cm}^{-2}$ seven times better than the planned RST design. We create these (L1) quarter mocks in order to explore BAO measurement when there is extremely good sampling.

The numbers of halos present in, and the mean redshifts of the three quarter-mock catalogues (L7, L3.5, L1) are given in Table 2.

In Fig. 2 we show galaxies with SFR more than $44.20 \text{yr}^{-1} M_{\odot}$ account for 15.9 per cent of the total galaxy counts. We assume that if a galaxy mass falls between $2 \times 10^{11} h^{-1} M_{\odot}$ and $5 \times 10^{11} h^{-1} M_{\odot}$, the probability of it being able to be detected is 15.9 per cent. We calculate this probability for different mass ranges by using data from BLUE TIDES, and match galaxies from simulated cuboid to decide which galaxies will be detected in the mock survey. Fig. 6 shows the difference after we apply this selection. Only 2 per cent galaxies is kept in the right figure compared with left one. And, Fig. 7 shows the relationship between probability and mass. The maximum mass in redshift 9 and 10 is much smaller than other redshift, but considering the maximum mass of the Outer Rim is $1.30 \times 10^{12} h^{-1} M_{\odot}$ in redshift 9.2 and $6.7 \times 10^{11} h^{-1} M_{\odot}$ in redshift 10. Fig. 7 is enough to cover the mass range in the Outer Rim. And Fig. 8 shows the distribution of galaxies along the z -axis under different line sensitivity. There are several sudden changes in the distribution, the reason is that we use snapshots from discrete redshift to construct mock data so there is discontinuity between two redshift slices.

3.3 Measuring the correlation function

We use the Landy Szalay estimator (Landy & Szalay 1993) to compute anisotropic two point correlation functions $\hat{\xi}(s)$ of galaxies in our data sets. We prefer the Landy Szalay estimator, since its performance has been proved to be better than other comparable two point correlation functions (Davis & Peebles 1983; Hamilton 1993) at large scales (Pons-Bordería et al. 1999; Kerscher, Szapudi & Szalay

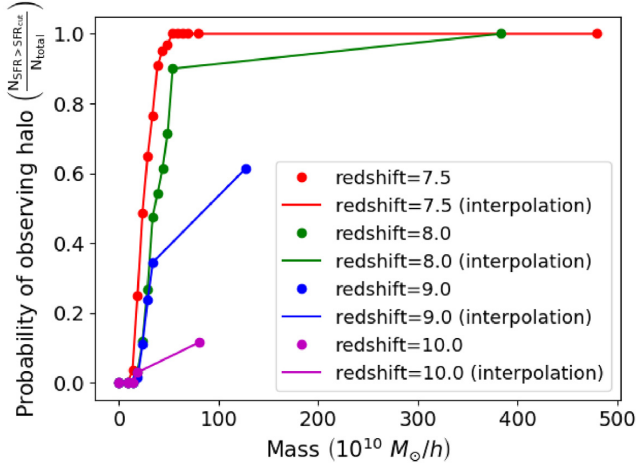


Figure 7. The relationship between the probability of a galaxy being observed and its mass. The solid lines are 1D interpolations between data points.

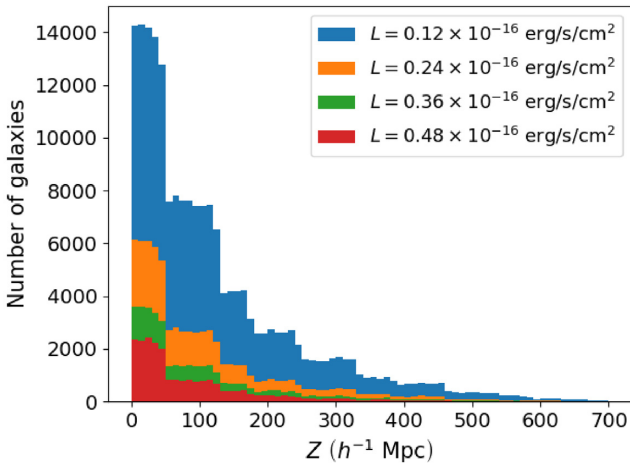


Figure 8. Here we show the distribution of galaxies along the z -axis for different line sensitivities.

2000). Equation (2) gives the formula for the calculation of the Landy Szalay estimator.

$$\hat{\xi}_{LS}(s, \mu) = \frac{DD(s, \mu) - 2DR(s, \mu) + RR(s, \mu)}{RR(s, \mu)}. \quad (2)$$

In equation (2), $\mu = \cos \theta$ (where θ is the angle between the line of sight and the pair separation vector for any given pair), $DD(s, \mu)$ depicts the pair count of galaxies with pair separation distance s and orientation μ . The symbol $DR(s, \mu)$ represents the cross-pair counts between galaxies and randoms which have pair separation s and angle θ , and $RR(s, \mu)$ is the number of pairs for a random distribution. We have computed the correlation functions as a function of s and μ as an additional check on our calculations, to see that the clustering is behaving as expected. We do not present or use the 2D results further because the use of the full 2D two point correlation functions in analyses is computationally expensive. One can use isotropized correlation functions ($\tilde{\xi}_f(s)$) to circumvent this. To obtain isotropized correlation functions, one uses specific kernels to marginalize out the μ coordinate in the two point correlation function. In Hamilton (1993) it was shown that one can use Legendre polynomials as a kernel ($f(\mu) = P_\ell(\mu)$) to find the isotropic correlation functions $\tilde{\xi}_\ell(s)$.

Different values of ℓ will lead to isotropic correlation functions of different orders. In our research, we use isotropic correlation functions with Legendre polynomials and $\ell = 0$ to obtain isotropized monopoles: ($\tilde{\xi}_0(s)$). For completeness, equation (3) shows the process by which anisotropic correlation function $\xi_\ell(s)$ is obtained from the marginalization of anisotropic two correlation function ($\hat{\xi}(s, \mu)$).

$$\tilde{\xi}_\ell(s) = \frac{2\ell + 1}{2} \int_{-1}^1 \hat{\xi}_{LS}(s, \mu) P_\ell(\mu) d\mu. \quad (3)$$

In equation (4), we show how we can compute $\tilde{\xi}_0(s)$ from equation (3).

$$\begin{aligned} \tilde{\xi}_0(s) &= \left(\frac{1}{2}\right) \int_{-1}^1 \hat{\xi}_{LS}(s, \mu) d\mu \\ &\approx \left(\frac{1}{2}\right) \sum_j \Delta\mu_j \hat{\xi}_{LS}(s, \mu_j). \end{aligned} \quad (4)$$

In the work presented in this paper, we have used 100 bins in μ in the computation of all 2D two-point correlation functions. Also, all the multipoles ($\xi_{0,2}(s)$) that we compute for BLUETIDES and Outer Rim galaxies have evenly spaced bins of width $3 h^{-1}$ Mpc in s . Apart from their use as an internal check here, we reserve the presentation of results for multipoles higher than the monopole for future work.

3.4 Fitting the correlation function

For fitting of the correlation functions, we use techniques which are outlined in Anderson et al. (2012), Xu et al. (2012, 2013), Vargas-Magaña et al. (2014), Ansarinejad & Shanks (2018), Sarpa et al. (2019). We briefly describe the methods that we have used in this section.

We use the following fiducial form to fit space correlation monopoles $\xi_0(s)$ over the range $33 h^{-1}$ Mpc $< s < 138 h^{-1}$ Mpc with bin sizes of $s = 3 h^{-1}$ Mpc.

$$\xi^{\text{fit}}(s) = B^2 \xi_m(\alpha s) + A(s). \quad (5)$$

Here, the symbol ‘ s ’ represents the pair separation between galaxies in redshift space. The quantity B is a multiplicative constant bias which adjusts the amplitude of the correlation functions (ξ), and lets the model account for any unknown large-scale bias. The term

$$A(s) = \frac{a_1}{s^2} + \frac{a_2}{s} + a_3 \quad (6)$$

is included so that the fitting model can marginalize over broadband effects due to scale-dependent bias, redshift space distortions, and errors during our assumption of the fiducial cosmology. The parameters a_1, a_2, a_3 in equation (6) are nuisance parameters. The choice of number of parameters in $A(s)$ in our work is inspired by Anderson et al. (2012), Xu et al. (2012, 2013), Vargas-Magaña et al. (2014), Ansarinejad & Shanks (2018), Sarpa et al. (2019). The inclusion of $A(s)$ can help us to mitigate the effects of assumption of a wrong cosmological model. At the same time, $A(s)$ can also be thought of as an effectual description of mode coupling not affecting the BAO scale but biasing its measurement if not considered properly (Crocco & Scoccimarro 2008).

The scale dilation parameter α measures the position of the BAO peak of the data relative to the model. One can also think of the parameter α as the extent to which the acoustic peak in the data is shifted relative to the model. This isotropic shift in the positions of the BAO peak in the data and the model occurs due to non-linear structure growth. A numerical definition of α can be obtained in

equation (7).

$$\alpha = \frac{D_V(z)}{r_d} \frac{r_{d,\text{fid}}}{D_{V,\text{fid}}(z)} \quad (7)$$

In equation (7) r_d represents the sound horizon at the drag epoch. The superscript ‘fid’ denotes the fiducial cosmology (given in Table 1.) A value of $\alpha < 1$ would represent a shift towards larger scales, whereas a value of $\alpha > 1$ denotes a shift towards smaller distance scales. The volume averaged distance D_V is related to the Hubble constant H , the angular diameter distance D_A , the speed of light c , and the redshift z by the following equation:

$$D_V = \left[(1+z)^2 c z \frac{D_A^2}{H} \right]^{1/3}. \quad (8)$$

The template power spectrum which we use is defined in the below mentioned relation.

$$P_m(k) = [P_{\text{lin}}(k) - P_{\text{smooth}}(k)] e^{-k^2 \Sigma_{\text{NL}}^2 / 2} + P_{\text{smooth}}(k). \quad (9)$$

In equation (9), $P_{\text{lin}}(k)$ represents the actual linear power spectrum at $z = 0$. The term $P_{\text{smooth}}(k)$ denotes the de-wiggled limit where the BAO feature is removed. More details of $P_{\text{smooth}}(k)$ can be found in Eisenstein & Hu (1998). The term Σ_{NL} helps to damp the BAO features in the linear power spectrum $P_{\text{lin}}(k)$. This parameter also helps us to account for the broadening and shift of the BAO feature due to non-linear structure evolution. We use $\Sigma_{\text{NL}} = 0 h^{-1} \text{Mpc}$ in our work, i.e. $P_m(k) = P_{\text{lin}}(k)$. Choice of this value of Σ_{NL} is justified given the high redshifts that we consider ($z_{\text{OuterRim}} = 9.4$) and the correspondingly low value of the growth factor, $g = 0.123$. More details of this are given in Section 3.5.

We use CAMB (Lewis & Bridle 2002) to compute the power spectrum $P_m(k)$ using the same cosmological model that is adopted in the Outer Rim simulation.

The model correlation function ξ_m is the Fourier transform of the the template power spectrum given in equation (9).

Xu et al. (2013) showed that variations in the bias like term B do not affect the position of the BAO peak in monopoles. At the same time, the parameter α , which gives us a quantification of the acoustic scale, affects the position of the BAO peak. As such, α is the parameter that we are most interested in obtaining from our fits. The models that we use are non-linear in α , so we can nest a linear least-squares fitter inside a non-linear fitting routine. The best-fitting value of α is obtained by minimizing the χ^2 goodness-of-fit indicator (non-linear fitter). To obtain an optimum value of α , we compute χ^2 and shift models in the range $0.8 < \alpha < 1.2$ with intervals of $\alpha = 0.01$. The choice of range within which values of α are examined, viz. [0.8, 1.2], is inspired by Vargas-Magaña et al. (2014) and Ansarinejad & Shanks (2018).

The best-fitting value of α is obtained by minimizing the χ^2 goodness-of-fit indicator given in equation (10).

$$\chi^2(\alpha) = [\xi^{\text{mock}} - \xi^{\text{fit}}(\alpha)]^T \Sigma^{-1} [\xi^{\text{mock}} - \xi^{\text{fit}}(\alpha)]. \quad (10)$$

In equation (10), $\xi^{\text{fit}}(\alpha)$ represents a model vector at a given value of α , ξ^{mock} denotes the correlation function measured from mock cuboids corresponding to a given line detection sensitivity, and Σ represents a diagonal covariance matrix which is obtained from all mock cuboids in the considered line detection sensitivity. In an ideal case, one would have preferred to use a variance-covariance matrix from a large number (preferably ~ 100 or higher) of mock catalogues. But, in our work, we are dealing with high redshifts ($z > 7.6$) and we only have four mock cuboids for each of the three line detection sensitivities. We use these cuboids to construct separate covariance matrices for each line detection sensitivity. Hereafter, we

shall use the symbol α^* to denote the best-fitting values of the scale dilation parameters α . To obtain robust results for best-fitting values of scale dilation parameters and best-fitting model vectors ($\xi^{\text{fit}}(\alpha^*)$), we construct diagonal covariance matrices from mocks in two steps. In the first step, we compute a diagonal covariance matrix (say Σ_0) from cuboids which correspond to the same line detection sensitivity. (Separate Σ_0 are constructed for catalogues which correspond to without patchy reionization and with patchy reionization.) In the second step, we construct a new diagonal covariance matrix (say Σ) by using a moving average (with a window size of three) over the diagonal elements of Σ_0 . That is to say,

$$\Sigma^{\{i,i\}} = \frac{[\Sigma_0^{\{(i-1),(i-1)\}} + \Sigma_0^{\{i,i\}} + \Sigma_0^{\{(i+1),(i+1)\}}]}{3}. \quad (11)$$

To compute $\Sigma^{\{0,0\}}$, we use equation (12)

$$\Sigma^{\{0,0\}} = \frac{[\Sigma_0^{\{0,0\}} + \Sigma_0^{\{1,1\}}]}{2}. \quad (12)$$

Similarly, to find the last diagonal element of Σ , i.e. $\Sigma^{\{N,N\}}$ we use equation (13). Here N denotes the index of the largest distance scale ($138 h^{-1} \text{Mpc}$) in our fitting range.

$$\Sigma^{\{N,N\}} = \frac{[\Sigma_0^{\{(N-1),(N-1)\}} + \Sigma_0^{\{N,N\}}]}{2}. \quad (13)$$

The approach highlighted in equations (11), (12), and (13) help in getting better agreement in positions of BAO peaks in mock and best-fitting model correlation functions.

Detailed description of the BAO fitting methodologies can also be found in Anderson et al. (2012), Xu et al. (2012, 2013), Vargas-Magaña et al. (2014).

3.5 Semi-analytic model for patchy reionization

We follow the semi-analytic model outlined in Battaglia et al. (2013) to construct reionization fields from our density fields. The density fluctuation field is supposed to be at the midpoint of reionization, which is $z = 9.5$ for BLUE TIDES. So, we use data from Outer Rim simulation which is closest to $z = 9.5$ (i.e. $z_{\text{OuterRim}} = 9.4$). The fluctuation field for haloes from the Outer Rim catalogue at $z_{\text{OuterRim}} = 9.4$ is defined as

$$\delta_{\text{halo}}(\mathbf{x}) \equiv \frac{\delta_{\text{halo}}(\mathbf{x}) - \bar{\delta}_{\text{halo}}}{\bar{\delta}_{\text{halo}}}. \quad (14)$$

Here, $\bar{\delta}_{\text{halo}}$ is the mean density of haloes. The required linear bias, b is obtained by using equation (15).

$$b = \frac{\sigma_8}{g} \cdot \sqrt{\frac{\xi_0^{\xi z=9.4 : \text{OuterRim}}}{\xi_0^{\xi z=0.0 : \text{CDM}}}}. \quad (15)$$

In equation (15), g is the growth factor between $z = 9.4$ and $z = 0.0$, i.e. $g = 0.123$. The symbol $\xi_0^{\xi z=9.4 : \text{OuterRim}}$ denotes the correlation function monopole computed from Outer Rim haloes at $z = 9.4$. The symbol $\xi_0^{\xi z=0.0 : \text{CDM}}$ represents the monopole from cold dark matter (CDM) at $z = 0.0$. We use a value of $\sigma_8 = 0.8$. After the aforementioned steps, one can obtain the matter overdensity (δ_m) from equation (16)

$$\delta_m(\mathbf{x}) = \delta_{\text{halo}}(\mathbf{x})/b. \quad (16)$$

Before we outline the steps for the computation of the reionization redshift field, $z_{\text{RE}}(\mathbf{x})$, it is important to define the fluctuation field

$\delta_z(\mathbf{x})$.

$$\delta_z(\mathbf{x}) \equiv \frac{[1 + z_{\text{RE}}(\mathbf{x})] - [1 + \bar{z}]}{1 + \bar{z}}. \quad (17)$$

In equation (17), \bar{z} is the mean value of the z_{RE} field. At the same time, \bar{z} sets the midpoint of reionization which is $z = 9.4$ for Outer Rim data.

From here, we can get the reionization redshift field, $z_{\text{RE}}(\mathbf{x})$, by following the below mentioned steps.

(1) We Fourier transform the density fluctuation field δ_m from Outer Rim simulation at the midpoint of reionization (i.e. $z = \bar{z} = 9.4$) and obtain $\tilde{\delta}_m(k)$.

(2) We multiply the field $\tilde{\delta}_m(k)$ with the bias factor given in equation (18) to obtain $\tilde{\delta}_z(k)$.

$$b_{\text{mz}}(k) = \frac{b_0}{(1 + k/k_0)^\alpha}. \quad (18)$$

The above-mentioned bias factor uses three parameters, viz. b_0 , which is the bias amplitude on the largest scales, k_0 , which is the scale threshold and α , which is the asymptotic exponent. For the three parameters b_0, k_0, α , we use the best-fitting values which were used in Battaglia et al. (2013). More specifically, we use values of $b_0 = 0.593$, $k_0 = 0.185 \text{ Mpc}^{-1}h$, and $\alpha = 0.564$ in our work. Since we do not focus much on small distance scales in our work, we ignore the top hat filters that are used in Battaglia et al. (2013).

(3) We inverse Fourier transform $\tilde{\delta}_z(k)$ to real space to get the δ_z field.

(4) Now, we can convert the δ_z field to z_{RE} using equation (17).

Once we have the $z_{\text{RE}}(\mathbf{x})$ field, we can use that to modulate the galaxy distribution. Basically, we can use the same grid for all snapshots, and if a galaxy is in a cell that has reionized by that snapshot redshift, we treat it as being visible. If a galaxy is in a cell which has not reionized (these will be the lower density cells, based on how the patchy reionization model works), then we say that it won't be visible, and we remove it from the sample. As with the previous mock catalogues, we include the effect of redshift-space distortions in the galaxy positions: we use real space position to locate the galaxy in the reionization field, and then position the visible galaxies in redshift space.

We follow the following prescription to distinguish between lower and higher density cells. We use a Gaussian filter with $\sigma = 10 h^{-1} \text{ Mpc}$ to smooth the $z_{\text{RE}}(\mathbf{x})$ field. In the smoothed $z_{\text{RE}}(\mathbf{x})$ field, any cell whose value is less than the mean value of the smoothed $z_{\text{RE}}(\mathbf{x})$ field is considered as a low density cell. Similarly, any cell in the smoothed $z_{\text{RE}}(\mathbf{x})$ field whose value is bigger than the mean is considered as a higher density cell. Fig. 9 shows a slice ($r_z = 1500 h^{-1} \text{ Mpc}$) of the three dimensional grid of cells with ionized and non-ionized regions which is obtained from the aforesaid method. We compare all snapshots (redshifts) with this grid of high and low density cells and construct a new catalogue of galaxies which correspond to scenarios after patchy reionization. In Fig. 10, we see slices (at $r_z = 1500 h^{-1} \text{ Mpc}$) from Outer Rim data at $z = 7.4$ with and without patchy reionization.

We measure correlation function monopoles in these galaxy catalogues which correspond the cases with and without patchy reionization. In the non-patchy case, we assume that the intervening intergalactic medium does not affect the visibility of the Lyman α line from galaxies (i.e. the universe is optically thin at the time of Lyman α emission). Results of monopoles and the ensuing BAO fitting are presented in Section 4.

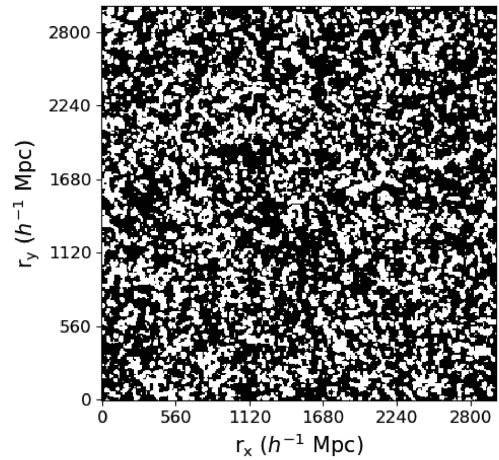


Figure 9. This plot shows a slice at $r_z = 1500 h^{-1} \text{ Mpc}$ obtained from the 3D grid of cells with ionized and non-ionized regions determined from the reionization redshift field, $z_{\text{RE}}(\mathbf{x})$. The slice thicknesses in this 3D grid are $10 h^{-1} \text{ Mpc}$. The white patches correspond to regions which are ionized at $z = 9.4$, while the black patches denote non-ionized regions.

4 RESULTS

We show the fitting results of correlation functions from L7, L3.5, and L1 quarter mocks for scenarios with and without patchy reionization. The L7, L3.5, and L1 represent different observational sensitivities (with L7 being the fiducial RST HLS choice.) For each quarter mock (and for each scenario), we have results for the variation of χ^2 with scale dilation parameter α for the four cuboids. At the same time, we show monopoles computed from the Landy Szalay estimator and best-fitting monopoles for each cuboid in each quarter mock (for each of the two aforementioned scenarios). The best-fitting monopoles correspond to the best-fitting value of α (i.e. α^*). The error on the angular diameter distance to the redshift $z = 7.7$ Universe is given by the error on the mean of the α^* values from the four quarter mocks.

4.1 L7 mocks

The top four panels in Fig. 11 show plots of variation of χ^2 with α for the four mock cuboids in L7 quarter-mock catalogue which correspond to distribution of haloes without patchy reionization. The plots in the bottom four panels of Fig. 11 show monopoles computed from Landy Szalay estimators and best-fitting monopoles ($\xi(\alpha^*)$) for the aforementioned mock cuboids. For the cuboids which correspond to distribution of haloes without patchy reionization, the minimum values of χ^2 are 27.41, 31.32, 28.17, 45.86. For these cuboids, the best-fitting values of α are 1.08, 1.16, 0.86, 1.08, respectively. Consequently, for the cuboids from L7 quarter mocks the mean value of minimum χ^2 is 33.19, and the mean value of α^* is 1.04. We estimate the fractional error on the distance to $z = 7.8$ (the mean galaxy redshift) by computing the sample standard deviation of the α^* results for the four quarter mocks, and then dividing by $\sqrt{4}$ to give the error on the mean. The fractional error computed this way is 0.064, i.e. a distance error accurate to 6.4 per cent. We can see from the bottom panel of Fig. 11 however that one of the BAO peaks is not visibly detected.

One thing that is worth noting at this point is the fact that for a given mock catalogue (L7 or L3.5 or L1), plots at corresponding positions in the figures which illustrate results with and without patchy reionization are results gotten from the same cuboid. That is, if the leftmost and topmost subplot of Fig. 11 corresponds to Cuboid

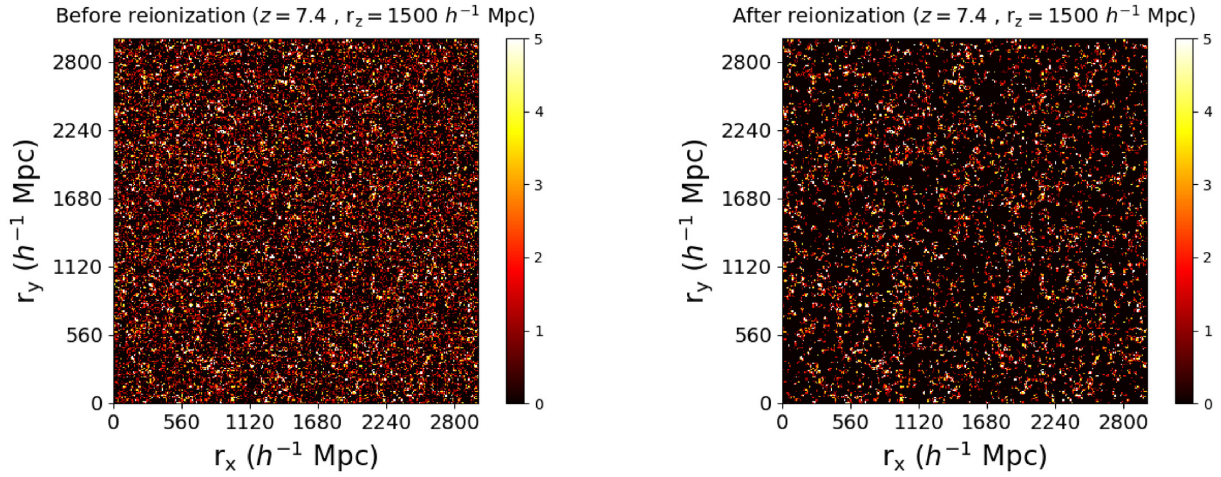


Figure 10. The plot on the left-hand side shows a slice ($r_z = 1500h^{-1}$ Mpc) of Outer Rim data at $z = 7.4$ before patchy reionization. In the right-hand side, we show a slice ($r_z = 1500h^{-1}$ Mpc) of Outer Rim data at $z = 7.4$ after reionization. The plot on the right-hand side is obtained after Outer Rim data at $z = 7.4$ is analysed using the grid of ionized and non-ionized regions described in section 3.5 and illustrated in Fig. 9.

1, then leftmost and topmost subplot of Fig. 12 also corresponds to Cuboid 1. The same holds for all mock catalogues (L7 or L3.5 or L1).

The left four panels in Fig. 12 illustrate the variations of χ^2 with α for the four mock cuboids in L7 quarter-mock catalogue which correspond to distribution of haloes after applying patchy reionization. In the right four panels in Fig. 12, we show monopoles computed from Landy Szalay estimators and best-fitting monopoles ($\xi(\alpha^*)$) for these four mock cuboids. The cuboids from L7 quarter-mock catalogue which correspond to distribution of haloes with patchy reionization have the following minimum χ^2 values: 24.41, 26.68, 34.96, 26.59. These cuboids have best-fitting values of α as 1.04, 0.81, 1.05, 1.19, respectively. As such, the mean value of minimum χ^2 for these cuboids is 28.16, and the mean value of α^* is 1.02. Again, as for the previous figure, one of the BAO peaks (in the top right-hand panel) is not detected (although there is a chi-squared minimum and an estimate of α^*). From the scatter of α^* values between quarter mocks, we find the fractional error on the mean distance to be 0.079.

4.2 L3.5 mocks

The left four panels of Fig. 13 show plots of variation of χ^2 with α for the four mock cuboids in L3.5 quarter-mock catalogue which correspond to distribution of haloes without patchy reionization. In the right four panels of Fig. 13, we show monopoles computed from Landy Szalay estimators and best-fitting monopoles ($\xi(\alpha^*)$) for the aforementioned mock cuboids. For the cuboids from L3.5 quarter mock which correspond to the distribution of haloes without patchy reionization, the values of minimum χ^2 are 21.59, 20.03, 16.20, 26.42, and the values of α^* are 0.99, 1.08, 1.00, 1.12, respectively. Hence, the mean values of minimum χ^2 and α^* are 21.06 and 1.05, respectively. The error on the BAO distance measurement is 0.032.

The plots in the left four panels of Fig. 14 illustrate the variations of χ^2 with α for the four mock cuboids in L3.5 quarter-mock catalogue which correspond to distribution of haloes after applying patchy reionization. In the bottom two rows of Fig. 14, we show monopoles computed from Landy Szalay estimators and best-fitting monopoles ($\xi(\alpha^*)$) for these four mock cuboids. These cuboids have minimum values of χ^2 of 38.80, 22.52, 43.14, 32.26 and best-fitting values of

α as 0.95, 0.99, 1.02, 0.98. As such, the mean values of minimum χ^2 and α^* are 34.18 and 0.98, respectively. The error on the BAO distance measurement is 0.014.

4.3 L1 mocks

The left four panels of Fig. 15 show plots of variations of χ^2 with α for the four mock cuboids in L1 quarter-mock catalogue which correspond to distribution of galaxies visible without patchy reionization. In the right four panels of Fig. 15, we show monopoles computed from Landy Szalay estimators and best-fitting monopoles ($\xi(\alpha^*)$) for the aforementioned mock cuboids. These cuboids have the following minimum values of χ^2 : 19.49, 18.58, 11.79, 13.34. They have the following values of α^* : 1.01, 1.03, 0.99, 0.95. Consequently, the mean values of minimum χ^2 and α^* for these mock cuboids are 15.80 and 1.00, respectively. The fractional distance error is 0.015.

The left four panels of Fig. 16 illustrate the variations of χ^2 with α for the four mock cuboids in L1 quarter-mock catalogue which correspond to distribution of galaxies visible with patchy reionization. In the right four panels of Fig. 16, we show monopoles computed from Landy Szalay estimators and best-fitting monopoles ($\xi(\alpha^*)$) for these four mock cuboids. The cuboids L1 quarter-mock catalogue which correspond to distribution of galaxies after patchy reionization have the following values of minimum χ^2 : 12.33, 14.26, 15.67, 14.26. The values of α^* for these cuboids are as follows: 0.96, 1.06, 1.01, 0.96. Hence, the mean values of minimum χ^2 and α^* for these mock cuboids are 14.13 and 1.00. The fractional distance error is 0.024.

Table 3 gives summaries of our best-fitting results for all mocks in all three quarter-mock catalogues.

4.4 Analysis of errors on the BAO distance measurements

We now carry out an analysis of the BAO distance errors and their scatter, in order to check that they are statistically consistent. For the L7 quarter mock catalogues, we found the following values of best-fitting α (α^*) from the four cuboids that represent the distribution of haloes without patchy reionization: 1.08, 1.16, 0.86, 1.08. These values of α^* lead to an overall fractional error of 0.064. At the same time, we observed the following values of α^* in the four cuboids

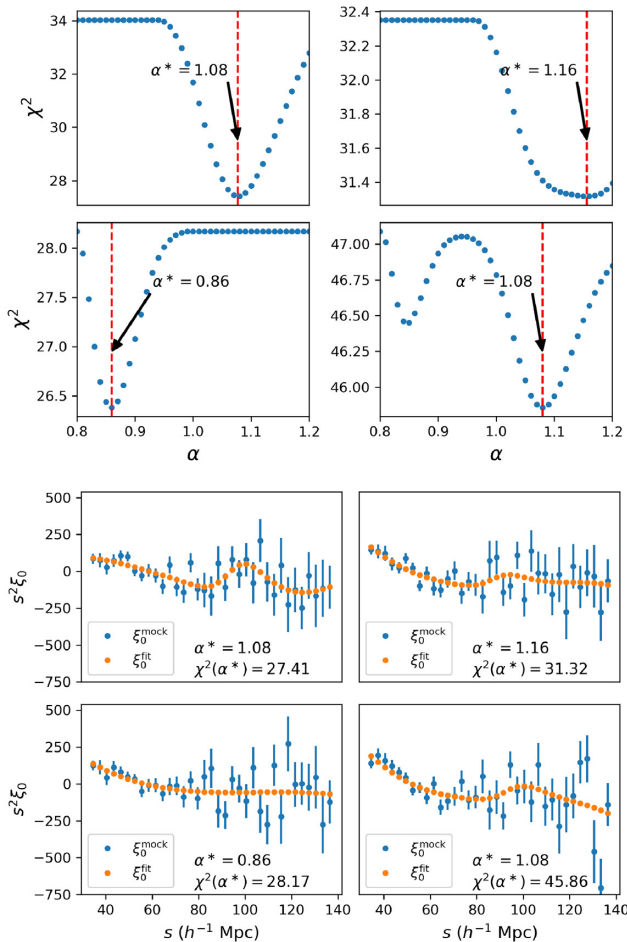


Figure 11. BAO fitting without patchy reionization: In the top two panels, we show plots of the variations of χ^2 with α for the four mock cuboids corresponding to $L = 7 \times 10^{-17} \text{ erg s}^{-1} \text{ cm}^{-2}$ (without patchy reionization). In the plots in the top four panels, the red dashed vertical lines denote the positions of the best fit α , viz. α^* for different mocks. In the figures in the bottom four panels, we show the monopoles computed from Landy Szalay estimator (ξ_0^{mock}) and the best-fitting monopoles ($\xi_0^{\text{fit}}(\alpha^*)$) obtained after fitting of ξ_0^{mock} for the four mock cuboids corresponding to $L = 7 \times 10^{-17} \text{ erg s}^{-1} \text{ cm}^{-2}$ (without patchy reionization). In the plots in the bottom panels, the blue dots represent ξ_0^{mock} while the orange dots denote $\xi_0^{\text{fit}}(\alpha^*)$.

in the L7 quarter mock catalogues with patchy reionization : 1.04, 0.81, 1.05, 1.19. These values of α^* yield an error of 0.079 on BAO distance measurement. To reconcile the scatter in the values of α^* that we see for cuboids with and without patchy reionization, we carry out Fisher’s exact test (Fisher 1922; Agresti 1992). For each category of cuboids (with reionization or without), we are working with only four cuboids. This is a classic case of data with small sample sizes. Fisher’s exact test enables us to conduct accurate statistical significance tests on small samples. In this test, our null hypothesis is that there is no significant deviation in the distributions of the values of α^* for cases with and without patchy reionization. A two-tailed Fisher exact hypothesis test for comparison of the distributions of the aforesaid values, gives us a p -value which is approximately 0.99. This p value is very high and it strongly suggests that we cannot ignore the null hypothesis.

Similarly, for L3.5 mocks, we have the following values of α^* for haloes without patchy reionization: 0.99, 1.08, 1.00, 1.12. For

the four cuboids with patchy reionization, we have: 0.95, 0.99, 1.02, 0.98. We again conduct Fisher’s exact test. Our null hypothesis in this test is that there is no significant deviation in distributions of values of α^* . From a two-tailed Fisher’s exact test, this time we get a p -value of approximately 1. This strongly suggests that there is significant evidence in support of our null hypothesis.

Finally, for L1 mocks, we get the following values of α^* for haloes without patchy reionization: 1.01, 1.03, 0.99, 0.95. When cuboids have patchy reionization, we get the following values of α^* : 0.96, 1.06, 1.01, 0.96. A two-tailed Fisher’s exact test gives us a p -value of approximately 1, again strong evidence of the fact that there is no significant deviation in distributions of values of α^* for cases with and without patchy reionization.

The results that we get from Fisher’s exact tests carried out on the mocks also suggest that we cannot make significant inferences from comparisons of scatter in values of α^* in cuboids with patchy reionization and without patchy reionization.

4.5 Linear bias between cases with and without reionization

The results that we see in Figs 12, 11, 14, 13, 16, and 15 suggest that the clustering amplitude changes somewhat significantly with and without reionization and also depends on the line strength. Equation (19) below defines the linear bias due to reionization compared to no reionization for a given cuboid. It is defined using a ratio of two galaxy correlation functions and is of course different from the usual linear bias of galaxies with respect to mass.

$$b_{\text{reion}} = \sqrt{\frac{\xi_0^{\text{reionization}}}{\xi_0^{\text{no reionization}}}}. \quad (19)$$

We tabulate the values of b_{reion} for cuboids in the L7, L3.5, and L1 mocks in Table 4. For the L7 mock catalogue, we obtain a mean b_{reion} of 1.32. We estimate the fractional errors on b_{reion} by computing the sample standard deviation of the b_{reion} results for the four quarter mocks, and then dividing by $\sqrt{4}$ to give the error on the mean. The fractional error computed this way is 0.136. In the case of the L3.5 mock catalogue, we get a mean b_{reion} of 1.24 and a fractional error of 0.089. Finally, for the four quarter mocks in L1 mock catalogue, we obtain a mean b_{reion} of 2.41 and fractional error of 0.514. It is as we might expect: in all cases the effects of patchy reionization lead to higher amplitude clustering for the observable galaxies. The fainter galaxies (the L1 mocks) are also relatively much more strongly affected than the brighter galaxies.

5 SUMMARY AND DISCUSSION

5.1 Summary

Using mock catalogues, we have investigated the prospect of utilizing the BAO feature at redshifts $z > 7.6$ as a standard ruler. We have utilized four mock catalogues, each representing approximately one quarter of the volume that will be surveyed by RST with grism spectroscopy.

We see evidence for BAO peaks in three of the four mocks when we use the planned RST HLS Lyman α line detection sensitivity of $L = 7 \times 10^{-17} \text{ erg s}^{-1} \text{ cm}^{-2}$. Hence, the BAO feature is likely detectable for this line detection sensitivity (measured in three out of four survey quadrants analysed independently). When we make mocks with better line detection sensitivities, we see evidence of BAO features in all mocks. For the standard line detection sensitivity, $L = 7 \times 10^{-17} \text{ erg s}^{-1} \text{ cm}^{-2}$, we find a root mean square

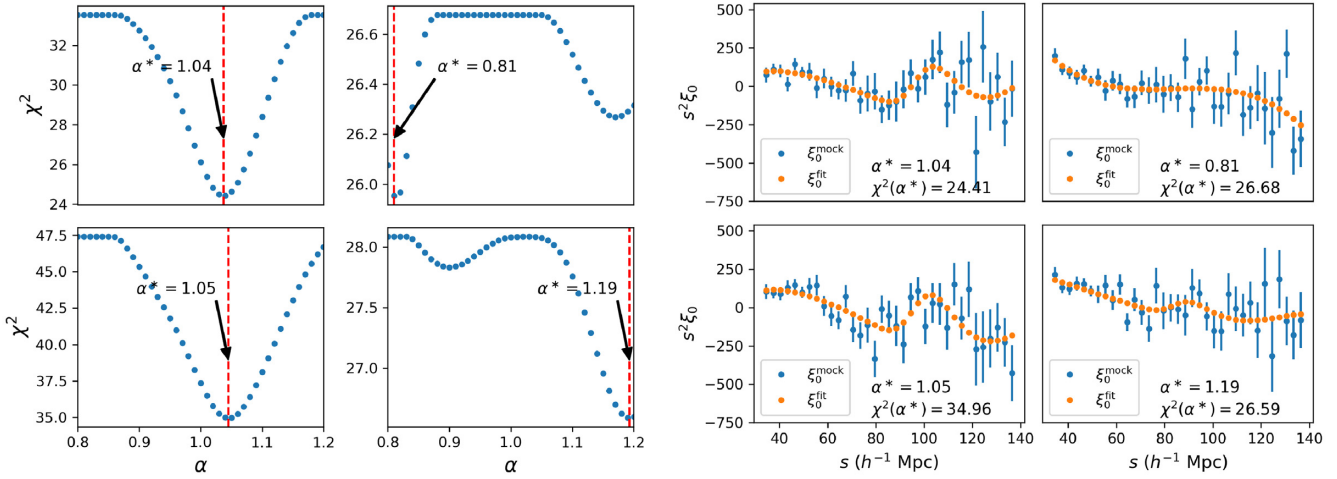


Figure 12. BAO fitting with patchy reionization: In the left four panels of this figure we show the variations of χ^2 with α for the four mock cuboids corresponding to $L = 7 \times 10^{-17} \text{ erg s}^{-1} \text{ cm}^{-2}$ (after reionization). In the plots in the left four panels, the red dashed vertical lines denote the positions of the best fit α , viz. α^* for different mocks. In the right four panels, we show the monopoles computed from Landy Szalay estimator (ξ_0^{mock}) and the best-fitting monopoles ($\xi_0^{\text{fit}}(\alpha^*)$) obtained after fitting of ξ_0^{mock} for the four mock cuboids corresponding to $L = 7 \times 10^{-17} \text{ erg s}^{-1} \text{ cm}^{-2}$ (with patchy reionization). In all the plots in the right-hand panels, the blue dots represent ξ_0^{mock} while the orange dots denote $\xi_0^{\text{fit}}(\alpha^*)$.

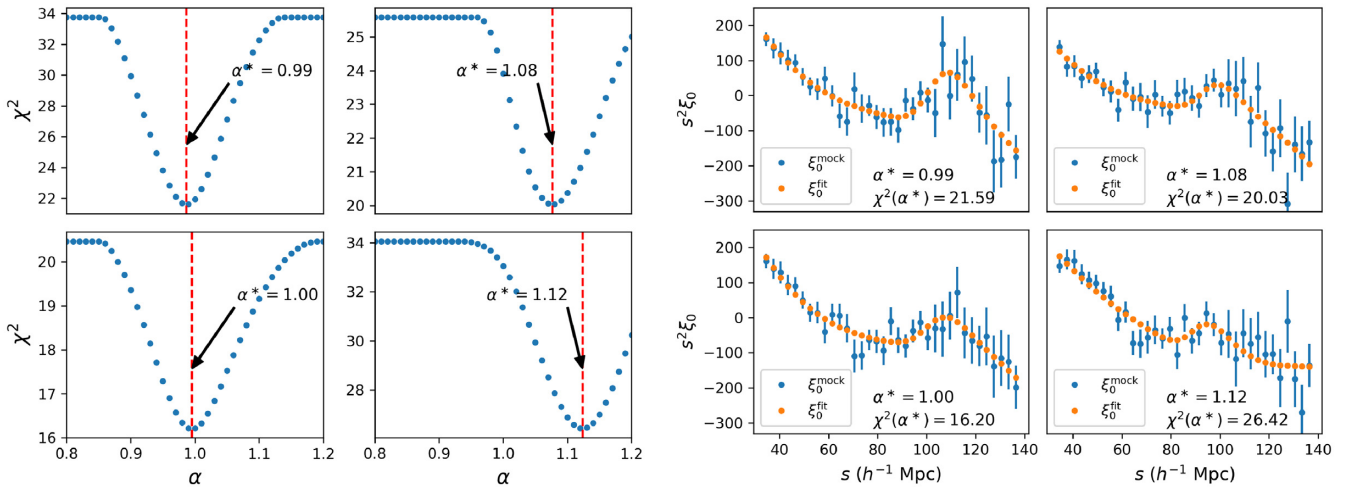


Figure 13. In the left four panels of this figure we show the variations of χ^2 with α for the four mock cuboids corresponding to $L = 3.5 \times 10^{-17} \text{ erg s}^{-1} \text{ cm}^{-2}$ (without patchy reionization). In the plots in the left four panels, the red dashed vertical lines denote the positions of the best-fitting α , viz. α^* for different mocks. In the plots in the right four panels, we show the monopoles computed from Landy Szalay estimator (ξ_0^{mock}) and the best-fitting monopoles ($\xi_0^{\text{fit}}(\alpha^*)$) obtained after fitting of ξ_0^{mock} for the four mock cuboids corresponding to $L = 3.5 \times 10^{-17} \text{ erg s}^{-1} \text{ cm}^{-2}$ (without patchy reionization). In each plot in the right-hand panels, the blue dots represent ξ_0^{mock} while the orange dots denote $\xi_0^{\text{fit}}(\alpha^*)$.

fractional distance error of 0.068 after comparison between mocks and fitted correlation function monopoles. For a factor of two fainter line detection sensitivity, $L = 3.5 \times 10^{-17} \text{ erg s}^{-1} \text{ cm}^{-2}$, we find a distance error of 0.013. We find a distance error of 0.021 for mocks with line detection sensitivity $L = 1 \times 10^{-17} \text{ erg s}^{-1} \text{ cm}^{-2}$.

5.2 Discussion

Our analysis of the use of BAO at high redshifts has included some of the relevant features and difficulties, such as non-linear clustering, redshift distortions, intrinsic galaxy line luminosities, and the effect of reionization bubbles. The results are promising, as we have summarized in Section 5.1 above. We have however left much for future work, including the visibility of the Lyman α line at these

redshifts (including the effects of dust and the Lyman α damping wing), and the effects of interlopers and low redshift foregrounds. These factors could easily derail the possibility of measuring large-scale galaxy clustering at redshift $z > 7.6$, and should be investigated in detail. On the other hand, there are examples of galaxies being detected in Lyman α emission at these redshifts, for example the $z = 7.51$ object by Tilvi et al. (2016), and $z = 7.452$ from Larson et al. (2018). Even if the Lyman α emission line itself cannot be seen, the Lyman α break can also be used to spectroscopically identify galaxies, as was done using the *HST* grism at $z = 11.1$ by Oesch et al. (2016). This could also be investigated to see whether it could enable clustering measurements at the BAO scale.

We have seen that the detectability of the BAO peak is just at the level of being possible with the fiducial flux limit of the RST HLS.

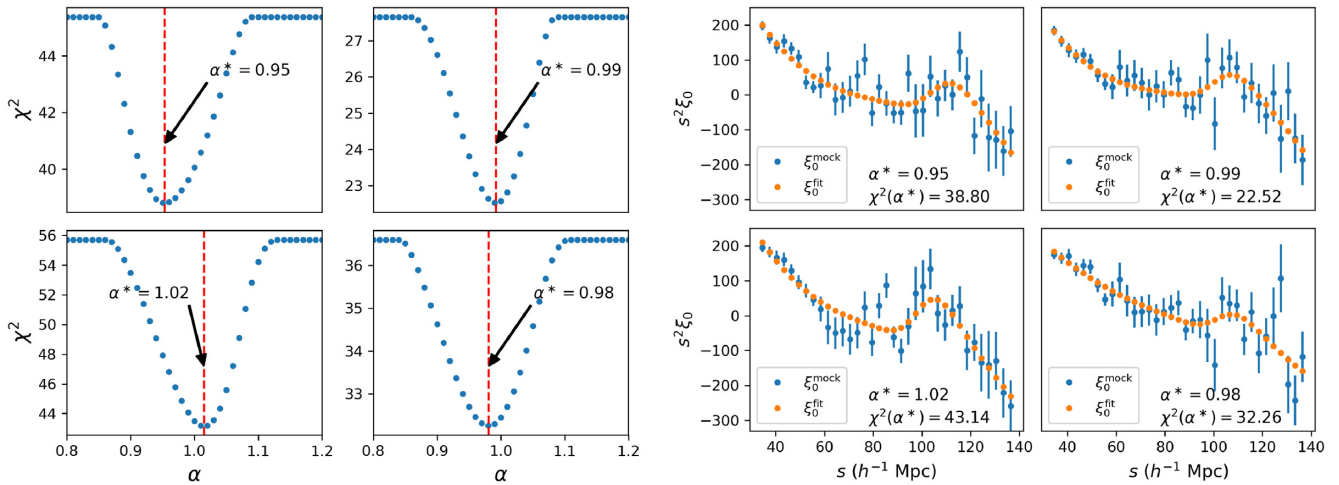


Figure 14. In the left four panels of this figure, we show the variations of χ^2 with α for the four mock cuboids corresponding to $L = 3.5 \times 10^{-17} \text{ erg s}^{-1} \text{ cm}^{-2}$ (with patchy reionization). In the plots in the left four panels, the red dashed vertical lines denote the positions of the best-fitting α , viz. α^* for different mocks. In the right four panels, we show the monopoles computed from Landy Szalay estimator (ξ_0^{mock}) and the best-fitting monopoles ($\xi_0^{\text{fit}}(\alpha^*)$) obtained after fitting of ξ_0^{mock} for the four mock cuboids corresponding to $L = 3.5 \times 10^{-17} \text{ erg s}^{-1} \text{ cm}^{-2}$ (with patchy reionization). In the plots in the right four panels, the blue dots represent ξ_0^{mock} while the orange dots denote $\xi_0^{\text{fit}}(\alpha^*)$.

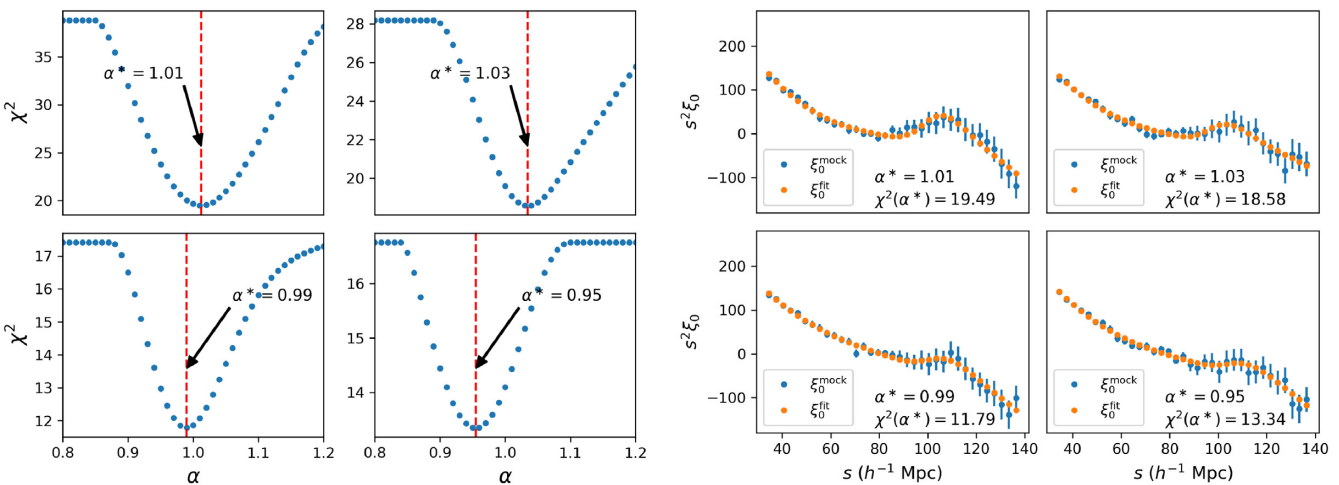


Figure 15. In the left four panels of this figure, we show the variations of χ^2 with α for the four mock cuboids corresponding to $L = 1 \times 10^{-17} \text{ erg s}^{-1} \text{ cm}^{-2}$ (without patchy reionization). In the plots in the left four panels, the red dashed vertical lines denote the positions of the best-fitting α , viz. α^* for different mocks. In the right-hand panels, we show the monopoles computed from Landy Szalay estimator (ξ_0^{mock}) and the best-fitting monopoles ($\xi_0^{\text{fit}}(\alpha^*)$) obtained after fitting of ξ_0^{mock} for the four mock cuboids corresponding to $L = 1 \times 10^{-17} \text{ erg s}^{-1} \text{ cm}^{-2}$ (without patchy reionization). In the plots in the right-hand panels, the blue dots represent ξ_0^{mock} while the orange dots denote $\xi_0^{\text{fit}}(\alpha^*)$.

To do this, we have already reduced the nominal 7σ detection limit to 5σ ($7 \times 10^{-17} \text{ erg s}^{-1} \text{ cm}^{-2}$). The exact survey design choices made when the RST mission is finalized will therefore be extremely relevant to whether the BAO measurement will be feasible. We have found that with a sensitivity factor of two better the BAO can be seen clearly in four survey quadrants. This means that it could be advantageous to make use of smaller and deeper areas.

The analysis presented in this work was based on only four quarter mocks for each of the tested line detection sensitivities. In the future, large number of mock catalogues should be generated to allow for the build up of a full covariance matrix for the correlation function. Because of the small number of mocks here, we were only able to use the diagonal elements of the covariance matrix in model fitting. As a result, the distance error estimates, which were derived from

the scatter between mocks, were necessarily conservative. In a future work with large numbers of mocks, the extra information available from the off diagonal covariance matrix elements will mean that the fits to the BAO peaks will be improved and the distance errors will be somewhat reduced. The uncertainty in the distance error will also be better estimated. In the present case, the error on the mean from only four quarter mocks will be large, and this is why the distance scale error on the L1 mocks increases to 2.4 per cent from 1.4 per cent for the L3.5 mocks, in spite of a factor of 3 increase in sensitivity.

In our treatment of reionization, we have assumed that galaxies which lie within ionized bubbles will be visible in Lyman α (consistent with the interpretation of Bagley et al. 2017). We have compared the effects of including this patchy reionization, or not including it (all galaxies visible), and have found that the effect on BAO peak

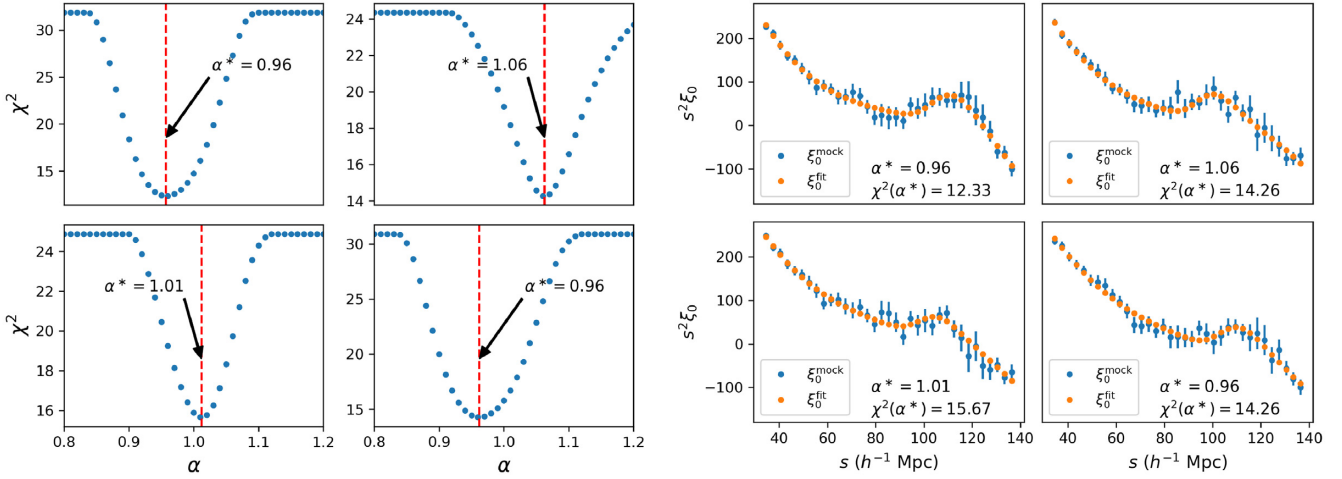


Figure 16. In the left four panels of this figure, we show the variations of χ^2 with α for the four mock cuboids corresponding to $L = 1 \times 10^{-17} \text{ erg s}^{-1} \text{ cm}^{-2}$ (with patchy reionization). In the plots on the left, the red dashed vertical lines denote the positions of the best-fitting α , viz. α^* for different mocks. In the right four panels, we show the monopoles computed from Landy Szalay estimator (ξ_0^{mock}) and the best-fitting monopoles ($\xi_0^{\text{fit}}(\alpha^*)$) obtained after fitting of ξ_0^{mock} for the four mock cuboids corresponding to $L = 1 \times 10^{-17} \text{ erg s}^{-1} \text{ cm}^{-2}$ (with patchy reionization). In the plots in the right-hand panels, the blue dots represent ξ_0^{mock} while the orange dots denote $\xi_0^{\text{fit}}(\alpha^*)$.

Table 3. This table shows best-fitting results of scale dilation parameter (i.e. α^*) and $\chi^2(\alpha^*)$ in the three quarter-mock catalogues (L7, L3.5, L1).

Line detection sensitivity	α^* (Mock 1)	α^* (Mock 2)	α^* (Mock 3)	α^* (Mock 4)
Before patchy reionization				
7 $\text{erg s}^{-1} \text{ cm}^{-2}$	1.08	1.16	0.86	1.08
3.5 $\text{erg s}^{-1} \text{ cm}^{-2}$	0.99	1.08	1.00	1.12
1 $\text{erg s}^{-1} \text{ cm}^{-2}$	1.01	1.03	0.99	0.95
With patchy reionization				
7 $\text{erg s}^{-1} \text{ cm}^{-2}$	1.04	0.81	1.05	1.19
3.5 $\text{erg s}^{-1} \text{ cm}^{-2}$	0.95	0.99	1.02	0.98
1 $\text{erg s}^{-1} \text{ cm}^{-2}$	0.96	1.06	1.01	0.96
Line detection sensitivity	$\chi^2(\alpha^*)$ (Mock 1)	$\chi^2(\alpha^*)$ (Mock 2)	$\chi^2(\alpha^*)$ (Mock 3)	$\chi^2(\alpha^*)$ (Mock 4)
Before patchy reionization				
7 $\text{erg s}^{-1} \text{ cm}^{-2}$	27.41	31.32	28.17	45.86
3.5 $\text{erg s}^{-1} \text{ cm}^{-2}$	21.59	20.03	16.20	26.42
1 $\text{erg s}^{-1} \text{ cm}^{-2}$	19.49	18.58	11.79	13.34
With patchy reionization				
7 $\text{erg s}^{-1} \text{ cm}^{-2}$	24.41	26.68	34.96	26.59
3.5 $\text{erg s}^{-1} \text{ cm}^{-2}$	38.80	22.52	43.14	32.26
1 $\text{erg s}^{-1} \text{ cm}^{-2}$	12.33	14.26	15.67	14.26

Table 4. The relative clustering linear bias b_{reion} between mocks with and without reionization in the different quarter-mock catalogues (L7, L3.5, L1).

Line detection sensitivity	Mock 1 b_{reion}	Mock 2 b_{reion}	Mock 3 b_{reion}	Mock 4 b_{reion}
7 $\text{erg s}^{-1} \text{ cm}^{-2}$	1.71	1.10	1.30	1.17
3.5 $\text{erg s}^{-1} \text{ cm}^{-2}$	1.13	1.38	1.41	1.05
1 $\text{erg s}^{-1} \text{ cm}^{-2}$	3.76	1.52	2.65	1.70

detection is quite small. For example, for the fiducial L7 sensitivity case, the effect of patchy reionization only increases the distance error from 6.4 per cent to 7.9 per cent. This is likely to be because in the patchy model for reionization employed, overdense regions reionize

first, and this is where most of the galaxy clustering signal lies. The reionization bubbles are also on smaller scales, than the BAO signal, and have a range of sizes, and therefore there is no masking of the BAO peak by bubbles. We note however that we have carried out no

Lyman α radiative transfer on the scales of individual galaxies, and that the visibility or not of the line in different environments is likely to be much more complex than our simple bubble model.

One significant difference between the BAO measurements at redshifts $z > 7$ and those for $z < 1$ is the lack of non-linearities in gravitational clustering. For example, we can see that in the right-hand panels of Fig. 15, there is no obvious smoothing of the BAO peak compared to the linear CDM fit. This is an advantage of working at higher redshifts, as no reconstruction techniques (Sherwin & White 2019) are needed to recover the maximum benefit. The high redshift galaxies do have a high bias (Bhowmick et al. 2019) and so their clustering is detectable where clustering of the density field would not be.

Measuring BAO in a new regime would at the very least be useful as a consistency check on measurements that have been made at low redshifts and in the CMB. It is important to fill in the large gap between the Lyman α forest measurements at $z \sim 2.2$ (Aubourg et al. 2015) and the features seen at $z \sim 1100$ by Planck and WMAP (Bennett et al. 2013; Planck Collaboration VI 2020b). Such measurements might shed some light on the tension between BAO and other Hubble parameter measurements (Cuceu et al. 2019), and whether early dark energy may be acting before the recombination epoch (Poulin et al. 2019; Verde, Treu & Riess 2019).

One of the targets of cosmological analysis with the RST satellite is large-scale structure measurements from galaxies at redshifts $z \sim 1-3$. This will extend the successful redshift space distortion, BAO, and other techniques to significantly earlier epochs than current large-scale structure measurements from ground-based surveys. Although much attention has rightly been focused on this redshift range, we have seen in this paper that it may be possible to make an even greater leap, and use the large-scale structure traced out by the very first massive galaxies as a cosmological probe. Although many potential pitfalls remain to be investigated, it is worth bearing this in mind when considering the RST survey strategies and parameters.

ACKNOWLEDGEMENTS

The BLUETIDES simulation was run on the BlueWaters facility at the National Center for Supercomputing Applications. TDM acknowledges funding from National Science Foundation (NSF) ACI-1614853, NSF AST-1517593, NSF AST-1616168, and National Aeronautics and Space Administration (NASA) ATP 19-ATP19-0084. TDM and RACC also acknowledge funding from NASA ATP 80NSSC18K101, and NASA ATP NNX17AK56G. RACC was supported by a Lyle Fellowship from the University of Melbourne, and TDM by a Shimmins Fellowship and a Lyle Fellowship from the University of Melbourne. GR acknowledges support from the National Research Foundation of Korea (NRF) through Grant No. 2017077508 funded by the Korean Ministry of Education, Science and Technology (MoEST), and from the faculty research fund of Sejong University in 2018. The work by KH was supported under the U.S. Department of Energy contract DE-AC02-06CH11357. This research used resources of the Argonne Leadership Computing Facility at the Argonne National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-06CH11357.

DATA AVAILABILITY

The data underlying this article will be shared on reasonable request to the corresponding author.

REFERENCES

- Agresti A., 1992, *Stat. Sci.*, 7, 131
 Anderson L. et al., 2012, *MNRAS*, 427, 3435
 Ansarinejad B., Shanks T., 2018, *MNRAS*, 479, 4091
 Ata M. et al., 2018, *MNRAS*, 473, 4773
 Aubourg É. et al., 2015, *Phys. Rev. D*, 92, 123516
 Bagley M. B. et al., 2017, *ApJ*, 837, 11
 Battaglia N., Trac H., Cen R., Loeb A., 2013, *ApJ*, 776, 81
 Bautista J. E. et al., 2018, *ApJ*, 863, 110
 Bennett C. L. et al., 2013, *ApJS*, 208, 20
 Bertin E., Arnouts S., 1996, *A&AS*, 117, 393
 Bhowmick A. K., Somerville R. S., DiMatteo T., Wilkins S., Feng Y., Tenneti A., 2019, *MNRAS*, 496, 754
 Bouwens R. J. et al., 2011, *Nature*, 469, 504
 Busca N. G. et al., 2013, *A&A*, 552, A96
 Casertano S., Brammer G. B., Dixon V., Mackenty J. W., Pirzkal N., Ravindranath S., Ryan R. E., 2015, Technical report, *Slitless Grism Spectroscopy with WFIRST: Observing Modes and Strategies*. Available at: http://www.stsci.edu/files/live/sites/www/files/home/scientific-community/wfirst/_documents/WFIRST-STScI-TR1506A.pdf. http://www.stsci.edu/files/live/sites/www/files/home/scientific-community/wfirst/_documents/WFIRST-STScI-TR1506A.pdf
 Cassata P. et al., 2011, *A&A*, 525, A143
 Chaikin E. A., Tyulneva N. V., Kaurov A. A., 2018, *ApJ*, 853, 81
 Colbert J. et al., 2018, in *American Astronomical Society Meeting Abstracts #231*, American Astronomical Society, Gaylord National Resort & Convention Center in National Harbor, MD. p. 355
 Cole S. et al., 2005, *MNRAS*, 362, 505
 Comparat J. et al., 2016, *A&A*, 592, A121
 Crocce M., Scoccimarro R., 2008, *Phys. Rev. D*, 77, 023533
 Cuceu A., Farr J., Lemos P., Font-Ribera A., 2019, *J. Cosmol. Astropart. Phys.*, 2019, 044
 Davis M., Peebles P. J. E., 1983, *ApJ*, 267, 465
 Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, *ApJ*, 292, 371
 de Sainte Agathe V. et al., 2019, *A&A*, 629, A85
 du Mas des Bourboux H. et al., 2017, *A&A*, 608, A130
 Eftekhazadeh S. et al., 2015, *MNRAS*, 453, 2779
 Eisenstein D. J., Hu W., 1998, *ApJ*, 496, 605
 Eisenstein D. J. et al., 2005, *ApJ*, 633, 560
 Ellis R. S. et al., 2013, *ApJ*, 763, L7
 Eyles L. P., Bunker A. J., Stanway E. R., Lacy M., Ellis R. S., Doherty M., 2005, *MNRAS*, 364, 443
 Feng Y., Matteo T. D., Croft R., Tenneti A., Bird S., Battaglia N., Wilkins S., 2015, *ApJ*, 808, L17
 Feng Y., Di-Matteo T., Croft R. A., Bird S., Battaglia N., Wilkins S., 2016, *MNRAS*, 455, 2778
 Finkelstein S. L. et al., 2013, *Nature*, 502, 524
 Fisher R. A., 1922, *J. R. Stat. Soc.*, 85, 87
 Fossati M. et al., 2017, *ApJ*, 835, 153
 Francis P. J., Dopita M. A., Colbert J. W., Palunas P., Scarlata C., Teplitz H., Williger G. M., Woodgate B. E., 2012, *MNRAS*, 428, 28
 Gehrels N., Spergel D., WFIRST SDT Project, 2015, in *Journal of Physics Conference Series*, Vol. 610. p. 012007
 Gronwall C. et al., 2007, *ApJ*, 667, 79
 Habib S., Morozov V., Frontiere N., Finkel H., Pope A., Heitmann K., 2013, in *SC '13: Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis*, Association for Computing Machinery, New York, NY, United States. p. 1
 Habib S. et al., 2016, *New Astron.*, 42, 49
 Hamilton A. J. S., 1993, *ApJ*, 417, 19
 Heitmann K. et al., 2019, *AAS*, 245, 16
 Hill J. C., Baxter E. J., 2018, *J. Cosmol. Astropart. Phys.*, 8, 037
 Hu W. et al., 2017, *ApJ*, 845, L16
 Kakiichi K., Dijkstra M., Ciardi B., Graziani L., 2016, *MNRAS*, 463, 4019
 Kerscher M., Szapudi I., Szalay A. S., 2000, *ApJ*, 535, L13
 Komatsu E. et al., 2011, *ApJS*, 192, 18
 Landy S. D., Szalay A. S., 1993, *ApJ*, 412, 64

- Laporte N. et al., 2017, *ApJ*, 837, L21
 Larson R. L. et al., 2018, *ApJ*, 858, 94
 Lehnert M. D. et al., 2010, *Nature*, 467, 940
 Leitherer C. et al., 1999, *ApJS*, 123, 3
 Lewis A., Bridle S., 2002, *Phys. Rev. D*, 66, 103511
 Li N. et al., 2019, *ApJ*, 878, 122
 Malhotra S., 2019, in American Astronomical Society Meeting Abstracts, American Astronomical Society. p. 315.01
 McCarthy P. J. et al., 1999, *ApJ*, 520, 548
 Mobasher B. et al., 2005, *ApJ*, 635, 832
 Oesch P. A. et al., 2016, *ApJ*, 819, 129
 Pentericci L. et al., 2018, *A&A*, 619, A147
 Planck Collaboration et al., 2020a, *A&A*, 641, A6
 Planck Collaboration et al., 2020b, *A&A*, 641, A6
 Pons-Bordería M.-J., Martínez V. J., Stoyan D., Stoyan H., Saar E., 1999, *ApJ*, 523, 480
 Poulin V., Smith T. L., Karwal T., Kamionkowski M., 2019, *Phys. Rev. Lett.*, 122, 221301
 Raichoor A. et al., 2017, *MNRAS*, 471, 3955
 Riess A. G. et al., 2018, *ApJ*, 855, 136
 Sarpa E., Schimd C., Branchini E., Matarrese S., 2019, *MNRAS*, 484, 3818
 Sherwin B. D., White M., 2019, *J. Cosmol. Astropart. Phys.*, 2019, 027
 Slosar A. et al., 2013, *J. Cosmol. Astropart. Phys.*, 4, 026
 Sobacchi E., Mesinger A., 2015, *MNRAS*, 453, 1843
 Song M., Finkelstein S. L., Livermore R. C., Capak P. L., Dickinson M., Fontana A., 2016, *ApJ*, 826, 113
 Spergel D. et al., 2015, preprint(arXiv:1503.03757)
 Tilvi V. et al., 2010, *ApJ*, 721, 1853
 Tilvi V. et al., 2016, *ApJ*, 827, L14
 Vanderlinde K., Chime Collaboration, 2014, in Exascale Radio Astronomy.
 Vargas-Magaña M. et al., 2014, *MNRAS*, 445, 2
 Verde L., Treu T., Riess A. G., 2019, *Nat. Astron.*, 3, 891
 Wang M.-Y., Peter A. H. G., Strigari L. E., Zentner A. R., Arant B., Garrison-Kimmel S., Rocha M., 2014, *MNRAS*, 445, 614
 Waters D., Di Matteo T., Feng Y., Wilkins S. M., Croft R. A. C., 2016, *MNRAS*, 463, 3520
 Weinberg D. H., Mortonson M. J., Eisenstein D. J., Hirata C., Riess A. G., Rozo E., 2013, *Phys. Rep.*, 530, 87
 Wilkins S. M. et al., 2013, *MNRAS*, 435, 2885
 Wilkins S. M., Feng Y., Di-Matteo T., Croft R., Stanway E. R., Bouwens R. J., Thomas P., 2016a, *MNRAS*, 458, L6
 Wilkins S. M., Feng Y., Di-Matteo T., Croft R., Stanway E. R., Bunker A., Waters D., Lovell C., 2016b, *MNRAS*, 460, 3170
 Wilkins S. M., Feng Y., Di Matteo T., Croft R., Lovell C. C., Waters D., 2017, *MNRAS*, 469, 2517
 Wilkins S. M. et al., 2020, *MNRAS*, 493, 6094
 Xu X., Padmanabhan N., Eisenstein D. J., Mehta K. T., Cuesta A. J., 2012, *MNRAS*, 427, 2146
 Xu X., Cuesta A. J., Padmanabhan N., Eisenstein D. J., McBride C. K., 2013, *MNRAS*, 431, 2834
 Yajima H., Sugimura K., Hasegawa K., 2018, *MNRAS*, 477, 5406
 Yan H. et al., 2011, *ApJ*, 728, L22
 Zarrouk P. et al., 2018, *MNRAS*, 477, 1639
 Zhai Z. et al., 2017, *ApJ*, 848, 76

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