

Correcting for small-displacement interlopers in BAO analyses

Setareh Foroozan,^{a,b} Elena Massara^{a,b} and Will J. Percival^{a,b,c}

^aWaterloo Centre for Astrophysics, University of Waterloo,
200 University Ave W, Waterloo, ON N2L 3G1, Canada

^bDepartment of Physics and Astronomy, University of Waterloo,
200 University Ave W, Waterloo, ON N2L 3G1, Canada

^cPerimeter Institute for Theoretical Physics,
31 Caroline St. North, Waterloo, ON N2L 2Y5, Canada

E-mail: s2forooz@uwaterloo.ca, elena.massara.cosmo@gmail.com,
will.percival@uwaterloo.ca

Received August 22, 2022

Revised September 19, 2022

Accepted October 5, 2022

Published October 21, 2022

Abstract. Due to the low resolution of slitless spectroscopy, future surveys including those made possible by the Roman and Euclid space telescopes will be prone to line mis-identification, leading to interloper galaxies at the wrong redshifts in the large-scale structure catalogues. The most pernicious of these have a small displacement between true and false redshift such that the interloper positions are correlated with the target galaxies. We consider how to correct for such contaminants, focusing on $H\beta$ interlopers in $[O III]$ catalogues as will be observed by Roman, which are misplaced by $\Delta d = 97 h^{-1}$ Mpc at redshift $z = 1$. Because this displacement is close to the BAO scale, the peak in the interloper-target galaxy cross-correlation function at the displacement scale can change the shape of the BAO peak in the auto-correlation of the contaminated catalogue, and lead to incorrect cosmological measurements if not accounted for properly. We consider how to build a model for the monopole and quadrupole moments of the contaminated correlation function, including an additional free parameter for the fraction of interlopers. The key input to this model is the cross-correlation between the population of galaxies forming the interlopers and the main target sample. It will be important to either estimate this using calibration data or to use the contaminated small-scale auto-correlation function to model it, which may be possible if a number of requirements about the galaxy populations are met. We find that this method is successful in measuring the BAO dilation parameters without significant degradation in accuracy, provided the cross-correlation function is accurately known.

Keywords: baryon acoustic oscillations, cosmological parameters from LSS, cosmological simulations, redshift surveys

ArXiv ePrint: [2208.05001](https://arxiv.org/abs/2208.05001)

Contents

1	Introduction	1
2	Displacement	3
3	Simulations	3
4	Modelling the contaminated correlation function	6
4.1	Correlation function with interlopers	7
4.2	Interloper-target cross-correlation	8
4.3	Building a cosmological model	10
5	Results	12
6	Discussions	16

1 Introduction

Space-based Large Scale Structure (LSS) spectroscopic surveys made possible by the Roman Space Telescope [1] and Euclid [2] missions will soon be available. These surveys will make use of slitless spectroscopy to measure galaxy redshifts meaning that, while they have the potential to constrain the cosmological parameters with sub percent levels of accuracy, this requires a thorough understanding of potential sources of systematic error arising from the slitless method. One possible systematic effect, that we study in this paper, is the contamination of the sample by interlopers resulting from line confusion.

To estimate the redshifts of a set of galaxies using slitless spectroscopy, we measure the wavelengths of one or more emission lines for each. In this paper we will primarily be concerned with samples where only one emission line is used to estimate the redshift. Generally, emission lines with different rest-frame wavelengths λ_1 and λ_2 , from two different galaxies with true redshifts $z_1 = \lambda_{\text{obs}}/\lambda_1 - 1$ and $z_2 = \lambda_{\text{obs}}/\lambda_2 - 1$, can appear at the same observed wavelength, λ_{obs} . The measured redshift for a galaxy given only the observed position of this single line will depend on the rest-frame wavelength assumed. In the presence of a secondary emission line, one can distinguish between different options. However, if we are measuring redshifts from a single emission line, we are prone to line misidentification, if we assume, for a sample of galaxies, that the observed line has the same rest-frame wavelength for all galaxies. The galaxies in the sample for which the lines are misidentified resulting in them being assigned the wrong redshift, are called interlopers. We refer to the galaxies whose observed emission line matches the assumed rest-frame wavelength when measuring redshifts as the target galaxies.

For Roman, for example, it is likely that many redshifts will be measured using 2 or more emission lines [3]. At $2 < z < 3$, [O III] is the primary line for Roman observations, which is actually a doublet. Resolving this doublet as two separate lines, which is possible at given the resolution of Roman, or observing a secondary line from [O II] emission would allow a secure redshift to be measured. However, to provide a bonus sample to higher number densities, it will be useful to analyse galaxies where only one unresolved line is observed, assumed

to be [O III], but where this identification is less secure. For this, we will have interlopers, where the $H\beta$ line is mistaken for [O III]. The increased size of this bonus sample will help reduce the statistical errors on measured cosmological parameters, at the cost of increasing the systematic errors. It is these systematic errors, and methods to mitigate them, that we consider in this paper.

In general, there are three different ways that a catalogue can be contaminated:

1. First, due to noise spikes being misidentified as emission lines. The noise is uncorrelated with all the other galaxies, and has a distribution different from that of the true observed lines. Thus, this population of contaminants Poisson samples a volume that has a window different from that of the target galaxy population. It can trivially be corrected in a clustering measurement as it is absorbed into (and is cancelled by) part of the expected density (often quantified by a random catalog) and simply adds to the shot noise and changes the amplitude of the signal measured.
2. Second, misidentified emission lines can arise from interloper galaxies displaced by a large radial distance. In this case, correlations with the target galaxy population are small and we only need to model the auto-correlation of the interloper sample (e.g. [4–7]). Because the displacement is large, there is a significant change in $H(z)$ between the true redshift of the interlopers and the incorrect redshift inferred, such that the Baryon Acoustic Oscillation (BAO) signal in the auto-correlation of the interlopers is shifted with respect to its true value. This needs to be included in any analysis but, as the shift is known from the line emission wavelengths, this contribution to the clustering of the contaminated sample can be easily modeled provided the fraction of interlopers is known.
3. The third class of contaminant is a more pernicious type of interloper, which is one only shifted by a small radial displacement by the incorrect rest-frame wavelength assignment. Such interlopers will be correlated with other such objects *and* with the target galaxy sample. This type of interloper can occur for samples obtained using the Roman Space Telescope at high redshifts, where the target is the primary [O III] line, and $H\beta$ line is the source of interlopers. The impact of these interlopers on the BAO peak position was studied by [8]. They showed that misidentifying $H\beta$ line as [O III] line in an Emission Line galaxy will lead to underestimating its distance by $\sim 90 h^{-1}$ Mpc. This would distort the shape of the galaxy correlation function and therefore introduce a shift in the BAO peak position. We call this class of interlopers, which are displaced by $\lesssim 150 h^{-1}$ Mpc, “small-displacement” interlopers (see section 5).

We focus on the third class of contaminant. Following [8], we present a model for the monopole and the quadrupole moments of the correlation function of a Roman-like catalogue, with an [O III] target line sample contaminated by $H\beta$ interlopers. With this model, we investigate how to find an unbiased estimation of the isotropic and anisotropic dilation parameters α and ϵ from the BAO signal, as well as the fraction of interlopers f_i . Our method is based on standard BAO modeling, additionally allowing for the interlopers. The model correlation function has two terms: the galaxy auto-correlation that we estimate using CAMB [9], and a non-negligible cross-correlation between the galaxies and the interlopers that we consider a number of ways of estimating. If we can measure this from the contaminated auto-correlation, or use another way of accurately calibrating this statistic, our pipeline gives an unbiased estimation of all the three parameters. If we do not take into account the interlopers, the measurements would be highly biased.

The plan of our paper is as follows: in section 2, we derive the interloper shift value. In section 3, we describe simulations that were used for our analysis. In section 4, we present our model for the correlation function of a catalogue contaminated by some small-displacement interlopers. We show the outcomes of our pipeline in section 5. We discuss our results and their implications in section 6.

2 Displacement

Let us assume that in a survey, potential interlopers are at redshift z^{true} , and they emit a primary emission line with rest-frame wavelength $\lambda_{\text{emit}}^{\text{true}}$, that is observed at wavelength λ_{obs} . The observed wavelength can be misidentified for another photon as if it was emitted with rest-frame wavelength $\lambda_{\text{emit}}^{\text{false}}$ at redshift z^{false} . This redshift misidentification would effectively lead us to a wrong estimation of the comoving proper distance to the galaxies by Δd :

$$\Delta d = d^{\text{true}} - d^{\text{false}} = \int_{z^{\text{false}}}^{z^{\text{true}}} \frac{cdz}{H(z)} \quad (2.1a)$$

$$\approx \frac{c}{H(z^{\text{true}})} (z^{\text{true}} - z^{\text{false}}). \quad (2.1b)$$

$$\approx \frac{c}{H(z^{\text{true}})} \left[1 - \frac{\lambda_{\text{emit}}^{\text{true}}}{\lambda_{\text{emit}}^{\text{false}}} \right] (1 + z^{\text{true}}). \quad (2.1c)$$

In this paper we consider $\text{H}\beta$ interlopers misidentified as $[\text{O III}]$ emitters. While $[\text{O III}]$ is a doublet, here we assume that $\text{H}\beta$ is misidentified as the primary line in the doublet at $\lambda_{\text{false}}^{\text{OIII}} = 500.7 \text{ nm}$, rather than the secondary line at $\lambda_{\text{false}}^{\text{OIII}} = 495.9 \text{ nm}$. The rest-frame wavelength of $\text{H}\beta$ is $\lambda_{\text{true}}^{\text{H}\beta} = 486.1 \text{ nm}$, leading to an offset between the true and wrongly inferred interloper comoving position equal to

$$\Delta d \approx 87.41 \frac{1 + z^{\text{true}}}{\sqrt{\Omega_{\Lambda} + \Omega_{\text{m}}(1 + z^{\text{true}})^3}} h^{-1} \text{ Mpc}. \quad (2.2)$$

Figure 1 displays the distance Δd as a function of the true galaxy redshift and with the parameters Ω_{Λ} and Ω_{m} evaluated at the cosmology of the simulations used in this paper and described in section 3. In that cosmology, the displacement Δd is equal to $97 h^{-1} \text{ Mpc}$ at $z = 1$ (the redshift of the N-body snapshots considered in this work), and $90, 85 h^{-1} \text{ Mpc}$ at $z = 1.7, 2.1$, which are the redshifts of interest for the Roman Space Telescope.

Note that measurements stemming from rest-frame wavelength differences cannot be used as standard rulers in the same way as BAO — what is fixed is the difference in redshift, not the distance. The displacement in physical units depends on the assumed fiducial cosmology used to convert redshifts to distances not the true cosmology, and we do not have to vary the displacement when fitting different cosmological models to the data.

3 Simulations

In this study, our focus is on building and testing a model framework that describes the effect of interloper galaxies rather than predicting the actual expected fraction of interlopers. While some progress have been made in modelling the population of $[\text{O III}]$ emitters (e.g., [10]), we need a thorough comparison with observations to know if we have the accuracy required. The population of $\text{H}\beta$ emitters is less well understood. Therefore, we will use halos in place of

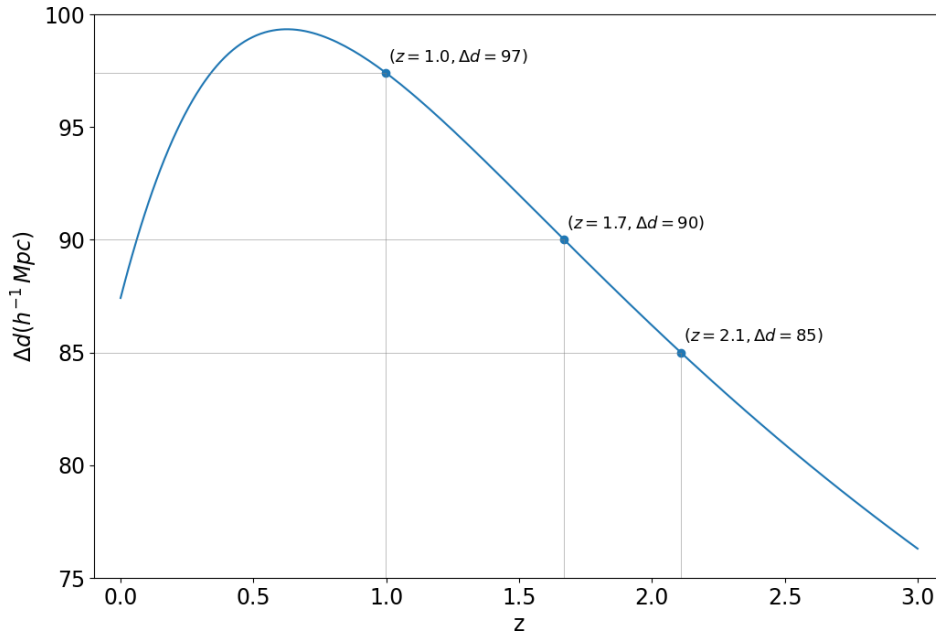


Figure 1. Displacement of interlopers as a function of their true redshift as shown in eq. (2.2).

galaxy catalogs to test our model, for simplicity. We employ the halo catalogs from 1,000 N-body simulations of the Quijote suite [11]. These simulations have been run in a box with size $1h^{-1}\text{Gpc}$ and follow the evolution of 512^3 cold dark matter particles in the fiducial cosmology of the suite, which is a flat ΛCDM cosmology with parameters $\Omega_m = 0.3175$, $\Omega_b = 0.049$, $n_s = 0.9624$, $h = 0.6711$, $\sigma_8 = 0.834$, $M_\nu = 0.0\text{eV}$. Halos have been identified using the Friends-of-Friends algorithm with linking length parameter equal to 0.2, and only those with 20 or more cold dark matter particles have been stored in the catalogue. This means that considered halos have minimum mass equal to $1.31 \cdot 10^{13}h^{-1}M_\odot$.

Even though [O III] emission galaxies are the primary target at redshifts beyond $z = 1.8$ in the High-Latitude Survey of the Roman telescope, we use halo catalogs at redshift $z = 1$ to have a larger number density of objects, $\bar{n}_h = 2 \cdot 10^{-4}h^3\text{Mpc}^{-3}$, consistent with the expected [O III] mean number density in Roman between redshifts 1.8–2.8 [8, 10]. At the redshift of the simulation, interlopers appear to be displaced by $97h^{-1}\text{Mpc}$ along the line of sight. To simulate interlopers at higher redshift, we also consider displacements equal 90 or $85h^{-1}\text{Mpc}$ that correspond to catalogs at redshifts $z = 1.7$ and $z = 2.1$, respectively. We build the contaminated halo catalogs by randomly selecting objects that will be interlopers and displacing them at the wrongly inferred distance along the z direction, that we assume to be the line-of-sight (hereafter, LOS). We implement four different percentages of interlopers (2, 5, 10, 15 %) and three different values for the displacement, obtaining 12 different contaminated halo catalogs for each of the initial uncontaminated ones.

We measure the correlation function from all these mocks, by making use of Nbodykit [12] adopting the Landy-Szalay estimator [13]. To build a better understanding of the problem, the monopole and quadrupole of this measured correlation function are shown in figure 2 and figure 3, respectively. In these plots, we show catalogues contaminated by different fractions of interlopers in different colours, and with different displacements in different panels: $\Delta d = 85$ (top), 90 (middle), and $97h^{-1}\text{Mpc}$ (bottom). The effects of the interlopers are clear

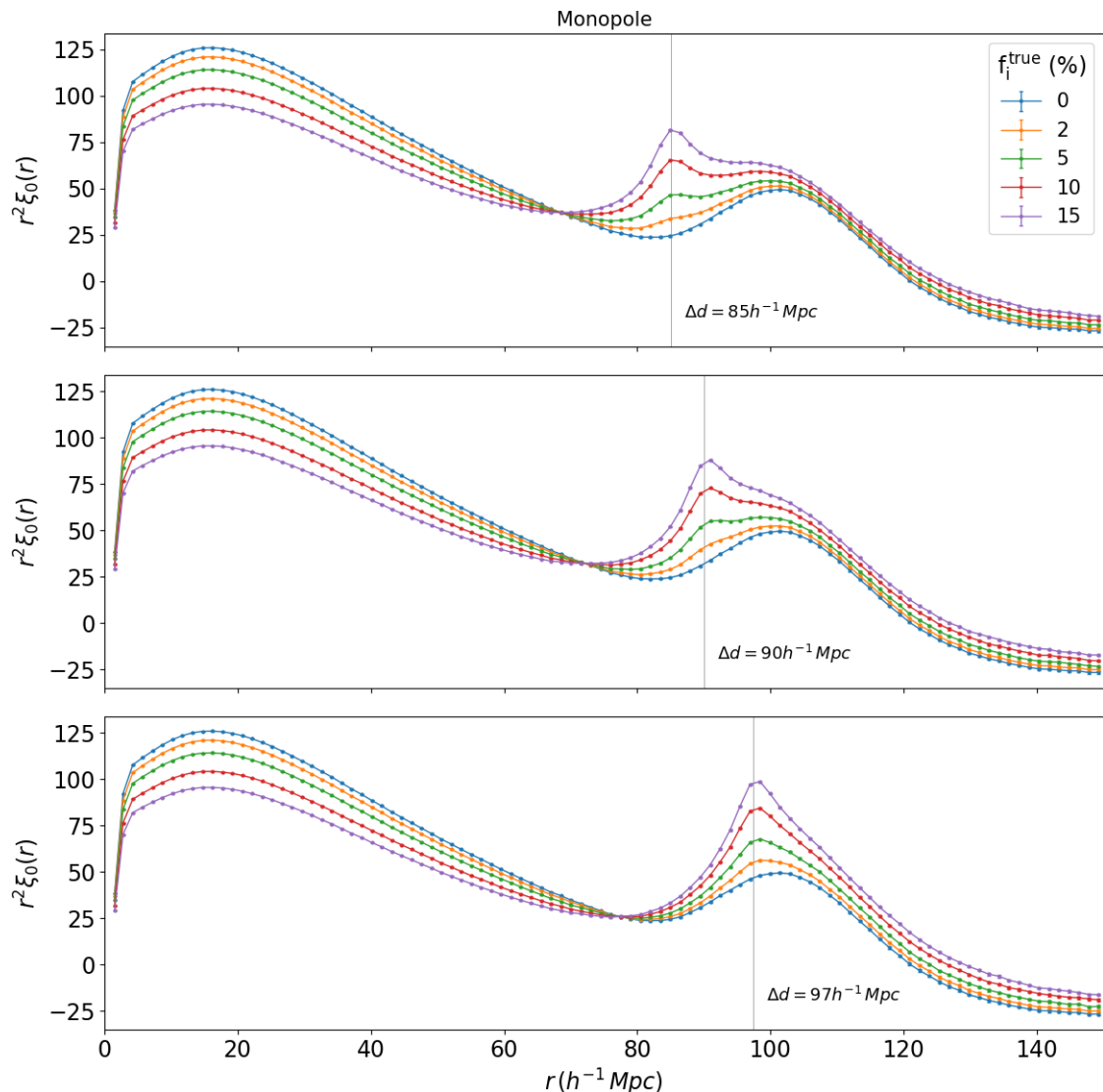


Figure 2. Monopole of the correlation function as calculated from the mean of 1000 mocks that are contaminated by different fractions of interlopers (different colours), and different interloper displacements: 85 (top row), 90 (middle row), and $97 h^{-1} \text{Mpc}$ (bottom row). As can be seen, the BAO peak is skewed towards Δd , and amplified more as the number of interlopers increase.

to see in both moments. The amplitude of the small-scale monopole is suppressed by the presence of interlopers, and the suppression increases with the fraction of interlopers. The missing small-scales correlation is converted into larger scales, and in particular it enhances the amplitude of the correlation around the displacement scale, which is close to the BAO peak position. As a result, the BAO peak appears to be enhanced, broadened and shifted towards smaller separations. The quadupole is also modified by the interlopers: its amplitude is enhanced on all scales with increasing interloper fraction. Moreover, it also exhibits a peak around the scales corresponding to the displacement of the interlopers, as a consequence of the increased number of interloper-target galaxy pairs along the LOS.

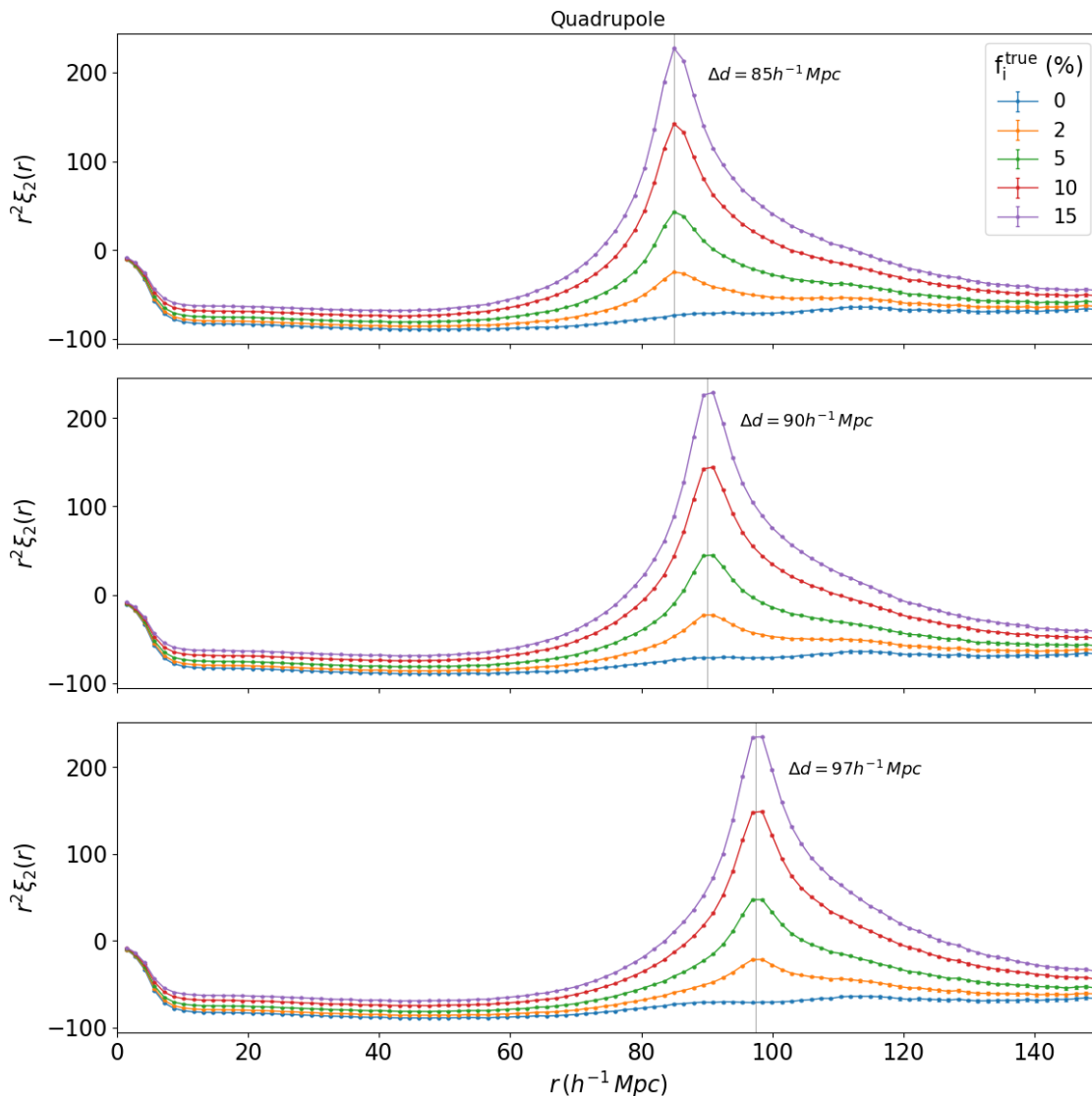


Figure 3. Quadrupole of the correlation function as calculated from the mean of 1000 mocks that are contaminated by different fractions of interlopers (different colours), and different interloper displacements: 85 (top row), 90 (middle row), and $97 h^{-1} \text{Mpc}$ (bottom row). As can be seen, the quadrupole peaks at Δd , which demonstrates the anisotropy caused by the shift of interlopers along the LOS.

4 Modelling the contaminated correlation function

In this section, we show how to model the monopole and quadrupole moments of the correlation function for a catalogue that is contaminated by interlopers misplaced by relatively small displacements such that the interloper-galaxy cross-correlation term is not negligible. We build a model for the auto-correlation function of a contaminated catalogue in section 4.1, and in section 4.2 we introduce our method to estimate the interloper-target cross-correlation function. Finally, in section 4.3, we discuss our completed model.

4.1 Correlation function with interlopers

We consider a catalogue contaminated by a given fraction of interlopers, f_i , and define ξ_{gg} the auto-correlation function of the target galaxies, ξ_{gi} the cross-correlation between these galaxies and interlopers and ξ_{ii} the auto-correlation function of the interlopers. Throughout this paper, indices “i”, and “g”, stand for interlopers, and galaxies respectively.

We denote the observed comoving distance of an object with \vec{x} : for target galaxies, this is the same as their true position, whereas for interlopers it is the wrong measured position. We denote the true position of an interloper with \vec{y} . Following [14], the contaminated galaxy overdensity is given by

$$\delta(\vec{x}) = (1 - f_i)\delta_g(\vec{x}) + f_i\delta_i(\vec{x}). \quad (4.1)$$

We can use this equation to calculate the contaminated correlation function at any observed separation $\vec{r} = \vec{x}_1 - \vec{x}_2$,

$$\begin{aligned} \xi(\vec{x}_1 - \vec{x}_2; f_i) &= \langle \delta(\vec{x}_1)\delta^*(\vec{x}_2) \rangle \\ &= \langle [(1 - f_i)\delta_g(\vec{x}_1) + f_i\delta_i(\vec{x}_1)] [(1 - f_i)\delta_g^*(\vec{x}_2) + f_i\delta_i^*(\vec{x}_2)] \rangle \\ &= (1 - f_i)^2\xi_{gg}(\vec{r}) + f_i^2\xi_{ii}(\vec{r}) + 2f_i(1 - f_i)\xi_{gi}(\vec{r}). \end{aligned} \quad (4.2)$$

This equation can be significantly simplified if we can make two key assumptions.

First, we assume that the displacement Δd in eq. (2.2) between the true position of interlopers and the wrongly inferred one is a constant within a redshift bin when calculating the correlation function at a fixed cosmology. Indeed, we only consider correlation functions at separations r smaller than $150 h^{-1}$ Mpc in the following analysis and in this range the displacement only varies by at most 0.8% (see figure 1). Therefore, we can assume that each object in the pair is displaced by the same amount. Moreover, it is commonly assumed that within a redshift bin $H(z)$ is a constant, therefore, we can assume that the displacement is fixed across the redshift bin considered for all the pairs. Thus, the wrongly inferred position of each interloper can be written in terms of a constant displacement along the LOS \hat{z} : $\vec{x}_i = \vec{y}_i - \Delta d \hat{z}$, where \vec{y}_i is the true position.

The interloper auto-correlation function can then be written as $\xi_{ii}(\vec{x}_1 - \vec{x}_2) = \xi_{ii}(\vec{y}_1 - \vec{y}_2)$ (where we have assumed that the LOS is the same for both galaxies in a pair), because two point correlators are invariant under translations — they depend only on the relative distance between two objects. This means that the correlation function of the interlopers at the wrong redshifts is equal to the correlation function of the interlopers at their true position.

Interlopers and targets belong to two different populations of galaxies and have in principle different bias schemes. Consequently, we either need calibration data to measure $\xi_{gi}(\vec{r})$ and $\xi_{ii}(\vec{r})$ (we discuss this particular case in section 6), or we need a way to model this function. On the other hand, we note that since displacements are small, interlopers and targets both trace a matter field with similar amplitude of clustering. It is also likely that interlopers have similar properties and intrinsic clustering as the target galaxies, meaning that they share the same galaxy bias. We thus make a second assumption: targets and interlopers share the same galaxy bias. In this case, together with the assumption of constant displacement for the interlopers, we have: $\xi_{ii}(\vec{r}) = \xi_{gg}(\vec{r})$. We assume that this is true in the rest of the paper. Our results and conclusions will be similar if this is not true, but instead we are able to accurately estimate or model the bias scheme of the interlopers. Substituting into eq. (4.2) yields

$$\xi(\vec{r}; f_i) = (1 + 2f_i^2 - 2f_i)\xi_{gg}(\vec{r}) + 2f_i(1 - f_i)\xi_{gi}(\vec{r}). \quad (4.3)$$

In the next section, we will further consider methods to estimate the cross-correlation between interlopers and galaxies, $\xi_{\text{gi}}(r)$.

4.2 Interloper-target cross-correlation

As mentioned before, the cross-correlation between galaxies and interlopers plays a key role in determining the shape of the contaminated correlation function. While the galaxy-galaxy and interloper-interloper terms remain unchanged after misplacing interlopers along the LOS, the galaxy-interloper cross-correlation changes greatly. This change can easily be understood considering that ξ_{gi} at $\vec{r} = \vec{x}_1 - \vec{x}_2$, which is the separation between the two objects after misplacing interlopers by Δd , should be equal to ξ_{gi} at the true separation $\vec{r}' = \vec{x}_1 - \vec{y}_2$. Thus,

$$\xi_{\text{gi}}(\vec{r}) = \xi_{\text{gi}}(\vec{r}'). \quad (4.4)$$

There is a simple trigonometry transformation between \vec{r} and \vec{r}'

$$\begin{aligned} r'(r, \mu) &= \sqrt{r^2 + \Delta d^2 - 2 \Delta d r \mu}, \\ \mu'(r, \mu) &= \frac{r\mu - \Delta d}{\sqrt{r^2 + \Delta d^2 - 2 \Delta d r \mu}}. \end{aligned} \quad (4.5)$$

where we have defined $r = |\vec{r}|$ being the modulus of the separation vector and $\mu = \vec{r} \cdot \hat{z}/r$ being the cosine of the angle between the separation vector and the LOS.

The auto-correlation to cross-correlation mapping effectively pushes pair separations to larger scales. Nearby objects are more correlated than more distant objects — the galaxy correlation function roughly decays by r^2 , so the displacement of interlopers effectively moves the strong correlation between nearby objects around $< 20 h^{-1}$ Mpc to larger scales in the galaxy-interloper cross-correlation, leaving a peak. The location of the center of this peak is at the interloper displacement. We show this in figure 4 with blue points obtained by measuring the cross-correlation between galaxies and interlopers in a Quijote simulation box at $z = 1$, after misplacing interlopers by $97 h^{-1}$ Mpc.

In practice, if we have a small survey with higher signal-to-noise ratio (calibration data), it will be possible to estimate which galaxies would appear as interlopers in a survey with lower signal-to-noise ratio. Using calibration data we can directly measure the cross-correlation between galaxies and interlopers at their *true* distance, and then infer the cross-correlation component at the wrong distance via eq. (4.5). If the calibration data is not available, there are two other possible approaches that we discuss in the following to estimate the cross-correlation assuming that the bias scheme of interlopers is the same as main target galaxies: analytical and observational approaches.

Analytical approach. We considered taking the Fourier transform of the analytical power spectrum from CAMB linear and non-linear (from HALOFIT [15, 16]) to estimate the auto-correlation function. The results from the linear CAMB model are shown in red in figure 4. The top and bottom panels of this plot show that the linear CAMB model roughly describes the shape of the measured monopole and quadrupole of the cross-correlation, but it underestimates the peak height by $\sim 30\%$. This is because the prediction of CAMB for the monopole at small scales is significantly lower than the non-linear reality, and, as we discussed earlier, this directly impacts the height of the peak at $97 h^{-1}$ Mpc. We have also used HALOFIT to create a non-linear model. The corresponding results are shown in green in the same figure. The outcome remains the same with a slightly better prediction of the height. In this case, while

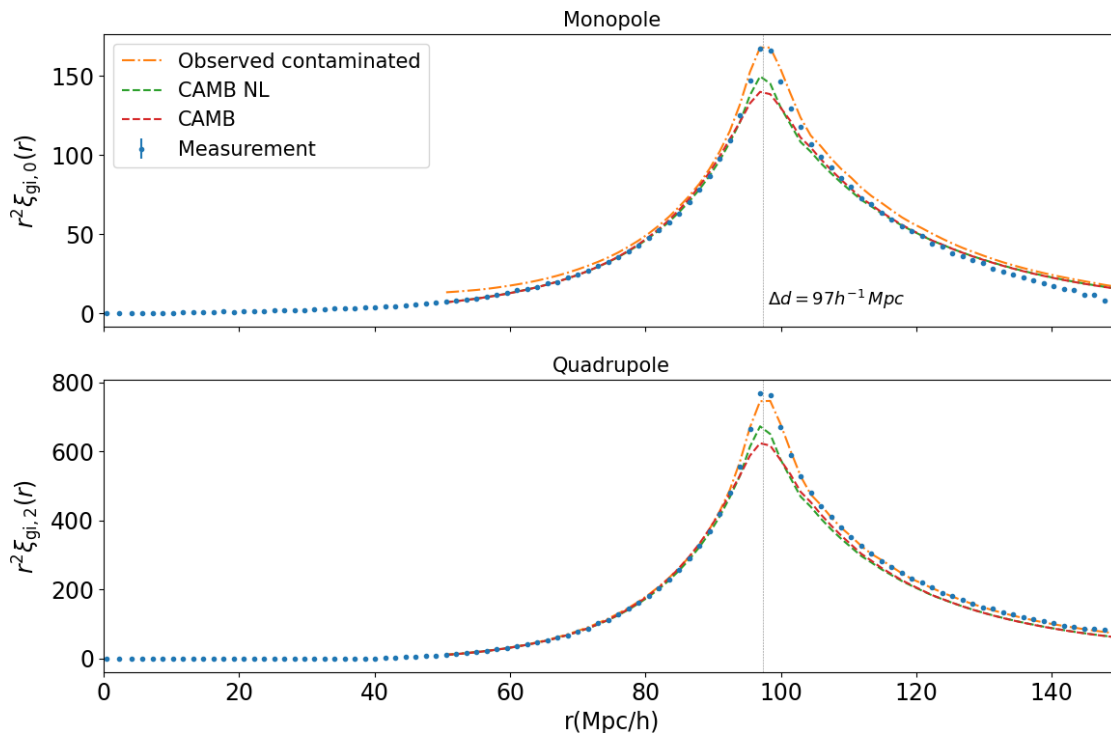


Figure 4. The importance of accurately modelling the small-scale cross-correlation in order to match the large-scale clustering around the BAO feature. Monopole (top) and quadrupole (bottom) of the galaxy-interloper cross-correlation function correlation in real-space, calculated from mapping of auto-correlation described in eq. (4.4). We considered four different auto-correlations described in more detail in text. The interloper displacement, Δd , used in mapping equals 97 for $z = 1$.

the displacement depends on the fiducial cosmology and it determines the position of the peak in the target-interloper correlation function, the shape of the cross-correlation depends on the true cosmology since it is a remapping of the auto-correlation of target galaxies via eq. (4.11). This can be easily implemented in the model for the correlation function. The inability of this model to accurately fit the cross-correlation function led us to consider our second approach, or the observational approach.

Observational approach. If we assume that interlopers and targets are drawn from the same population of galaxies, we can estimate their cross-correlation directly from the measured auto-correlation of the contaminated catalogue. As shown in figure 4, the monopole of cross-term vanishes on small scales $r < 40 h^{-1}$ Mpc. Therefore, one can rewrite eq. (4.3) as

$$\xi(\vec{r}; f_i) \approx (1 + 2f_i^2 - 2f_i) \xi_{gg}(\vec{r}) \quad \text{for small } r, \quad (4.6)$$

indicating that the contaminated and uncontaminated catalogues have small-scale correlation functions that differ by an overall amplitude only, and this amplitude depends on the fraction of interlopers in the contaminated catalogue. We can therefore infer the galaxy-galaxy small-scale correlation function directly from the measured contaminated correlation function, modulo an overall normalization. The large-scale feature in the cross-correlation seen in figure 4 is mainly determined by the small-scale correlation of galaxies and interlopers at the true position, or equivalently the galaxy auto-correlation since we assumed that galaxies and interlopers

have the same bias (see eq. (4.4)). Therefore, using eqs. (4.6) and (4.4) we can estimate the cross-correlation term on the scales of interest ($r < 150 h^{-1} \text{Mpc}$) directly from the measured small-scale contaminated correlation function as

$$\xi_{\text{gi}}(\vec{r}) = \frac{\xi(\vec{r}'; f_i)}{(1 + 2f_i^2 - 2f_i)}, \quad (4.7)$$

where the mapping between \vec{r} and \vec{r}' is described in eq. (4.5). Using this observational method, the model for the cross-correlation is already built from a correlation function in the true cosmology but measured in the fiducial one, and we will not need to account for differences between the two cosmologies when performing the fit.

Using this method to determine the cross-correlation, one can immediately see the improvement in the peak height estimation of the monopole and quadrupole moments of the cross-correlation function in figure 4 (orange line). Note that this method assumes that interlopers and targets have the same bias scheme, and this assumption is satisfied exactly in our mocks. On the other hand, the general shape of the monopole of the cross-correlation, especially at larger scales, deviates from that of the measurement. We believe that this is due to the approximation that we made to derive eq. (4.6); that of a vanishing cross-correlation on small scales. The quadrupole is less affected by this approximation. These effects do not alter the measured parameters (see section 4.3). Finally, using this approximation, we can use eq. (4.7) for the galaxy-interloper cross-correlation and rewrite eq. (4.3)

$$\xi(\vec{r}; f_i) = (1 + 2f_i^2 - 2f_i)\xi_{\text{gg}}(\vec{r}) + \frac{2f_i(1 - f_i)}{(1 + 2f_i^2 - 2f_i)} \times \mathcal{M}[\xi(f_i)](\vec{r}), \quad (4.8)$$

where for simplicity, we use \mathcal{M} to denote the mapping of scales in eq. (4.5) that describes the deformation due to interlopers.

4.3 Building a cosmological model

Thus far, we have introduced a model for the contaminated correlation function that consists of the galaxy-galaxy correlation function term, $\xi_{\text{gg}}(\vec{r})$, and the cross-correlation term, $\xi_{\text{gi}}(\vec{r}; f_i)$. There is still one part missing: how does our model depend on the assumptions about the fiducial cosmology? In this section, we will address this question based on the methodology and notation of ref. [17] and [18].

To measure the correlation function from data, we need to calculate the separation between pairs by assuming a fiducial cosmology to translate from redshifts to distances. The BAO peak in the correlation function can be distorted if the assumed cosmology is not the same as the true cosmology. This distortion can be parameterized by Alcock-Paczynski parameters (hereafter AP; [19]): α , the isotropic dilation parameter, and ϵ , the anisotropic warping parameter:

$$\alpha = \left[\frac{D_{\text{A,true}}^2(z) H_{\text{fid}}(z)}{D_{\text{A,fid}}^2(z) H_{\text{true}}(z)} \right]^{1/3} \frac{r_{\text{s,fid}}}{r_{\text{s,true}}}, \quad (4.9)$$

$$1 + \epsilon = \left[\frac{D_{\text{A,fid}}(z) H_{\text{fid}}(z)}{D_{\text{A,true}}(z) H_{\text{true}}(z)} \right]^{1/3}.$$

And we can relate the fiducial and true distances using α and ϵ using the following equations:

$$\begin{aligned} r_{\parallel,\text{true}} &= \alpha(1 + \epsilon)^2 r_{\parallel}, \\ r_{\perp,\text{true}} &= \alpha(1 + \epsilon)^{-1} r_{\perp}, \end{aligned} \quad (4.10)$$

where we have dropped subscript “fid” for simplicity. The separation between two objects can be described in Cartesian $(r_{\perp}, r_{\parallel})$ or polar coordinates (r, μ) , which are related by

$$\begin{aligned} r^2 &= r_{\perp}^2 + r_{\parallel}^2, \\ \mu &= \cos(\theta) = \frac{r_{\parallel}}{r}. \end{aligned} \quad (4.11)$$

Therefore eq. (4.11) for the true separations would be

$$\begin{aligned} r_{\text{true}} &= \alpha \sqrt{(1 + \epsilon)^4 r_{\perp}^2 + (1 + \epsilon)^{-2} r_{\parallel}^2} \\ \mu_{\text{true}} &= \frac{(1 + \epsilon)^2 r_{\parallel}}{\sqrt{(1 + \epsilon)^4 r_{\perp}^2 + (1 + \epsilon)^{-2} r_{\parallel}^2}} \end{aligned} \quad (4.12)$$

The true correlation function can be decomposed into its Legendre multipoles. In the following, we assume that monopole, quadrupole, and hexadecapole would suffice in the summation below, as the higher order moments can be ignored.

$$\xi(r_{\text{true}}, \mu_{\text{true}}) = \sum_{l=0,2,4} \xi_l(r_{\text{true}}) \mathcal{L}_l(\mu_{\text{true}}) \quad (4.13)$$

What we measure in reality, assuming a wrong fiducial cosmology, is shifted multipole moments of the correlation function,

$$\xi_l(r; \alpha, \epsilon) = \frac{2l+1}{2} \int_{-1}^1 d\mu L_l(\mu) \xi(r_{\text{true}}(\alpha, \epsilon), \mu_{\text{true}}(\alpha, \epsilon)). \quad (4.14)$$

For a BAO-only measurement, we are not interested in any feature of the correlation function other than where the BAO is. Typically, one adds an additive polynomial term $A(r)$ to account for the scale-dependent bias and redshift space distortions, and a multiplicative $B(r)$ factor to adjust the amplitude,

$$\xi_l^{\text{model}}(r; \alpha, \epsilon) = B(r) \xi_l(r; \alpha, \epsilon) + A(r) \quad (4.15)$$

Using the right form of polynomials is important, because even though $A(r)$ and $B(r)$ are nuisance parameters not containing any BAO information, they may bias our estimation of the BAO peak if not properly fitted for. We use three orders of polynomials: $A(r) = a_1/r^2 + a_2/r + a_3$, and a constant for the amplitude $B(r) = B_0$. Substituting eq. (4.15) into eq. (4.8) gives our final model for the monopole and quadrupole of a catalogue contaminated by some fraction of interlopers f_i

$$\begin{aligned} \xi_0^{\text{model}}(r; \{a_1, a_2, a_3, B_0, \alpha, \epsilon, f_i\}) \\ = (1 + 2f_i^2 - 2f_i) \times \left[\left(\frac{a_1}{r^2} + \frac{a_2}{r} + a_3 \right) + B_0 \xi_0(r; \alpha, \epsilon) \right] \\ + \frac{2f_i(1 - f_i)}{(1 + 2f_i^2 - 2f_i)} \times \mathcal{M}_0[\xi^{\text{measured}}](r), \end{aligned} \quad (4.16)$$

$$\begin{aligned} \xi_2^{\text{model}}(r; \{w_1, w_2, w_3, B_0, \alpha, \epsilon, f_i\}) \\ = (1 + 2f_i^2 - 2f_i) \times \left[\left(\frac{w_1}{r^2} + \frac{w_2}{r} + w_3 \right) + B_0 \xi_2(r; \alpha, \epsilon) \right] \\ + \frac{2f_i(1 - f_i)}{(1 + 2f_i^2 - 2f_i)} \times \mathcal{M}_2[\xi^{\text{measured}}](r). \end{aligned} \quad (4.17)$$

Throughout this paper, when we refer to a monopole only fit, we use eq. (4.16) assuming $\epsilon = 0$, which simplifies $\xi_0(r; \alpha, \epsilon)$ to just $\xi_0(\alpha r)$. In that case, there are six free parameters that we fit for. When we refer to simultaneous monopole and quadrupole fit, we use eq. (4.16) and (4.17) together. In that case, there are four additional free parameters (ten in total) that should be fitted for.

In the following, we describe the method used to obtain the templates of the monopole and quadrupole, $\xi_0(r)$ and $\xi_2(r)$. We model the matter power spectrum in real space as

$$P(k, \mu) = [P_{\text{lin}}(k) - P_{\text{smooth}}(k)] e^{-\frac{1}{2}k^2(\mu^2\Sigma_{\parallel}^2 + (1-\mu^2)\Sigma_{\perp}^2)} + P_{\text{smooth}}(k), \quad (4.18)$$

where $P_{\text{lin}}(k)$ is the linear matter power spectrum obtained from CAMB and $P_{\text{smooth}}(k)$ is the no-wiggle counterpart computed as in [20], where the BAO oscillations have been removed. The Gaussian term accounts for the damping of the BAO wiggles due to non-linear evolution, with $\Sigma_{\perp} = 4.8 h^{-1} \text{Mpc}$ and $\Sigma_{\parallel} = 7.3 h^{-1} \text{Mpc}$ at $z = 1$ and the considered cosmology. In redshift space, the power spectrum is distorted by the Kaiser [21] and the Finger-of-God (FoG) [22] effects, which lead us to the following equation for the redshift space power spectrum [23]

$$P_s(k, \mu) = b^2(1 + \beta\mu^2)^2 F(k, \mu) P(k, \mu). \quad (4.19)$$

In our analysis, we use the streaming model for FoG, $F(k, \mu) = (1 + k^2\mu^2\Sigma_s^2)^{-2}$, with $\Sigma_s = 3 h^{-1} \text{Mpc}$. Subsequently, we calculate the correlation function by taking the Fourier transform of the redshift-space power spectrum.

5 Results

In this section, we present the results obtained by fitting the mean of the multipole moments of the correlation functions from 1,000 mocks in redshift space. To find the best fit parameters, we carried out a Markov-chain Monte Carlo (hereafter, MCMC) analysis (emcee package [24]) to minimize the posterior function introduced by Percival et al. [25]:

$$f(\xi^{\text{model}} | \xi^{\text{data}}, C) \propto \left[1 + \frac{\chi^2}{n_s - 1} \right]^{-\frac{m}{2}}, \quad (5.1)$$

where m is given by eq. 54 of [25]. We also assumed the following uniform priors on our parameters:

$$f(\alpha, \epsilon, f_i) = \begin{cases} 1 & \text{if } 0.6 < \alpha < 1.4, \\ & \& -1 < \epsilon < 1, \\ & \& 0 < f_i < 0.5 \\ 0 & \text{otherwise} \end{cases} \quad (5.2)$$

The best fit values and the statistical errors presented in this section (i.e., figure 5, 6, table 1, 2) are respectively calculated from the mean and standard deviation calculated using a Monte Carlo Markov Chain used to explore the posterior with 3000 steps. To calculate χ^2 , we estimated the covariance matrix of $n_s = 1000$ independent contaminated mocks using

$$C_{ij}[\xi(r_i)\xi(r_j)] = \frac{1}{(n_s - 1)} \sum_{n=1}^{n_s} [\xi_n(r_i) - \bar{\xi}(r_i)][\xi_n(r_j) - \bar{\xi}(r_j)]. \quad (5.3)$$

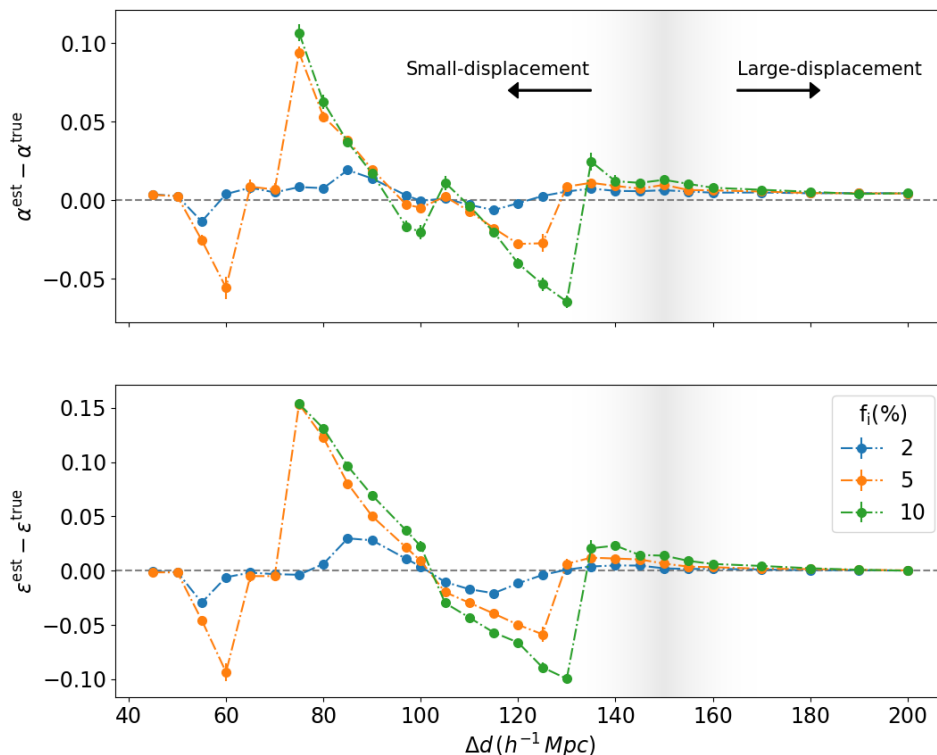


Figure 5. The effect of small-displacement interlopers on dilation parameters when interlopers are not included in the model, shown for 2, 5, and 10% fraction of interlopers. As can be seen, for small-displacements, the estimation of α (upper panel) and ϵ (lower panel) are notably affected by interlopers, and therefore, small-displacement interloper corrections need to be implemented in the fitting model to avoid systematic bias. For large-displacements, the effect of the cross-correlation on the recovered parameters rapidly decreases.

To fit the model to the mean of n_s mocks, eq. (5.3), which gives the covariance matrix used for an individual mock, is divided by n_s . For the purpose of this paper, we only need to fit the data around the BAO peak. Therefore, in our analyses, we fit our model to data on scales between $r_{\min} = 50 h^{-1} \text{ Mpc}$ and $r_{\max} = 150 h^{-1} \text{ Mpc}$.

Let us begin with highlighting the necessity of correcting for the small-displacement interloper effect, and also roughly quantifying what we mean by “small” displacements. For this purpose, we repeated the process of making contaminated catalogues described in section 3, but for a wider range of displacements, from 45 to $200 h^{-1} \text{ Mpc}$. Without correcting for the interloper effect, we found the best fit estimations on α and ϵ . In figure 5, the difference between the estimation and true is plotted against displacement, and from this figure we find that:

- How much the estimated dilation parameters deviate from their true value depends on the level of contamination of the catalogue (different colors). In the small-displacement regime, estimations are worse for higher fractions of interlopers. For 10% interlopers, the MCMC code does not even converge for displacements smaller than $75 h^{-1} \text{ Mpc}$ (hence why not shown in this plot).
- For $\Delta d \approx 100 h^{-1} \text{ Mpc}$, the best fit does not differ from 1 significantly, since the two peaks (Δd and BAO) overlap, and there is no need to change the peak location in the model by varying α . As we make Δd slightly larger (smaller) than the BAO scale, until

around $120 h^{-1}$ Mpc ($80 h^{-1}$ Mpc), α becomes smaller (larger) than 1 to allow the model without the interloper signal to compromise between fitting the BAO and the interloper cross-correlation term.

- There is a smooth transition around $\Delta d \approx 140\text{--}160 h^{-1}$ Mpc such that, at smaller displacements, the estimated dilation parameters deviate from the truth, while on larger scales they become less affected by the presence of interlopers. This is because as we increase the displacement, the cross-correlation peak phases out of the maximum fitting range of $150 h^{-1}$ Mpc, and the tail that remains within the fitting range gets corrected by polynomials, rather than the dilation parameters.
- The extrema reflect a jump in the best-fit position in the posterior between two isolated solutions with different Δd : when the cross-correlation becomes too far from the position of the BAO, the best fit solution jumps from being a compromise between fitting both BAO and spike to a solution that fits the BAO (only) better, ignoring the cross-correlation spike.

We should emphasize that large-displacements can still bias the BAO measurement, even though the cross-correlation would be negligible, as mentioned in section 1.

Now, we take into account the effect of interlopers in the fitting model by using the formalism described in section 4. Figure 6 demonstrates how much the estimation of our pipeline for α (top row), ϵ (middle row), and f_i (lower row) deviates from the corresponding true values when considering catalogs contaminated by different fractions of interlopers (x-axis), and three different values for the displacements Δd : $85 h^{-1}$ Mpc (left column), $90 h^{-1}$ Mpc (middle column), and $97 h^{-1}$ Mpc (right column). The orange symbols corresponds to the case where we account for the presence of interlopers in our analysis using eq. (4.8). We also consider green symbols which show the results when ignoring interlopers (as if f_i was set to zero in eq. (4.8), and as assumed in figure 5). We considered fits to the monopole alone (shown in triangle) and to the monopole and quadrupole simultaneous (rectangles). From figure 6, we deduce that:

- *Monopole only with no correction (green triangles)*. At a fixed value for the displacement Δd , the estimated isotropic dilation parameter α_{est} increasingly deviates from 1 when considering larger interloper fractions. This means that the bias in α introduced by the interlopers increases with the amount of contamination, as previously noted in [8]. How much the bias increases with f_i depends on the value of the displacement: the smaller the displacement, the faster the bias increases with the amount of contamination. Indeed, when the displacement is close to the BAO position, e.g. $\Delta d = 97 h^{-1}$ Mpc, the peak in the galaxy-interloper cross-correlation is very close to the BAO location and it causes a mild distortion and shift of the BAO peak. On the other hand, smaller displacements present a peak in the cross-correlation that is further away from the BAO peak location, causing a much larger distortion of the BAO feature (see also figure 2).
- *Monopole+Quadrupole with no correction (green rectangles)*. Compared to the previous case (monopole only with no correction) and for each combination of Δd and f_i^{true} , α_{est} is closer to 1 when including the quadrupole in the fit. This happens because the contamination is anisotropic: by including the quadrupole, the parameter ϵ can mimic and absorb the anisotropy, leaving α less biased. However, the ϵ measurements are strongly biased, with deviations from the fiducial value increasing with the amount of contamination and decreasing with increasing Δd .

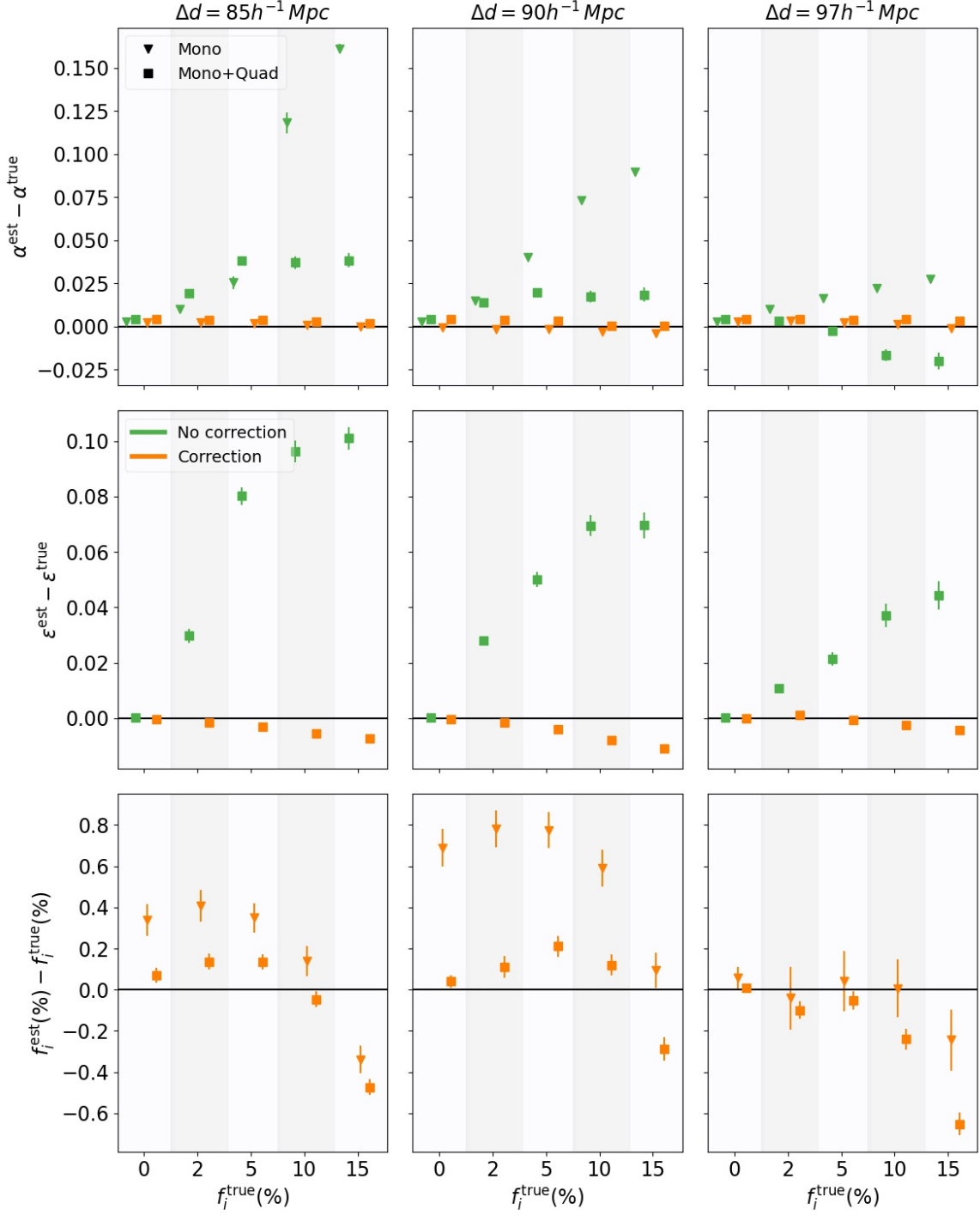


Figure 6. The difference between estimation and true for α (top row), ϵ (middle row), and f_i (bottom row), as plotted against the true fraction of interlopers. For comparison, the green lines show the fit without taking into account interlopers and the orange shows the fit with interlopers (our analysis).

- *Monopole only with interloper correction (orange triangles)*. Compared to the no correction case, the values of α_{est} are now much closer to the true value and they exhibit a systematic bias smaller than 3.9×10^{-3} for all considered level of contamination and displacements. This model allows us to obtain an estimation for the fraction of interlopers f_i^{est} with a bias that is below 8×10^{-3} for all levels of contamination and displacements. (More details in table 2)
- *Monopole+Quadrupole with interloper correction (orange rectangles)*. The values of α_{est} and ϵ_{est} are much closer to the truth than in the no correction case. The residual systematic biases are smaller than 5×10^{-3} for α and 0.01 for ϵ . The percentage of contaminants can be estimated with better precision in this case than when fitting for the monopole alone: the statistical errors decrease by a factor of two on average. The accuracy on f_i is also generally improved when the displacement is small, while no significant amelioration appears if $\Delta d = 97 h^{-1}$ Mpc.

A more detailed description of our results of the monopole+quadrupole and monopole only analyses are given in table 1 and table 2 respectively, where the systematic and statistical errors on α , ϵ and f_i are reported for all the displacements and fractions considered. We find that the systematic error on α with interlopers using our model is 4.4×10^{-3} which is consistent with pre-reconstruction BAO analyses on SDSS data without interlopers ([26–28]). This shows that slitless spectroscopy does not degrade BAO measurements.

6 Discussions

We have investigated how to extract the BAO scale from a contaminated galaxy catalogue containing a sub-sample of interlopers that are correlated with the main target galaxies. This case is particularly important for the future Roman Space Telescope, where a population of [O III] emitting galaxies extracted thanks to the observation of a single line is likely to be contaminated by $H\beta$ interlopers. In this study, we assumed that the $H\beta$ line is misidentified as the primary line at 500.7 nm of the [O III] doublet. Depending on how the doublet is fitted, one might need to take into account the weighted average of the primary and the secondary lines of [O III]. Nonetheless, this will only impact the value of Δd as a function of redshift in eq. (2.2), and will not affect our conclusions. In the following, we highlight some of the main outcomes of our analysis.

Generally, the contaminated correlation function has three terms (eq. (4.2)): the galaxy-galaxy term, the interloper-interloper term, and the galaxy-interloper term. We have split interlopers into three categories: random, large-displacement and small-displacement interlopers. We have argued that random contaminants, uncorrelated with either themselves or the target galaxies are trivial to account for.

For both small-displacement and large-displacement interlopers, we need to model their auto-correlation, but this is generally easy to do. In addition, the effect of the interloper-interloper auto-correlation on the contaminated signal is a simple linear super-position of two correlation functions: in the extremes, either the change in $H(z)$ between true redshifts of the targets and interlopers is small and so the BAO position is unaffected, or it is large and the superimposed BAO peak is far from that of the targets. In either case and in between, to measure the BAO position, we only need a model on large scales where linear theory is sufficient to describe the BAO feature and we know the position through the relative line positions. If interlopers and galaxies have different bias, an extra (bias) parameter

$f_i^{\text{true}} (\times 10^{-2})$	Error on $\alpha (\times 10^{-3})$		Error on $\epsilon (\times 10^{-3})$		Error on $f_i (\times 10^{-3})$	
	systematic	statistical	systematic	statistical	systematic	statistical
$\Delta d = 85h^{-1}\text{Mpc}$						
0	4.1	0.7	-0.5	0.9	0.7	0.4
2	3.9	0.8	-1.5	1.1	1.4	0.4
5	3.5	0.8	-3.1	1.1	1.4	0.4
10	2.5	0.8	-5.4	1.0	-0.5	0.4
15	1.6	0.9	-7.3	1.1	-4.7	0.4
$\Delta d = 90h^{-1}\text{Mpc}$						
0	4.1	0.7	-0.5	1.0	0.4	0.3
2	3.6	0.9	-1.5	1.4	1.1	0.5
5	3.0	0.8	-4.1	1.3	2.1	0.5
10	0.3	0.8	-7.9	1.3	1.2	0.5
15	0.3	0.9	-10.8	1.5	-2.9	0.6
$\Delta d = 97h^{-1}\text{Mpc}$						
0	4.3	0.7	0.1	0.8	0.1	0.1
2	4.4	0.7	1.1	0.9	-1.0	0.4
5	3.5	0.8	-0.7	1.1	-0.5	0.5
10	3.9	1.1	-2.5	1.5	-2.4	0.5
15	3.4	1.2	-4.3	1.7	-6.5	0.6

Table 1. Systematic and statistical errors on α , ϵ , and f_i using monopole and quadrupole. Analysis is performed on the mean of 1000 mocks in redshift space, and different displacements.

can be added to the second term of eq. (4.2) and can be fitted for along other parameters introduced in our analysis. It also may be possible to measure this using calibration data, by comparing the interloper-interloper and target-target correlation functions. In contrast, for small-displacement interlopers, it is critical to have a good estimate of the interloper-target cross-correlation.

For our simulations, the interlopers are drawn from the same underlying distribution as the target galaxies. This allowed us to simplify the equations, merging the galaxy-galaxy and interloper-interloper terms together. In practice, the populations of [O III] and H β emission line galaxies are different, and this can change our equations requiring us to keep these terms separate. We do not expect this to significantly alter our conclusions, or to present a difficulty for modeling.

For small-displacement interlopers, the interloper shift can be considered as being constant, and is dependent on the fiducial cosmology used to convert redshifts to distances. Thus its position does not provide cosmological information. We made the assumption of a constant shift because the displacement only changes by less than one percent across the bin considered, similar to the case for the BAO scale being fixed within an bin of observational data. Hence, the BAO signal in $\xi_{ii}(\vec{r})$ will not be shifted as a correlation function is not

$f_i^{\text{true}} (\times 10^{-2})$	Error on $\alpha (\times 10^{-3})$		Error on $f_i (\times 10^{-3})$	
	systematic	statistical	systematic	statistical
$\Delta d = 85h^{-1}\text{Mpc}$				
0	2.2	0.7	3.4	0.8
2	2.1	0.7	4.1	0.8
5	1.9	0.7	3.5	0.7
10	0.8	0.7	1.4	0.7
15	-0.1	0.8	-3.4	0.7
$\Delta d = 90h^{-1}\text{Mpc}$				
0	-0.6	0.8	6.9	0.9
2	-1.4	0.8	7.8	0.9
5	-1.5	0.9	7.7	0.9
10	-3.0	0.9	5.9	0.9
15	-3.9	0.9	1.0	0.8
$\Delta d = 97h^{-1}\text{Mpc}$				
0	2.7	0.7	0.6	0.5
2	3.2	0.9	-0.4	1.5
5	2.4	0.9	0.4	1.5
10	1.3	1.0	0.1	1.4
15	-0.9	1.2	-2.4	1.5

Table 2. Systematic and statistical errors on α , and f_i using the monopole only analysis performed on the mean of 1000 mocks in redshift space, and different displacements.

modified by a translation, while it is modified in $\xi_{gi}(\vec{r})$. Thus, the key to understanding and modelling the contaminated correlation function is through accurately modelling the cross-correlation between the main galaxy targets and the interlopers that are misidentified as main targets, which is described in eq. (4.4) and required a mapping between true and wrongly inferred positions for the interlopers. Again, with calibration data, it is possible to apply this mapping (refer to (4.5)) to the measured cross-correlation at the true distance. We leave it to future work to determine the size of calibration data required to determine this function with sufficient accuracy that it does not impact on BAO measurements. The good news is that we need to know this function on small scales, where small volumes of the Universe will be sufficient.

For our study, we used the same assumptions used to create the simulations (the [O III] and H β emitting galaxies are drawn from the same population) to estimate the cross-correlation with the auto-correlation function of the main galaxy targets. We considered models based on CAMB and HALOFIT to derive an estimation of the cross-correlation. We found that they both fail in precisely modelling the peak of the monopole and quadrupole of the cross-correlation. These models also need to account for the difference between fiducial and true cosmologies via the AP parameters, that would need to be added in the cross-correlation

function of eq. (4.16), and eq. (4.17) as well. Therefore, we tested another possibility that is more powerful on smaller scales: using the measured contaminated correlation functions to estimate cross-correlation. We argued that using the latter is the most precise for the Quijote mocks used in this paper, since it matches the way that the mocks were constructed (interlopers and target galaxies have the same galaxy bias). Indeed, the resulting cross-correlation is able to better model the peak of the monopole of the cross-correlation, and both peak and the shape of the quadrupole of the cross-correlation (refer to figure 4).

With all of the mentioned assumptions made, we finally presented a model for monopole and quadrupole of the contaminated correlation function. We tested our model by fitting it to the mean of 1,000 mocks for five different interloper fractions and three different displacements. We found that for all these cases, our method is successful at making an unbiased prediction of the dilation parameters (recovering the BAO peak), with a systematic error of less than 4.4×10^{-3} , which is consistent with previous pre-reconstruction analyses. It also enables us to have a good estimation of the fraction of interlopers. Whereas if we did not take interlopers into account in the model, for 15% interlopers we would have an order of magnitude higher systematic errors — as high as -0.020 and 0.044 for α and ϵ . Moreover, the signal in the correlation function from interlopers is sufficiently clear that it allows us to measure the fraction of interlopers with a systematic error of at most few percent. This will be useful for cosmological analysis other than measuring BAO where the fraction of interlopers can change the amplitude of the signal, and knowing the fraction is important.

Since our analysis is specifically useful for future Roman [O III] targets that are contaminated by $H\beta$ interlopers, we predict the statistical error that our method would have for Roman telescope data. To do so, we scaled the statistical errors on the mean calculations from 1,000 mocks at redshift 1 ($\Delta d = 85 h^{-1} \text{Mpc}$). These 1,000 mocks have a cumulative volume of $1,000(h^{-1}\text{Gpc})^3$. The approximate volume of Roman survey between redshift 1.8 and 2.8 (where [O III] galaxies live) is $V \approx 10(h^{-1}\text{Gpc})^3$. Thus, the statistical error is scaled by $\sqrt{1000/10} = 10$. For the range of 0–15% interlopers, we find that the statistical error on α and ϵ is 0.01 which is larger than their corresponding systematic errors.

Small-displacement interlopers give rise to a strong signal in the correlation function on similar scales to the cosmological signal. Even so, this signal has key differences from the cosmological signal such that a joint model fitting both the cosmological signal and the effect of interlopers can be constructed and we do not see a degradation in the cosmological measurements in the tests we have performed. Furthermore, the interloper signal itself allows excellent diagnostics on the contamination, allowing a simultaneous fit to the contaminant fraction. For random interlopers however, it is more difficult to measure the contaminant fraction, which has to be done before the amplitude of clustering can be measured. Thus, while small-displacement interlopers are definitely the most pernicious in terms of the correlation function, for fits to the data this is not the case.

Acknowledgments

We thank Yun Wang for useful comments and discussions. Research at Perimeter Institute is supported in part by the Government of Canada through the Department of Innovation, Science and Economic Development Canada and by the Province of Ontario through the Ministry of Colleges and Universities. This research was enabled in part by support provided by Compute Ontario (<https://www.computeontario.ca>) and the Digital Research Alliance of Canada (<https://alliancecan.ca>).

References

- [1] D. Spergel et al., *Wide-Field InfraRed Survey Telescope-Astrophysics Focused Telescope Assets WFIRST-AFTA Final Report*, [arXiv:1305.5422](#) [INSPIRE].
- [2] EUCLID collaboration, *Euclid Definition Study Report*, [arXiv:1110.3193](#) [INSPIRE].
- [3] Y. Wang et al., *The High Latitude Spectroscopic Survey on the Nancy Grace Roman Space Telescope*, *Astrophys. J.* **928** (2022) 1 [[arXiv:2110.01829](#)] [INSPIRE].
- [4] D.J. Farrow et al., *Correcting correlation functions for redshift-dependent interloper contamination*, *Mon. Not. Roy. Astron. Soc.* **507** (2021) 3187 [[arXiv:2104.04613](#)] [INSPIRE].
- [5] H.S. Grasshorn Gebhardt et al., *Unbiased Cosmological Parameter Estimation from Emission Line Surveys with Interlopers*, *Astrophys. J.* **876** (2019) 32 [[arXiv:1811.06982](#)] [INSPIRE].
- [6] Y. Gong, H. Miao, P. Zhang and X. Chen, *Self-calibrating Interloper Bias in Spectroscopic Galaxy-clustering Surveys*, *Astrophys. J.* **919** (2021) 12 [[arXiv:2107.04745](#)] [INSPIRE].
- [7] A.R. Pullen, C.M. Hirata, O. Doré and A. Raccanelli, *Interloper bias in future large-scale structure surveys*, *Publ. Astron. Soc. Jap.* **68** (2016) 12 [[arXiv:1507.05092](#)] [INSPIRE].
- [8] E. Massara, S. Ho, C.M. Hirata, J. DeRose, R.H. Wechsler and X. Fang, *Line confusion in spectroscopic surveys and its possible effects: shifts in Baryon Acoustic Oscillations position*, *Mon. Not. Roy. Astron. Soc.* **508** (2021) 4193 [[arXiv:2010.00047](#)] [INSPIRE].
- [9] A. Lewis, A. Challinor and A. Lasenby, *Efficient computation of CMB anisotropies in closed FRW models*, *Astrophys. J.* **538** (2000) 473 [[astro-ph/9911177](#)] [INSPIRE].
- [10] Z. Zhai, A. Benson, Y. Wang, G. Yepes and C.-H. Chuang, *Prediction of $H\alpha$ and $[OIII]$ emission line galaxy number counts for future galaxy redshift surveys*, *Mon. Not. Roy. Astron. Soc.* **490** (2019) 3667 [[arXiv:1907.09680](#)] [INSPIRE].
- [11] F. Villaescusa-Navarro et al., *The Quijote simulations*, *Astrophys. J. Suppl.* **250** (2020) 2 [[arXiv:1909.05273](#)] [INSPIRE].
- [12] N. Hand et al., *nbbodykit: an open-source, massively parallel toolkit for large-scale structure*, *Astron. J.* **156** (2018) 160 [[arXiv:1712.05834](#)] [INSPIRE].
- [13] S.D. Landy and A.S. Szalay, *Bias and variance of angular correlation functions*, *Astrophys. J.* **412** (1993) 64 [INSPIRE].
- [14] A.R. Pullen, C.M. Hirata, O. Doré and A. Raccanelli, *Interloper bias in future large-scale structure surveys*, *Publ. Astron. Soc. Jap.* **68** (2015) 12.
- [15] VIRGO CONSORTIUM collaboration, *Stable clustering, the halo model and nonlinear cosmological power spectra*, *Mon. Not. Roy. Astron. Soc.* **341** (2003) 1311 [[astro-ph/0207664](#)] [INSPIRE].
- [16] R. Takahashi, M. Sato, T. Nishimichi, A. Taruya and M. Oguri, *Revising the Halofit Model for the Nonlinear Matter Power Spectrum*, *Astrophys. J.* **761** (2012) 152 [[arXiv:1208.2701](#)] [INSPIRE].
- [17] N. Padmanabhan and M.J. White, *Constraining Anisotropic Baryon Oscillations*, *Phys. Rev. D* **77** (2008) 123540 [[arXiv:0804.0799](#)] [INSPIRE].
- [18] X. Xu, A.J. Cuesta, N. Padmanabhan, D.J. Eisenstein and C.K. McBride, *Measuring D_A and H at $z = 0.35$ from the SDSS DR7 LRGs using baryon acoustic oscillations*, *Mon. Not. Roy. Astron. Soc.* **431** (2013) 2834 [[arXiv:1206.6732](#)] [INSPIRE].
- [19] C. Alcock and B. Paczyński, *An evolution free test for non-zero cosmological constant*, *Nature* **281** (1979) 358.
- [20] Z. Vlah, U. Seljak, M.Y. Chu and Y. Feng, *Perturbation theory, effective field theory, and oscillations in the power spectrum*, *JCAP* **03** (2016) 057 [[arXiv:1509.02120](#)] [INSPIRE].

- [21] N. Kaiser, *Clustering in real space and in redshift space*, *Mon. Not. Roy. Astron. Soc.* **227** (1987) 1 [INSPIRE].
- [22] J.C. Jackson, *A critique of rees's theory of primordial gravitational radiation*, *Mon. Not. Roy. Astron. Soc.* **156** (1972) 1P.
- [23] C. Blake et al., *The WiggleZ Dark Energy Survey: the growth rate of cosmic structure since redshift $z = 0.9$* , *Mon. Not. Roy. Astron. Soc.* **415** (2011) 2876 [arXiv:1104.2948] [INSPIRE].
- [24] D. Foreman-Mackey, D.W. Hogg, D. Lang and J. Goodman, *emcee: The MCMC Hammer*, *Publ. Astron. Soc. Pac.* **125** (2013) 306 [arXiv:1202.3665] [INSPIRE].
- [25] W.J. Percival, O. Friedrich, E. Sellentin and A. Heavens, *Matching Bayesian and frequentist coverage probabilities when using an approximate data covariance matrix*, *Mon. Not. Roy. Astron. Soc.* **510** (2022) 3207 [arXiv:2108.10402] [INSPIRE].
- [26] J.E. Bautista et al., *The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: measurement of the BAO and growth rate of structure of the luminous red galaxy sample from the anisotropic correlation function between redshifts 0.6 and 1*, *Mon. Not. Roy. Astron. Soc.* **500** (2020) 736 [arXiv:2007.08993] [INSPIRE].
- [27] M. Vargas Magaña et al., *SDSS-III Baryon Oscillation Spectroscopic Survey: Analysis of Potential Systematics in Fitting of Baryon Acoustic Feature*, arXiv:1312.4996 [INSPIRE].
- [28] BOSS collaboration, *The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: cosmological analysis of the DR12 galaxy sample*, *Mon. Not. Roy. Astron. Soc.* **470** (2017) 2617 [arXiv:1607.03155] [INSPIRE].