

# Analysis of the response of void BAO to systematic effects in the SDSS observations using mock datasets

Daniel Forero-Sánchez,<sup>1</sup><sup>★</sup>

<sup>1</sup>École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland.

Accepted XXX. Received YYY; in original form ZZZ

## ABSTRACT

Baryon Acoustic Oscillations (BAO) were produced due to (baryonic) matter – photon plasma oscillations in the early Universe and they are therefore encoded in its spatial 3-dimensional distribution. The signal imprinted recognizable in configuration space as a peak in the 2-point correlation function and in Fourier space as wiggles in the power spectrum. The BAO feature provides a standard scale that can be measured at different redshifts. In recent years, large volume galaxy surveys have dedicated efforts to the mapping of the large-scale structure by using different kinds of matter tracers (such as galaxies), however, these observations are always clouded by systematical effects coming from observational difficulties and instrument limitations. In the present work, we aim to evaluate the impact of these systematical effects in the void BAO signal in configuration space (as obtained from the discrete distribution of Emission Line Galaxies (ELG) using a Delaunay Triangulation algorithm) as opposed to the galaxy BAO. To do this we compute the optimal radius cut to separate voids-in-voids (bigger, true voids) from voids-in-clouds (smaller, dark matter-rich regions) and find it to be  $R_c = 15.5 h^{-1} \text{Mpc}$ . We measure the BAO scale by fitting the dilation parameter to different models. Measurements hint towards a difference in the impact of systematics on voids vs. galaxies. Nonetheless the error is still relatively high to draw definitive conclusions. We measure, in the mean case,  $\alpha_{\text{all}} - \alpha_{\text{none}} = (3.34 \pm 2.89) \times 10^{-3}$  for galaxies and  $\alpha_{\text{all}} - \alpha_{\text{none}} = (-9.9 \pm 72.9) \times 10^{-4}$  for voids. Such small differences suppose this analysis should be done with other tracers with better signal-to-noise ratio in the BAO feature such as LRGs.

**Key words:** methods: data analysis – cosmology: large-scale structure of Universe – cosmology: distance scale

## 1 INTRODUCTION

Baryon Acoustic Oscillations (BAO) provide a *standard ruler* or scale that can be tracked at different redshifts, its measurement is then of This is due to the fact that it allows us to further improve the constraints on cosmological parameters (Bassett & Hlozek 2010). The signal is encoded in the spatial distribution of matter in the Universe. Traditionally, different kinds of massive cosmic objects or *tracers*, as well as the anisotropies in the Cosmic Microwave Background (CMB) are used to measure the scale.

While projects like COBE, WMAP and Planck have deeply studied the characteristics of the CMB, redshift surveys have gathered data regarding the distribution of matter tracers in the observable universe. Projects such as the 2dF survey (Colless et al. 2001), the Sloan Digital Sky Survey (SDSS) (Dawson et al. 2015), the Dark

Energy Spectroscopic Instrument<sup>1</sup> (DESI) and Euclid<sup>2</sup> pretend to generate precise 3D maps of the universe (or have done so). To do this, they identify different tracers via their spectroscopic signature which also allows them to compute their redshift  $z$ . Ideally, different tracer populations will allow for the measurement of the BAO signature at different redshifts, allowing us to track the expansion history of the Universe. These surveys are expensive both in resources and in time.

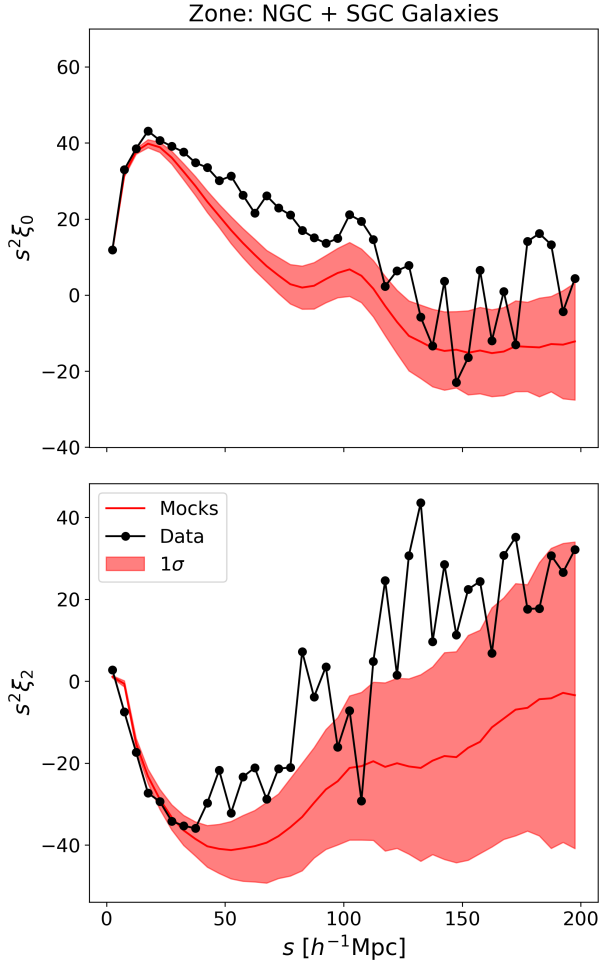
Recent work by Zhao et al. (2018) has shown that the cosmic void population can be used as a tracer for the measurement of the signal and that when combined with the galaxy sample, it can yield to a  $\sim 10\%$  improvement on the error of the BAO peak position. This is comparable to a 20% enlargement in the galaxy survey size at no significant extra cost.

Furthermore, work by Kitaura et al. (2016) and Liang et al. (2016)

\* E-mail: daniel.forerosanchez@epfl.ch (DFS)

<sup>1</sup> <https://www.desi.lbl.gov/>

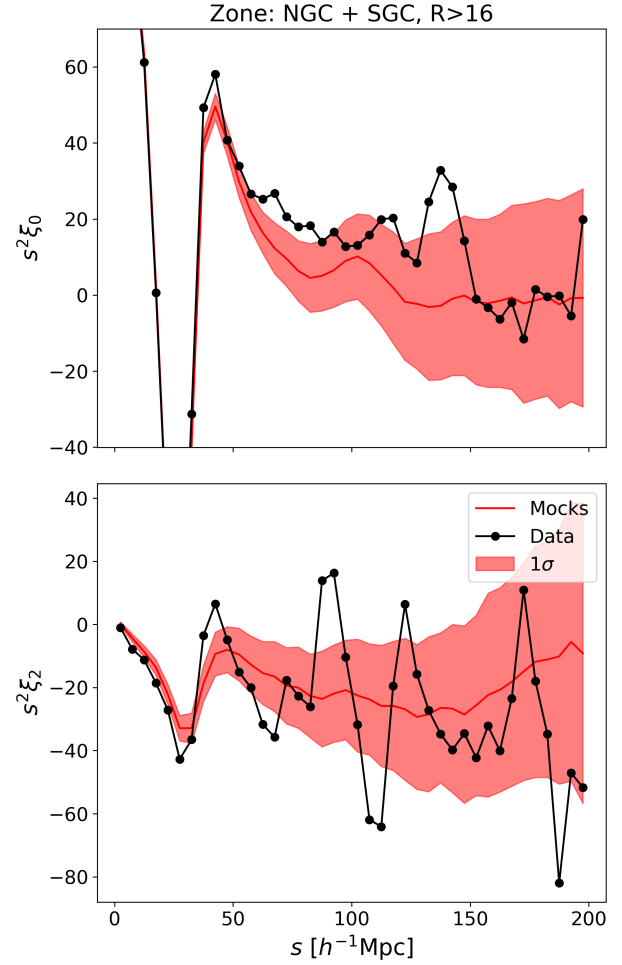
<sup>2</sup> <https://www.euclid-ec.org/>



**Figure 1.** Galaxy 2PCF from eBOSS ELG sample (black line) and ELG mocks. Top panel shows the monopole, while the bottom one shows the quadrupole. The latter shows an unexpected trend towards large scales since theories predict a negative 2PCF quadrupole in this case.

has hinted that voids can be less sensitive to observational and systematical effects unavoidably present in the galaxy sample. To illustrate this, we can see the ELG 2PCF from eBOSS data (figure 1); we would expect to have a negative quadrupole at large scales as predicted by various models (see figure 3 in White et al. (2015)), nonetheless both mocks and observations seem to show otherwise. On the contrary, figure 2 shows that this bias is not present in the void case and the large-scale behavior in the quadrupole is as expected. The difference is possibly due to how these tracers are affected by systematical effects, therefore, this work aims towards the study of the impact that these effects have on the void BAO measured from eBOSS Emission Line Galaxy (ELG) EZmocks (Chuang et al. 2015; Zhao & Chuang 2020).

This report is structured as follows. In section 2 we discuss some of the basics of modern cosmology to properly introduced the quantities to be measured and the BAO phenomenon itself. We briefly introduce a linear approximation to the evolution of the large-scale structure in the Universe. In section 3, we briefly introduce the concept of cosmic voids, mention how they are extracted from the matter distribution and how they are classified. In section 4 we briefly explain the concepts behind the EZmocks and introduce the sur-



**Figure 2.** Void 2PCF from eBOSS ELG sample (black line) and ELG mocks. Top panel shows the monopole, while the bottom one shows the quadrupole. The bias towards positive values of  $\xi_2$  observed in the galaxy 2PCF is not seen here, hinting that this tracer could be less sensitive to systematics.

vey itself. Later, section 5 discusses the method used to analyze the data, including the creation of the matter catalogs used in this study, as well as the procedure to obtain void catalogs from a given galaxy one. We then introduce the fitting procedure. Section 6 shows the obtained results and discusses their meaning. Finally, we state the conclusions in section 7 of the study and mention some future directions.

## 2 BACKGROUND THEORY

### 2.1 FLRW Metric

Modern Cosmology is greatly based on the works of Friedman, Lemaître, Robertson and Walker, who proposed the Universe and its evolution are described by the now called FLRW metric:

$$ds^2 = dt^2 - R^2(t) \left[ \frac{d\vec{r}^2}{1 - \kappa \vec{r}^2} + \vec{r}^2 d\Omega^2 \right]. \quad (1)$$

The function  $R(t)$  describes the time evolution of the spatial part of the metric, and the parameter  $\kappa$  defines its curvature. The FLRW

metric generalizes the geometry of an isotropic, homogeneous universe that can either be open ( $\kappa = -1$ , like an hyperboloid), closed ( $\kappa = +1$ , like a sphere) or flat ( $\kappa = 0$ ).

We should then define the dimensionless function  $a(t)$ , which describes the expansion of the universe too and is defined as  $a(t) \equiv \frac{R(t)}{R_0}$  where  $R_0 \equiv R(t_0)$ , the scale of the universe in the present,  $t_0$ . This choice is made such that  $a(t_0) \equiv a_0 = 1$ . Note that in equation 1, the coordinate  $\bar{r}$  is dimensionless.

To cast the metric (eq. 1) into a more insightful form, we define a new coordinate

$$\chi \equiv \frac{d\bar{r}}{\sqrt{1-\kappa\bar{r}^2}}. \quad (2)$$

The new definition implies that the form of  $\bar{r}$  depends on the curvature of the Universe. It is trivial to show that

$$\bar{r} = S_\kappa(\chi) = \begin{cases} \sinh(\chi), & \kappa = -1 \Rightarrow \text{open universe} \\ \chi, & \kappa = 0 \Rightarrow \text{flat universe} \\ \sin(\chi), & \kappa = +1 \Rightarrow \text{closed universe} \end{cases} \quad (3)$$

In this way, the (spatial) length element is

$$dl^2 = R^2(t) [d\chi^2 + S_\kappa(\chi)^2 d\Omega^2], \quad (4)$$

so we can easily see that the cases in equation 3 correspond, from top to bottom, to a 3D hyperboloid, plane and sphere.

### 2.1.1 Friedmann Equations

Given a metric,  $g_{\mu\nu}$ , the gravitational evolution of the system is described by the Einstein equations

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}\tilde{R}g_{\mu\nu} - \Lambda g_{\mu\nu} = 8\pi G T_{\mu\nu}, \quad (5)$$

where  $G_{\mu\nu}$  is the Einstein tensor,  $R_{\mu\nu}$  is the Ricci curvature tensor,  $\tilde{R}$  is the Ricci scalar,  $T_{\mu\nu}$  is the stress-energy tensor,  $G$  is the gravitational constant and  $\Lambda$  is the cosmological constant.

The Friedmann equations are precisely the Einstein equations when the metric is FLRW coupled with a dust-type (no shear) stress-energy tensor

$$T_{\mu\nu} = (\rho + p)u_\mu u_\nu + p g_{\mu\nu}, \quad (6)$$

with  $\rho$  the energy density,  $p$  the pressure of the fluid and  $u_\mu$  the 4-velocity of the particles.

Given that this metric describes a homogeneous universe, the spatial Einstein equations all yield the same information and we are left with two relations

$$\left(\frac{\dot{R}}{R}\right)^2 = \frac{8\pi G}{3}\rho - \frac{\kappa}{R^2} + \frac{\Lambda}{3} \quad (7)$$

$$\left(\frac{\dot{R}}{R}\right)^2 = -8\pi G p - 2\frac{\ddot{R}}{R} - \frac{\kappa}{R^2} + \Lambda. \quad (8)$$

One last useful relation is given by the energy conservation. The Bianchi identity ensures that  $\nabla_\mu T^{\mu\nu} = 0$ , this implies that

$$\frac{\partial \rho}{\partial t} + 3\frac{\dot{R}}{R}(\rho + p) = 0. \quad (9)$$

Some definitions are now in order:

(i) The Hubble Parameter:

$$H(t) \equiv \frac{\dot{R}}{R} = \frac{\dot{a}}{a}$$

**Table 1.** Values for  $w$  and  $n$  for the different species. Matter comprises both dark and baryonic matter. Even though curvature is most certainly not a physical species, the definition of curvature abundance is a convenient mathematical trick. Last two columns represent the evolution of  $a$  and  $\Omega$ , details in text.

Species	$w_i$	$n_i$	$a(t)$	$\Omega_i$
Matter	0	3	$t^{2/3}$	$a^{-3}$
Radiation	1/3	4	$t^{1/2}$	$a^{-4}$
Vacuum	-1	0	$\exp(Ht)$	$a^0$
“Curvature”	-1/3	2	$t$	$a^{-2}$

(ii) The Hubble constant:

$$H_0 \equiv H(t_0) = 100 h \text{ km s}^{-1} \text{ Mpc}^{-1}.$$

$h = 0.676$  throughout this work (Planck Collaboration et al. 2018).

(iii) The critical density, a.k.a. the density at which the Universe would be flat.

$$\rho_c = \frac{3H_0^2}{8\pi G}.$$

(iv) Density parameter or Abundance:

$$\Omega = \frac{\rho}{\rho_c}$$

(v) Equation of state:

$$\rho = p/w.$$

### 2.1.2 Evolution of the Universe

The Universe consists in various species such as radiation, matter, dark matter and even dark (or vacuum) energy, each identified by its own equation of state,  $p_i = w_i \rho_i$ . The density  $\rho$  used in the Friedmann equations must be the sum over all species, which means that the abundances are also summed to obtain a total abundance  $\Omega_0 = \sum_i \Omega_{0,i}$ . We define these quantities with subscript 0 to make clear that they are measured in the present. Indeed, since the densities for different species evolve in time in a different way, the relative abundances will do so. To see this we can solve equation 9 for a particular species  $i$  to get

$$\rho_i = \rho_0 a^{-n_i}; \quad n_i = 3(1 + w_i). \quad (10)$$

We summarize the values for  $w_i$  in table 1. This derivation allows us to compute the evolution of the density of species  $i$  in the  $i$ -th domination regime (where we assume all others densities to be negligible).

In addition, the first Friedmann equation (eq. 7) can be written as

$$H(t)^2 = H_0^2 \sum_{i \in \text{species, curvature}} \Omega_{0,i}(t) \equiv H_0^2 E^2(t), \quad (11)$$

where we defined the curvature abundance,

$$\Omega_{0,c} = -\frac{\kappa}{H^2 R^2},$$

as a purely mathematical trick to write the equations more compactly.

Using equations 7 and 10, we can solve for the time dependence of the scale factor  $a$  (or equivalently,  $R$ ). It can be easily seen that

$$a(t) \propto \begin{cases} t^{2/n_i}, & n_i \neq 0 \\ \exp(Ht), & n_i = 0 \end{cases} \quad (12)$$

making it clear that the expansion of the universe is strongly related to which species is dominant. It is then clear that due to the decrease in densities for all physical species but vacuum, the universe is eventually dominated by it.

## 2.2 Distance measurement in Cosmology

When observing the sky we can see right-ascension (RA) and declination (DEC) on the sphere and the third, depth, coordinate is the redshift  $z$ . The latter describes the shift in wavelength between identical signals emitted at two different moments in cosmological time:

$$z = \frac{\lambda_0 - \lambda}{\lambda} \quad (13)$$

If we identify an initial wavelength,  $\lambda$  at some initial time  $t$  with the scale  $R(t)$  of the universe at that time, and track how it expands as a product of the universe's expansion, the wavelength  $\lambda_0$  at a posterior time  $t_0$  (say, now) is  $R_0$ . The (cosmological) redshift can then be related to the scale of the universe as

$$z = \frac{R_0 - R(t)}{R(t)} = \frac{R_0}{R(t)} - 1 \Rightarrow a(t) = \frac{1}{1+z}. \quad (14)$$

However, the redshift is not a physical distance, it can of course mix with the “usual” redshift so the depth measurements are affected by the peculiar velocities of the objects along the line of sight.

The relation between distance and redshift is non-trivial, indeed, the structure of the spacetime makes it difficult to properly define *one* distance, so multiple definitions are widely used in cosmology.

### 2.2.1 Luminosity distance, $D_L$

This definition comes from the possibility of measuring the energy flux,  $F$  [ $\text{W m}^{-2}$ ], of certain objects of known intrinsic luminosity,  $L$  [ $\text{W}$ ], the *standard candles* (SC). The light coming from the SC will lose power in the shell around it so, the measured quantity is

$$4\pi D_L^2 = \frac{L}{F}, \quad (15)$$

The effect of the universe expansion in the flux is twofold. Each photon emitted in the SC is redshifted in the way to the telescope, this amounts to a change  $F \rightarrow F/(1+z)$ . The flux also depends on the frequency with which photons hit the telescope (not the light's frequency, but how many photons hit the detector in a given time). Due to the expansion, the interval  $\delta t$  between hits is redshifted too, so  $\delta t \rightarrow \delta t(1+z)$  which in turn implies a further  $F \rightarrow F/(1+z)$ . In this framework, the area in which the initial luminosity is diffused is  $A = 4\pi R^2 \bar{r}^2$  so

$$4\pi R_0^2 \bar{r}^2 = \frac{F}{L(1+z)^2}. \quad (16)$$

By comparing equations 15 and 16 we conclude that the luminosity distance is

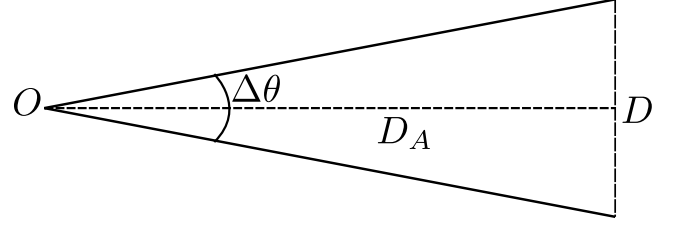
$$D_L = (1+z)R_0\bar{r} = (1+z)R_0S_\kappa(\chi). \quad (17)$$

The photons traveling from the SC do so along null geodesics, so taking  $ds = 0$  and assuming radial such trajectories ( $d\Omega = 0$ ) in equation 1 we see that

$$\chi = \int_{t'}^{t_0} \frac{dt'}{R(t')} = \frac{1}{R_0} \int_a^1 \frac{da'}{a'^2 H(a')} = \frac{1}{R_0} \int_0^z \frac{dz'}{H(z')}. \quad (18)$$

Here, it is more convenient to write the first Friedmann equation as

$$H(z) = H_0 E(z) = H_0 \sqrt{\sum_i \Omega_{0,i} (1+z)^{n_i}} \quad (19)$$



**Figure 3.** Scheme representing the diameter distance  $D_A$ . The angle  $\Delta\theta$  is the angle subtended by the proper distance  $D$ . The point labeled  $O$  represents the observer.

### 2.2.2 Angular diameter distance, $D_A$

Alternatively to  $D_L$  the diameter distance  $D_A$  can be measured based on the proper diameter  $D$  of an object as seen by the observer, this is, the subtended angle  $\Delta\theta$ . Figure 3 shows that  $\tan(\Delta\theta/2) = D/D_A$ , for small angles, as the ones subtended by objects in the sky, it is enough to use the first order  $\Delta\theta = D/D_A$ .

From equation 4, noticing that  $d\chi = d\phi = 0$ , we have

$$dl^2 = R^2(t) S_\kappa^2(\chi) d\theta^2. \quad (20)$$

It is now straightforward to identify  $D = dl$  and  $d\theta = \Delta\theta$  and conclude that

$$D_A = R(t) S_\kappa(\chi) = \frac{R_0}{1+z} S_\kappa(\chi) = \frac{1}{(1+z)^2} D_L \quad (21)$$

### 2.2.3 Instantaneous physical distance, $D_P$

From equation 3 we know that the dimensionless comoving coordinate  $\bar{r} = S_\kappa(\chi)$ . We define the instantaneous physical distance as

$$D_P = R(t)\bar{r} = R(t)S_\kappa(\chi). \quad (22)$$

For small  $\chi$ ,  $S_\kappa(\chi) \approx \chi \forall \kappa$ . This implies too that equation 18 can be approximated in a Riemann fashion as  $\chi \approx \frac{t_0 - t}{R(t_0)}$ . By linearly approximating  $R(t)$  around  $t_0$  in the definition of  $z$  (eq. 14), one finds that

$$z \approx R(t)\chi H(t) = D_P H(t). \quad (23)$$

Notice how all these distances strongly rely on the definition of a cosmology, this is, the definition of the cosmological parameters  $\Omega_{0,i}$ ,  $H_0$ , etc.

### 2.2.4 The dilation scale, $D_V$

The dilation scale is an important quantity in the measurement of cosmological parameters. The dilation scale  $D_V$  describes how much a distance measurement changes when the underlying cosmology does so. In general, this relation is non-trivial and requires detailed re-computation of all distances. For small  $z$ , however, the change in the measurements can be accounted for by a simple scaling by  $D_V$  (Eisenstein et al. 2005; Tegmark et al. 2006; Bassett & Hlozek 2010). Moreover, the scaling must be different in the directions perpendicular and parallel to the line-of-sight (LOS) (Alcock & Paczyński 1979). It is then convenient to define  $D_V$  to take into account the differences in scaling as

$$D_V = \left[ (1+z)^2 D_A^2(z) \frac{z}{H(z)} \right]^{1/3}. \quad (24)$$

Now let  $r$  be the distance measured with a given cosmology,  $\mathcal{C}$ , and  $r'$  the one given by a different, yet similar cosmology,  $\mathcal{C}'$ . The scaling  $\alpha$  is given by (Anderson et al. 2014)

$$\alpha = \frac{D_V}{D'_V} \frac{r'}{r}. \quad (25)$$

### 2.2.5 Separations to distances

In the sky, we are able to observe angular and radial separations  $\Delta\theta$  and  $\Delta z$  respectively. These require, as stated before, a fiducial cosmology to be converted into distances. This can be via the relations 21 and 23 as

$$\Delta\theta = \frac{r_\perp}{(1+z)D'_A(z)}, \quad (26)$$

$$\Delta z = r_\parallel H'(z), \quad (27)$$

where the primes denote quantities extracted from the fiducial cosmology and the symbols “ $\parallel$ ” and “ $\perp$ ” denote quantities parallel and perpendicular to the LOS, respectively.

These can be combined into a single, isotropic, scale  $r \equiv (r_\perp^2 r_\parallel)^{1/3}$  by taking into account the same scaling arguments used to define  $\alpha$  in equation 25. We see then that

$$r = \left[ \{(1+z)D'_A(z)\Delta\theta\}^2 \frac{\Delta z}{H'(z)} \right]^{1/3}. \quad (28)$$

### 2.3 Baryon Acoustic Oscillations

The behavior of small matter density anomalies,  $\delta$ , in an expanding Universe is described by equation 29. Notice that it has the form of an oscillator, these acoustic oscillations generate the BAO and depend on the coupling between baryons and photons (whose relation is hidden in the speed of sound,  $c_s$ ).

$$\ddot{\delta} + 2H\dot{\delta} + \left\{ \frac{c_s^2 k^2}{a^2} - 4\pi G\rho \right\} \delta = 0 \quad (29)$$

Notice the factor in braces has a *pressure term*  $\left(\frac{c_s^2 k^2}{a^2}\right)$  and a *gravitational term*,  $(4\pi G\rho)$ . The oscillatory behavior is expected as long as  $\frac{c_s^2 k^2}{a^2} > 4\pi G\rho$ . Even though the oscillatory behavior would be possible in the absence of gravity, its presence implies that the oscillations will be dampened and that if  $k = a\frac{\sqrt{4\pi G\rho}}{c_s} \equiv k_J$  (Jeans' wavenumber), the oscillation frequency would be zero. This tells us that at small scales ( $k > k_J$ ), pressure dominates and drives the oscillatory behavior, while at large scales ( $k < k_J$ ), the gravitational term will dominate and fluctuations will simply grow (Baumann & Baumann).

Additionally, it is imperative that the baryonic plasma is coupled (electromagnetically) to the photon gas for this process to be carried on, this phenomenon will stop when such coupling disappears. This happens at the age of recombination, when the hydrogen is formed (and not destroyed by high-energy photons) and the baryonic gas becomes neutral, hence, transparent to the photons. This, in turn, implies that these oscillations were “frozen” at time  $t_{rec} \approx 4 \times 10^5$  yr, while the previously decoupled photon gas was free to leave, forming the CMB. This, rather well-known time, combined with the speed at which this pressure waves moved,  $c_s$ , can be used to define a *standard ruler* or a characteristic scale (Eisenstein & Hu 1998; Eisenstein et al. 2007), that is, a standard angular distance that can be accurately measured at different redshifts.

It is possible to compute the distance at which these oscillations were frozen, a.k.a. the *sound horizon*, more formally defined as the comoving distance the sound waves travel from  $t = 0$  to  $t = t_{rec}$  (Weinberg et al. 2013). This quantity can be easily computed knowing that the comoving distance travelled with speed  $c_s$  is

$$r_s = \int_0^{t_{rec}} \frac{c_s(t)}{a(t)} dt. \quad (30)$$

In our case,  $c_s = c/\sqrt{3(1+\mathcal{R})}$  with  $\mathcal{R} \equiv 3\rho_b/\rho_\gamma$  (Eisenstein & Hu 1998; Dodelson 2003) and the result is  $r_s \approx 150$  Mpc (Weinberg et al. 2013). Once the photon gas decoupled, the baryon gas continued to interact gravitationally with the dark matter, so we expect the latter to also show such characteristic, see figure 1 in Eisenstein et al. (2007).

Then, this BAO signature is imprinted both in the characteristic size of the temperature anomalies of the CMB and in the clustering of galaxies. The analysis of the former yields values of  $r_s(z_{rec}) = (146.8 \pm 1.8)$  Mpc and  $r_s(z_{drag}) = (153.3 \pm 2.0)$  Mpc (Komatsu et al. 2009). In the latter case, the sound horizon can be statistically extracted from the 3D distribution of matter by means of the *two-point correlation function* (2PCF).

### 2.4 Two-point correlation function and the power spectrum

The 2PCF is formally defined as the probability over random that two particles within volume elements  $dV_{1,2}$  be separated a distance  $s$  (Bernardeau et al. 2002). This is,

$$dP = n^2 [1 + \xi(s)] dV_1 dV_2. \quad (31)$$

If there is no interaction among particles, we expect  $\xi = 0$  and a completely random distribution. In the presence of interactions, correlations arise and  $\xi \neq 0$ . If the particles are correlated, we will see an excess probability  $\xi > 0$ , while anticorrelation would imply a probability deficiency  $\xi < 0$ .

Naturally, we can write the 2PCF as the correlator of the density (anomaly) field,  $\delta(\mathbf{x})$ :

$$\xi(s) = \langle \delta(\mathbf{x})\delta(\mathbf{x}+\mathbf{s}) \rangle. \quad (32)$$

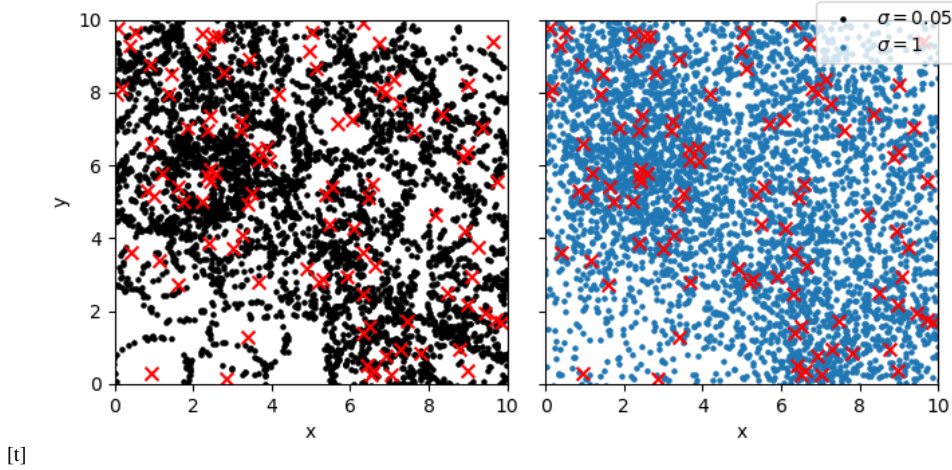
We can easily regard this function as the histogram of the separations between pairs of galaxies (Peebles 1980; Eisenstein et al. 2005).

The freezing of the oscillatory behavior described by equation 29 at decoupling, implies that matter would be clustered at the sound horizon scale, meaning that the 2PCF would show a peak at  $r_s$ . If the peak were very narrow, we would be able to see the circles in the sky, however, for a wider peak, the circles are indistinguishable and the peak position can only be retrieved statistically (Bassett & Hlozek 2010). Figure 4 illustrates this. A narrow peak (left panel) in the distribution of separations can lead to a tracer distribution showing the preferred separation  $s$ . On the other hand, a wider peak smears the signal until it is impossible to recognize (right panel in the same figure). The 2PCF was modeled as a simple gaussian  $\mathcal{N}(\mu, \sigma)$  in these cases.

An equivalent description of the system can be made in Fourier space (as opposed to the *conformal* space we have been using). We can start by Fourier transforming the density distribution (following the Fourier transform convention used by Bernardeau et al. (2002))

$$\delta(\mathbf{k}) = \int d^3\mathbf{x} \delta(\mathbf{x}) \exp(i\mathbf{k} \cdot \mathbf{x}) \quad (33)$$





**Figure 4.** Illustration of the effect of the width of the peak in the distribution of tracers. Left panel shows the distribution of tracers (dots) at a distance  $s \sim \mathcal{N}(1, 0.05)$  from the centers (cross). Right panel shows the distribution of tracers when the distances are  $s \sim \mathcal{N}(1, 1)$ . The narrow width allows for the easy identification of the preferred radius  $s$ , while this is impossible for a wider distribution.

This implies that the Fourier space correlator is given by

$$\langle \delta(\mathbf{k})\delta(\mathbf{k}') \rangle = \int d^3\mathbf{x} d^3\mathbf{s} \langle \delta(\mathbf{x})\delta(\mathbf{x}+\mathbf{s}) \rangle \exp(i\mathbf{k} \cdot \mathbf{x} + i\mathbf{k}' \cdot [\mathbf{x}+\mathbf{s}]) \quad (34)$$

$$\langle \delta(\mathbf{k})\delta(\mathbf{k}') \rangle = \int d^3\mathbf{x} d^3\mathbf{s} \xi(s) \exp(i\mathbf{k}' \cdot \mathbf{s}) \exp(i\mathbf{x} \cdot [\mathbf{k}+\mathbf{k}']) \quad (35)$$

$$\langle \delta(\mathbf{k})\delta(\mathbf{k}') \rangle \equiv \delta_D(\mathbf{k}+\mathbf{k}')P(\mathbf{k}'), \quad (36)$$

where we have defined the *power spectrum*  $P(\mathbf{k})$ , often noted  $|\sigma_k|^2$  in the literature owing to the relation between the variance of the fluctuations and the two-point correlator.

In general, the power spectrum can be parametrized as (Particle Data Group et al. 2018)

$$P(k) \propto k_s^n T^2(k), \quad (37)$$

where  $n_s \approx 1$  is the spectral index and  $T(k)$  is the transfer function. This function encodes the highly complex evolution of the density perturbations  $\delta(k)$ . It is usually computed numerically to grasp details in the evolution. However, fitting formulae can be used in simplified cases (like a baryonless evolution). See Eisenstein & Hu (1998) for such formulae.

The most important feature in  $T(k)$  is perhaps the matter domination - radiation domination threshold, where the dependence of the energy density  $\rho$  changed due to (roughly)  $w_{\text{rad}} = 1/3 \rightarrow w_{\text{matter}} = 0$ . In fact, the power spectrum has a maximum at  $k_{\text{eq}} \propto a_{\text{eq}} H_{\text{eq}} \approx 10^{-2} h \text{Mpc}^{-1}$ .

Considering baryons, the transfer function shows oscillatory behavior which is, of course, observed in the power spectrum. This oscillations are the BAO feature. It is expected that the BAO peak in  $\xi(s)$  be translated into a series of (baryon-acoustic) oscillations in  $P(k)$ .

The BAO signal, is inherent to the matter distribution in the Universe. This means that a variety of different tracers can be used to extract it. This is rather convenient when tracking the peak position at different redshifts, allowing to evaluate the expansion history of the Universe (see Anderson et al. 2014, figure 21). Some previously used tracers are quasi-stellar-objects (QSO) (McDonald & Eisenstein 2006; Ata et al. 2018), galaxies (Eisenstein et al. 2005) and galaxy clusters (Hong et al. 2012, 2016). Recent studies (Kittaura et al. 2016; Liang et al. 2016; Zhao et al. 2018) evaluate the possibility to use cosmic voids as tracers, too.

(van de Weygaert & Schaap 2009; Zhao et al. 2016).

## 2.5 Zeldovich Approximation

Formally, the Zeldovich Approximation (ZA, Zeldovich (1970)) is the first-order solution to the particle motion in a Lagrangian Perturbation Theory (LPT) approach. In LPT, the particle positions  $\mathbf{x}$  are defined by their initial Lagrangian positions  $\mathbf{q}$  and the displacement field  $\Psi(\mathbf{q}, t)$ , this is

$$\mathbf{x}(\mathbf{q}, t) = \mathbf{q} + \Psi(\mathbf{q}, t). \quad (38)$$

If we further impose mass conservation under this transformation,  $\rho(\mathbf{x}, t) d^3\mathbf{x} = \bar{\rho} d^3\mathbf{q}$  we see that

$$\rho(\mathbf{x}, t) = \frac{\bar{\rho}}{\det(J)}; \quad J_{ij} \equiv \delta_{ij} + \frac{\partial \Psi_i(\mathbf{q}, t)}{\partial q_j}, \quad (39)$$

where  $\bar{\rho}$  is the average density (which remains constant in the Lagrangian space). And  $J$  is the Jacobian matrix, which in this case also receives the name of *deformation tensor*. In an appropriate basis, the tensor is diagonal and we can see that

$$\det(J) = \prod_i J_{ii} \approx 1 + \sum_i \frac{\partial \Psi_i(\mathbf{q}, t)}{\partial q_i} = 1 + \nabla \cdot \Psi(\mathbf{q}, t), \quad (40)$$

where the “ $\approx$ ” symbol denotes a linear approximation. We can further exploit the approximation by replacing the rhs of equation 40 in equation 39

$$\rho(\mathbf{x}, t) = \frac{\bar{\rho}}{1 + \nabla \cdot \Psi(\mathbf{q}, t)} \approx \bar{\rho} [1 - \nabla \cdot \Psi(\mathbf{q}, t)]. \quad (41)$$

If we impose a linearization in the mass density  $\rho(\mathbf{x}, t) = \bar{\rho} + \delta\rho(\mathbf{x}, t)$  it is trivial to see that

$$\delta(\mathbf{x}, t) = \frac{\delta\rho(\mathbf{x}, t)}{\bar{\rho}} = -\nabla \cdot \Psi(\mathbf{q}, t). \quad (42)$$

Then, the displacement field is related to the (Fourier space) overdensity,  $\delta(\mathbf{k})$  as

$$\Psi(\mathbf{q}) = \int \frac{d^3\mathbf{k}}{2\pi^3} \exp(i\mathbf{k} \cdot \mathbf{q}) \frac{i\hat{\mathbf{k}}}{k} \delta(\mathbf{k}). \quad (43)$$

The ZA consists on using this linear approximation, for which we assumed  $\nabla \cdot \Psi(\mathbf{q}, t) \ll 1 \Leftrightarrow \delta(\mathbf{x}, t) \ll 1$ , in a regime of large density

perturbations. From equation 40, Zeldovich (1970) predicted that, due to the low probability of having three identical eigenvalues for the deformation tensor, the density would start diverging first due to compression along one axis as the eigenvalue approached zero, leading to a planar collapse (Bernardeau et al. 2002). Filaments would then form as the second eigenvalue approaches zero and spherical collapse would need for the three eigenvalues to cause  $\rho$  to diverge in a similar manner.

The ZA then yields remarkable results at large scales, where the gravitational effects are well approximated in a linear fashion. On the contrary, when gravity affects the evolution in a highly non-linear way, the ZA and other LPT methods break down.

### 3 VOIDS

The BAO feature is certainly encoded in the matter distribution, however, we could also use the strongly related distribution of *cosmic voids*. The definition of voids is rather difficult and depends strongly on the matter tracer used (van de Weygaert & Schaap 2009). The many possible definitions of these objects has produced different kinds of void-finding algorithms. As stated by Zhao et al. (2016), we can outline some common definitions of voids as follows.

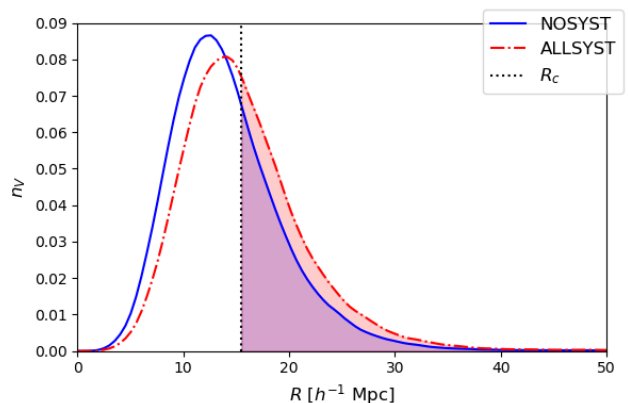
Colberg et al. (2005) define voids as regions with mean overdensity of at most  $-0.8$ . Such regions are minima of the smoothed initial density field, where spheres are located and subsequently merged according to a set of rules (see reference for details) to form the final voids. A second approach is that of Hahn et al. (2007); Forero-Romero et al. (2009), in which the definition of void is given by the local gravitational evolution as given by halo dynamics. Abel et al. (2012) analyze the cosmic web in phase space and classify regions according to the DM stream density. Voids are defined as particles that record being part of only their primordial stream (see figure 5, top-left panel in the reference). Finally, voids can be regarded as simply the empty regions in the spatial distribution of tracers. Among these, the first three depend on free parameters such as the smoothing scale of the density field (Hahn et al. 2007). Either way, all definitions agree on voids being underdensities. It is this what makes them less vulnerable to non-linearities in the gravitational evolution of the Universe.

In the present work we will be relying on the latter definition of voids. In particular, we will be using spherical voids obtained from a discrete (matter) tracer distribution via Delaunay Triangulation (DT) (Delaunay 1934). It divides the space in simplices, with each vertex defined by a tracer position, and defines voids as the circumspheres within, which do not contain tracers, by construction (Zhao et al. 2016). It is worth mentioning that these voids can overlap given their constrained spherical shape. The density of void centers can, however, be a proxy of the size of a void region.

#### 3.1 DIVE

The DIVE code (Zhao et al. 2016) is based on the DT algorithm (Delaunay 1934).

The code returns a void catalog, composed of the comoving coordinates of the centers of the voids and the respective radii. Additionally, the efficiency allows for an easy scaling to a large number of galaxy catalogs, for instance, mock catalogs from simulations. This is useful for estimating the covariance matrices of clustering measurements. Unlike some of the methods briefly presented in the



**Figure 5.** Void number density distribution for a single mock realization. Dash-dotted line shows the distribution in the case where all systematical effects have been applied (see 4.3 for details). The vertical line shows the radius cut  $R_c = 15.5 h^{-1} \text{Mpc}$ . The area to the right of the radius cut i.e. the fraction of used voids is  $\sim 30\%$  for no systematics and  $\sim 40\%$  for all systematics.

previous section, this approach does not depend on free parameters. This code is also convenient for the study of voids as tracers, since the output is a large catalog of void-centers that favors the statistical analysis of the population. Figure 5 shows the radial distribution of DT voids in one EZmock realization.

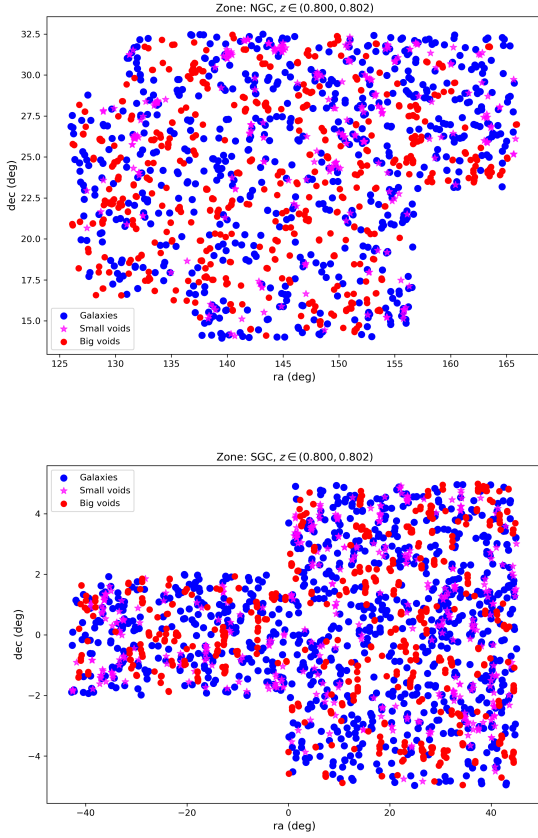
#### 3.2 Types of voids

A study (Zhao et al. 2016) of the properties of DT voids, shows that the resulting catalog can be divided, with respect to the local dark matter density anomaly,  $\delta_m$  into two distinct populations. Zhao et al. (2016) show that, in particular, small spheres ( $R < 8 h^{-1} \text{Mpc}$ ) are correlated to the population of haloes, while big voids seem anticorrelated (see figure 6). The former are called *voids-in-clouds* owing to their presence where matter density is higher, while big voids are called *voids-in-voids* since they are predominantly in empty regions of the cosmic web (Sheth & van de Weygaert 2004). Studies on the use of cosmic voids as tracers for BAO peak detection (see Liang et al. 2016) have reached the conclusion that the *signal-to-noise ratio* (SNR) in DT voids extracted from PATCHY mocks (Kitaura et al. 2014) simulating the BOSS CMASS DR11 LRG sample, is maximized for a radius cut of  $R \approx 16 h^{-1} \text{Mpc}$ . Moreover, the number density of big voids has been shown to be independent on the galaxy density (Kitaura et al. 2016; Zhao et al. 2016), which has the most significant impact on the BAO measurements as shown by Ross et al. (2017). On the contrary, small voids, seem to provide redundant information on the galaxy distribution, given that they are mostly located in high matter density regions (Kitaura et al. 2016) as mentioned previously.

### 4 DATA

#### 4.1 The eBOSS survey

The Extended Baryon Oscillation Spectroscopic Survey (eBOSS) started taking data in 2014, aiming towards a precise measurement of the BAO signal from various different matter tracers such as *Luminous Red Galaxies* (LRG), *Emission Line Galaxies* (ELG)



**Figure 6.** Spatial distribution of matter and void tracers in a  $z \in (0.800, 0.802)$  slice of the NGC (top) and SGC (bottom) eBOSS DR14 ELG sample. Big voids ( $R_c > 16 h^{-1} \text{Mpc}$  for this plot) are clustered in void regions while small voids are mainly in matter-rich regions.

and *Quasi-Stellar Objects* (QSO). These allow for the exploration of redshifts  $z \in (0.6, 2.2)$  (Dawson et al. 2015).

Figure shows the geometry of the NGC zone of the eBOSS survey. In terms of depth, the redshift coverage in the survey is divided as follows: LRG targets cover  $z \in (0.6, 1)$ , ELGs cover  $z \in (0.6, 1.1)$  with 300 plates and QSO targets cover  $z \in (0.8, 2.2)$ . Observations are expected to take a similar time to complete as the BOSS, about 5400 hr (Dawson et al. 2015).

As discussed in Raichoor et al. (2017), the target selection is dependent on some systematic effects such as stellar density, galactic extinction, or fiber collisions.

## 4.2 Mock catalogs

Mock catalogs are of paramount importance in cosmological studies involving tracer clustering, given that they allow for the estimation of the errors on the fitted parameters. Ideally, these catalogs would be generated via large n-body simulations that could properly emulate the tracer distribution regardless of the scale. However, this approach is unsustainable for a large number of tracers due to the huge computational time and memory requirements by the algorithms. This fact, added up to the increasing volume of the ongoing (e.g. DESI) and future (Euclid) surveys makes

it imperative to generate mock catalogs in a more time-efficient manner and with small trade-offs in terms of accuracy.

A number of attempts to either speed-up n-body simulations (e.g. Tashev et al. 2013) or to take gravitational effects only up to a linear regime by using linear perturbation theory (at the cost of neglecting small-scale non-linear gravitational effects) have been made.

### 4.2.1 EZmocks

The present work uses Effective Zeldovich approximation mock catalogs (EZmocks, Chuang et al. (2015)). As their name suggests, the catalogs are generated based on the Zeldovich approximation (Zeldovich 1970), which is known to accurately approximate the large-scale behavior of the tracer clustering, while performing rather less satisfactorily at small scales as briefly discussed earlier. The EZmocks, however, effectively model low-scale, non-linear behavior by imposing corrections over the initial ZA computations.

The generation of these catalogs consists on seven steps.

(i) Computation of the Zeldovich displacement field according to equation 43.

(ii) Even though, EZmocks are not n-body simulations, one such catalog is necessary to extract the probability density function (PDF) of its halo population and map it onto the previously generated displacement field. Once the PDF is extracted, a random integer is assigned to each grid point. This number will represent the number of halos to assign around the point. To populate the halo density field around the grid point (vertex) at position  $\mathbf{x}_0$ , the position  $\mathbf{x}$  of the halo is randomly generated following

$$\mathbf{x}_i - \mathbf{x}_0 = \text{sgn}(R_i) \left\{ 1 - \sqrt{|R_i|} \right\} \Delta \mathbf{x}. \quad (44)$$

The index  $i = 1, 2, 3$  runs over the grid dimensions,  $\Delta \mathbf{x}$  is the cell size and  $R_i \in (-1, 1)$  is a random number independently generated for every  $i$ . Notice that this definition only allows for placement within the eight cells around the grid point.

(iii) A density threshold and saturation is applied after:

$$\rho'_0 = \begin{cases} 0, & \rho_0 < \rho_{\text{low}}; \\ \rho_{\text{high}}, & \rho_0 > \rho_{\text{high}}, \end{cases} \quad (45)$$

with  $\rho_{\text{low}}, \rho_{\text{high}}$  the low and high thresholds. These are necessary to properly model three-point statistics (Chuang et al. 2015).

(iv) While completely sampling a new catalog from a single PDF will yield different results due to the random assignments (equation 44), it does not take into account the variance over possible PDFs. To model this effect, the halo density is scattered following

$$\rho_s = \begin{cases} \rho_0(1 + G(\lambda)), & G(\lambda) \geq 0; \\ \rho_0 \exp(G(\lambda)), & G(\lambda) < 0, \end{cases} \quad (46)$$

where  $\rho_s$  denotes the scattered density and  $\rho_0$  the ZA density.  $G(\lambda)$  is a random number sampled from a Gaussian distribution of width  $\lambda$ .

(v) For further correction of non-linear effects, the power spectrum is tuned by enhancing the BAO feature. This is done by isolating the feature and enhancing it by a (“damping”) factor  $\exp(k^2/k_*^2)$  before reintroducing the broad shape of the power



**Table 2.** Fiducial cosmology parameters used to produce the EZmocks used in this work.

Parameter	Value
$\Omega_m$	0.307115
$\Omega_b$	0.048206
$h$	0.6777
$\sigma_8$	0.8225
$n_s$	0.9611

spectrum. This is better shown by the relation

$$P_{\text{BAO}}(k) = [P_{\text{lin}}(k) - P_{\text{nw}}(k)] \exp\left(\frac{k^2}{k_*^2}\right) + P_{\text{nw}}(k), \quad (47)$$

where  $P_{\text{lin}}$  is the linear power spectrum and  $P_{\text{nw}}$  is the smooth power spectrum. An overall correction to small scales is also applied in the form

$$P_{\text{lin}}(k) = P_{\text{BAO}}(k)(1 + Ak), \quad (48)$$

with the parameter  $A$ , free.

(vi) Peculiar velocities in the  $i$ -th direction are modeled with

$$v_i(\mathbf{r}) = B' \Psi_i(\mathbf{r}) + G_i(\lambda'). \quad (49)$$

As expected, the peculiar velocities depend on the displacement field obtained from the ZA.  $B'$  is a parameter related to the linear growth ( $B' = af(a)H(a)$ , with  $f(a)$  the dimensionless linear growth rate (Zhao & Chuang 2020)) and  $G_i$  is other random number generated from a Gaussian of width  $\lambda'$ .

(vii) The particle mass assignment is done by statistically dividing the halos in the  $n$ -body simulation in sub-catalogs depending on their mass. There is a given nonzero probability that a halo will be “claimed” by either of 2 sub-catalogs. Then, EZmocks are generated for each sub-catalog with a mass randomly picked from the ones collected.

In the end, EZmocks provide a fast and accurate alternative to  $n$ -body simulations, showing a 1% agreement in terms of clustering statistics. The present work uses an updated version of the catalogs (Zhao & Chuang 2020, in preparation) which are calibrated using clustering measurements from BOSS and eBOSS data and avoids the BAO enhancement.

The ELG mocks comprise a redshift range  $z \in (0.6, 1.1)$  with a raw number density of  $n = 6.4 \times 10^{-4} h^3 \text{Mpc}^{-3}$ . The fiducial cosmology used for producing the mocks is given in table 2. They are generated from a  $5h^{-1}\text{Gpc}$  boxes cast into light-cone mocks by applying the eBOSS survey geometry. In particular, mocks are generated for each of the eboss21, eboss22, eboss23 and eboss25 chunks.

### 4.3 Systematical Effects

#### 4.3.1 Fiber collisions

For the targets’ spectra to be observed, each should be assigned a fiber. When two targets are too close (less than  $r_{cp} \equiv 62''$ ), they can’t be observed by the same plate, at the same time. In this case, only one of them is selected for measurement, however, the others’ spectra could be measured by another plate. The *tiling success rate* (TSR), defined over each sector as the number of measured targets over the number of targets quantifies this effect.

Collision groups are defined as (de Mattia 2020) sets of targets within the fiber collision radius,  $r_{cp}$  that couldn’t be observed by any plate. The *close-pair weight*,  $w_{cp}$  is defined (per collision

group) as the total number of targets over the number of observed ones. By applying this weight to the resolved objects, unobserved targets are accounted for in the final catalog.

#### 4.3.2 Redshift Failures

The success rate of the spectroscopy pipeline (SSR) is not 100%. It can be affected by factors such as observation conditions and position of the fiber in the focal plane. de Mattia (2020) defines the redshift failure weight,  $w_{noz}$  to take into account both these effects by estimating the SSR based on the median signal to noise ratio,  $\text{SSR}_{\text{obs}}$ , and the position in the focal plane  $\text{SSR}_{\text{pos}}$ . In this way,

$$w_{noz} \equiv (\text{SSR}_{\text{obs}} \text{SSR}_{\text{pos}})^{-1}.$$

#### 4.3.3 Angular photometric systematics

Various effects are grouped under this category. A HEALPIX map contains all the photometric parameters,  $p_i$ , such as galactic extinction, H I column density, Gaia stellar density, DECaLS depth and seeing parameters (de Mattia 2020), hereon  $p_i$ . A multilinear fit to the model

$$y^k = \varepsilon + \sum_i c_i p_i^k$$

is performed for all  $p_i^k$  using a squared error loss. The fit is done per chunk in the survey, this is eboss21, eboss22, eboss23 and eboss25. The first two are the SGC, while the other are in the NGC. The index  $k$  then labels the pixels in the chunk.

The weights designed to partially correct for these effects are defined as

$$w_{\text{systot}} = (y^k)^{-1}.$$

This correction is performed after applying  $w_{noz}$  and  $w_{cp}$ , so that the fit for  $y^k$  takes these corrections into account.

#### 4.3.4 Normalization and FKP weights

To effectively apply the weights, the mean of  $w_{\text{systot}}$  over all ELG objects in the chunk is normalized to one.

The completeness weights are then computed as

$$w_{\text{comp}} = w_{\text{systot}} w_{cp} w_{noz}.$$

The redshift failure weight is normalized too, such that the mean of  $w_{\text{comp}}$  over objects with reliable spectroscopic measurement and stars (with  $w_{noz} = 1$ ) is the same as the mean of  $w_{\text{comp}}$  over valid ELGs.

Invalid objects such as those with no reliable redshift or with a collided fiber are zero-weighted by setting  $w_{noz} = 0$  and  $w_{cp}$  respectively.

The final clustering sample is obtained after selecting objects with  $\text{SSR} \geq 0$ ,  $z \in (0.6, 1.1)$  and completeness (number of resolved fibers over number of objects, per sector) greater than 50%.

Due to the dependence of the tracer number density  $n(z)$  on the redshift, we must apply too a final weight

$$w_{\text{FKP}} \equiv \frac{1}{1 + n(z)P_0}; \quad P_0 = 4000 h^{-3} \text{Mpc}^3.$$

**Table 3.** Fiducial cosmology parameters used in our analysis. Given that these values differ from the fiducial cosmology used in the production of the EZmocks (which would then take the role of the “real” cosmology), we expect shifts in the BAO scale.

Parameter	Value
$\Omega_m$	0.31
$h$	0.67

#### 4.3.5 Applying systematic weights

Through the application of the weighing schemes previously mentioned, we are able to control the systematic effects present in the final catalog. To partially apply systematics we proceed as follows:

(i) We compute  $w_{\text{comp}}^{\text{ALLSYST}} = w_{\text{systot}} w_{\text{cp}} w_{\text{noz}} w_{\text{FKP}}$  and the effective number of tracers in each chunk

$$n_{\text{eff}} = \sum_{i=1}^{N_{\text{chunk}}} w_{\text{comp},i}^{\text{ALLSYST}},$$

where  $N_{\text{chunk}}$  is the number of tracers in the chunk considered.

(ii) We then compute

$$w_{\text{comp}}^{\text{PARTIAL}} = w_{\text{FKP}} \prod_{s \in \mathcal{S}'} w_s,$$

where  $\mathcal{S}' \subseteq \mathcal{S}$  is the subset of systematic effects to be considered and  $\mathcal{S} = \{\text{systot}, \text{noz}, \text{cp}\}$ . In this case we compute the new effective number of tracers

$$n'_{\text{eff}} = \sum_{i=1}^{N_{\text{chunk}}} w_{\text{comp},i}^{\text{PARTIAL}}.$$

(iii) Objects in the catalog with  $w_{\text{comp}}^{\text{PARTIAL}} = 0$  or  $\text{veto} = 0$  (such as objects outside the survey area or masked by bright stars) are removed.

(iv) We then normalize the completeness weights in each chunk by the corresponding  $n_{\text{eff}}/n'_{\text{eff}}$  to keep the original effective number of tracers and avoid recomputing  $n(z)$ .

## 5 METHOD

As stated in earlier sections, this work relies of 1000 EZmocks emulating the eBOSS ELG sample (the analysis of LRG and QSO samples will be the focus of coming work), with and without systematics. Mocks with partial systematics are generated but are not yet used in this analysis. Our computations use the cosmological parameters in table 3. These do not match the parameters used in the construction of the EZmocks, therefore we expect a shift in the BAO signal.

### 5.1 Pipeline

We rely in the EZmock ELG catalogs provided by [de Mattia \(2020\)](#). From these, catalogs with applied (partial or none) systematics are generated following the procedure described in section 4.3. The output is a catalog listing the sky position (RA, DEC,  $z$ ), the final weight  $w_{\text{ALL}} = w_{\text{cp}} w_{\text{FKP}}$  completeness weight,  $w_{\text{cp}}$ , FKP weight,  $w_{\text{FKP}}$  and redshift distribution,  $n(z)$ . To extract the DT void catalog using DIVE, the galaxy catalog is converted to comoving coordinates. Once the void catalog is extracted, it is converted to sky coordinates to apply the appropriate veto masks (e.g. survey geometry, bright stars, etc.). Final void catalogs are then generated by

selecting unmasked voids. It contains the position in the first three columns and the radius  $R$  in the fourth.

Once the voids inside the survey region were selected, the random void catalogs are generated in the following way, as sketched in [Liang et al. \(2016\)](#):

- (i) Combine 100 mock void catalogs for each cap.
- (ii) Divide each file in  $z$ -bins with boundaries 0.6, 0.7, 0.8, 0.9, 1, 1.1.
- (iii) Divide each of the  $z$ -bins into  $R$ -bins with linearly spaced boundaries from 0 to  $21 h^{-1} \text{Mpc}$  and bin size of 1. The last three bins are taken from 21 to 25, 25 to 30 and 30 to  $50 h^{-1} \text{Mpc}$ .
- (iv) Now, divide each of the resulting bins in two, one containing the RA, DEC and the other containing the  $z$ ,  $R$  columns.
- (v) Shuffle in-place, the lines of one of the two halves.
- (vi) Undo all the splitting, that is, combine the halves, then the  $R$ -bins and then the  $z$ -bins.

Up to this step we ignore the radius of the voids.

(vii) Finally randomly choose some lines of the big random catalog left. In this case, We chose 2 700 000. When choosing lines we perform the radius cut, choosing objects with  $R \in (R_c, 50 h^{-1} \text{Mpc})$ , with  $R_c = 15.5 h^{-1} \text{Mpc}$  (see section 5.2). This step is repeated for each test with different  $R_c$ .

Now, with the random catalog built, it is possible to compute the 2PCFs for voids. In order to do this, we use the Landy-Szalay estimator ([Landy & Szalay 1993](#))

$$\xi = \frac{DD - 2DR + RR}{RR}. \quad (50)$$

When we consider the combined galaxy and void sample, this estimator conserves the form 50, however, each term has to be recomputed to account for both types of tracers as follows ([Zhao et al. 2018](#))

$$DD = \frac{D_g D_g n_g^2 + 2D_g D_v n_g n_v w + D_v D_v n_v^2 w^2}{(n_g + n_v w)^2}, \quad (51)$$

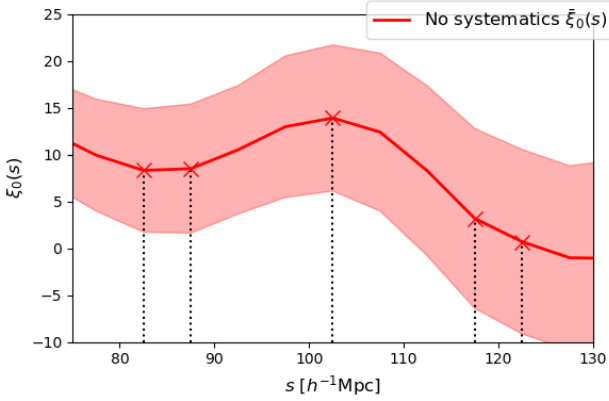
$$DR = \frac{D_g R_g n_g^2 + D_v R_v n_v^2 w^2 + (D_g R_v + D_v R_g) n_g n_v w}{(n_g + n_v w)^2}, \quad (52)$$

$$RR = \frac{R_g R_g n_g^2 + 2R_g R_v n_g n_v w + R_v R_v n_v^2 w^2}{(n_g + n_v w)^2}. \quad (53)$$

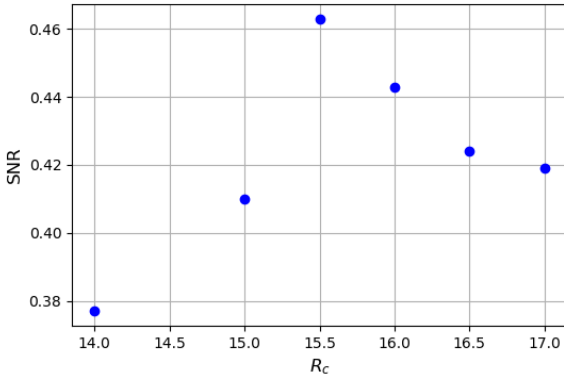
The new  $w$  (do not confuse with the equation of state) parameter is introduced to prevent the negative BAO peak coming from the void-galaxy cross correlation 2PCF to weaken the galaxy BAO signal ([Zhao et al. 2018](#)). As described in [Liang et al. \(2016\)](#), [Zhao et al. \(2016\)](#) and [Zhao et al. \(2018\)](#) and mentioned earlier in this text, there are two types of DT voids, we will use  $R_{\text{big}} > 15.5 h^{-1} \text{Mpc}$ . The small voids  $R_{\text{small}} < 8 h^{-1} \text{Mpc}$  are clearly voids in the geometrical sense, nonetheless, the local dark matter density contrast around them shows that these are actually rich in dark matter, while big voids are restricted to low local DM density contrast regions (see figure 9 in [Zhao et al. 2016](#)). These two cases should therefore be analyzed independently.

### 5.2 SNR analysis for BAO feature

In order to construct the void catalogs, we analyzed the Signal-to-noise ratio (SNR) of the BAO feature in the 2PCF. To do this, we followed the work done by [Liang et al. \(2016\)](#). To be precise, we generated 100 correlation functions for the catalogs with no sys-



**Figure 7.** BAO feature on the mean void 2PCF of mocks with no systematics applied, along with the respective standard deviation (shaded region). Vertical lines show the values of  $s_1^{\text{dl}} = 82.5$ ,  $s_2^{\text{dl}} = 87.5$ ,  $s_1^{\text{BAO}} = 102.5$ ,  $s_1^{\text{dr}} = 117.5$  and  $s_2^{\text{dr}} = 122.5 h^{-1} \text{Mpc}$  respectively from left to right.



**Figure 8.** SNR vs low radius cut  $R_c$ . Upper cut is always set to  $50 h^{-1} \text{Mpc}$ . Maximum SNR is obtained for  $R_c = 15.5 h^{-1} \text{Mpc}$ .

tematics and computed the signal  $S$  defined as

$$S = \xi_0(s^{\text{BAO}}) - \frac{\xi_0(s_1^{\text{dl}}) + \xi_0(s_2^{\text{dl}}) + \xi_0(s_1^{\text{dr}}) + \xi_0(s_2^{\text{dr}})}{4}, \quad (54)$$

where we have used  $s_1^{\text{BAO}} = 102.5$ ,  $s_1^{\text{dl}} = 82.5$ ,  $s_2^{\text{dl}} = 87.5$ ,  $s_1^{\text{dr}} = 117.5$  and  $s_2^{\text{dr}} = 122.5 h^{-1} \text{Mpc}$  as originally done by Liang et al. (2016). The location of these points can be seen in figure 7. The SNR is then

$$\text{SNR}(R_c) = \frac{\langle S \rangle}{\sigma_S}, \quad (55)$$

with  $\langle S \rangle$  the average signal over the 100 mocks used and  $\sigma_S$  the corresponding standard deviation. We did the test for different values of the low radius cut  $R_c$ . Figure 8 shows that the SNR peaks at  $R_c = 15.5 h^{-1} \text{Mpc}$ . Note that this is different from the optimal radius-cut given by the analysis by Liang et al. (2016). Their SNR was also much higher, with a maximum at  $\text{SNR} \sim 10^1$ .

### 5.3 Fitting procedure

#### 5.3.1 The model

In section 5.1 we briefly outlined the necessary steps to generate the masked mock catalogs which include, for instance, coordinate conversions and computing distances. For these, a *fiducial cosmology* is necessary. The introduction of this cosmology implies that our measurements depend on it in such a way that we will actually look at how the BAO feature behaves in our data with respect to what is expected from the fiducial cosmology.

To fit the BAO feature we use the model proposed by Xu et al. (2012) to compute the 2PCF of the fiducial cosmology,  $\xi_t$ . This model relies on a generated template power spectrum,  $P_t(k)$ , which is then related to the model 2PCF via Fourier transform. As was done in Zhao et al. (2018), we include a Gaussian damping such that the model correlation function is

$$\xi_t(s) = \int \frac{k^2 dk}{2\pi^2} \frac{\sin ks}{ks} P_t(k) \exp(-k^2 a^2), \quad (56)$$

where the parameter  $a = 1 h^{-1} \text{Mpc}$  (Xu et al. 2012; Zhao et al. 2018). The function  $P_t(k)$  is the template power spectrum. It is obtained via the relation

$$P_t(k) = [P_{\text{lin}}(k) - P_{\text{nw}}(k)] \exp\left(\frac{-\Sigma_{\text{nl}}^2 k^2}{2}\right) + P_{\text{nw}}(k). \quad (57)$$

The linear power spectrum  $P_{\text{lin}}$  is generated by CAMB software<sup>3</sup>, while the *non-wiggle* power spectrum is computed from the baryonless transfer function as shown by Eisenstein & Hu (1998). The first term in equation 57 describes the Gaussian damping of the BAO feature (the *wiggles*) while the second term just reconstructs the power spectrum with the dampening taken into account. Given that  $\xi_t$  describes the clustering of the fiducial cosmology, to perform our measurements we introduce the *model* 2PCF,

$$\xi_{\text{model}}(s) \equiv B^2 \xi_t(\alpha s) + A(s), \quad (58)$$

which is the curve that should actually fit our data, given that it contains the parameters  $B$  and  $\alpha$  along with the *nuisance parameters*,  $a_i$ ,  $i = 1, 2, 3$  contained in the polynomial

$$A(s) = \frac{a_1}{s^2} + \frac{a_2}{s} + a_3, \quad (59)$$

meant to fit the broadband spectrum of the data (Zhao et al. 2018). The other free parameters in the model are then  $B$ , for normalization,  $\Sigma_{\text{nl}}$  to account for the dampening of the BAO peak; and  $\alpha$ , the dilation parameter. The latter contains the information on how the peak is shifted with respect to the fiducial cosmology used to build  $\xi_t$ .

Nonetheless, to fit voids it is necessary to modify this model. Zhao et al. (2018) found that the ratio between the tracer and linear non-wiggled power spectra,  $P_{\text{tracer,nw}}(k)/P_{\text{lin,nw}}(k)$ , showed a nontrivial behavior in the case of voids, while for halos it is almost flat (see figure 7 in Zhao et al. 2018). The non-linear behavior of such quantity is modeled by them as an extra quadratic term in  $k$  (see also figure 10 in Tegmark et al. 2006). The modified template spectrum is, in general

$$P_t^v(k) = \left\{ [P_{\text{lin}}(k) - P_{\text{nw}}(k)] \exp\left(\frac{-\Sigma_{\text{nl}}^2 k^2}{2}\right) + P_{\text{nw}}(k) \right\} \frac{P_{\text{t,nw}}(k)}{P_{\text{lin,nw}}(k)}, \quad (60)$$

<sup>3</sup> <https://camb.info/>

with

$$\frac{P_{\text{lin},\text{nw}}(k)}{P_{\text{lin},\text{nw}}(k)} = 1 + ck^2, \quad (61)$$

where  $c$  is a seventh free parameter for the void fit.

### 5.3.2 The $\alpha$ parameter

Among the free parameters in the fitting model,  $\alpha$  deserves special attention. Equation 25 shows how this parameter describes the change in a given scale  $r$  when changing the cosmology. We can recover this equation by considering the distance  $r'$  calculated from angular and radial separations  $\Delta\theta'$ ,  $\Delta z'$  using the fiducial cosmology:

$$r' = (r_{\perp}'^2 r_{\parallel}')^{1/3} = \left[ \left\{ (1+z) D_A'(z) \Delta\theta' \right\}^2 \frac{\Delta z'}{H'(z)} \right]^{1/3}. \quad (62)$$

Now let

$$\alpha = \frac{r'}{r} = \frac{\left[ \left\{ (1+z) D_A'(z) \Delta\theta' \right\}^2 \frac{\Delta z'}{H'(z)} \right]^{1/3}}{\left[ \left\{ (1+z) D_A(z) \Delta\theta \right\}^2 \frac{\Delta z}{H(z)} \right]^{1/3}} = \left[ \frac{\Delta\theta'}{\Delta\theta} \right]^{2/3} \left[ \frac{\Delta z'}{\Delta z} \right]^{1/3}. \quad (63)$$

It is now natural to split  $\alpha$  into  $\alpha_{\parallel}$  and  $\alpha_{\perp}$  as

$$\alpha = \alpha_{\perp}^{2/3} \alpha_{\parallel}^{1/3}. \quad (64)$$

We will, however, focus on the isotropic case,  $\alpha$  (where we assume that  $r_{\parallel} = r_{\perp}$ ), in the present work.

The separations can be also computed with respect to the appropriate cosmologies:

$$\Delta\theta = \frac{\tilde{r}}{(1+z) D_A(z)}, \quad (65)$$

$$\Delta z = \tilde{r} H(z), \quad (66)$$

and the primed separations as given by equation 62. By replacing these into the rhs of equation 63 we recover equation 25 (with the replacement  $r \rightarrow \tilde{r}$ ).

If we add the subscript  $s$  to indicate that we measure the sound horizon, the coordinate  $r_s'$  then represents the argument of the template correlation function and  $r_s$  the BAO peak position we want to measure. By virtue of  $r_s' = \alpha r_s$ , we see that if  $\alpha > (<) 1$ , then  $r_s' > (<) r_s$  and the BAO scale has shifted to smaller (larger) scales with respect to the fiducial cosmology.

### 5.3.3 Parameter inference

We rely in Bayesian inference to compute the posterior distribution,  $p(\Theta|X)$ , for each parameter,  $\Theta$ :

$$p(\Theta|X) = \frac{p(X|\Theta)p(\Theta)}{p(X)}, \quad (67)$$

where we denote,  $X$ , the data. As usual, the distribution  $p(X|\Theta) \equiv \mathcal{L}$  is the *likelihood* a.k.a. how likely it is that the data  $X$  has been generated from a model with parameters  $\Theta$ , in other words, it is related to how good the model fits the data. We parameterize this distribution as

$$\mathcal{L}(\Theta) \equiv A \exp\left(-\frac{\chi(\Theta)^2}{2}\right), \quad (68)$$

where  $A$  is a normalization constant  $A = \int \mathcal{L}(\Theta) d\Theta$  and  $\chi^2(\Theta)$  is defined as usual

$$\chi^2(\Theta) = [\xi_{\text{data}} - \xi_{\text{model}}(\Theta)]^T \mathbf{C}^{-1} [\xi_{\text{data}} - \xi_{\text{model}}(\Theta)], \quad (69)$$

where  $\xi_{\text{data}}$  is the measured correlation function and  $\xi_{\text{model}}$  is the model 2PCF as defined in equation 58. The matrix  $\mathbf{C}^{-1}$  is the inverse unbiased covariance matrix as estimated from the mocks (Hartlap et al. 2007)

$$\mathbf{C}^{-1} = \mathbf{C}_s^{-1} \frac{N_m - N_{\text{bins}} - 2}{N_m}. \quad (70)$$

The matrix  $\mathbf{C}_s^{-1}$  is the (biased) sample covariance matrix,

$$\mathbf{C}_{s,ij} = \frac{1}{N_m - 1} \sum_{k=1}^{N_m} [\xi_k(s_i) - \bar{\xi}(s_i)][\xi_k(s_j) - \bar{\xi}(s_j)], \quad (71)$$

$N_m$  is the number of mocks used,  $N_{\text{bins}}$  is the number of bins ( $s$  points) used for the fit,  $\xi_k(s)$  is the 2PCF as obtained from the  $k$ -th catalog,  $\bar{\xi}(s)$  is the average 2PCF over all mocks and the indices  $i, j = 1, \dots, N_{\text{bins}}$ .

Back to equation 67, the distribution  $p(\Theta)$  is the *prior* for the parameter and it must be set by us. Finally,  $p(X)$  is often called the evidence and is often denoted  $\mathcal{Z}(X)$ . It can be computed as the normalization of the posterior

$$p(X) \equiv \mathcal{Z} = \int d\Theta p(\Theta) p(\Theta|X). \quad (72)$$

This definition shows that the evidence is the average of the likelihood over the prior, thus, it will be higher for models that yield a well distributed likelihood over the parameter space than for narrow ones, making it a good proxy for model selection (Feroz & Hobson 2008). The MULTINEST algorithm (Feroz & Hobson 2008) based on *Nested Sampling* (Skilling 2004) is used to compute the evidence and the posteriors in this work. The reported values of  $\alpha$  and  $\sigma_{\alpha}$  are computed as half the sum and difference (respectively) of the lower and upper bounds as obtained from the marginalized posterior, which were in turn calculated using the GETDIST Python package (Lewis 2019).

## 5.4 Priors

From the 6 (7) free parameters in the galaxy (void) model, we only use MULTINEST to fit the non-nuisance parameters. The  $a_i$  are fitted by a simple least squares procedure.

Following the results in Zhao et al. (2018), we use the fitting range  $s \in [60, 150]$  with 1000 live points and tolerance 0.1.

### 5.4.1 Galaxies

In the case of galaxies, we set the following priors

$$p(\Sigma_{\text{nl}}) = \mathcal{U}(5, 17) \quad (73)$$

$$p(B) = \mathcal{N}(1.4, 1.6, 0.12) \quad (74)$$

$$p(\alpha) = \mathcal{U}(0.8, 1.2) \quad (75)$$

### 5.4.2 Voids

For voids we choose the following priors

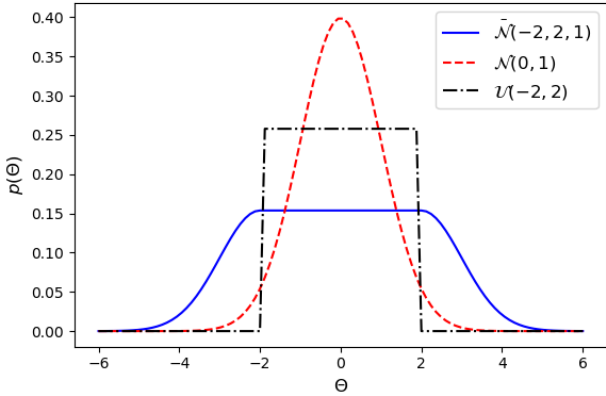
$$p(\Sigma_{\text{nl}}) = \mathcal{U}(0, 20) \quad (76)$$

$$p(B) = \mathcal{N}(2, 0.15) \quad (77)$$

$$p(\alpha) = \mathcal{U}(0.8, 1.2) \quad (78)$$

$$p(c) = \mathcal{N}(-500, 1000, 100) \quad (79)$$





**Figure 9.** Examples of the different kinds of priors used in the fitting.

Here,  $\mathcal{U}(\Theta_{\min}, \Theta_{\max})$  denotes a uniform distribution between  $\Theta_{\min}$  and  $\Theta_{\max}$ ;  $\mathcal{N}(\mu_{\Theta}, \sigma_{\Theta})$  is as (as before) a Gaussian distribution with mean  $\mu_{\Theta}$  and width  $\sigma_{\Theta}$  and  $\tilde{\mathcal{N}}(\Theta_{\text{flat min}}, \Theta_{\text{flat max}}, \sigma_{\Theta})$  is a flat distribution in the range  $[\Theta_{\text{flat min}}, \Theta_{\text{flat max}}]$  but with Gaussian tails of width  $\sigma_{\Theta}$ :

$$\tilde{\mathcal{N}}(\Theta_{\text{flat min}}, \Theta_{\text{flat max}}, \sigma_{\Theta}) = \begin{cases} A', & \Theta \in [\Theta_{\text{flat min}}, \Theta_{\text{flat max}}] \\ A' \exp\left(-\frac{(\Theta - \Theta_{\text{flat min}})^2}{2\sigma_{\Theta}^2}\right), & \Theta < \Theta_{\text{flat min}} \\ A' \exp\left(-\frac{(\Theta - \Theta_{\text{flat max}})^2}{2\sigma_{\Theta}^2}\right), & \Theta > \Theta_{\text{flat max}}, \end{cases} \quad (80)$$

where  $A'$  is a normalization constant. Figure 9 shows examples of these distributions.

## 6 RESULTS AND DISCUSSION

### 6.1 Fit to the mean 2PCF

We evaluate our fitting procedure by fitting the mean of 1000 mocks. Figure 10 shows the posterior distributions for the matter tracers and figure 11 does so for void tracers. Left panel shows the distributions in the case where no systematics have been applied, while the right panel shows the result where all of them have. Table 4 shows the fitting results.

The evidence has the same order of magnitude  $\sim e^{-4}$  in all cases, that is, the average likelihood over the priors is similar in all cases. However, we should note that the evidence is slightly higher in the galaxy case, though still consistent up to a  $1\sigma_{\ln \mathcal{Z}}$  level, the systematic offset could point towards a more effective model in the case of voids. This could be expected in the light of the discussion in section 5.3.1, where we introduced an extra parameter,  $c$ , to the void model.

Usually, the inclusion of more parameters, widening of the priors or in general any form of making the model more flexible, past a certain complexity, leads to smaller evidence (Knuth et al. 2015). Given that the inclusion of the  $c$  parameter does not show this effect in our tests, we assert that the modified model could still be improved in terms of careful prior selection. Even though more parameters could be added, they should be physically justified too, making it more difficult than tuning the priors.

Nonetheless, Zhao et al. (2018) have analyzed the impact of the

inclusion of the  $c$  parameter in the halo BAO fit using 100 PATCHY mocks. They observe that it results in a much smaller evidence (about  $\exp(4)$  smaller) and a fitted  $c \approx 0$ , leading them to the conclusion that the modified model does not improve the halo fit.

As a goodness-of-fit parameter, we also report in the same table the reduced  $\chi^2$  ( $\chi^2/\text{dof}$ ) for the best fit model. We see that in the case of matter tracers, this quantity is one order of magnitude smaller than for voids. Indeed, the fit in the former is very good. For voids, the model can be considered good too, given its closeness to 1, however the posterior distribution of  $\alpha$  (see figure 11) is not completely satisfactory owing to its asymmetric shape and various local maxima, which make it multimodal-like. Our tests with various different priors and fit intervals  $s \in [s_{\min}, s_{\max}]$  showed no further improvement in this regard. This could imply that the dataset itself may be noisy. The  $\alpha$  posterior for galaxies is, on the other hand, quite symmetric and with a single local maximum. The best-fit curves obtained are shown, for the cases considered, in figures A1 and A2. Notice that the fit is indeed well within the error (shaded region) for both tracers, yet the one for galaxies is much closer to the mean curve as demonstrated by  $\chi^2$ . We also observe oscillations towards larger scales in the fit range, specially in the voids without systematics, this can be due to the low SNR of the signal.

For the other parameters, we see that the  $B$  distribution is seemingly unchanged from the prior while the  $c$  posterior is a narrow, tailed distribution incompatible with 0 (once the error is scaled) and further confirms the importance of this extra parameter in the fitting of void BAO. In addition, the  $\Sigma_{\text{nl}}$  posterior shows, in general, a skewed distribution with higher values than LRGs (e.g. Zhao et al. (2018)) in both galaxies and voids, due to the smaller BAO signal obtained using ELGs.

Regarding the main objective of this study, we compare the fitted  $\alpha$  in the cases with all and no systematics affecting the original galaxy mock catalogs. For the fit to the mean, we plot the results in figure 12. Void fit shows a bigger error due to the width of the posterior shown in figure 11. Given that our analysis didn't use the "real" cosmology (the one used to create the EZmocks, see table 2), we expect peak position deviations. Indeed, we see that for the fit to the mean the peak is shifted towards smaller scales.

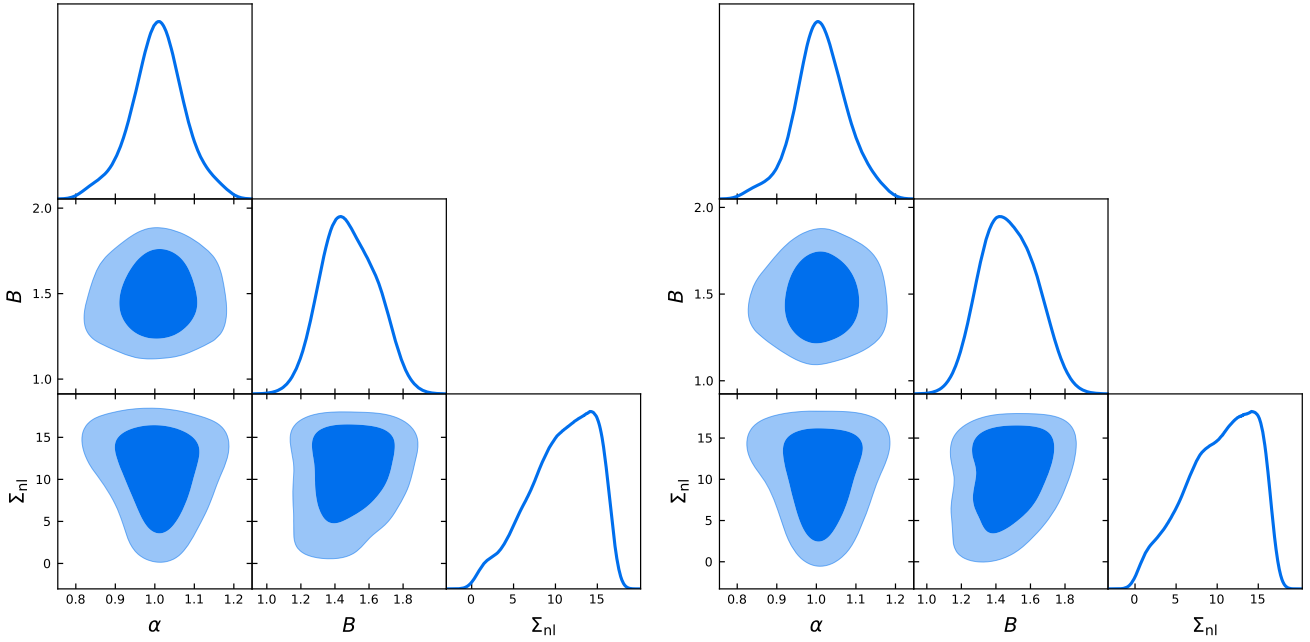
Additionally, it is worth noting that the  $\alpha$  values obtained in the cases with all and no systematics differ. Even though the uncertainty in the estimation of the parameter for voids is relatively high, the mean values point towards a higher discrepancy in the values for the matter tracers. As stated before, a better  $\alpha$  posterior might be possible in the case of voids, in which case we could be more certain about these results.

### 6.2 Fit to the individual mocks

After fitting the mean 2PCF, the individual, 1000 correlation functions were also fitted using the same covariance matrix defined in equation 70. For this test we extracted the  $\alpha$  and  $\sigma_{\alpha}$  in the same way as described before. We plot these results in figure 13. The top panel of the figure shows the obtained distributions for  $\alpha$  in every case (as obtained using `scipy's stats.gaussian_kde`). The lower panel shows only the mean and error as reported in table 5. In this case, the final values are evidently more sparse than the fit to the mean. This is to be expected since averaging removes some noise from the signal. In addition, the values reported in table 5 are clearly divided into matter tracers and voids (better seen in figure 13). The values within each group are clearly compati-

**Table 4.** Fitting results for the mean of 1000 mocks for both kinds of tracers with and without systematic effects. The Bayesian evidence  $\mathcal{Z}$  and the best fit  $\chi^2/\text{dof}$  (degrees of freedom) are also presented. The reported errors (except for  $\ln \mathcal{Z}$ ) correspond to the scaled standard deviations extracted from the posteriors in figures 10 and 11, taking into account that the fit was done on the mean of 1000 mocks.

$\Theta$	No syst.		All. syst.	
	Gal.	Void	Gal.	Void
$\alpha - 1$	$0.00744 \pm 0.00207$	$0.00915 \pm 0.00522$	$0.01078 \pm 0.00202$	$0.00816 \pm 0.00509$
$B$	$1.484 \pm 0.005$	$1.975 \pm 0.005$	$1.473 \pm 0.005$	$1.974 \pm 0.005$
$\Sigma_{\text{nl}}$	$10.840 \pm 0.126$	$15.537 \pm 0.121$	$10.892 \pm 0.134$	$15.822 \pm 0.108$
$c$	-	$-81.744 \pm 6.819$	-	$-84.956 \pm 6.106$
$\ln \mathcal{Z}$	$-4.286 \pm 0.040$	$-4.225 \pm 0.040$	$-4.136 \pm 0.030$	$-4.109 \pm 0.050$
$\chi^2_{\text{best}}/\text{dof}$	1.38	11.29	2.02	13.25



**Figure 10.** Posterior distributions for the model parameters in the galaxy BAO fit. Left panel shows the case in which no systematical effects were applied. Right panel shows the results for mocks with all systematics applied.

**Table 5.** Fitting results for  $\alpha$  when each mock is fitted individually. Mean and error are as given by the distributions in figure 13, however the error is scaled according to the number of mocks used.

		$\alpha - 1$
No syst.	Gal.	$0.00831 \pm 0.00166$
	Void	$0.00303 \pm 0.00166$
All. syst.	Gal.	$0.01020 \pm 0.00172$
	Void	$0.00196 \pm 0.00172$

ble among them at  $1\sigma$  level, while incompatible with the rest. It is worth noting that in this case, given that we are reporting  $\alpha = \langle \alpha_k \rangle$ ,  $k = 1, \dots, N_{\text{mocks}}$ , the reported error has to be scaled too with  $N_{\text{mocks}}^{-1/2}$ .

We must also observe that the mean values for the individual fits also show the relation seen in the fit to the mean regarding the difference between applying all systematics and none of them for each tracer.

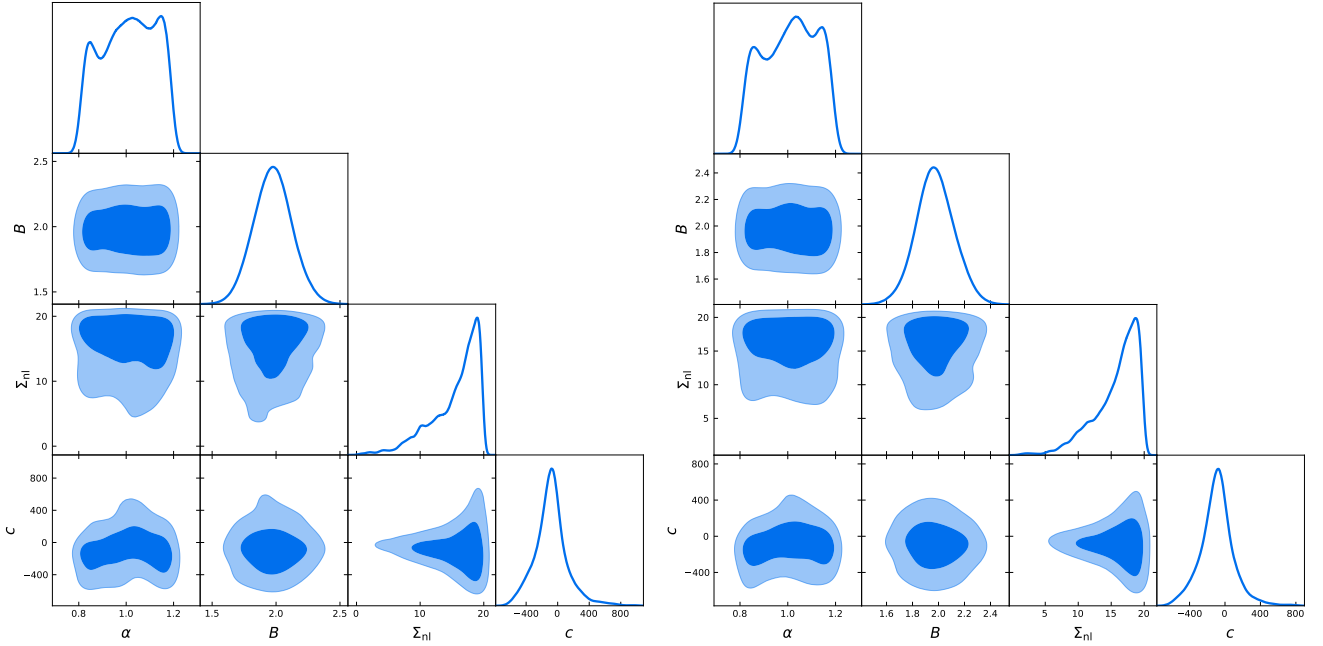
## 7 CONCLUSION

In this work we studied the impact of the observational systematic effects on the BAO feature from two different tracers: galaxies and voids-in-voids drawn from their (discrete) distribution by means of DT the algorithm. Following the works of [Kitauro et al. \(2016\)](#) and [Liang et al. \(2016\)](#), we hypothesize that the void sample is less sensitive to such effects than the galaxy one. As a first approach to testing this, we use ELG tracers from 1000 EZmock catalogs emulating the eBOSS survey properties.

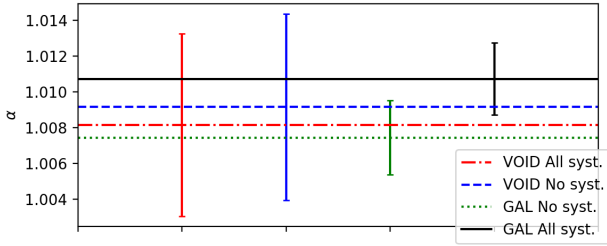
From the “raw” catalogs we generated “clean” ones by applying various different combinations of systematics, among which we study the cases with all (a.k.a. photometric effects, redshift failures and fiber collisions) and none of them. The systematic weights applied conserve the original effective number of tracers, this is, the number of tracers a telescope would be able to observe (all systematics applied).

We computed the void catalogs using the DIVE software. The correlation functions were subsequently computed in the void case. The galaxy ones were precomputed and provided to us.

We start by analyzing the signal-to-noise ratio for the BAO feature in void 2PCF to extract the optimal radius cut separating voids-in-voids from voids-in-clouds as explained by [Liang et al. \(2016\)](#);



**Figure 11.** Posterior distributions for the model parameters in the void BAO fit. Left panel shows the case in which no systematical effects were applied. Right panel shows the results for mocks with all systematics applied.



**Figure 12.** Fit results for the dilation parameter  $\alpha$  when using the mean of the mocks. Error bars show the scaled error as reported in table 4.

Zhao et al. (2016). We conclude that the optimal radius cut for ELGs is  $R_c = 15.5 h^{-1} \text{Mpc}$ . We must observe too that the SNR for these voids (extracted from ELGs) is much smaller (at least one order of magnitude) than the one for voids extracted from LRGs. This shows our dataset is more noisy, which affects our measurements of cosmological parameters.

We first consider the mean 2PCF from 1000 mocks and fit the model to it. We conclude that the BAO model fits very well to the data. Even more, we find that the void tracers show a smaller discrepancy than the galaxies in the mean dilation parameter,  $\alpha$ , when comparing the cases with no and all systematics applied. However, we observe too that the width of the posterior distributions for the same parameter is relatively high, asymmetrical and shows various different local maxima. Given that the measured Bayesian evidences are similar in all cases, we could think of yet improving the model, preferably by tuning the prior distributions. We explored this direction but saw no improvement regarding the shape of the distributions. Also, the goodness-of-fit parameter  $\chi^2/\text{dof} \approx 10^1$  is not close to one for the void case, showing a rather poor fit, consistent with the posterior distributions.

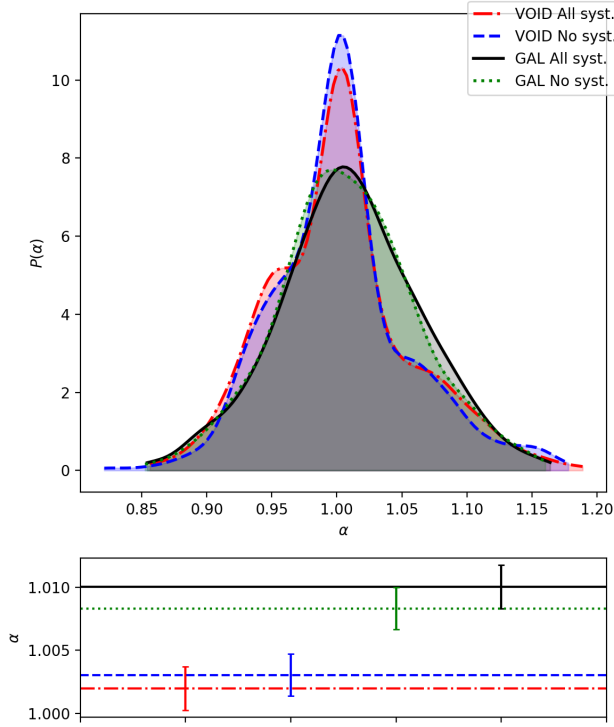
The fact that  $\chi^2$  is large, yet difficult to improve, may suggest that

the large widths in the posteriors are related not to the model but to the dataset itself. Indeed, the SNR measurements done in the void case show  $\text{SNR} < 1$  consistently, which means that the estimated noise is higher than the signal itself, rendering it non-significant. This implies that future work must extend the present analysis to other tracers with better SNR for the BAO feature such as LRGs, where  $\text{SNR} \approx 10^1$  were observed (see Liang et al. 2016) and, to evaluate the different tracers in the survey, QSOs.

On the other hand, we tested the same hypothesis by fitting each of the individual 1000 2PCF. Somewhat consistent with our previous result, we see that the distribution of  $\alpha$  values over our dataset is mostly symmetric in the case of galaxies and asymmetric for voids. Even more, the fit results are clearly split into voids with values  $\alpha \approx 1.002$  and galaxies with  $\alpha \approx 1.010$  this makes the groups' measurements incompatible with each other. Furthermore, the difference in the all vs. no systematics case is seen in this test as well as our first one. Given that the error bars are yet smaller in this second test, we conclude that the cosmic voids are indeed less vulnerable to systematic effects. This, as stated by Zhao et al. (2018), could be due to their much lower density, which makes them less susceptible to nonlinear effects in the gravitational evolution of structure in the Universe.

As was expected, the obtained dilation parameters deviate from 1 due to the use of different cosmologies in the creation of the EZ-mocks and our analysis.

As stated before, further analysis needs to be done to evaluate this effect on other tracers. This work, however, points towards the conclusion that voids are not as sensitive to systematics as galaxies are. In addition, measurements of the parameters on real data are to be done.



**Figure 13.** Fit results for the  $\alpha$  parameter from fitting each of the mocks individually. Top panel shows the distributions obtained. Bottom panel shows the mean values and the (scaled) standard deviation as error, both reported in table 5.

## ACKNOWLEDGEMENTS

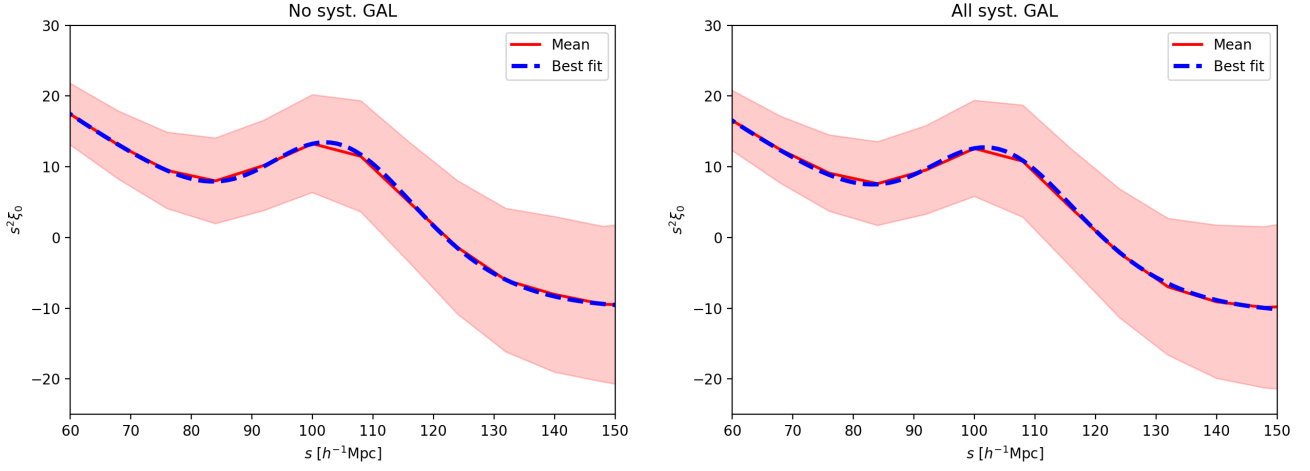
The authors acknowledge the contributions of Amélie Tamone, who provided essential data to this analysis. Computations were performed in the Cori cluster from NERSC and in Lesta, the cluster of the Geneva Observatory.

## REFERENCES

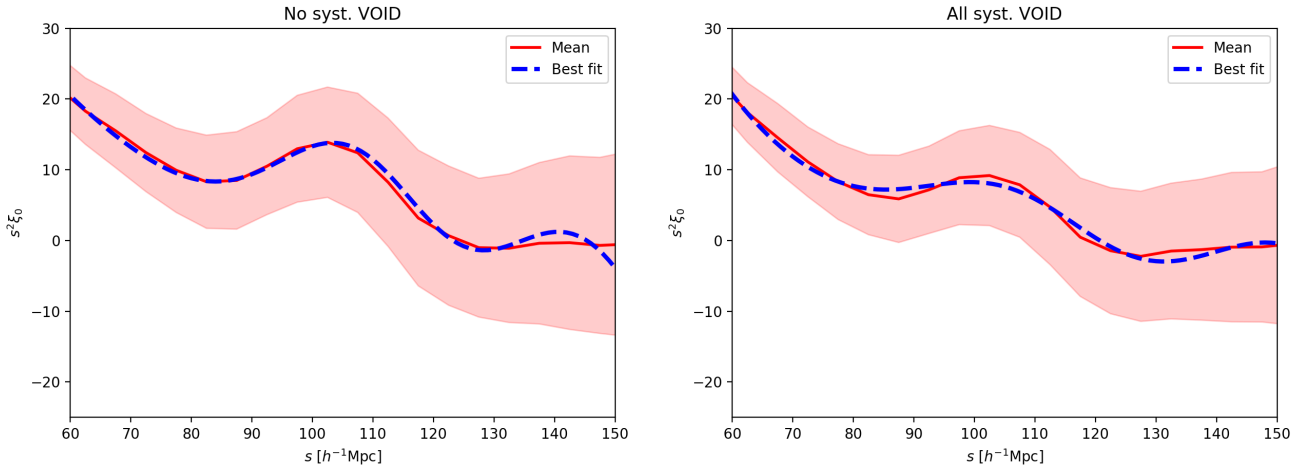
Abel T., Hahn O., Kaehler R., 2012, *Monthly Notices of the Royal Astronomical Society*, 427, 61  
 Alcock C., Paczyński B., 1979, *Nature*, 281, 358  
 Anderson L., et al., 2014, *Monthly Notices of the Royal Astronomical Society*, 441, 24  
 Ata M., et al., 2018, *Monthly Notices of the Royal Astronomical Society*, 473, 4773  
 Bassett B. A., Hlozek R., 2010, in Ruiz-Lapuente P., ed., *Dark Energy: Observational and Theoretical Approaches*. Cambridge University Press, pp 246–276 ([arXiv:0910.5224](https://arxiv.org/abs/0910.5224)), <http://arxiv.org/abs/0910.5224>  
 Baumann D., , *Cosmology - Mathematical Tripos*, <http://www.damtp.cam.ac.uk/user/db275/Cosmology/Lectures.pdf>  
 Bernardeau F., Colombi S., Gaztañaga E., Scoccimarro R., 2002, *Physics Reports*, 367, 1  
 Chuang C.-H., Kitaura F.-S., Prada F., Zhao C., Yepes G., 2015, *Monthly Notices of the Royal Astronomical Society*, 446, 2621  
 Colberg J. M., Sheth R. K., Diaferio A., Gao L., Yoshida N., 2005, *Monthly Notices of the Royal Astronomical Society*, 360, 216  
 Colless M., et al., 2001, *Monthly Notices of the Royal Astronomical Society*, 328, 1039  
 Dawson K. S., et al., 2015, *The Astronomical Journal*, 151, 44

Delaunay B., 1934, *Bulletin of the Academy of Sciences of the U. S. S. R. Classe des Sciences Mathematiques et Naturelles*, 7, 793  
 Dodelson S., 2003, *Modern cosmology*. Academic Press, <https://cds.cern.ch/record/1282338/export/hx>  
 Eisenstein D., Hu W., 1998, *The Astrophysical Journal*, 496, 605  
 Eisenstein D., et al., 2005, *The Astrophysical Journal*, 633, 560  
 Eisenstein D., Seo H.-j., White M., 2007, *The Astrophysical Journal*, 664, 660  
 Feroz F., Hobson M. P., 2008, *Monthly Notices of the Royal Astronomical Society*, 384, 449  
 Forero-Romero J. E., Hoffman Y., Gottlöber S., Klypin A., Yepes G., 2009, *Monthly Notices of the Royal Astronomical Society*, 396, 1815  
 Hahn O., Carollo C. M., Porciani C., Dekel A., 2007, *Monthly Notices of the Royal Astronomical Society*, 381, 41  
 Hartlap J., Simon P., Schneider P., 2007, *Astronomy & Astrophysics*, 464, 399  
 Hong T., Han J. L., Wen Z. L., Sun L., Zhan H., 2012, *The Astrophysical Journal*, 749, 81  
 Hong T., Han J. L., Wen Z. L., 2016, *The Astrophysical Journal*, 826, 154  
 Kitaura F. S., Yepes G., Prada F., 2014, *Monthly Notices of the Royal Astronomical Society: Letters*, 439  
 Kitaura F. S., et al., 2016, *Physical Review Letters*, 116  
 Knuth K. H., Habeck M., Malakar N. K., Mubeen A. M., Placek B., 2015, *Digital Signal Processing*, 47, 50  
 Komatsu E., et al., 2009, *Astrophysical Journal, Supplement Series*, 180, 330  
 Landy S. D., Szalay A. S., 1993, *The Astrophysical Journal*, 412, 64  
 Lewis A., 2019, pp 1–14  
 Liang Y., Zhao C., Chuang C.-H., Kitaura F.-S., Tao C., 2016, *Monthly Notices of the Royal Astronomical Society*, 459, 4020  
 McDonald P., Eisenstein D., 2006, *Physical Review D*, 76, 063009  
 Particle Data Group et al., 2018, *Physical Review D*, 98, 30001  
 Peebles P. J. E. P. J. E., 1980, *The large-scale structure of the universe*. Princeton University Press, <https://press.princeton.edu/titles/724.html>  
 Planck Collaboration et al., 2018  
 Raichoor A., et al., 2017, *Monthly Notices of the Royal Astronomical Society*, 471, 3955  
 Ross A. J., et al., 2017, *Monthly Notices of the Royal Astronomical Society*, 464, 1168  
 Sheth R. K., van de Weygaert R., 2004, *Monthly Notices of the Royal Astronomical Society*, 350, 517  
 Skilling J., 2004, in *AIP Conference Proceedings*. AIP, pp 395–405, [doi:10.1063/1.1835238](https://doi.org/10.1063/1.1835238), <http://aip.scitation.org/doi/abs/10.1063/1.1835238>  
 Tashev S., Zaldarriaga M., Eisenstein D. J., 2013, *Journal of Cosmology and Astroparticle Physics*, 2013  
 Tegmark M., et al., 2006, *Physical Review D - Particles, Fields, Gravitation and Cosmology*, 74, 1  
 Weinberg D. H., Mortonson M. J., Eisenstein D. J., Hirata C., Riess A. G., Rozo E., 2013, *Physics Reports*, 530, 87  
 White M., Reid B., Chuang C. H., Tinker J. L., McBride C. K., Prada F., Samushia L., 2015, *Monthly Notices of the Royal Astronomical Society*, 447, 234  
 Xu X., Padmanabhan N., Eisenstein D. J., Mehta K. T., Cuesta A. J., 2012, *Monthly Notices of the Royal Astronomical Society*, 427, 2146  
 Zeldovich Y., 1970, *Astron. Astrophys.*, 5, 84  
 Zhao C., Chuang C.-h., 2020, in preparation, 10, 1  
 Zhao C., Tao C., Liang Y., Kitaura F.-S., Chuang C.-H., 2016, *Monthly Notices of the Royal Astronomical Society*, 459, 2670  
 Zhao C., et al., 2018, 19, 1  
 de Mattia A., 2020, in preparation  
 van de Weygaert R., Schaap W., 2009, *Lecture Notes in Physics*, 665, 291





**Figure A1.** Best fit curves for galaxy tracers with (right) and without (left) systematics. The model fits the mean curve very well indeed, as shown by the  $\chi^2$  reported in table 4. Shaded region shows the  $1\sigma$  band obtained from the mock sample.



**Figure A2.** Best fit curves for void tracers with (right) and without (left) systematics. The model fits the mean curve well within the  $1\sigma$  band (shaded region). This is confirmed by the  $\chi^2$  reported in table 4. We observe, however, that the model does not follow the mean curve as close as for the galaxy case, specially at larger scales.

## APPENDIX A: BEST-FIT CURVES

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