

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/340997058>

# Does jackknife scale really matter for accurate large-scale structure covariances?

Preprint · April 2020

CITATIONS

0

READS

20

4 authors, including:



**Ginevra Favole**

École Polytechnique Fédérale de Lausanne

40 PUBLICATIONS 1,731 CITATIONS

SEE PROFILE



**Domenico Sapone**

University of Chile

60 PUBLICATIONS 2,616 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



Testing extended Jordan-Brans-Dicke theories with future cosmological observations [View project](#)

# Does jackknife scale really matter for accurate large-scale structure covariances?

Ginevra Favole,<sup>1</sup>\* Benjamin R. Granett,<sup>2</sup> Javier Silva Llafaurie,<sup>3</sup> Domenico Sapone<sup>3</sup>

<sup>1</sup>*Institute of Cosmology & Gravitation, University of Portsmouth, Dennis Sciama Building, Portsmouth, PO13FX, UK*

<sup>2</sup>*Università degli Studi di Milano, Via Celoria 16, Milan 20133, Italy*

<sup>3</sup>*Grupo de Cosmología y Astrofísica Teórica, Departamento de Física, FCFM, Universidad de Chile, Blanco Encalada 2008, Santiago, Chile*

Accepted XXX. Received YYY; in original form ZZZ

## ABSTRACT

The jackknife method gives an internal covariance estimate for large-scale structure surveys and allows model-independent errors on cosmological parameters. Using the SDSS-III BOSS CMASS sample, we study how the jackknife size and number of resamplings impact the precision of the covariance estimate on the correlation function multipoles and the error on the inferred baryon acoustic scale. We compare the measurement with log-normal mock galaxy catalogues with the same survey geometry. We build several jackknife configurations that vary in size and number of resamplings. We find that it is useful to apply the tapering scheme to estimate the precision matrix from a limited number of jackknife resamplings. The results from CMASS and mock catalogues show that the error estimate of the baryon acoustic scale does not depend on the jackknife scale. For the shift parameter  $\alpha$ , we find an average error of 1.6% and 1.2%, respectively from CMASS and mock jackknife covariances, consistent with pre-reconstruction analyses. However, when relatively few resamplings are used, the jackknife error estimate becomes unreliable. Future large-scale structure surveys will map even greater volumes allowing percent-level estimation of the covariance matrix using a jackknife approach.

**Key words:** cosmology: large-scale structure of Universe; cosmological parameters; observations; theory – galaxies: statistics; haloes

## 1 INTRODUCTION

The most popular methods to estimate the uncertainties on the galaxy two-point correlation function (2PCF) internally in a survey are bootstrap (Efron 1979; Davison & Hinkley 1997; Norberg et al. 2009, 2011) and jackknife (Queiroz 1956; Miller 1974; Turkey 1958; Norberg et al. 2009, 2011) resampling. Bootstrap resampling is carried out by randomly selecting  $N_b$  sub-volumes, with replacement, from the original sample. Then the galaxy clustering measurement is performed in each resampling, which is associated a constant weight equal to the number of times that the sub-volume has been selected (Norberg et al. 2009). Similarly, the jackknife method uses  $N_{jk}$  regions in the survey footprint, each with approximately the same galaxy number density. The correlation function is measured on the survey multiple times, each time removing a different jackknife region (Norberg et al. 2009). The covariance matrix is finally inferred from the 2PCF measurements and the  $1\sigma$  errors are derived as the square root of the diagonal elements.

Internal methods for error estimation are computationally inexpensive and are derived directly from the measurements. Therefore, the analysis does not depend on an as-

sumed cosmological model, which is an attractive feature when testing alternative models such as dark energy or modified gravity. Indeed, jackknife resampling has been widely used to estimate the uncertainties on the galaxy clustering measurements from large-volume spectroscopic surveys (e.g. Zehavi et al. 2002, 2005, 2011; Guo et al. 2012; Ross et al. 2012; Anderson et al. 2012; Guo et al. 2015a,b; Favole et al. 2016a, 2017).

Jackknife resampling has two main disadvantages: (i) it tends to overestimate the errors due to the lack of independence between the  $N_{jk}$  copies; and (ii) it is necessary to balance the number and size of resamplings to be drawn in the survey footprint. The last issue is driven by several factors. First of all, in order to have covariances with reduced noise in their off-diagonal terms, we need a large number of jackknife resamplings. This limits the size of the jackknife regions and also the maximum scales that can be probed in the galaxy clustering observables. It is often assumed in the literature that the jackknife cell size  $S_{jk}$  should embed the maximum scale measured in the two-point correlation functions. At the same time, to have an invertible (i.e. non-singular) covariance matrix, the number of resamplings should be larger than the number of bins in the measured 2PCF. These conditions are difficult to satisfy in galaxy samples with limited area (e.g. Beutler et al. 2011; Hong et al. 2016). Due to the

\* ginevra.favole@port.ac.uk

finite size of any survey footprint, the more resamplings we draw, the smaller their size and the variation between one copy and the next one (Norberg et al. 2009, 2011).

The issues discussed above have discouraged some cosmologists to use jackknife resampling in favour of estimating the galaxy clustering uncertainties from large sets of independent synthetic galaxy catalogues. In the last years, the advent of efficient codes based on fast gravity solvers has considerably reduced the computational time needed for massive mock production, making available many different realisations of accurate, independent mocks for covariance estimates. Among these methods, PTHALOS (Scoccimarro & Sheth 2002; Manera et al. 2013), PINOCCHIO (Monaco et al. 2002, 2013), PATCHY (Kitaura et al. 2016) and HALOGEN (Avila et al. 2015) are all based on Lagrangian perturbation theory (LPT). Others, such as QPM (White et al. 2014), FastPM (Feng et al. 2016) or PPM-GLAM (Klypin & Prada 2018), use a quick particle mesh approach. Algorithms such as EZ-MOCKS (Chuang et al. 2015) adopt the effective Zel’dovich approximation, while COLA (Tassev et al. 2013; Koda et al. 2016), L-PICOLA (Howlett et al. 2015) or ICE-COLA (Izard et al. 2016) combine LPT with N-body solvers to speed up the computational time. Finally, high-fidelity mocks can also be obtained from multiple realisations of a log-normal density field (Coles & Jones 1991; Beutler et al. 2011; Hong et al. 2016; Agrawal et al. 2017; Lippich et al. 2019).

All these fast mocking approaches are extremely convenient compared to running a full N-body code, but they are generally limited to predicting the dark matter distribution. On top of the dark matter field, it is necessary to model the galaxy distribution by properly accounting for the different baryonic components and the complex physics of galaxy formation and evolution (e.g. Mo et al. 2010; Naab & Ostriker 2017).

Simulating baryons is a non-trivial task, which requires advanced computational techniques and resources. Semi-analytic models of galaxy formation and evolution (SAMs; White & Frenk 1991; Kauffmann et al. 1993; Baugh 2006; Benson 2010, 2012; Somerville & Davé 2015; Cora 2006; Cora et al. 2018; Croton et al. 2006, 2016; Gargiulo et al. 2015; Collacchioni et al. 2018) and hydrodynamical simulations (e.g. Yepes et al. 1997; Springel & Hernquist 2003; Yoshida et al. 2003; Springel 2010; Vogelsberger et al. 2014b,a; Genel et al. 2014; Crain et al. 2015; Schaye et al. 2015; Pillepich et al. 2018) are now able, with different degrees of accuracy, to incorporate the multitude of ingredients and physical processes that contribute to shape the formation and evolution of galaxies within their host dark matter haloes. Some of these processes are: gas accretion (Guo et al. 2011; Henriques et al. 2013; Hirschmann et al. 2016) and cooling (De Lucia et al. 2010; Monaco et al. 2014; Hou et al. 2017), star formation (Lagos et al. 2011), stellar winds (Lagos et al. 2013), stellar evolution Tonini et al. (2009); Henriques et al. (2011); Gonzalez-Perez et al. (2014), AGN feedback Bower et al. (2006); Croton et al. (2006) or environmental processes Weinmann et al. (2006); Font et al. (2008); Stevens & Brown (2017); Collacchioni et al. (2018).

Analogously, one should also account for the effect of massive neutrinos on the growth of cold dark matter perturbations, which are responsible of suppressing the matter power spectrum at intermediate and small scales (Ali-

Haïmoud & Bird 2013; Wright et al. 2017; Paribelli et al. 2019).

All of these assumptions and prescriptions have uncertainties which become significant on non-linear scales and limit the accuracy of the covariance estimate.

Upcoming surveys, such as the Dark Energy Spectroscopic Instrument<sup>1</sup> (DESI; Schlegel et al. 2015), Euclid<sup>2</sup> (Laureijs et al. 2011; Sartoris et al. 2016) and the Large Synoptic Survey Telescope<sup>3</sup> (LSST; Ivezić et al. 2019), will bring us to the era of high precision cosmology. In order to prepare to this new phase, it is imperative to improve and compare different methods to construct accurate covariances able to capture the hidden physical process of gravitational collapse. These methods have to carefully optimise the specific binning scheme adopted in order to minimise the noise in the measurements.

For the reasons above, in this work we aim to rehabilitate the use of jackknife resamplings versus mocks for estimating covariances. We explore how varying the size ( $S_{JK}$ ) and number ( $N_{JK}$ ) of jackknife regions impacts the precision in the error estimates of galaxy clustering and on the baryon acoustic oscillation scale through the  $\alpha$  shift parameter. In concrete, we measure the monopole and quadrupole two-point correlation functions of the BOSS CMASS DR12 galaxies and we compute their covariances using four different jackknife configurations, coupled with two binning schemes. We compare these results with those obtained from 200 independent log-normal mock light-cones with the same volume of CMASS, and with the covariance obtained by performing jackknife resampling on a single light-cone.

From these covariances, we build the precision matrices needed to estimate the  $\alpha$  shift parameter through a Monte Carlo Markov Chain (MCMC) algorithm. We reduce the noise in their off-diagonal terms by applying the tapering correction (Kaufman et al. 2008). We study the impact of a variation in the tapering parameter,  $T_p$ , on the  $\alpha$  results. These estimates of  $\alpha$  will be directly compared with the galaxy clustering pre-reconstruction results from the BOSS collaboration (Ross et al. 2016).

The paper is organised as follows: in Sec. 2, we introduce the observational galaxy sample used in our analysis, SDSS-III/BOSS CMASS DR12; in Sec. 3, we present the galaxy clustering measurements performed, together with the jackknife methodology and schemes used to estimate their uncertainties. Sec. 4 describes the models used to analyse the CMASS observations: the log-normal mock galaxy catalogues and light-cones (§ 4.1), and the analytic models used in the Monte Carlo runs (§ 4.2). In Sec. 5, we explain the Monte Carlo algorithm used to extract the  $\alpha$  BAO parameter. Sec. 6 presents our main results, which are discussed in Sec. 7 together with our conclusions.

Throughout the paper we adopt a Planck et al. (2014) cosmology with  $\Omega_m = 0.307115$ ,  $\Omega_\Lambda = 0.692885$ ,  $h = 0.6777$ ,  $n = 0.96$  and  $\sigma_8 = 0.8228$ .

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

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

<sup>3</sup> <https://www.lsst.org>

## 2 OBSERVED GALAXY SAMPLE: BOSS CMASS DR12

The SDSS-III BOSS survey observed about 1.2 million galaxies over an effective area of  $9329 \text{ deg}^2$  (Alam et al. 2017), using the 2.5m Sloan Telescope (Gunn et al. 2006) at the Apache Point Observatory in New Mexico. It used a drift-scanning mosaic CCD camera with five photometric bands, *ugriz* (Gunn et al. 1998; Fukugita et al. 1996), and two spectrographs covering the wavelength range  $3600\text{--}10000\text{\AA}$  with a resolving power of 1500 to 2600 (Smeed et al. 2013). Spectroscopic redshifts were measured using the minimum- $\chi^2$  template-fitting procedure by Aihara et al. (2011), with templates from Bolton et al. (2012).

BOSS targeted galaxies into two main samples: LOWZ at  $z < 0.43$  and CMASS at  $0.43 < z < 0.7$  (Ahn et al. 2012). For our analysis, we use the data from the BOSS CMASS DR12 sample (Alam et al. 2015; Reid et al. 2016; Alam et al. 2017), which is defined through a number of magnitude and colour cuts aimed at obtaining a selection of galaxies with approximately constant stellar mass. These cuts are:

$$\begin{aligned} 17.5 < i_{\text{cmod}} < 19.9 \\ r_{\text{mod}} - i_{\text{mod}} < 2 \\ i_{\text{fib2}} < 21.5 \\ d_{\perp} > 0.55 \\ i_{\text{cmod}} < 19.86 + 1.6(d_{\perp} - 0.8), \end{aligned} \quad (1)$$

where  $i_{\text{cmod}}$  represents the SDSS *i*-band cmodel magnitude,  $i_{\text{mod}}$  and  $r_{\text{mod}}$  are model magnitudes,  $i_{\text{fib2}}$  is the fibre magnitude within  $2''$  aperture, and  $d_{\perp}$  is the following colour combination:

$$d_{\perp} = r_{\text{mod}} - i_{\text{mod}} - (g_{\text{mod}} - r_{\text{mod}})/8. \quad (2)$$

In addition, the CMASS sample satisfies also the star-galaxy separation cuts:

$$\begin{aligned} i_{\text{psf}} - i_{\text{mod}} > 0.2 + 0.2(20 - i_{\text{mod}}) \\ z_{\text{psf}} - z_{\text{mod}} > 9.125 - 0.46z_{\text{mod}}, \end{aligned} \quad (3)$$

where the subscript ‘‘psf’’ stands for PSF magnitude.

## 3 MEASUREMENTS

### 3.1 Two-point correlation functions

We measure the two-point correlation function,  $\xi(s, \mu)$ , of the galaxy sample described in Sec. 2 using the code from Favole et al. (2017). This is based on the Landy-Szalay estimator (Landy & Szalay 1993),

$$\xi(s, \mu) = \frac{DD(s, \mu) - 2DR(s, \mu) + RR(s, \mu)}{RR(s, \mu)}, \quad (4)$$

where  $s$  is the redshift-space distance and  $\mu$  is the cosine of the angle between  $s$  and the line of sight.

In the expression above,  $DD$ ,  $DR$  and  $RR$  are respectively the data-data, data-random and random-random pair counts that we can form between the galaxy and random catalogues. The latter is built to have the same angular footprint and radial distribution of the CMASS observations. All the pairs above are properly normalised by the number of galaxies (randoms) and weighted to correct from different systematic effects (see e.g. Sánchez et al. 2012; Ross et al.

2012; Favole et al. 2016b). In particular, we weight the observed data for potential fibre collisions ( $w_{\text{fc}}$ ) and for redshift failures ( $w_{\text{zf}}$ ). We also account for possible variation in the galaxy (random) number densities assuming the FKP (Feldman et al. 1994) weight:

$$w_{\text{FKP}} = \frac{1}{1 + n(z)P_0}, \quad (5)$$

where  $n(z)$  is the galaxy (random) number density at redshift  $z$  and  $P_0$  is a constant quantity that roughly corresponds to the amplitude of the CMASS power spectrum at  $k = 0.1 \text{ hMpc}^{-1}$ . We assume  $P_0 = 20000 \text{ h}^3 \text{ Mpc}^{-3}$  as in Anderson et al. (2012).

From Eq. 4, we compute the multipoles of the CMASS correlation function as:

$$\xi_l(s) = \frac{2l+1}{2} \int_{-1}^1 \xi(s, \mu) P_l(\mu) d\mu, \quad (6)$$

where  $P_l(\mu)$  are the Legendre polynomials. In this study, we focus only on the first two even multipoles of the 2PCF, i.e. the monopole  $\xi_0(s)$  and the quadrupole  $\xi_2(s)$ . We explore two different binning schemes, both centered on the BAO distance scale, coupled with the jackknife configurations defined in Sec. 3.2: (i) 20 linear bins in  $24 < s < 184 \text{ h}^{-1} \text{ Mpc}$  and 120 linear bins in  $0 < \mu < 1$ ; (ii) 10 linear bins in  $24 < s < 184 \text{ h}^{-1} \text{ Mpc}$  and 120 linear bins in  $0 < \mu < 1$ .

### 3.2 Jackknife configurations and covariances

We implement jackknife resampling in the BOSS CMASS DR12 galaxy sample adopting four different configurations summarised in Table 1. We divide the survey footprint into 200, 100, 50 and 20 RA×DEC cells approximately containing the same number of galaxies (randoms). The CMASS covariance matrix for  $N_{\text{JK}}$  jackknife resamplings is (e.g. Ross et al. 2012; Favole et al. 2016a):

$$C_{ij} = \frac{N_{\text{JK}} - 1}{N_{\text{JK}}} \sum_{a=1}^{N_{\text{JK}}} (\xi_i^a - \bar{\xi}_i)(\xi_j^a - \bar{\xi}_j), \quad (7)$$

where  $\bar{\xi}_i$  is the mean jackknife correlation function in the  $i^{\text{th}}$  bin,

$$\bar{\xi}_i = \sum_{a=1}^{N_{\text{JK}}} \xi_i^a / N_{\text{JK}}. \quad (8)$$

The overall factor in Eq. 7 corrects from the lack of independence between the  $N_{\text{JK}}$  jackknife copies, which is the main limitation of the jackknife method. In fact, from one configuration to the next,  $N_{\text{JK}} - 2$  cells are the same (Norberg et al. 2011).

## 4 MODELS

### 4.1 Log-normal mock galaxy catalogues and light-cones

We generated 200 log-normal mock galaxy catalogues for the BOSS CMASS sample at mean redshift  $z \sim 0.56$ . The target power spectrum was computed with the code *CLASS*<sup>4</sup>

<sup>4</sup> [https://github.com/lesgourg/class\\_public](https://github.com/lesgourg/class_public)

$N_{\text{JK}}$	$A_{\text{JK}}$ [deg <sup>2</sup> ]	$S_{\text{JK}}$ [ $h^{-1}$ Mpc]
200	46.6	110.7
100	93.3	156.6
50	186.6	221.4
20	932.9	495.1

**Table 1.** Jackknife configurations adopted in our analysis. For each of the four cases implemented, we indicate the number of jackknife resamplings ( $N_{\text{JK}}$ ), the area ( $A_{\text{JK}}$ ) and comoving size ( $S_{\text{JK}}$ ) of the individual cell computed in Planck et al. (2014) cosmology at the mean redshift of CMASS,  $z = 0.56$ .

(Blas et al. 2011). We applied a linear bias and the Halofit (Takahashi et al. 2012) prescription to model the non-linear galaxy power spectrum:  $P(k) = b^2 P_m(k)$  with the value  $b = 2.1$ .

We present the *Synmock* code used to produce the log-normal catalogues in a public repository<sup>5</sup>. The implementation follows the standard approach for generating log-normal simulations (see also Beutler et al. 2011; Pearson et al. 2016). For each realization, first we generated a Gaussian density field  $\delta_G(\vec{x})$  on a cubic grid with dimension  $L = 4096 h^{-1}$  Mpc and step size  $h = 8 h^{-1}$  Mpc and transformed it to derive the target log-normal field:

$$\delta(\vec{x}) = \exp\left(\delta_G(\vec{x}) - \sigma^2/2\right) - 1, \quad (9)$$

where  $\sigma^2$  is the variance of the Gaussian field. In order to match the target power spectrum we made a Fourier transform of the power spectrum to compute the correlation function and using the relationship (Coles & Jones 1991)

$$\xi_G(|\vec{x} - \vec{x}'|) = \log(1 + \xi(|\vec{x} - \vec{x}'|)). \quad (10)$$

The log-normal density field was used to build a discrete galaxy field by Poisson sampling the number density  $n(\vec{x}) = \bar{n}(1 + \delta(\vec{x}))$ . We applied a uniform random offset to move the mock galaxies away from the grid points.

The velocity field was computed on the same  $\delta(x)$  grid using the linear continuity equation in Fourier space:

$$\vec{v}(\vec{k}) = i \frac{f a H}{b} \delta(\vec{k}) \frac{\vec{k}}{k^2}, \quad (11)$$

where  $f$  is the logarithmic growth rate. After a final Fourier transform, the velocity of each galaxy was assigned using the value at the nearest grid point.

We built 200 BOSS CMASS light-cones (LCs) with  $0.43 < z < 0.7$  by cutting the BOSS survey geometry in the log-normal simulation box above. The Cartesian galaxy coordinates were transformed to the spherical coordinates right ascension, declination and radial distance with the origin at the center of the simulation box. In order to transform to the redshift-space coordinates, the line-of-sight peculiar velocity component was computed and applied to the radial comoving distance:  $r_s = r + \vec{r} \cdot \vec{v}/(aH|\vec{r}|)$ . We constructed a coarse angular mask using the Healpix (Górski et al. 2005) scheme at resolution  $n_{\text{side}} = 64$  and discarded galaxies outside the mask. The catalog was further downsampled along the radial direction to match the target redshift distribution. We generated an unclustered random catalogue with

10 times the number density of the CMASS data that precisely corresponds to the mock construction using the same angular mask and radial selection function.

After computing the correlation functions of the  $N_{\text{LC}} = 200$  log-normal light-cones, we derive their covariance matrix as:

$$C_{ij} = \frac{1}{N_{\text{LC}} - 1} \sum_{a=1}^{N_{\text{LC}}} (\xi_i^a - \bar{\xi}_i)(\xi_j^a - \bar{\xi}_j), \quad (12)$$

where  $\bar{\xi}_i$  is their mean 2PCF in the  $i^{\text{th}}$  bin. The pre-factor properly accounts for the fact that the mock realisations are independent.

## 4.2 Analytic models

Besides the log-normal mocks, we also model the multipoles of the BOSS CMASS two-point correlation function using an analytic approach, which is required to run the Monte Carlo analysis (see Sec. 5). The 2PCF can be obtained from the Fourier transform of the matter power spectrum,  $P(k)$ , for which we assume the template from Padmanabhan & White (2008):

$$P(k) = [P_{\text{lin}}(k) - P_{\text{dw}}(k)] e^{-k^2 \Sigma_{\text{nl}}^2/2} + P_{\text{dw}}(k). \quad (13)$$

In the equation above,  $P_{\text{lin}}(k)$  is the linear matter power spectrum computed using the Boltzmann code CLASS (Lesgourgues 2011) assuming the Planck 2015 (Ade et al. 2016) fiducial cosmology. The  $P_{\text{dw}}(k)$  term is the de-wiggled power spectrum (Eisenstein & Hu 1998), while the  $\Sigma_{\text{nl}}$  parameter encodes the smoothing of the BAO peak due to non-linear effects (Crocco & Scoccimarro 2006). The multipoles of the analytic 2PCF are defined as:

$$\xi_l(s) = \frac{i^l}{2\pi^2} \int_0^\infty P_l(k) j_l(ks) k^2 dk, \quad (14)$$

from which we recover the monopole ( $l = 0$ ) and the quadrupole ( $l = 2$ ). In Eq. 14,  $j_l(x)$  represents the spherical Bessel function of first kind and order  $l$ , while  $P_l(k)$  are the multipoles of the power spectrum defined as:

$$P_l(k) = \frac{2l+1}{2} \int_{-1}^1 (1 + f\mu^2)^2 P(k) L_l(\mu) d\mu, \quad (15)$$

where  $L_l(x)$  is the Legendre polynomial of order  $l$  and  $P(k)$  is the template given in Eq. 13. By replacing Eq. 15 in Eq. 14, the analytic expressions for monopole ( $l = 0$ ) and quadrupole ( $l = 2$ ) are respectively (Xu et al. 2012):

$$\xi_{\text{model}}^{(0)}(s) = B_0 \xi_0(\alpha s) + a_0^{(0)} + \frac{a_1^{(0)}}{s} + \frac{a_2^{(0)}}{s^2}, \quad (16)$$

$$\xi_{\text{model}}^{(2)}(s) = B_2 \xi_2(\alpha s) + a_0^{(2)} + \frac{a_1^{(2)}}{s} + \frac{a_2^{(2)}}{s^2}, \quad (17)$$

where  $\alpha$  is the shift parameter, while  $(a_1^{(i)}, a_2^{(i)}, a_3^{(i)})$  are linear nuisance parameters.

The shift parameter  $\alpha$  in Eqs. 16 and 17 is usually defined as (Padmanabhan & White 2008):

$$\alpha = \frac{D_V}{r_s} \frac{r_s^{\text{fid}}}{D_V^{\text{fid}}}, \quad (18)$$

<sup>5</sup> <https://github.com/bengranett/synmock>

where  $r_s$  represents the sound horizon (Hu & Sugiyama 1996), and  $D_V$  the volume-averaged distance given by (Eisenstein et al. 2005):

$$D_V(z) = \left[ cz(1+z)^2 D_A^2(z) H^{-1}(z) \right]^{1/3}, \quad (19)$$

with  $D_A(z)$  being the angular diameter distance, and  $H(z)$  the Hubble parameter at redshift  $z$ . The  $\alpha$  shift parameter accounts for the observed distortion between distances due to the chosen fiducial cosmology, while the nuisance parameters ( $a_1^{(i)}, a_2^{(i)}, a_3^{(i)}$ ) and  $B_1, B_2$  incorporate those effects that are responsible of modulating the clustering amplitude, such as redshift-space distortions (Xu et al. 2012), linear bias, and the power spectrum normalisation,  $\sigma_8$ .

## 5 SHIFT PARAMETER ESTIMATION

Following the methodology presented in Favole et al. (2019), we analyse the BOSS CMASS covariances, both computed from jackknife and 200 light-cones, using a Monte Carlo Markov Chain based on a Metropolis-Hastings algorithm<sup>6</sup>. Our MCMC code is publicly available on GitHub<sup>7</sup>.

In order to find the optimal parameter values, we assume a likelihood function  $\mathcal{L} \propto \exp(-\chi^2/2)$ , with

$$\chi^2 = \left( \vec{\xi}_{\text{model}} - \vec{\xi}_{\text{obs}} \right)^T \hat{\Psi} \left( \vec{\xi}_{\text{model}} - \vec{\xi}_{\text{obs}} \right), \quad (20)$$

where

$$\vec{\xi}_{\text{model}} \equiv \left( \vec{\xi}_{\text{model}}^{(0)}, \vec{\xi}_{\text{model}}^{(2)} \right) \quad (21)$$

represents the theoretical correlation function whose components are given in Eqs. 16-17, while  $\vec{\xi}_{\text{obs}}$  corresponds to the observed monopole and quadrupole 2PCFs, both grouped in a vector depending on the comoving distance. The  $\Psi$  term above is the precision matrix defined as:

$$\hat{\Psi} = \left( 1 - \frac{n_b + 1}{N_{\text{res}} - 1} \right) \left( \hat{C} \circ T \right)^{-1} \circ T, \quad (22)$$

where  $\hat{C}$  is the total assembled covariance matrix:

$$\hat{C} = \begin{pmatrix} \hat{C}_{\xi_0 \xi_0} & \hat{C}_{\xi_0 \xi_2} \\ \hat{C}_{\xi_0 \xi_2}^T & \hat{C}_{\xi_2 \xi_2} \end{pmatrix}. \quad (23)$$

The first term in parenthesis in Eq. 22 is the Hartlap factor (Hartlap et al. 2006), which corrects from the bias introduced in the covariance matrix by the limited number of jackknife resamplings and 2PCF bins. In Tab. 2, we report the values of the Hartlap factor as a function of the number of jackknife resamplings and bins used in our analysis.

The quantity  $T$  in Eq. 22 is the tapering correction (Kaufman et al. 2008) that minimises the noise in the off-diagonal terms of the covariance matrix; for further details see also Favole et al. (2019). In this work, we assume a tapering parameter  $T_p = 500 h^{-1} \text{Mpc}$  to ensure that the entire covariance matrix is positive semi-definite and the noise in the off-diagonal terms is minimised. In Sec. 6, we test how a variation in the tapering parameter affects the results for  $\alpha$  and its uncertainty. Further details on the dependence of  $\alpha$  on  $T_p$  are addressed also in Paz & Sánchez (2015).

<sup>6</sup> <https://emcee.readthedocs.io/en/stable/>

<sup>7</sup> <https://github.com/javiersilvalafaurie/BTCosmo>

$n_b$	$N_{\text{JK}}$	Hartlap factor
10	20	0.42105
20	50	0.57143
20	100	0.78788
20	200	0.89447

**Table 2.** Values of the Hartlap factor (Hartlap et al. 2006) as a function of the number of bins  $n_b$  and jackknife resamplings  $N_{\text{JK}}$  used in our analysis.

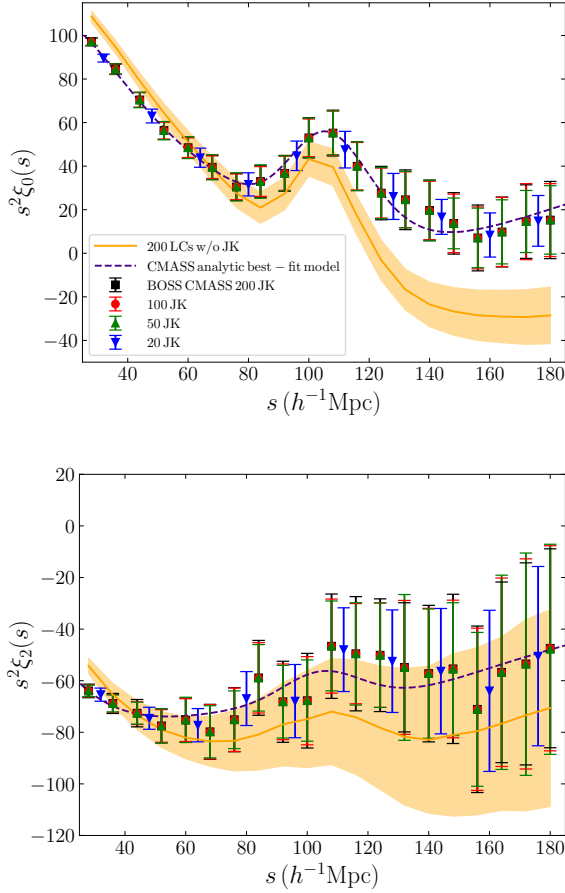
## 6 RESULTS

In Fig. 1 we present the BOSS CMASS monopole and quadrupole two-point auto-correlation functions compared to the mean predictions from the 200 log-normal light-cones. The CMASS error bars are inferred from the jackknife covariances based on the four configurations shown in Tab. 1 coupled with two different binning schemes (see Sec. 3.1). For the LCs, we show the dispersion obtained from the 200 realisations without jackknife resampling. The LCs reproduce the BAO peak well, but differ from the CMASS measurements in the broadband shape. The LC monopole prediction tends to overestimate the observed clustering amplitude at  $s \lesssim 60 h^{-1} \text{Mpc}$  and to underestimate it beyond BAO scales. The systematic difference in shape will be accounted for by the nuisance parameters in the model and so will not influence the analysis of the  $\alpha$  shift parameter. We also overplot the analytic 2PCF model used in our MCMC algorithm to estimate the  $\alpha$  BAO parameter (Sec. 5). The best-fit analytic model is in good agreement with the CMASS multipole measurements on all scales.

In Fig. 2 we display the normalised covariance of the monopole and quadrupole from the 200 log-normal light-cones, in 20  $s$  bins, built in Sec. 4.1, without jackknife resamplings. The matrix is normalised as  $C_{ij}^{\text{norm}} = C_{ij} / \sqrt{C_{ii} C_{jj}}$ , with  $C_{ij}$  given in Eq. 12.

Fig. 3 compares the normalised covariances obtained by performing jackknife resampling on a light-cone (upper triangles) versus BOSS CMASS data (lower triangles). From top to bottom, we present the 20, 50, 100 and 200 jackknife configurations listed in Tab. 1, respectively coupled with 10, 20, 20, 20 linear bins in  $s$  (see Sec. 3). The normalisation is calculated as described in Fig. 2, with  $C_{ij}$  given by Eq. 7 for CMASS data and by Eq. 12 for the light-cone covariance estimate without jackknife resampling. It is evident that the noise in the covariance estimate is reduced as the number of resamplings is increased. The covariances from 200 jackknife resamplings with 20 bins on a light-cone and on CMASS observations are consistent with the result from the 200 LCs without jackknife shown in Fig. 2. These covariances lead to consistent error bars on the galaxy clustering multipoles, as shown in Fig. 1.

The top panel of Fig. 4 shows, as a function of the scale, the ratio of the uncertainties obtained from CMASS jackknife covariances and from 200 light-cones without jackknife resampling. The results for the monopole and the quadrupole are shown as solid and dashed lines, respectively. We remind the reader that the 200, 100 and 50 JK configurations are measured in 20  $s$  bins, while the 20 JK case in 10 bins. The combined action of the jackknife size, number and binning is what determines the level of noise in the covariances. The 20 JK bins scheme leads to the largest

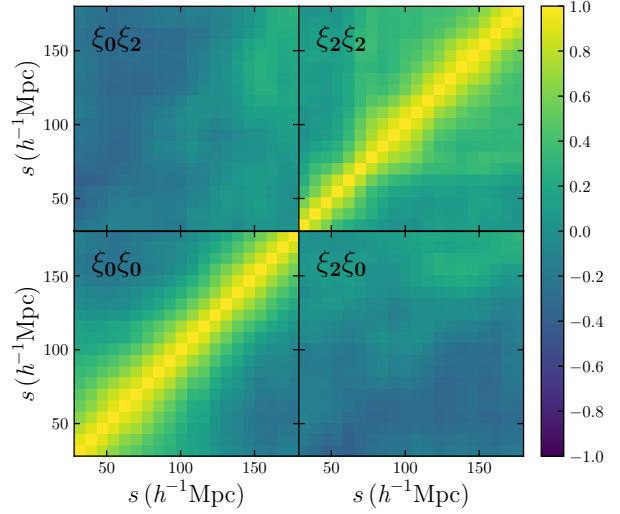


**Figure 1.** Monopole (top) and quadrupole (bottom) auto-correlation functions of the BOSS CMASS galaxies (markers) computed using two different binning schemes (20 and 10 linear bins in  $s$ ) coupled with the jackknife configurations given in Table 1 for the error estimation (200, 100, 50, 20 resamplings). We overplot the mean  $\pm\sigma$  values from the 200 log-normal light-cones (Sec. 4.1) as orange lines with the  $1\sigma$  uncertainty as shaded area. The analytic best-fit models to the CMASS measurements that we use to estimate the  $\alpha$  shift parameter (see Sec. 4.2) are shown as dashed purple curves.

fluctuations in the  $\sigma$  estimate due to having relatively few jackknife resamplings available. However, doing only 10 bins instead of 20 helps to partially mitigate these fluctuations.

On small scales, the errors from CMASS resamplings shown in the top panel of Fig. 4 are underestimated with respect to the LCs by up to  $\sim 40\%$  for both monopole and quadrupole. Around  $130 h^{-1}\text{Mpc}$ , the monopole errors from CMASS covariances are  $\sim 10 - 50\%$  larger than those from 200 light-cones, and the discrepancy increases with the number of resamplings. Compared to the monopole, the quadrupole shows smaller fluctuations in the  $1\sigma$  ratio shown in the top panel of Fig. 4.

The amplitude of the quadrupole 20 JK result is  $\sim 10 - 20\%$  lower than the others on scales below  $150 h^{-1}\text{Mpc}$ , while the monopole is lower than the rest only beyond  $100 h^{-1}\text{Mpc}$ . As expected, the 20 and 50 jackknife schemes return the largest fluctuations. Although larger jackknife re-



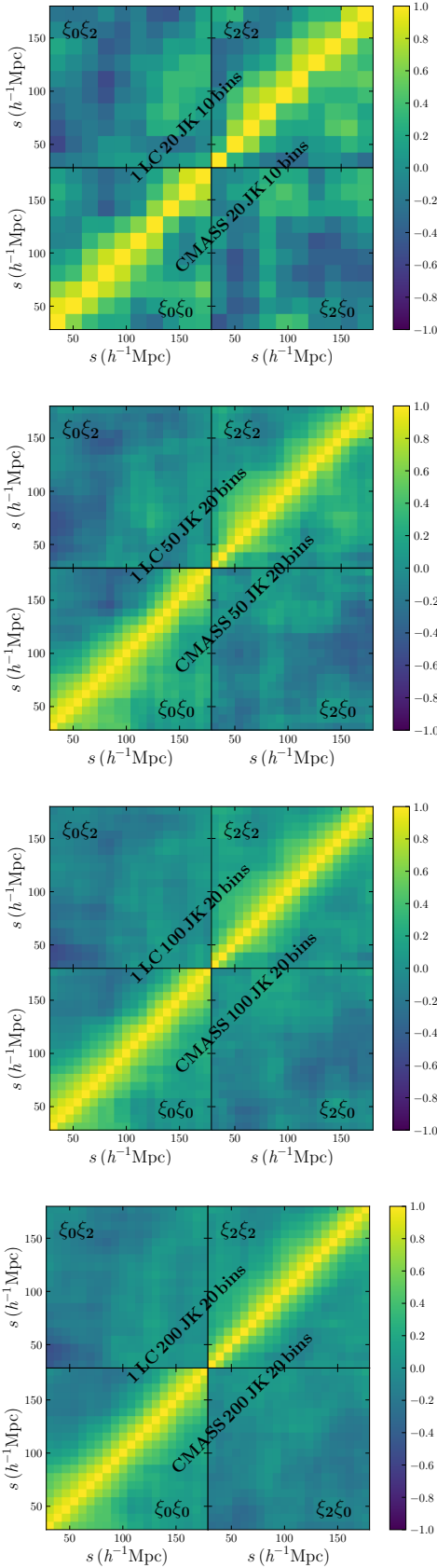
**Figure 2.** Normalised monopole and quadrupole auto- and cross-covariances obtained from the 200 log-normal light-cones without jackknife resampling. The normalisation is computed as  $C_{ij}^{\text{norm}} = C_{ij}/\sqrt{C_{ii}C_{jj}}$ , where  $C_{ij}$  is given in Eq. 12. The mean value and  $1\sigma$  dispersion of these mocks are shown in Fig. 1 as a solid line with the corresponding shaded region.

gions with greater independence may give a more accurate covariance estimate, the uncertainty on the covariance is large due to having few resamplings available. In the 20 JK scheme coupled with 10  $s$  bins, the large fluctuations due to a limited number of resamplings are partially mitigated by the smaller number of bins compared to the other cases.

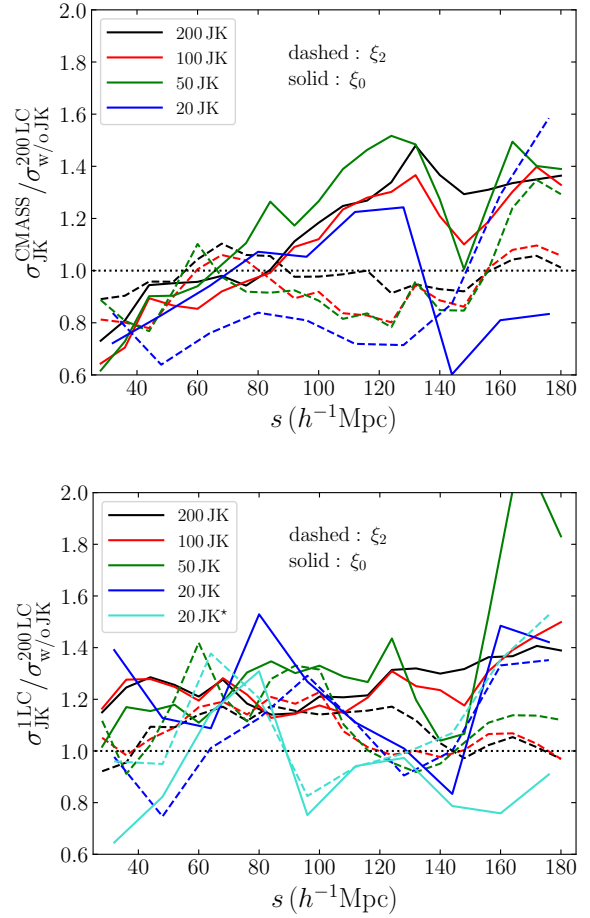
In the bottom panel of Fig. 4, we display the ratio of the uncertainties obtained by performing jackknife resampling on one of the light-cones and those from 200 LCs without jackknife. For the 20 JK case, we show the results from two different light-cone realisations (blue and turquoise lines) to highlight the significant uncertainty in the covariance estimate, which we expect to be high in this case. Compared to the ratio of CMASS jackknife to 200 LCs shown in the upper panel, overall here we find smaller fluctuations and a flatter trend. Again, the 20 and 50 JK configurations are the ones exhibiting the largest fluctuations due to the limited number of resamplings.

Out to BAO scales, the errors from jackknife resampling on the LC are overestimated by up to  $\sim 30\%$  for both the monopole and quadrupole. Beyond  $110 h^{-1}\text{Mpc}$ , the error estimates fluctuate significantly, but tend to indicate overestimation by  $\sim 50\%$ . Those from 50 and 20 JK also maintain a growing trend, but with even larger fluctuations due to the smaller number of resamplings. Beyond BAO scales, the quadrupole errors tend to decrease such that the JK estimates agree with the covariance computed from 200 independent realizations.

Overall, we find strong consistency between the uncertainties based on covariances computed either performing jackknife re-sampling on CMASS data, or on a light-cone with the same volume of CMASS, or from 200 LC realisations without jackknife. In general, with respect to the 200 LCs without jackknife, which represent the “ideal” case, the



**Figure 3.** Normalised covariances obtained from jackknife resampling performed on a light-cone (upper triangles) and on BOSS CMASS data (lower triangles). From top to bottom we display the 20, 50, 100 and 200 jackknife configurations coupled with two binning schemes.



**Figure 4.** *Top:* Ratios of the  $1\sigma$  uncertainties obtained from the CMASS jackknife covariances and the 200 LCs without JK. The solid (dashed) lines correspond to the monopole (quadrupole) measurements. *Bottom:* Ratios of the  $1\sigma$  errors obtained by applying jackknife to one of the light-cones and those from 200 LCs without jackknife. For the 20 JK LC scheme we show two different light-cone realisations (blue and turquoise lines). We remind the reader that the 200, 100 and 50 JK configurations are coupled with 20  $s$  bins, while the 20 JK case with 10 bins. The horizontal dotted lines are shown to help the comparison.

monopole errors from CMASS JK are underestimated on small scales and overestimated beyond  $\sim 80 h^{-1}\text{Mpc}$ , while those from jackknife on a light-cone are overestimated on all scales. For the quadrupole, the CMASS JK errors are overall underestimated compared to 200 LCs, while those from jackknife resampling on a LC are overestimated. All these discrepancies remain within 30% in most cases.

We next consider how the covariance of the correlation function propagates to the error on the  $\alpha$  shift parameter. Fig. 5 compares the values of the  $\alpha$  shift parameter and corresponding uncertainties inferred from (i) the BOSS CMASS jackknife covariances, (ii) the log-normal light-cone jackknife covariances, and (iii) the covariance from 200 LCs without jackknife. The specific values are reported in Tab. 3. All these results assume a tapering parameter  $T_p = 500$  which we found to be optimal. For the 20 JK/10 bins scheme applied to the light-cone we show two different LC realisations



	$\alpha$	
	BOSS CMASS	log-normal LC
200 JK, 20 bins:	$0.974^{+0.022}_{-0.020}$	$0.974^{+0.016}_{-0.014}$
100 JK, 20 bins:	$0.978^{+0.017}_{-0.017}$	$0.973^{+0.011}_{-0.011}$
50 JK, 20 bins:	$0.973^{+0.011}_{-0.011}$	$0.966^{+0.010}_{-0.010}$
20 JK, 10 bins:	$0.980^{+0.019}_{-0.018}$	$0.996^{+0.021}_{-0.016}$
20 JK, 10 bins*:		$0.985^{+0.013}_{-0.013}$
200 LCs w/o JK, 20 bins:		$0.992^{+0.010}_{-0.009}$

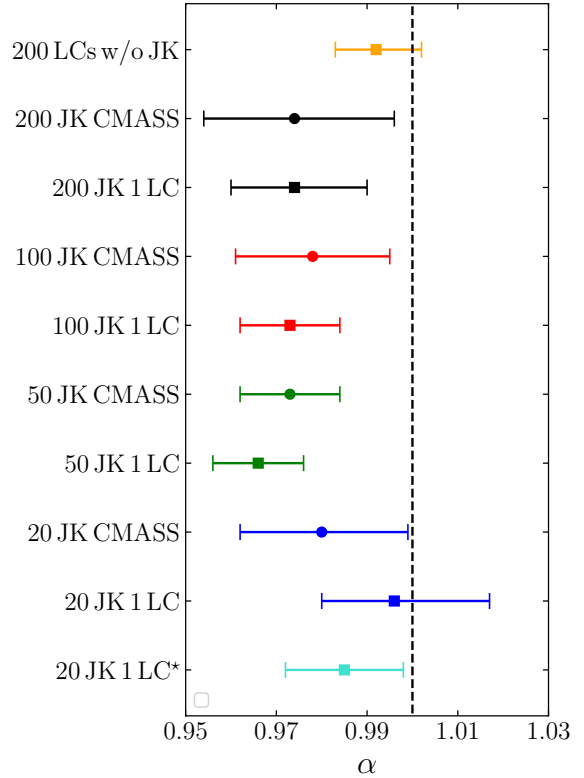
**Table 3.** Estimates of the  $\alpha$  shift parameter and its uncertainty obtained from the four jackknife configurations coupled with two binning schemes applied to both CMASS data and a log-normal lightcone. The last row shows the result obtained from the covariances of the 200 LCs without performing jackknife resampling. All these results here assume an optimal tapering parameter of  $T_p = 500$ . These results are shown in Fig. 5.

to highlight the fluctuations that the  $\alpha$  uncertainty can suffer due to the small number of resamplings. Overall, we find good agreement between the results based on covariances from jackknife, both applied to CMASS observations and LCs. For the LC results, we use the same realisation for all the jackknife configurations except the second 20 JK case, indicated with a star symbol, in which we test a different realisation. The 200 LC result without jackknife is consistent with the JK outcomes, despite the difference in the pre-factor of their covariances (Eqs. 7,12) and the lower level of noise in their off-diagonal terms (see Fig. 2).

The uncertainties on  $\alpha$  are all in agreement with each other, independently from the number/size of jackknife resamplings adopted. The average errors obtained from 200, 100 and 50 JK resamplings (i.e. the most robust ones) performed on CMASS data and a light-cone are  $\sim 1.6\%$  and  $\sim 1.2\%$ , respectively. That from the 200 LC covariances without jackknife is  $\sim 0.95\%$ . This latter case is “ideal” since the 200 LCs are all independent (but we do not expect the log-normal catalogs to capture the full covariance of the CMASS galaxy sample). The 50 and 20 JK schemes are the ones returning the largest fluctuations in the covariances, which can result in errors on  $\alpha$  as large as  $\sim 2.1\%$ . In order to precisely estimate the effect of the fluctuations in the 20 JK case, one should repeat the resamplings on many different LC realisations, but this goes beyond the scope of our analysis.

Despite the large fluctuations observed in the 20 JK configurations, the constraints on  $\alpha$  in some cases show that the error bars tend to reduce when the jackknife size increases. However, the trend depends on the tapering scheme and further trials on mocks are required to determine if this trend is real or not.

Fig. 6 shows the dependence of  $\alpha$  and its uncertainty on the tapering parameter,  $T_p$ . In the top panel, we show the results from the 200 light-cones and from jackknife applied to one of the LCs; in the bottom panel we show the CMASS jackknife outcomes. In both cases, we have run the MCMC chains assuming  $T_p = [50, 100, 300, 500, 700]$ . In the plots we



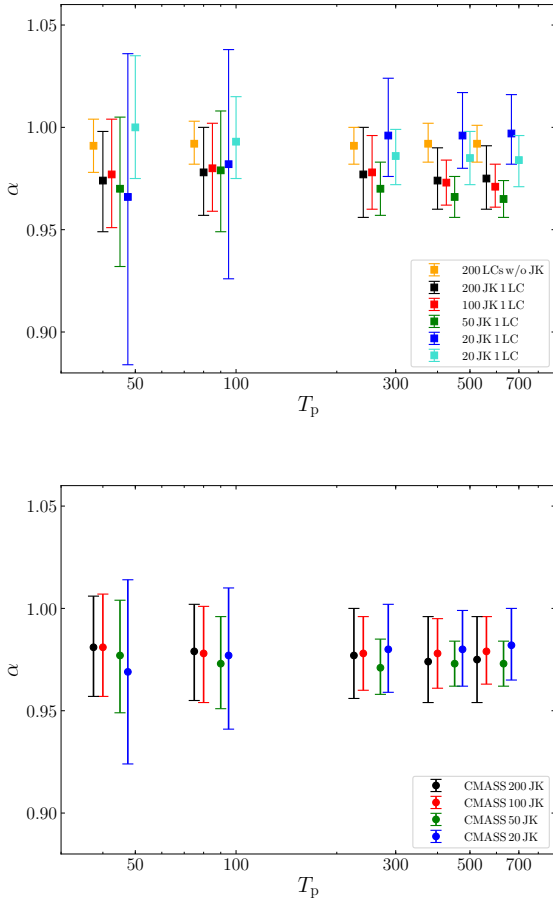
**Figure 5.** Summary of the  $\alpha$  shift parameters obtained from the covariances calculated using the jackknife configurations and binning schemes reported in Table 3. The points are color-coded as in Fig. 1, where each colour corresponds to a different jackknife/binning scheme. The results from CMASS are represented by dots, those from LCs by squares. The vertical line shows the value  $\alpha = 1$  to help the comparison. For the 20 JK case applied to a LC, we show two different LC realisations, one of them indicated with a star symbol (turquoise), to highlight how the  $1\sigma$  error can fluctuate due to the small number of resamplings. All these results are calculated assuming a tapering parameter  $T_p = 500$ .

offset the tapering values by a multiplicative factor to avoid crowding (see caption of Fig. 6). The optimal value, which provides errors on  $\alpha$  of  $\sim 1 - 2\%$ , turns out to be  $T_p = 500$ . In this way, our BOSS CMASS  $\alpha$  estimates are comparable with previous results in the literature (pre-reconstruction, e.g. Ross et al. 2016; Pearson & Samushia 2018).

The 20 JK configuration shows the largest fluctuations due to the limited number of resamplings. We have further tested our MCMC code without including any tapering correction and leaving only the Hartlap factor. In this case we find that the covariances from 20 JK resamplings are no longer semi-positive definite, meaning that they are not invertible, hence not useful for assembling the precision matrix needed to estimate  $\alpha$ . Such a result confirms that jackknife configurations with few cells tend to provide non robust covariance estimates.

## 7 DISCUSSION AND SUMMARY

We have studied the impact of choosing different sizes and numbers of jackknife resamplings on the accuracy of the co-



**Figure 6.** Shift parameter  $\alpha$  and its uncertainty as a function of the tapering parameter  $T_p$ . *Top*: results from covariances computed from 200 light-cones without jackknife and from jackknife performed on a LC. We have offset the  $T_p$  values on the x-axis by multiplying them, from left to right, by [0.75, 0.80, 0.85, 0.90, 0.95, 1.0]. *Bottom*: results from CMASS jackknife resampling. The  $T_p$  values have been offset by multiplying them, from left to right, by [0.80, 0.85, 0.90, 0.95, 1.0].

variance estimates and the  $\alpha$  shift parameter. To this purpose, we have measured the first two even multipoles of the BOSS CMASS DR12 galaxy sample at  $0.43 < z < 0.7$  and we have modelled the results both using a set of 200 log-normal light-cones (Sec. 4.1) and an analytic approach (Sec. 4.2). We have computed their covariances using 200, 100, 50 and 20 jackknife resamplings coupled with two binning schemes: 20 or 10 linear bins in  $24 < s < 184 h^{-1} \text{Mpc}$ , with 120 linear bins in  $0 < \mu < 1$  (see Sec. 3). We have compared the results with the covariances obtained from the 200 log-normal light-cones without jackknife. We have then applied the same jackknife configurations above on one of the light-cones to derive LC JK covariances directly comparable with the CMASS ones.

From these different covariance matrices we have derived corresponding precision matrices (Sec. 5), which we have used as inputs for our Monte Carlo Markov Chain to estimate the baryon acoustic scale through the  $\alpha$  shift parameter and its uncertainty. Our main findings are summarised in what follows:

- We find good consistency between the covariances obtained from CMASS and LC jackknife resamplings, and from 200 LCs without jackknife resampling. This leads to consistent error bars in both the galaxy clustering measurements and the  $\alpha$  shift parameter.
- We find no evidence for a bias in the inferred value of  $\alpha$  or its error when the jackknife cell size is smaller than the maximum 2PCF scale measured. However, with few resamplings available the error estimate becomes unreliable.
- We have demonstrated that it is useful to apply the Hartlap factor and the tapering scheme to estimate the precision matrix with jackknife resampling. The  $\alpha$  shift parameter estimated either from CMASS or LC jackknife covariances, or from 200 light-cones without jackknife, are all consistent between each other and with previous BOSS CMASS DR12 results from galaxy clustering pre-reconstruction analysis (Ross et al. 2016). We find uncertainties on  $\alpha$  of 1-2%, depending on the jackknife size/ 2PCF binning scheme adopted. This confirms that the jackknife methodology applied to both observations and mocks produces a comparable level of noise in the covariance estimates. This noise is then reduced in the precision matrix by applying the tapering correction (see Sec. 5).
- We have tested different values for the tapering parameter, in the range  $50 \leq T_p \leq 700$ , to maximise the accuracy in the  $\alpha$  shift parameter estimation. We find that the optimal value is  $T_p = 500$ . By lowering it, the noise in the precision matrix estimate is suppressed but the error on  $\alpha$  grows.

To summarise, performing jackknife resamplings either on BOSS CMASS DR12 data or on a log-normal light-cone with the same CMASS volume provides covariances that are consistent with those obtained from a set of 200 independent log-normal LCs and with previous results in the literature (Ross et al. 2016). These covariances lead to  $\alpha$  estimates with 1-2% uncertainties, depending on the jackknife size/2PCF binning scheme assumed.

The largest differences between covariance estimates from jackknife resampling and 200 log-normal light-cones without JK are visible in the off-diagonal terms. Here the jackknife results exhibit a higher level of noise. This difference is key for determining the accuracy of the  $\alpha$  shift parameter. The action of the tapering correction (Sec. 5) is to considerably reduce this noise returning comparable uncertainties on  $\alpha$  from all of the different covariance estimates tested.

Although previous works limit the jackknife scale to larger than the measured 2PCF scale (e.g. Beutler et al. 2011; Hong et al. 2016), we find that this is not essential. In fact, when using jackknife to estimate covariances, one should prioritise building a large number of resamplings rather than choosing a jackknife size larger than the maximum galaxy clustering scale measured. In fact, especially when studying BAO scales, by requiring  $S_{JK} \geq \max(s)$ , we are able to build only few wide jackknife regions, which leads to large uncertainties in the error estimates. In our results we do see a trend that the  $\alpha$  error bars tend to reduce as the jackknife size increases, but it is not seen in all tapering configurations and we do not have sufficient statistics to confirm whether it is real or not.

The new generation of cosmological surveys, such as DESI, Euclid or LSST, will span larger volumes compared to

SDSS-III/BOSS. The precision in their covariance estimates based on jackknife will be determined by the number of resamplings. We find that it is not essential to use jackknife sizes larger than the BAO scale, and so it will be possible to achieve  $N > 10^3$  resamplings to reach percent level precision on the error of cosmological parameters using the jackknife approach. In a followup work, we will address the feasibility of inferring accurate covariance estimates for a survey such as Euclid using a large number of jackknife resamplings.

## ACKNOWLEDGEMENTS

GF is funded through a Dennis Sciama fellowship at the Institute of Cosmology and Gravitation (ICG), at Portsmouth University.

DS acknowledges financial support from the Fondecyt Regular project number 1200171.

Funding for SDSS-III has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, and the U.S. Department of Energy Office of Science. The SDSS-III web site is <http://www.sdss3.org/>.

SDSS-III is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS-III Collaboration including the University of Arizona, the Brazilian Participation Group, Brookhaven National Laboratory, Carnegie Mellon University, University of Florida, the French Participation Group, the German Participation Group, Harvard University, the Instituto de Astrofísica de Canarias, the Michigan State/Notre Dame/JINA Participation Group, Johns Hopkins University, Lawrence Berkeley National Laboratory, Max Planck Institute for Astrophysics, Max Planck Institute for Extraterrestrial Physics, New Mexico State University, New York University, Ohio State University, Pennsylvania State University, University of Portsmouth, Princeton University, the Spanish Participation Group, University of Tokyo, University of Utah, Vanderbilt University, University of Virginia, University of Washington, and Yale University.

## REFERENCES

- Ade P. A. R., et al., 2016, *Astron. Astrophys.*, 594, A13
- Agrawal A., Makiya R., Chiang C.-T., Jeong D., Saito S., Komatsu E., 2017, *J. Cosmology Astropart. Phys.*, 2017, 10, 003
- Ahn C. P., Alexandroff R., Allende Prieto C., et al., 2012, *ApJS*, 203, 2, 21
- Aihara H., Allende Prieto C., An D., et al., 2011, *ApJS*, 193, 29
- Alam S., Albareti F. D., Allende Prieto C., et al., 2015, *ApJS*, 219, 1, 12
- Alam S., Ata M., Bailey S., et al., 2017, *MNRAS*, 470, 3, 2617
- Ali-Haïmoud Y., Bird S., 2013, *MNRAS*, 428, 4, 3375
- Anderson L., Aubourg E., Bailey S., et al., 2012, *MNRAS*, 427, 4, 3435
- Avila S., Murray S. G., Knebe A., Power C., Robotham A. S. G., Garcia-Bellido J., 2015, *MNRAS*, 450, 2, 1856
- Baugh C. M., 2006, *Reports on Progress in Physics*, 69, 3101
- Benson A. J., 2010, *Phys. Rep.*, 495, 2-3, 33
- Benson A. J., 2012, 17, 175
- Beutler F., Blake C., Colless M., et al., 2011, *MNRAS*, 416, 4, 3017
- Blas D., Lesgourgues J., Tram T., 2011, *J. Cosmology Astropart. Phys.*, 2011, 7, 034
- Bolton A. S., Schlegel D. J., Aubourg É., et al., 2012, *AJ*, 144, 5, 144
- Bower R. G., Benson A. J., Malbon R., et al., 2006, *MNRAS*, 370, 645
- Chuang C.-H., Kitaura F.-S., Prada F., Zhao C., Yepes G., 2015, *MNRAS*, 446, 3, 2621
- Coles P., Jones B., 1991, *MNRAS*, 248, 1
- Collacchioni F., Cora S. A., Lagos C. D. P., Vega-Martínez C. A., 2018, *MNRAS*, 481, 1, 954
- Cora S. A., 2006, *MNRAS*, 368, 1540
- Cora S. A., Vega-Martínez C. A., Hough T., et al., 2018, *MNRAS*, 479, 2
- Crain R. A., Schaye J., Bower R. G., et al., 2015, *MNRAS*, 450, 2, 1937
- Crocce M., Scoccimarro R., 2006, *Phys. Rev.*, D73, 063519
- Croton D. J., Springel V., White S. D. M., et al., 2006, *MNRAS*, 365, 11
- Croton D. J., Stevens A. R. H., Tonini C., et al., 2016, *ApJS*, 222, 22
- Davison A. C., Hinkley D. V., 1997, *Frontmatter, i-iv*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press
- De Lucia G., Boylan-Kolchin M., Benson A. J., Fontanot F., Monaco P., 2010, *Monthly Notices of the Royal Astronomical Society*, 406, 3, 1533
- Efron B., 1979, *Ann. Statist.*, 7, 1, 1
- Eisenstein D. J., et al., 2005, *Astrophys. J.*, 633, 560
- Eisenstein D. J., Hu W., 1998, *Astrophys. J.*, 496, 605
- Favole G., McBride C. K., Eisenstein D. J., et al., 2016a, *MNRAS*, 462, 2218
- Favole G., McBride C. K., Eisenstein D. J., et al., 2016b, *MNRAS*, 462, 2, 2218
- Favole G., Rodríguez-Torres S. A., Comparat J., et al., 2017, *MNRAS*, 472, 550
- Favole G., Sapone D., Silva Lafaunie J., 2019, arXiv:1912.06155
- Feldman H. A., Kaiser N., Peacock J. A., 1994, *ApJ*, 426, 23
- Feng Y., Chu M.-Y., Seljak U., McDonald P., 2016, *MNRAS*, 463, 3, 2273
- Font A. S., Johnston K. V., Ferguson A. M. N., et al., 2008, *ApJ*, 673, 215
- Fukugita M., Ichikawa T., Gunn J. E., Doi M., Shimasaku K., Schneider D. P., 1996, *AJ*, 111, 1748
- Gargiulo I. D., Cora S. A., Padilla N. D., et al., 2015, *MNRAS*, 446, 3820
- Genel S., Vogelsberger M., Springel V., et al., 2014, *MNRAS*, 445, 1, 175
- Gonzalez-Perez V., Lacey C. G., Baugh C. M., et al., 2014, *MNRAS*, 439, 264
- Górski K. M., Hivon E., Banday A. J., et al., 2005, *ApJ*, 622, 2, 759
- Gunn J. E., Carr M., Rockosi C., et al., 1998, *AJ*, 116, 3040
- Gunn J. E., Siegmund W. A., Mannery E. J., et al., 2006, *AJ*, 131, 2332
- Guo H., Zehavi I., Zheng Z., et al., 2012, ArXiv e-prints: 1212.1211
- Guo H., Zheng Z., Jing Y. P., et al., 2015a, *MNRAS*, 449, L95
- Guo H., Zheng Z., Zehavi I., et al., 2015b, *MNRAS*, 453, 4368
- Guo Q., White S., Boylan-Kolchin M., et al., 2011, *MNRAS*, 413, 101
- Hartlap J., Simon P., Schneider P., 2006, *Astron. Astrophys.*, [Astron. Astrophys.464,399(2007)]
- Henriques B., Maraston C., Monaco P., et al., 2011, *MNRAS*, 415, 3571
- Henriques B. M. B., White S. D. M., Thomas P. A., et al., 2013, *MNRAS*, 431, 3373
- Hirschmann M., De Lucia G., Fontanot F., 2016, *MNRAS*, 461, 1760

- Hong T., Han J. L., Wen Z. L., 2016, *ApJ*, 826, 2, 154
- Hou J., Lacey C. G., Frenk C. S., 2017, ArXiv e-prints
- Howlett C., Manera M., Percival W. J., 2015, *Astronomy and Computing*, 12, 109
- Hu W., Sugiyama N., 1996, *Astrophys. J.*, 471, 542
- Ivezić Ž., Kahn S. M., Tyson J. A., et al., 2019, *ApJ*, 873, 111
- Izard A., Crocce M., Fosalba P., 2016, *MNRAS*, 459, 3, 2327
- Kauffmann G., White S. D. M., Guiderdoni B., 1993, *MNRAS*, 264, 201
- Kaufman C. G., Schervish M. J., Nychka D. W., 2008, *Journal of the American Statistical Association*, 103, 484, 1545
- Kitaura F.-S., Rodríguez-Torres S., Chuang C.-H., et al., 2016, *MNRAS*, 456, 4, 4156
- Klypin A., Prada F., 2018, *MNRAS*, 478, 4602
- Koda J., Blake C., Beutler F., Kazin E., Marin F., 2016, *MNRAS*, 459, 2, 2118
- Lagos C. D. P., Baugh C. M., Lacey C. G., Benson A. J., Kim H.-S., Power C., 2011, *MNRAS*, 418, 1649
- Lagos C. d. P., Lacey C. G., Baugh C. M., 2013, *MNRAS*, 436, 1787
- Landy S. D., Szalay A. S., 1993, *ApJ*, 412, 64
- Laureijs R., Amiaux J., Arduini S., et al., 2011, arXiv e-prints, arXiv:1110.3193
- Lesgourgues J., 2011
- Lippich M., Sánchez A. G., Colavincenzo M., et al., 2019, *MNRAS*, 482, 2, 1786
- Manera M., Scoccimarro R., Percival W. J., et al., 2013, *MNRAS*, 428, 2, 1036
- Miller R. G., 1974, *Biometrika*, 61, 1
- Mo H., van den Bosch F. C., White S., 2010, *Galaxy Formation and Evolution*
- Monaco P., Benson A. J., De Lucia G., Fontanot F., Borgani S., Boylan-Kolchin M., 2014, *MNRAS*, 441, 2058
- Monaco P., Sefusatti E., Borgani S., et al., 2013, *MNRAS*, 433, 3, 2389
- Monaco P., Theuns T., Taffoni G., 2002, *MNRAS*, 331, 3, 587
- Naab T., Ostriker J. P., 2017, *Annual Review of Astronomy and Astrophysics*, 55, 1, 59
- Norberg P., Baugh C. M., Gaztañaga E., Croton D. J., 2009, *MNRAS*, 396, 19
- Norberg P., Gaztañaga E., Baugh C. M., Croton D. J., 2011, *MNRAS*, 418, 2435
- Padmanabhan N., White M. J., 2008, *Phys. Rev.*, D77, 123540
- Paribelli G., Viel M., Sefusatti E., 2019, *J. Cosmology Astropart. Phys.*, 2019, 1, 010
- Paz D. J., Sánchez A. G., 2015, *MNRAS*, 454, 4, 4326
- Pearson D. W., Samushia L., 2018, *Monthly Notices of the Royal Astronomical Society*, 478, 4, 4500
- Pearson D. W., Samushia L., Gagrani P., 2016, *MNRAS*, 463, 3, 2708
- Pillepich A., Springel V., Nelson D., et al., 2018, *MNRAS*, 473, 3, 4077
- Planck, Ade P. A. R., Aghanim N., et al., 2014, *A&A*, 571, A16
- Quenouille M. H., 1956, *Biometrika*, 43, 353
- Reid B., Ho S., Padmanabhan N., et al., 2016, *MNRAS*, 455, 2, 1553
- Ross A. J., Beutler F., Chuang C.-H., et al., 2016, *Monthly Notices of the Royal Astronomical Society*, 464, 1, 1168
- Ross A. J., Percival W. J., Sánchez A. G., et al., 2012, *MNRAS*, 424, 564
- Sánchez A. G., Scóccola C. G., Ross A. J., et al., 2012, *MNRAS*, 425, 1, 415
- Sartoris B., Biviano A., Fedeli C., et al., 2016, *MNRAS*, 459, 2, 1764
- Schaye J., Crain R. A., Bower R. G., et al., 2015, *MNRAS*, 446, 521
- Schlegel D. J., Blum R. D., Castander F. J., et al., 2015, in *American Astronomical Society Meeting Abstracts #225*, vol. 225 of *American Astronomical Society Meeting Abstracts*, 336.07
- Scoccimarro R., Sheth R. K., 2002, *MNRAS*, 329, 3, 629
- Smee S. A., Gunn J. E., Uomoto A., et al., 2013, *AJ*, 146, 32
- Somerville R. S., Davé R., 2015, *Annual Review of Astronomy and Astrophysics*, 53, 1, 51
- Springel V., 2010, *MNRAS*, 401, 2, 791
- Springel V., Hernquist L., 2003, *MNRAS*, 339, 2, 312
- Stevens A. R. H., Brown T., 2017, *MNRAS*, 471, 1, 447
- Takahashi R., Sato M., Nishimichi T., Taruya A., Oguri M., 2012, *ApJ*, 761, 2, 152
- Tassev S., Zaldarriaga M., Eisenstein D. J., 2013, *J. Cosmology Astropart. Phys.*, 2013, 6, 036
- Tonini C., Maraston C., Devriendt J., Thomas D., Silk J., 2009, *MNRAS*, 396, L36
- Turkey J., 1958, *The Annals of Mathematical Statistics*, 29, 1
- Vogelsberger M., Genel S., Springel V., et al., 2014a, *MNRAS*, 444, 1518
- Vogelsberger M., Genel S., Springel V., et al., 2014b, *Nature*, 509, 177
- Weinmann S. M., van den Bosch F. C., Yang X., Mo H. J., 2006, *MNRAS*, 366, 2
- White M., Tinker J. L., McBride C. K., 2014, *MNRAS*, 437, 3, 2594
- White S. D. M., Frenk C. S., 1991, *ApJ*, 379, 52
- Wright B. S., Winther H. A., Koyama K., 2017, *Journal of Cosmology and Astroparticle Physics*, 2017, 10, 054
- Xu X., Padmanabhan N., Eisenstein D. J., Mehta K. T., Cuesta A. J., 2012, *Mon. Not. Roy. Astron. Soc.*, 427, 2146
- Yepes G., Kates R., Khokhlov A., Klypin A., 1997, *MNRAS*, 284, 1, 235
- Yoshida N., Sokasian A., Hernquist L., Springel V., 2003, *ApJ*, 598, 1, 73
- Zehavi I., Blanton M. R., Frieman J. A., et al., 2002, *ApJ*, 571, 172
- Zehavi I., Zheng Z., Weinberg D. H., et al., 2005, *ApJ*, 630, 1
- Zehavi I., Zheng Z., Weinberg D. H., et al., 2011, *ApJ*, 736, 59

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