Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **484,** 3713–3730 (2019)
Advance Access publication 2019 January 12

doi:10.1093/mnras/stz051

Enhancing BOSS bispectrum cosmological constraints with maximal compression

Davide Gualdi [1]★ Héctor Gil-Marín [2,3] Robert L. Schuhmann,[4,5] Marc Manera,[6,7] Benjamin Joachimi[1] and Ofer Lahav[1]

[1]*Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK*
[2]*Sorbonne Université, Institut Lagrange de Paris (ILP), 98 bis Boulevard Arago, F-75014 Paris, France*
[3]*Laboratoire de Physique Nucléaire et de Hautes Energies, Université Pierre et Marie Curie, F-75252 Paris, France*
[4]*Institute for Astronomy, University of Edinburgh, Blackford Hill, Edinburgh EH9 3HJ, UK*
[5]*School of Physics and Astronomy, University of Manchester, Manchester M13 9PL, UK*
[6]*Institut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology, Campus UAB, E-08193 Bellaterra (Barcelona), Spain*
[7]*Centre for Mathematical Sciences, DAMTP, Cambridge University, Wilberforce Rd, Cambridge CB3 0WA, UK*

Accepted 2018 December 25. Received 2018 November 29; in original form 2018 June 7

ABSTRACT

We apply two compression methods to the galaxy power spectrum monopole/quadrupole and bispectrum monopole measurements from the Baryon Oscillation Spectroscopic Survey DR12 CMASS sample. Both methods reduce the dimension of the original data vector to the number of cosmological parameters considered, using the Karhunen–Loève (KL) algorithm with an analytic covariance model. In the first case, we infer the posterior through Markov chain Monte Carlo (MCMC) sampling from the likelihood of the compressed data vector (MC–KL). The second, faster option, works by first Gaussianizing and then orthogonalizing the parameter space before the compression; in this option (G-PCA) we only need to run a low-resolution preliminary MCMC sample for the Gaussianization to compute our posterior. Both compression methods accurately reproduce the posterior distributions obtained by standard MCMC sampling on the CMASS data set for a k-space range of 0.03–0.12 $h\,\mathrm{Mpc}^{-1}$. The compression enables us to increase the number of bispectrum measurements by a factor of $\sim$23 over the standard binning (from 116 to 2734 triangle bins used), which is otherwise limited by the number of mock catalogues available. This reduces the 68 per cent credible intervals for the parameters (b_1, b_2, f, σ_8) by (-24.8 per cent, -52.8 per cent, -26.4 per cent, -21 per cent), respectively. Using these methods for future redshift surveys such as DESI, Euclid, and PFS will drastically reduce the number of simulations needed to compute accurate covariance matrices and will facilitate tighter constraints on cosmological parameters.

Key words: methods: analytical – cosmological parameters – large-scale structure of Universe.

1 INTRODUCTION

Large data sets have recently become available from current cosmological surveys (*Planck*,[1] Ade et al. 2014; Sloan Digital Sky Survey,[2] Eisenstein et al. 2011; KiDS de Jong et al. 2013; DES, Dark Energy Survey Collaboration 2016[3]) and even larger ones will be provided in future (DESI,[4] Levi et al. 2013; Euclid,[5] Laureijs et al. 2011; PFS,[6] Takada et al. 2014; the LSST,[7] LSST Science Collaboration 2009). In order to exploit their full potential, it is desirable to go beyond standard two-points (2pt) statistics.

Three-points (3pt) statistics are a complementary probe that is possible to investigate both in configuration and Fourier space and have been used extensively in galaxy clustering analyses (Groth & Peebles 1977; Fry 1984; Fry & Gaztanaga 1993; Frieman &

★ E-mail: davide.gualdi.14@ucl.ac.uk
[1]http://sci.esa.int/planck/
[2]http://www.sdss3.org/surveys/boss.php
[3]https://www.darkenergysurvey.org
[4]http://desi.lbl.gov
[5]http://sci.esa.int/euclid/
[6]http://pfs.ipmu.jp
[7]https://www.lsst.org/

Gaztanaga 1994; Matarrese, Verde & Heavens 1997; Heavens, Matarrese & Verde 1998; Verde et al. 1998; Scoccimarro et al. 1998a; Scoccimarro 2000; Sefusatti et al. 2006). Deviations from General Relativity (Borisov & Jain 2009; Bernardeau & Brax 2011; Gil-Marín et al. 2011) and primordial non-Gaussianities (Fry & Scherrer 1994; Gangui et al. 1994; Verde et al. 2000; Liguori et al. 2010; Tellarini et al. 2016) have been investigated using 3pt statistics. Their potential in lifting degeneracies present at 2pt level has been shown by the most recent measurement on the Baryon Oscillation Spectroscopic Survey (BOSS) data set, for the bispectrum by Gil-Marín et al. (2017) and for the 3pt correlation function by Slepian et al. (2017a). Baryonic acoustic oscillations have also been measured using the 3pt correlation function by Slepian et al. (2017b) and detected using the bispectrum by Pearson & Samushia (2017).

Recently, 3pt statistics have been studied in the case of 21cm emission lines by Hoffmann et al. (2018). For what concerns weak lensing, its effect on 3pt galaxy clustering have been studied by Schmidt et al. (2008). Moreover, the weak-lensing bispectrum has been an object of several studies in recent years (Takada & Jain 2004; Joachimi, Shi & Schneider 2009; Kayo & Takada 2013; Kayo, Takada & Jain 2013). The skewness of mass aperture statistic was considered by Jarvis, Bernstein & Jain (2004), while the 3pt correlation function of cosmic shear was analysed by Schneider, Kilbinger & Lombardi (2005) and Kilbinger & Schneider (2005). Higher order statistics such as the bispectrum via gravitational lensing have been investigated also by Simon et al. (2013), Fu et al. (2014), Simon et al. (2015), and Pyne, Joachimi & Peiris (2017).

Besides being computationally more expensive than 2pt statistics, 3pt statistics present the drawback to be described by very large data vectors, which in turn require a high number of simulations to accurately estimate their covariance matrix (Hartlap, Simon & Schneider 2007). In Gualdi et al. (2018), Paper I from now on, we presented two methods to compress the redshift-space galaxy bispectrum, namely MC–KL (MCMC sampling + Karhunen–Loève/MOPED compression) and PCA + KL (principal component analysis transformation + Karhunen–Loève/MOPED compression). The MOPED algorithm was introduced in Heavens, Jimenez & Lahav (2000) and it achieves maximal compression of the original datas-vector by extending to the multiple parameter case the Karhunen-Loève method introduced in Tegmark, Taylor & Heavens (1997).

MC–KL consists in sampling via MCMC the compressed data vector's likelihood. PCA + KL reconstructs the multidimensional physical posterior distribution from the 1D posterior of orthogonalized parameters obtained by diagonalizing the Fisher information matrix. Modifications/improvements of the MOPED algorithm were introduced also recently by Heavens et al. (2017), Alsing & Wandelt (2018), and Alsing, Wandelt & Feeney (2018) also with the target of data compression.

In this work, we apply our compression methods to both the power spectrum monopole/quadrupole and to te bispectrum monopole measurements from the CMASS sample of BOSS DR12. While the MC–KL is more flexible than the PCA + KL method since it doesn't require the multidimensional Gaussian posterior assumption, the PCA + KL is much faster in terms of computational time and requires far fewer computational resources (it can be run on standard laptop). We compare both methods and test their convergence in terms of deriving equivalent posterior distributions.

In order to make the PCA + KL method applicable also to parameter spaces with strong degeneracies, for which the posterior Gaussianity approximation is no longer valid, we introduce a pre-Gaussianization step based on the algorithm developed by Schuhmann, Joachimi & Peiris (2016).

We measure the bispectrum monopole using the same code used for the BOSS DR12 analysis done by Gil-Marín et al. (2017). We vary the size of the triangle vectors by changing the bin size Δk for k, which returns different number of triangular shapes given the minimum and maximum scales. For the same number of triangle bins, the compression returns posterior distributions slightly larger than the MCMC counterparts. However, when compressing a much larger number of triangle bins (which cannot be done for the MCMC on the full data vector because of the limited number of mocks available constraint), the posterior distribution becomes more Gaussian and narrow. It eventually returns tighter constraints than the ones obtained by the standard analysis.

In Section 2, we present the analytical model used for the data vector considered and the analytical expression of the covariance matrix used to derive the weights for the compression. In Section 3, we describe the data set and the galaxy mocks used to estimate the covariance matrix together with the settings of our analysis. In Section 4, we recap the compression methods applied including the Gaussianization extension for the original PCA + KL method. We report the performance of the compression methods compared to the MCMC sampling for the cases in which it is possible to run it on the full data vector in Section 5. We describe the gain in parameter constraints as a function of the number of triangle bins used in the bispectrum monopole data vector component in Section 6. We test the flexibility and accuracy of the compression methods presented in Section 7. Finally, we conclude summarizing our results in Section 8. In Appendix A, we report the full derivation of all the analytic expressions used in the analysis. In Appendix B, additional validation tests are presented.

2 DATA VECTOR AND COVARIANCE MATRIX

In order to measure the power spectrum and bispectrum from the data and the mocks catalogues, we use the estimators described in Gil-Marín et al. (2016a,b). These are based on the weighted field of density fluctuations (Feldman, Kaiser & Peacock 1994):

$$F_\lambda(\boldsymbol{r}) = \frac{w_{\rm FKP}(\boldsymbol{r})}{I_\lambda^{1/2}}[w_{\rm c}(\boldsymbol{r})n(\boldsymbol{r}) - \alpha n_{\rm syn}(\boldsymbol{r})], \quad (1)$$

where $w_{\rm c}$ is the weight taking into account all the measurement systematics (redshift failure, fiber collision, and target density variations), $w_{\rm FKP}$ (Feldman, Kaiser, and Peacock) ensures the condition of minimum variance, n is the observed number density of galaxies, $n_{\rm syn}$ is the number density of objects in a synthetic catalogue, and I_λ is the normalization of the amplitude of the observed power ($\lambda = 2$, 3 for power spectrum and bispectrum, respectively). α is the ratio between weighted number of observed galaxies over the weighted number of objects in the synthetic catalogues.

2.1 Power spectrum monopole and quadrupole

The redshift-space galaxy power spectrum model adopted in this work is a linear one including redshift-space distortions (RSDs) plus a damping function taking into account the Finger-of-God (FoG) effect:

$$\mathrm{P}_{\rm g}^{\rm s}(k,\mu) = \mathrm{D}_{\rm FoG}^{\rm P}\left(k,\mu,\sigma_{\rm FoG}^{\rm P}[z]\right)\mathrm{Z}_1^{\rm s}(\boldsymbol{k})^2\,\mathrm{P}_{\rm m}^{\rm lin.}(k)\ , \quad (2)$$

where k is the module of the wave vector $\boldsymbol{k}$, and μ is the cosine of the angle between the wave vector and the line of sight. The standard RSD kernels $\mathrm{Z}_i^{\rm s}$ are reported in the Appendix of Gil-Marín et al.

(2014) together with the FoG damping function expression. $\sigma_{\rm FoG}^{\rm P}[z]$ is the FoG free parameter for the power spectrum. For the range of scales considered in this work, the linear RSD model has been tested on N-body simulations and proved to be a good approximation (Taruya, Nishimichi & Saito 2010, fig. 2). The redshift-space galaxy power spectrum can be expanded in terms of Legendre polynomials using its dependence on μ:

$$\mathrm{P_g^s}\,(k,\mu) = \sum_{\ell=0}^{\infty} \mathrm{P_g^{(\ell)}}\,(k)\,L_\ell\,(\mu)\ , \tag{3}$$

where $L_\ell(\mu)$ is the ℓ-order Legendre polynomial. Almost all the signal is contained in the first two even multipoles, the monopole, and the quadrupole ($\ell = 0, 2$). These can be found by inverting the above expression:

$$\mathrm{P_g^{(\ell)}}\,(k) = \frac{2\ell+1}{2}\int_{-1}^{+1} \mathrm{d}\mu\,\mathrm{P_g^s}\,(k,\mu)\,L_\ell\,(\mu)\ . \tag{4}$$

2.2 Analytical expression for $\mathrm{P_g^{(0,2)}}$ covariance matrices

Defining an estimator as in Appendix A1, it is possible to derive the expression for the Gaussian term of the power spectrum monopole and quadrupole covariance matrices (Appendix A2):

$$\mathrm{C_G^{P_g^{(\ell)}P_g^{(\ell)}}}\,(k_1;k_2) = \left(\frac{2\ell+1}{2}\right)^2 \frac{2\delta_{12}^{\rm K}}{N_{\rm p}\,(k_1)}\quad \mathrm{P_g^{(\ell)}}\,(k_1)^2\,, \tag{5}$$

where $\delta_{12}^{\rm K}$ is the Kronecker delta between k_1 and k_2, while $N_{\rm p}(k_1)$ is the number of pairs of grid points inside the estimator integration volume in Fourier space $V_k = 4\pi k^2\Delta k$ (Scoccimarro et al. 1998b), and it is proportional to an effective survey volume $V_{\rm e}$. The $V_{\rm e}$ normalization is used to obtain a closer match between the analytic and mocks covariance matrices (please refer to equations A2 and A11 for more details). We set the cross-covariance between power spectrum monopole and quadrupole to zero since it is negligible with respect to the other terms, as can be seen from fig. 3 in Gil-Marín et al. (2017).

2.3 Bispectrum monopole

For the redshift-space galaxy bispectrum, we adopt the effective model presented in Gil-Marín et al. (2014), which modifies the RSD kernels derived from perturbations theory in order to better fit the data at non-linear scales (see the Appendix of the paper above for the full expressions). This effective model includes 18 parameters that have been calibrated using simulations (Gil-Marín et al. 2012, 2014). The model has been applied to both BOSS DR11 and DR12 data sets (Gil-Marín et al. 2015, 2017). The tree level has also been corrected to take into account the FoG damping effect:

$$\begin{aligned}&\mathrm{B_g^s}\,(\boldsymbol{k}_1,\boldsymbol{k}_2,\boldsymbol{k}_3) = \mathrm{D_{FoG}^B}\left(\boldsymbol{k}_1,\boldsymbol{k}_2,\boldsymbol{k}_3,\sigma_{\rm FoG}^{\rm B}[z]\right)\\ &\times\left[Z_1^{\rm s}\,(\boldsymbol{k}_1)\,Z_1^{\rm s}\,(\boldsymbol{k}_2)\,Z_{2,\rm eff.}^{\rm s}\,[\boldsymbol{k}_1,\boldsymbol{k}_2]\,\mathrm{P_m^{lin.}}\,(k_1)\,\mathrm{P_m^{lin.}}\,(k_2) + \mathrm{cyc.}\right]\ ,\end{aligned} \tag{6}$$

where $\sigma_{\rm FoG}^{\rm B}[z]$ is the FoG free parameter for the bispectrum. The monopole of the bispectrum corresponds to the average of all the possible orientations of a certain triangle, given by three wave-vectors' moduli, with respect to the line of sight. It can therefore be obtained by integrating over two angular coordinates:

$$\begin{aligned}\mathrm{B_g^{(0)}}\,(k_1,k_2,k_3) &= \frac{1}{4}\int_{-1}^{1}\mathrm{d}\mu_1\int_{-1}^{1}\mathrm{d}\mu_2\,\mathrm{B_g^s}\,(\boldsymbol{k}_1,\boldsymbol{k}_2,\boldsymbol{k}_3)\\ &= \frac{1}{4\pi}\int_{-1}^{1}\mathrm{d}\mu_1\int_{0}^{2\pi}\mathrm{d}\phi\,\mathrm{B_g^s}\,(\boldsymbol{k}_1,\boldsymbol{k}_2,\boldsymbol{k}_3)\ ,\end{aligned} \tag{7}$$

where μ_i is the cosine of the angle between the $\boldsymbol{k}_i$ vector and the line of sight. The angle ϕ is defined as $\mu_2 \equiv \mu_1 x_{12} - \sqrt{1-\mu_1^2}\sqrt{1-x_{12}^2}\cos\phi$ and where x_{12} is the cosine of the angle between $\boldsymbol{k}_1$ and $\boldsymbol{k}_2$. More details are given in Appendix A.

2.4 Analytical expression for $\mathrm{B_g^{(0)}}$ covariance matrix

In order to apply the compression methods presented in Paper I, we need an analytical expression for the bispectrum monopole covariance matrix. This allows us to compress a data vector with an arbitrarily large number of triangle bins, which on the contrary wouldn't be possible using a covariance matrix estimated from the galaxy mock catalogues. That is because in order to obtain an accurate numerical estimate of the covariance matrix, the number of simulations used must be much greater than the data vector's dimension (Hartlap et al. 2007; Percival et al. 2014).

As it has been shown in Paper I, compressing the power spectrum together with the bispectrum, or leaving it uncompressed, does not make any substantial difference in terms of recovered parameter constraints. However, it makes a huge difference in terms of complexity of the covariance matrix that one has to model analytically in order to compress the data vector. Compressing the power spectrum as well (monopole and quadrupole) also requires modelling their covariance matrices together with the cross-covariance with the bispectrum monopole. Leaving them uncompressed just requires to model the bispectrum monopole covariance matrix.

The covariance terms for the bispectrum monopole below reported are original of this work. Expressions for the matter bispectrum were derived also by Scoccimarro et al. (1998b), Sefusatti et al. (2006), and Chan & Blot (2017), however, in order to compute covariance matrix we proceed similar to what done in Kayo et al. (2013).

The expression for the Gaussian term of $\mathrm{C^{B_g^0B_g^0}}$ is derived in Appendix A3 and reads

$$\begin{aligned}&\mathrm{C_G^{B_g^0B_g^0}}\,(k_1,k_2,k_3;k_4,k_5,k_6) =\\ &= \frac{\mathrm{D}_{123456}}{16\pi^2}\frac{V_{\rm e}}{N_{\rm t}\,(k_1,k_2,k_3)}\mathrm{P_g^{(0)}}\,(k_1)\,\mathrm{P_g^{(0)}}\,(k_2)\,\mathrm{P_g^{(0)}}\,(k_3)\,,\end{aligned} \tag{8}$$

where D_{123456} stands for all the possible permutations for which each side of the first triangle is equal to a side of the second one; it has the values 6, 2, and 1, respectively, for equilateral, isosceles, and scalene triangles. $N_{\rm t}(k_1, k_2, k_3)$ is the number of independent triplets of grid points in the integration volume in Fourier space $V_{k_{123}} \simeq 8\pi^2 k_1 k_2 k_3 \Delta k_1 \Delta k_2 \Delta k_3$. For the values of the effective survey volume and the average galaxy density number used in computing the analytical covariance matrix, we adopt the values $V_{\rm e} = 2.43\times10^9\,\mathrm{Mpc}^3$ and $\bar{n}_{\rm g} = 1.14\times10^{-4}\,\mathrm{Mpc}^{-3}$ used by Slepian et al. (2017a) for both power spectrum monopole/quadrupole and bispectrum monopole analytical covariance matrices. In practice, we use the analytic expression of the covariance matrix only to determine the weights for the compression. Since all the terms considered scale as $V_{\rm e}^{-1}$, the effective volume acts only as a scaling factor not affecting the compression performance.

In order to describe the correlation between different triangle bins in our analytical model of the covariance matrix, we include also a non-Gaussian term of the bispectrum monopole covariance matrix. In the expansion of the bispectrum covariance matrix presented in the Appendix of Paper I, for the bispectrum monopole this corresponds to a term proportional to the product of two bispectra

monopoles as shown in Appendix A4:

$$\mathrm{C}_{\mathrm{NG}}^{\mathrm{B}_{\mathrm{g}}^{0}\mathrm{B}_{\mathrm{g}}^{0}}(k_1,k_2,k_3;k_4,k_5,k_6) = \\ = \frac{\delta_{34}^{\mathrm{K}}}{16\pi^2}\frac{k_f^3}{4\pi k_3^2\Delta k_3}\mathrm{B}_{\mathrm{g}}^{(0)}(k_1,k_2,k_3)\,\mathrm{B}_{\mathrm{g}}^{(0)}(k_3,k_5,k_6) \quad +8 \quad \text{perm.} \tag{9}$$

It is important to include a term modelling the correlation between different triangle bins since the number of possible configurations increases very quickly as the bin size decreases. For simplicity, we only used this term to model the correlation between different triangle bins. However, it is important to notice that a better approximation of the analytical covariance matrix can be obtained by including the expressions corresponding to all the terms present in the expansion given in the Appendix of Paper I.

We do not include a corresponding non-Gaussian term into the power spectrum monopole and quadrupole covariances since the number of data points considered is relatively low, thus the separation between the k modules values is more than sufficient to assume that the correlation between two different modes k_i and k_j is negligible with respect to their variance (approximated by the Gaussian term on the diagonal of the covariance matrix). From fig. 3 in Gil-Marín et al. (2017), it can be seen that the cross-covariance between different data points for the monopole and quadrupole of the power spectrum is much weaker than their variance.

2.5 Analytical expression for $\left[\mathrm{P}_{\mathrm{g}}^{(0,2)}, \mathrm{B}_{\mathrm{g}}^{(0)}\right]$ cross-covariance matrix

Finally we also model the cross-covariance between power spectrum multipoles and bispectrum monopole as described in Appendix A5:

$$\mathrm{C}^{\mathrm{P}_{\mathrm{g}}^{(\ell)}\mathrm{B}_{\mathrm{g}}^{0}}(k_1;k_2,k_3,k_4) = \\ = \frac{1}{2\pi}\left(\frac{2\ell+1}{2}\right)\frac{\delta_{12}^{\mathrm{K}}}{N_{\mathrm{p}}(k_2)}\mathrm{P}_{\mathrm{g}}^{(\ell)}(k_2)\,\mathrm{B}_{\mathrm{g}}^{(0)}(k_2,k_3,k_4) \quad +2 \quad \text{perm.}. \tag{10}$$

As done in Paper I, we made the assumption that the shot noise is well approximated by a Gaussian distribution (which is reasonable if the galaxy number density is fairly high). Therefore, we just modify the galaxy power spectrum expressions by adding a $\bar{n}_{\mathrm{g}}^{-1}$ term. We did not take into account the effect of the survey geometry in the theoretical covariance matrix expression, which would affect the large scales inducing an extra correlation among the modes. We leave the inclusion of this correction for future work. Please refer to Howlett & Percival (2017) for a more detailed study on how to include this effect in the covariance matrix.

3 DATA, MOCKS, AND ANALYSIS

3.1 DR12 BOSS data and mocks catalogues

In this paper, we use the CMASS galaxy sample ($0.43 \leq z \leq 0.70$) of the BOSS (Dawson et al. 2013), which is part of the Sloan Digital Sky Survey III (Eisenstein et al. 2011). In the final data release DR12, the CMASS sample contains the spectroscopic redshift of 777202 galaxies (see Gil-Marín et al. 2017 and Alam et al. 2017 for more details).

In order to accurately numerically estimate the covariance matrix, it is necessary to employ a large suite of mock galaxy catalogues. These are different realizations of the same region of the Universe based on methods such as second-order Lagrangian perturbation theory (Scoccimarro & Sheth 2002; Manera et al. 2013) or augmented Lagrangian perturbation theory as described in Kitaura & Heß (2013). By measuring the data vector of interest on each one of these catalogues, we can numerically estimate the covariance matrix, which will be used in the likelihood evaluation. In this work, we use subsets of the 2048 realizations of the Multidark Patchy BOSS DR12 mocks by Kitaura et al. (2016). This set of mocks has been run using the underlying cosmology: $\Omega_\Lambda = 0.693$, $\Omega_{\mathrm{m}}(z=0) = 0.307$, $\Omega_{\mathrm{b}}(z=0) = 0.048$, $\sigma_8(z=0) = 0.829$, $n_{\mathrm{s}} = 0.96$, and $h_0 = 0.678$.

3.2 Analysis settings

For the power spectrum monopole and quadrupole, the bin size was fixed to $\Delta k = 0.01h\,\mathrm{Mpc}^{-1}$. We measured the bispectrum monopole from both data and mocks using different multiples of the fundamental frequency defined as $k_f^3 = \frac{(2\pi)^3}{V_{\mathrm{s}}}$, where V_{s} is the survey volume that in this case was the cubic box volume $V_{\mathrm{s}} = L_{\mathrm{b}}^3 = (3500\,\mathrm{Mpc}/h)^3$ used to analyse the galaxy mocks. In particular, the considered bin sizes for the bispectrum are $\Delta k = (6, 5, 4, 2) \times k_f$, respectively, corresponding to 116, 195, 404, and 2734 triangle bins used between $0.02 < k_i\,(h\,\mathrm{Mpc}^{-1}]) < 0.12$. For every Δk_i bin size, all the measured triangle bins bispectra, which depends on the chosen bin size, are regrouped in the number of triangle bins above specified. The largest bin size $\Delta k = 6 \times k_f$ corresponds to the one used in the BOSS collaboration analysis done by Gil-Marín et al. (2017). For the k-range considered in the BOSS analysis, the Δk_6 ($\Delta k = 6 \times k_f$) binning case corresponded to 825 triangle bins (triplets of wave-vector's modulus), while Δk_2 would have corresponded to more than ∼7000 triangle bins.

In all the parameter estimation analyses that we are going to perform, we use the covariance matrix derived from the galaxy catalogues described above (see Section 3.1). In particular, we use 1400 mocks to estimate the covariance matrix when running the MCMC sampling on the full data vector. We use 700 when the analysis is performed using the compressed data vector.

The largest scales considered in this work are $k_{\mathrm{min}} = 0.03\,h\,\mathrm{Mpc}^{-1}$ for both power spectrum monopole and quadrupole and $k_{\mathrm{min}} = 0.02\,h\,\mathrm{Mpc}^{-1}$ for the bispectrum monopole. The choice of $k_{\mathrm{min}} = 0.03\,h\,\mathrm{Mpc}^{-1}$ reduces the impact of the large-scale systematic errors on the analysis (Gil-Marín et al. 2016a). In the case of the bispectrum, since the model is more accurate and very similar to the one that it has been already applied and tested on data in the BOSS analysis, we preferred to use a lower k_{min} in order to be able to use a wider range of triangle bins.

The smallest scales considered are $k_{\mathrm{max}} = 0.09\,h\,\mathrm{Mpc}^{-1}$ and $k_{\mathrm{max}} = 0.12\,h\,\mathrm{Mpc}^{-1}$ for power spectrum (monopole and quadrupole) and bispectrum monopole, respectively. The lower k_{max} used for the power spectrum is due to the fact that we did not include 1-loop corrections for it, hence we consider only scales belonging to the quasi-linear regime. We chose a higher k_{max} for the bispectrum since we implemented the effective model developed by Gil-Marín et al. (2014), which works up to non-linear scales.

The fiducial cosmology chosen for the analysis corresponds to a flat-ΛCDM model close to the one reported in Planck Collaboration XIII (2016). In particular, we set $\Omega_{\mathrm{m}}(z=0) = 0.31$, $\Omega_{\mathrm{b}}(z=0) = 0.049$, $A_{\mathrm{s}} = 2.21 \times 10^{-9}$, $n_{\mathrm{s}} = 0.9624$, and $h_0 = 0.6711$. In order to compute the covariance terms and the derivatives of the model necessary for the compression, we fix the fiducial value of the galaxy bias model parameters, the growth rate, and the amplitude of dark

matter fluctuations to the ones obtained by running a preliminary low-resolution MCMC ($b_1 = 2.5478$, $b_2 = 1.2127$, $f = 0.7202$, $\sigma_8 = 0.4722$). The FoG parameters for both power spectrum, $\sigma^{\rm P}_{\rm FoG}$, and bispectrum, $\sigma^{\rm B}_{\rm FoG}$, have been set to zero after checking that for the range of scales considered (quasi-linear regime) they were compatible with zero. In Section 7, we check that the choice of fiducial Parameters used to compute the derivatives of the mean of the data vector and the analytical covariance matrix does not significantly influence the results of the compression.

4 COMPRESSION METHODS

In Paper I, we presented two compression methods and applied them to the galaxy bispectrum and power spectrum: MC–KL and PCA + KL. Both methods rely on the MOPED presented in Heavens, Jimenez & Lahav (2000) which achieves maximal compression by extending to the multiple parameters case the algorithm introduced in cosmology by Tegmark, Taylor & Heavens (1997). Using this MOPED/KL compression, it is possible to shrink an arbitrarily large data vector $\boldsymbol{x}$ to a compressed one $\boldsymbol{y}$ having dimension equal to the number of model parameters considered preserving Fisher information. This is obtained by deriving a set of weights for the full data vector for each model parameter. Taking the scalar product between the weighting vectors and the original full data vector $\boldsymbol{x}$ gives the elements y_i of the compressed data vector. Here, we report only the main equations, please refer to Paper I for more details. The weighting vector for each parameter θ_i is given by

$$\boldsymbol{b}_i = \mathbf{Cov}_{\boldsymbol{x}}^{-1} \langle \boldsymbol{x} \rangle_{,i} \,, \tag{11}$$

where $\mathbf{Cov}^{-1}$ is the inverse of the original full data vector covariance matrix, and $\langle \boldsymbol{x} \rangle_{,i}$ is the derivative with respect to the model parameter θ_i of the mean of the modelled data vector $\boldsymbol{x}$, computed at a fiducial parameter vector $\boldsymbol{\theta}_{\rm fid.}$. In our case, the fiducial values are reported in Section 3.2. Therefore, the elements of the compressed data vector $\boldsymbol{y}$ are given by

$$y_i \;=\; \langle \boldsymbol{x} \rangle_{,i}^{\intercal} \, \mathbf{Cov}_{\boldsymbol{x}}^{-1} \, \boldsymbol{x} \;\equiv\; \boldsymbol{b}^{\intercal} \, \boldsymbol{x}. \tag{12}$$

In the MC–KL method, an MCMC sampling algorithm using $\boldsymbol{y}$ as data vector is ran after compression. An estimate of the compressed covariance matrix from the mock catalogues can be obtained as shown in the Appendix of Paper I:

$$\mathbf{Cov}_{\boldsymbol{y},ij} \;=\; \mathbf{Cov}\left[y_i, y_j\right] \;=\; \boldsymbol{b}_i^{\intercal} \cdot \mathbf{Cov}_{\boldsymbol{x}} \cdot \boldsymbol{b}_j \,, \tag{13}$$

where $\mathbf{Cov}_{\boldsymbol{x}}$ is the original covariance matrix.

4.1 PCA + KL

As described in Paper I, instead of orthogonalizing the weights as in Zablocki & Dodelson (2016), we perform a PCA transformation of our parameter space before applying the KL compression. This is done by diagonalizing the Fisher information matrix using the eigenvalue decomposition:

$$\mathsf{F}_{\boldsymbol{\theta}_{\rm phys.}} \;=\; \mathsf{P} \; \mathsf{F}_{\boldsymbol{\theta}_{\rm PCA}} \; \mathsf{P}^{\intercal} \quad , \text{ where } \quad \boldsymbol{\theta}_{\rm PCA} \;=\; \mathsf{P}^{\intercal} \, \boldsymbol{\theta}_{\rm phys.} \tag{14}$$

and $\boldsymbol{P}$ is the linear transformation matrix. After having diagonalized the Fisher matrix, we compress the data vector with respect to this new set of parameters $\boldsymbol{\theta}_{\rm PCA}$. The effect of a PCA decomposition is to rotate the parameter space to the axes corresponding to the degeneracies between the original set of parameters. Therefore, taking the outer product of the 1D posteriors of the parameters $\boldsymbol{\theta}_{\rm PCA}$ in order to get the multidimensional posterior distribution should return a good approximation to the one sampled by the MCMC code.

Since the $\boldsymbol{\theta}_{\rm PCA}$ are uncorrelated, one can randomly sample the 1D posteriors and then rotate the resulting parameter vector using $\boldsymbol{P}$ back into the physical space. Doing this avoids the use of the MCMC sampling altogether.

As shown in Paper I, this works only for those parameter sets that have a sufficiently low degree of degeneracy such that the approximation of Gaussianity for the multidimensional posterior can be assumed to be valid (no or very weak 'banana-shaped' contours). Since this is not always the case, as for our choice of parameters, an additional Gaussianization pre-step is required.

4.2 Gaussianization pre-step

In Paper I, the PCA + KL method assumed that it was possible to rotate through a linear transformation the physical parameter space into a new one where the new parameters are orthogonal/uncorrelated between each other. In order to be able to deal with distributions containing non-linear degeneracies (e.g. 'banana-shaped' contours), we add a pre-Gaussianization transformation of the parameter space using the procedure described in Schuhmann et al. (2016). In their work, they introduced an extension of the Box–Cox transformations, which are functions of two parameters (a, λ):

$$\tilde{\theta}^i = BC_{(a,\lambda)}(\theta^i) = \begin{cases} \lambda^{-1}[(\theta^i + a)^{\lambda} - 1] & (\lambda \neq 0) \\ \log(\theta^i + a) & (\lambda = 0) \end{cases}, \tag{15}$$

where $\tilde{\theta}^i$ is the transformed ith model parameter, while θ^i is the original ith model parameter. Their method was labelled Arcsinh–Box–Cox transformation (ABC). For each of the model parameters, a set of three ABC transformation parameters (a, λ, t) is computed by the algorithm that are then used in the following way:

$$\theta^i_{\rm Gauss.} = {\rm ABC}(\theta^i_{\rm phys.}) = \begin{cases} t^{-1} \sinh[t \, {\rm BC}_{(a,\lambda)}(\theta^i_{\rm phys.})] & (t > 0) \\ {\rm BC}_{(a,\lambda)}(\theta^i_{\rm phys.}) & (t = 0) \\ t^{-1} {\rm arcsinh}[t \, {\rm BC}_{(a,\lambda)}(\theta^i_{\rm phys.})] & (t < 0) \end{cases}, \tag{16}$$

where $\theta^i_{\rm Gauss.}$ is the Gaussianized ith model parameter, while $\theta^i_{\rm phys.}$ is the original ith physical model parameter. We then relabel this compression as G-PCA. In order to obtain the transformation parameters of the Gaussianizing transformations, it is necessary to run a preliminary MCMC sampling using the full data vector. What we want to prove is that once the transformation parameters have been obtained for the standard number of triangle bins corresponding to the Δk_6 binning case, these are valid also for a higher number of triangle bins included in the bispectrum.

4.3 Analytical covariance matrix: usage

In the following analysis, we are going to use two different options for the analytical covariance matrices. For the MC–KL method, we compress only the bispectrum monopole part of the data vector. To derive the weights in equation (11), we use the analytical covariance matrix of the bispectrum monopole given by the sum of the Gaussian term in equation (8) and the non-Gaussian one given in equation (9). For the G-PCA method, the full data vector needs to be compressed since the computation of the 1D posteriors

of the $\boldsymbol{\theta}_{PCA}$ parameters requires each data vector element to be sensitive to the variation of just one $\theta_{\rm PCA}$ parameter, as explained in Paper I. Therefore, for the power spectrum monopole/quadrupole we use equation (5) as our analytical covariance matrix; similarly for the bispectrum monopole we use equation (8) for the covariance matrix (the same as the one we used for the MC–KL case), and finally, we use equation (10) for our cross-covariance matrix.

5 RECOVER MCMC-DERIVED POSTERIOR DISTRIBUTION

For MCMC sampling, we use EMCEE[8] (Foreman-Mackey et al. 2013). All the likelihoods have been corrected as suggested by Sellentin & Heavens (2016) in order to take into account the bias induced by estimating the inverse of the real covariance matrix from a limited number of mocks. In order to check whether our analytical estimate of the covariance matrix is good enough to be used for deriving the weights as explained in Section 4, we compare to the full MCMC 1D posterior distributions in the left-hand panels of Figs 1 and 2 with results from the MCMC + MC–KL and G-PCA methods, respectively.

The violin plots include the standard binning case Δk_6 (116 triangle bins) and the Δk_5 case (195 triangle bins). For these two cases, we compare the MCMC (grey and purple) with the compression results (cyan and orange). From each point, we subtract the mean of the model parameters obtained using the MCMC. This makes it easier to check that the shift in the mean of the compression results with respect to the MCMC ones is small when compared to the size of the inner quartiles of the distribution. This concept is also quantified in the bottom half of Table 1, which shows that the shifts in the mean values are relative to the 1D 68 per cent credible intervals. In the top half of Table 1, we report the precise values of both the means and the 68 per cent credible intervals for all model parameters. Additionally, Fig. B1 in Appendix B shows the comparison between the 2D MCMC posterior distributions and the MC–KL and G-PCA ones for both Δk_6 and Δk_5 cases. We conclude that even if a small part of the constraining power is lost (see the Δk_6 columns in Table 2 for details), both compression methods return posterior distributions that well agree with the MCMC distribution for all model parameters under consideration.

6 INFORMATION CONTENT AND NUMBER OF TRIANGLE BINS

The right-hand panels of Figs 1 and 2 show how using a ∼23 times larger number of triangle bins tightens the posterior contours of the four model parameters considered and reduces the degeneracies between them. At the same time, the maxima of the 2D posterior distributions converge to the same values for each compression method as the number of triangle bins is increased.

Note that the shift in the posterior distribution between binning cases is not an artefact of the compression: It is also present when we fit using the standard MCMC method. This can be seen when comparing the location and shape of the 2D contour regions in Figs B1 and B2 in Appendix B for the Δk_6 and Δk_5 binning cases. Quantitatively, it can be observed by comparing means and standard deviations in Table 1. Thus, both compression algorithms reproduce posterior distributions very similar to the ones derived via MCMC sampling for the relevant binning cases Δk_6 and Δk_5. The observed shift between binning cases is due to the strong degeneracy between the model parameters. In particular, the shift happens along the degeneration direction of b_1, b_2, and f with σ_8. It may have a statistical origin. Further checks on this effect may be performed using the galaxy mocks, for example, by fitting several different realizations for both the Δk_6 and Δk_5 binning cases using the G-PCA method (which would be much faster than doing parameter estimation via MCMC or MC–KL). We reserve to do these tests in future work. Additionally, the practically identical (compared to the error bars amplitude) residuals plots for the different models in Fig. 4 show that the shifts in the best-fitting parameters as a function of the number of triangle bins used are an effect of the strong degeneracy present in the parameter space. Even if employing more triangle bins partially lifts this, the degree of how well the models for the different number of triangle bins fit the data does not change.

The main result of this paper is that the variance of the parameters is substantially reduced when the number of triangle bins used is increased up to ∼23 times the original number. In terms of percentages of the original 1D 68 per cent credible intervals obtained running an MCMC on the full data vector for the parameters (b_1, b_2, f, σ_8) in the BOSS Δk_6 case, the Δk_2 MC–KL and G-PCA analyses obtain tighter constraints by (−35 per cent, −45.3 per cent, −22.6 per cent, −22.6 per cent) and (−24.8 per cent, −52.8 per cent, −26.4 per cent, −21 per cent), respectively. These optimal constraints as obtained by the compression methods are also shown in summary in Fig. 5. The gain in parameter constraints is due to the fact that when we increase the number of triangle bins, by decreasing the k-bins size, the information is less 'washed out' than when using larger k-bins.

For future surveys, the compression can be then used to maximize the constraining power of the main analysis and also to find out the minimum number of triangle bins for a given k-range needed to fully capture the non-Gaussian information contained in 3pt statistics such as the bispectrum. The later will indicate how many mock catalogues/simulations are required in order to accurately estimate the covariance matrix. In our analysis, the saturation seems to be reached already for the Δk_4 binning case (404 triangle bins).

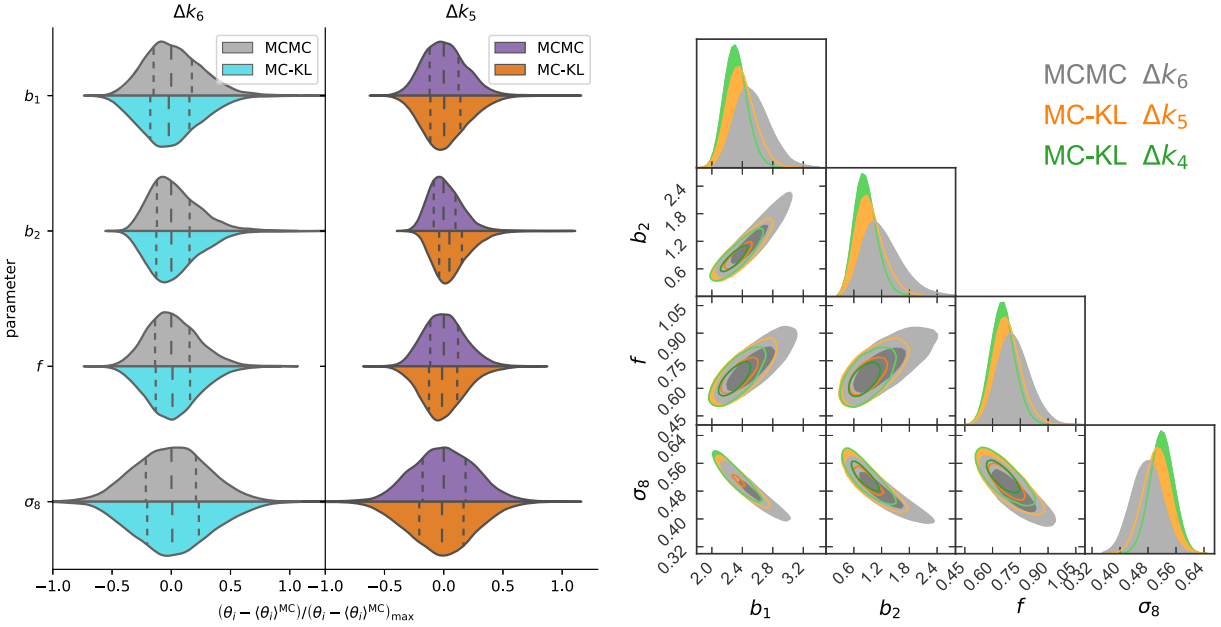
For what concerns Δk_2, the smallest k-bin size considered (2734 triangle bins), Tables 1 and 2 show that the Δk_2 posterior distribution is very similar to the Δk_4 case.

The trend in the information content in terms of the 1D 68 per cent credible intervals as a function of the triangle number used is shown in the left-hand panel of Fig. 3, and the improvement quantified in Table 2. From Fig. 3, it appears that the parameters constraint improvement as a function of the number of triangle bins reaches the saturation already for the Δk_4 case. For the chosen k-range, the additional triangle bins (and bispectra) included in the Δk_2 with respect to the Δk_4 one do not substantially add new features to the bispectrum data vector therefore the constraining power results weakly improved.

7 CONSISTENCY CHECK

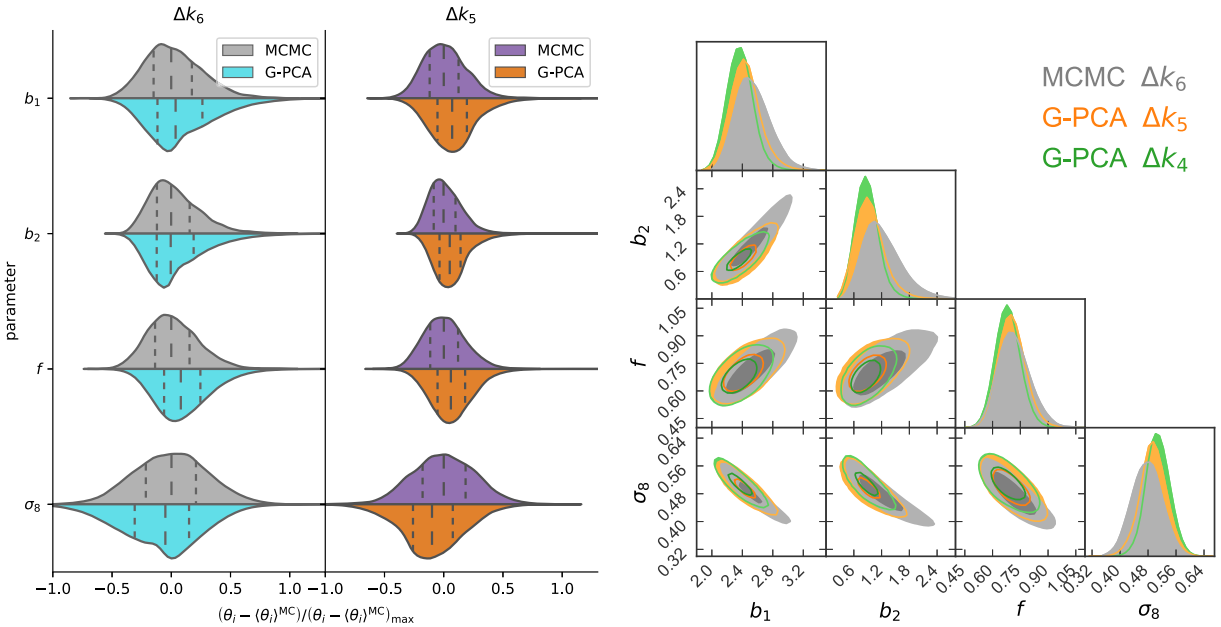
In order to test the validity of our analysis, we compute the reduced χ^2 and corresponding p-value for each set of parameters obtained using either the MCMC sampling or the compression methods. For all parameter vectors (compressed and uncompressed), this has been done using the data vector corresponding to the standard Δk_6

[8]We use 192 walkers, 1100 burn-in steps, and 1700 steps. For the low-resolution MCMC, we use half of the previous quantities.



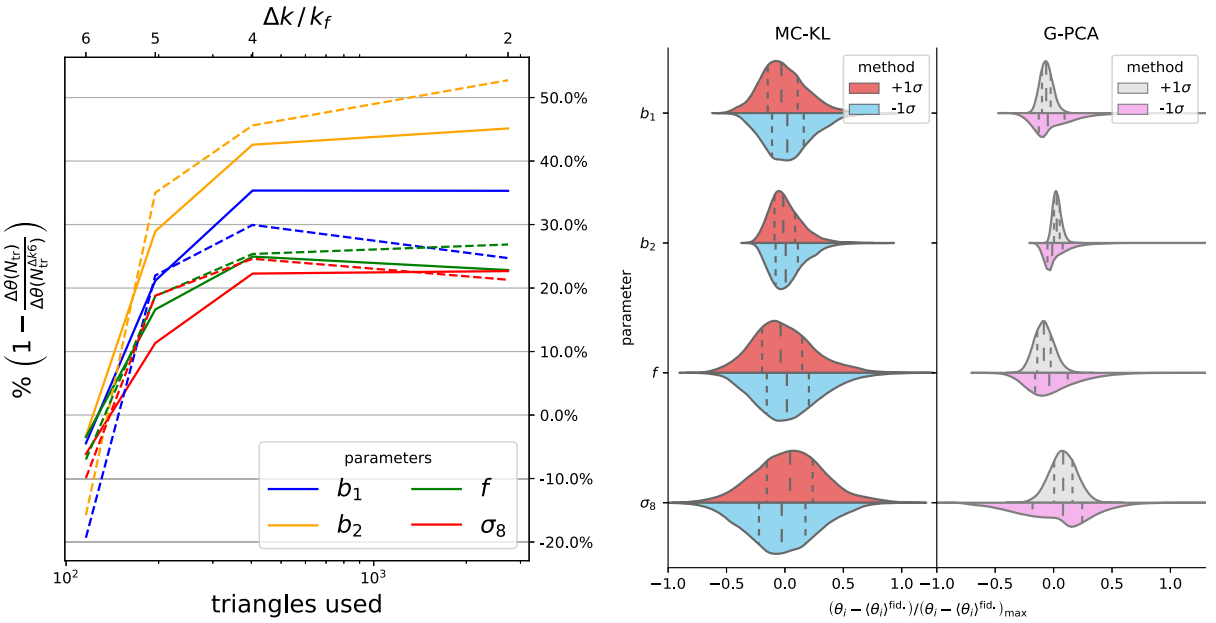
(a) consistency check (b) constraints improvement

Figure 1. Joint data vector $\left[P_g^{(0)}, P_g^{(2)}, B_g^{(0)}\right]$ posteriors: MC–KL four-parameter case. (a) The violin plots show for two test cases (Δk_6 and Δk_5 binning) the comparison between the 1D posterior densities obtained via MCMC and MC–KL for all parameters. The vertical lines represent the 25 per cent, 50 per cent, and 75 per cent quartiles. All distributions have been centred by subtracting the mean value obtained from the MCMC analysis, and they have been normalized by dividing by the maximum difference between the parameter value of each sample and the mean of the distribution. Even if the 1D distributions are not Gaussian, the agreement between MCMC and MC–KL results is qualitatively good. For a quantitative comparison, see Table 1 and additionally Figs B1 and B2 in Appendix B. (b) The 2D 68 per cent and 95 per cent credible regions are shown in order to highlight the improved constraints and reduced parameter degeneracies obtained by employing a higher number of triangle bins in the data vector because of the compression with respect to the standard MCMC for the full data vector. In particular, the grey contours correspond to the standard binning Δk_6 used to run the MCMC for the full data vector. The orange and green contours correspond to the distributions for the compressed data vector for the binnings Δk_5 and Δk_4 (which corresponds to $N_{\rm triangles} = 195, 404$, the number of triangle bins increases as the k-bin size approaches the fundamental frequency). See also Table 2.



(a) consistency check (b) constraints improvements

Figure 2. Joint data vector $\left[P_g^{(0)}, P_g^{(2)}, B_g^{(0)}\right]$ posteriors: G-PCA four-parameters case. Same as Fig. 1 but for the G-PCA method.



(a) information content vs. triangle number

(b) shifted fiducial parameter set

Figure 3. (a) The 1D 68 per cent credible intervals as a function of the number of triangle bins used in the bispectrum monopole data vector. Continuous lines represent the MC–KL results, while the dashed ones are given by the G-PCA compression method. (b) The compression results for the MC–KL and G-PCA cases when the fiducial parameter set used to compute the analytical covariance matrix and the derivatives of the mean are shifted by $\pm 1\sigma$ credible intervals. The violin plots show, for the test case of the Δk_6-binning, the comparison between the 1D posterior distributions for all parameters, using shifts by $+1\sigma$ (red/grey) and -1σ (blue/pink) for the MC–KL / G-PCA methods. The vertical lines represent the 25 per cent, 50 per cent, and 75 per cent quartiles. All distributions are mean-subtracted using the fiducial parameter set for the compression, and they have been normalized by the maximum difference between the parameter value of each sample and the mean of the distribution. Even if the 1D distributions are not Gaussian, the effect of compressing with respect to a shifted cosmology is qualitatively negligible for the MC–KL method, while it affects the G-PCA performance more. Nevertheless, the modifications to the fiducial parameter sets are substantial (∼10–40 per cent varations) given the broad posteriors due to the strong degeneracy in the parameter set.

Table 1. Four-parameter case, check consistency. *Upper half:* Mean values of the posterior distributions and 68 per cent credible intervals for the MCMC and the MC–KL / G-PCA compression methods. We report the values for a range of k-binnings. From the largest bin Δk_6, the size used in the BOSS analysis, corresponding to the lowest number of triangle bins (116), to the thinnest binning Δk_2 corresponding to the highest number of triangle bins (2734). The observed shift in the mean values as a function of the number of triangle bins considered is due to the strong degeneracy present between the model parameters. As can be seen in Fig. 4, the shift does not have any effect on the goodness of fit. *Lower half:* In the compression columns, we report the relative difference between the posterior modes obtained via MCMC and the ones obtained via compression (MC–KL or G-PCA). In the MCMC columns, the relative size of the 68 per cent credible intervals obtained via MCMC sampling is shown. Comparing the MCMC columns to the compression ones, the difference between the mean parameter values obtained via MCMC and the ones obtained via compression (MC–KL or G-PCA) is evidently within the 68 per cent credible intervals given by the MCMC on the full data vector.

	Δk_6			Δk_5			Δk_4		Δk_2	
	MCMC	MC–KL	G-PCA	MCMC	MC–KL	G-PCA	MC–KL	G-PCA	MC–KL	G-PCA
b_1	2.41 ± 0.22	2.41 ± 0.23	2.49 ± 0.27	2.34 ± 0.17	2.38 ± 0.18	2.42 ± 0.17	2.27 ± 0.14	2.38 ± 0.16	2.28 ± 0.14	2.31 ± 0.17
b_2	1.00 ± 0.40	1.04 ± 0.42	1.08 ± 0.47	0.82 ± 0.26	0.83 ± 0.29	0.85 ± 0.26	0.79 ± 0.23	0.81 ± 0.22	0.68 ± 0.22	0.77 ± 0.19
f	0.69 ± 0.08	0.72 ± 0.09	0.72 ± 0.09	0.67 ± 0.07	0.67 ± 0.07	0.70 ± 0.07	0.65 ± 0.06	0.68 ± 0.06	0.68 ± 0.06	0.67 ± 0.06
σ_8	0.50 ± 0.04	0.48 ± 0.05	0.48 ± 0.05	0.51 ± 0.04	0.50 ± 0.04	0.49 ± 0.03	0.53 ± 0.03	0.51 ± 0.03	0.52 ± 0.03	0.51 ± 0.03
	$\frac{\Delta\theta^{\rm mc}_{\Delta k_6}}{\theta^{\rm mc}_{\Delta k_6}}$ [%]	$\frac{\theta^{\rm comp.}-\theta^{\rm mc}_{\Delta k_6}}{\theta^{\rm mc}_{\Delta k_6}}$ [%]		$\frac{\Delta\theta^{\rm mc}_{\Delta k_5}}{\theta^{\rm mc}_{\Delta k_5}}$ [%]	$\frac{\theta^{\rm comp.}-\theta^{\rm mc}_{\Delta k_5}}{\theta^{\rm mc}_{\Delta k_5}}$ [%]		$\frac{\theta^{\rm comp.}-\theta^{\rm mc}_{\Delta k_5}}{\theta^{\rm mc}_{\Delta k_5}}$ [%]		$\frac{\theta^{\rm comp.}-\theta^{\rm mc}_{\Delta k_5}}{\theta^{\rm mc}_{\Delta k_5}}$ [%]	
b_1	9.2	−0.3	3.3	7.3	1.9	3.5	−2.7	1.9	-2.7	−1.1
b_2	40.3	3.5	7.5	32.2	1.9	4.4	−3.6	−1.2	-16.5	−5.7
f	12.1	4.4	4.4	10.1	−1.3	3.8	−3.3	0.2	0.5	−1.1
σ_8	8.5	−5.1	−5.5	7.3	−1.1	−3.6	4	−0.3	2.2	−1.2

binning. The results can be seen in Fig. 4. This test proves that the shift observed in the parameters as the number of triangle bins is increased is simply due to the strong degeneracy present between b_1, b_2, f, and σ_8. Indeed both the reduced χ^2 and p-values show that all these models fit the data very well. In Fig. 4, we did not show the lines and statistics for the Δk_5 cases just for the sake of clarity and because the results are equivalent to those of the other binnings. From the same figure, it can also be noticed that the tightest error bars are those from the power spectrum case.

Table 2. Four-parameter case, constraints improvement. Below are shown the relative variations in percentage of the size of the 68 per cent credible intervals as a function of the k-binning considered (number of triangle bins used for the bispectrum monopole). In orange and green are highlighted, respectively, the improvements achieved via compression for the Δk_5 and at the saturation level (404 triangle bins – Δk_4) of the bispectrum monopole constraining power case for the considered set of parameters (e.g. left-hand panel of Fig. 3). Finally, in blue and red are highlighted the improvements obtained via compression for the highest number of triangle bins considered (2734 triangle bins – Δk_2 binning) for MC–KL and G-PCA, respectively.

	Δk_6			Δk_5			Δk_4		Δk_2	
	$\Delta\theta^{\rm mc}_{\Delta k_6}$	$\frac{\Delta\theta^{\rm comp.}-\Delta\theta^{\rm mc}_{\Delta k_6}}{\Delta\theta^{\rm mc}_{\Delta k_6}}$ [%]		$\Delta\theta^{\rm mc}_{\Delta k_5}$	$\frac{\Delta\theta^{\rm comp.}-\Delta\theta^{\rm mc}_{\Delta k_6}}{\Delta\theta^{\rm mc}_{\Delta k_6}}$ [%]		$\frac{\Delta\theta^{\rm comp.}-\Delta\theta^{\rm mc}_{\Delta k_6}}{\Delta\theta^{\rm mc}_{\Delta k_6}}$ [%]		$\frac{\Delta\theta^{\rm comp.}-\Delta\theta^{\rm mc}_{\Delta k_6}}{\Delta\theta^{\rm mc}_{\Delta k_6}}$ [%]	
	MCMC	MC–KL	G-PCA	MCMC	MC–KL	G-PCA	MC–KL	G-PCA	MC–KL	G-PCA
Δb_1	0.22	4.4	18.8	0.17	– 21.3	– 22.0	– 35.3	– 30.0	– 35.3	– 24.8
Δb_2	0.40	2.9	16.2	0.26	– 28.9	– 35.0	– 42.6	– 46.0	– 45.3	– 52.8
Δf	0.08	3.7	7.0	0.07	– 16.5	– 18.5	– 24.7	– 25.1	– 22.6	– 26.4
$\Delta\sigma_8$	0.04	6.5	10.0	0.04	– 11.3	– 18.7	– 22.3	– 24.5	– 22.6	– 21.0

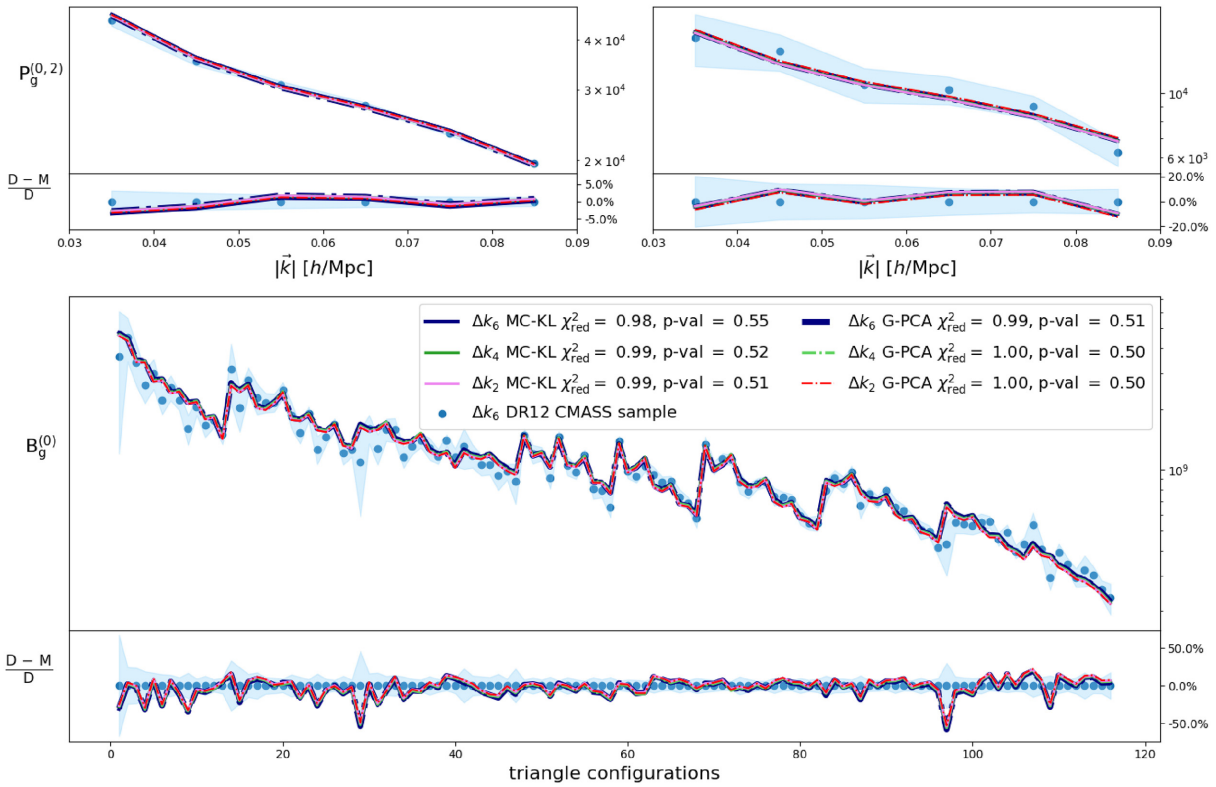


Figure 4. Reduced χ^2 and p-values for the best-fitting models obtained using the MCMC, MC–KL, and G-PCA compression methods. The k-binnings shown are, respectively, the standard Δk_6 (navy), an intermediate size Δk_4 (green), and the smallest one Δk_2 (pink for MC–KL and red for G-PCA) corresponding to the highest number of triangle used in the bispectrum monopole. The two upper panels are for the power spectrum monopole (left) and quadrupole (right), while the bottom panel refers to the bispectrum monopole. The lower part of each panel shows the relative difference between the data measurements and the different models. Even if for example b_1 and σ_8 values are shifted between the cases of Δk_6 and Δk_2, the strong degeneracy has the result of making the two models practically identical.

To demonstrate the flexibility of the compression methods, we check their performance when the fiducial parameter set is shifted by $\pm 1\sigma$ credible intervals in the Δk_6 case. The effect of this is shown in the right-hand panel of Fig. 3. For this plot, we centre each 1D distribution by subtracting the mean obtained by running the compression pipelines using the fiducial parameters values. In this way, it is possible to observe by how much the posterior distributions derived via MC–KL or G-PCA shift as a function of the chosen fiducial parameter set. In Appendix B, the precise numbers are reported in Table B1.

MC–KL appears to be more stable than the G-PCA when the fiducial parameter set is shifted. The explanation of this could be the fact that G-PCA involves several transformations of the parameter space, including a diagonalization of the Fisher information matrix that is computed from the analytical model of the covariance matrix.

Nevertheless, it should be noted that we are testing the performances of the compression in a regime of strong degeneracy of the parameter space and therefore shifting the fiducial parameter set by $\pm 1\sigma$ credible intervals actually means increasing/reducing the individual values by ∼10–40 per cent (second panel Table 1).

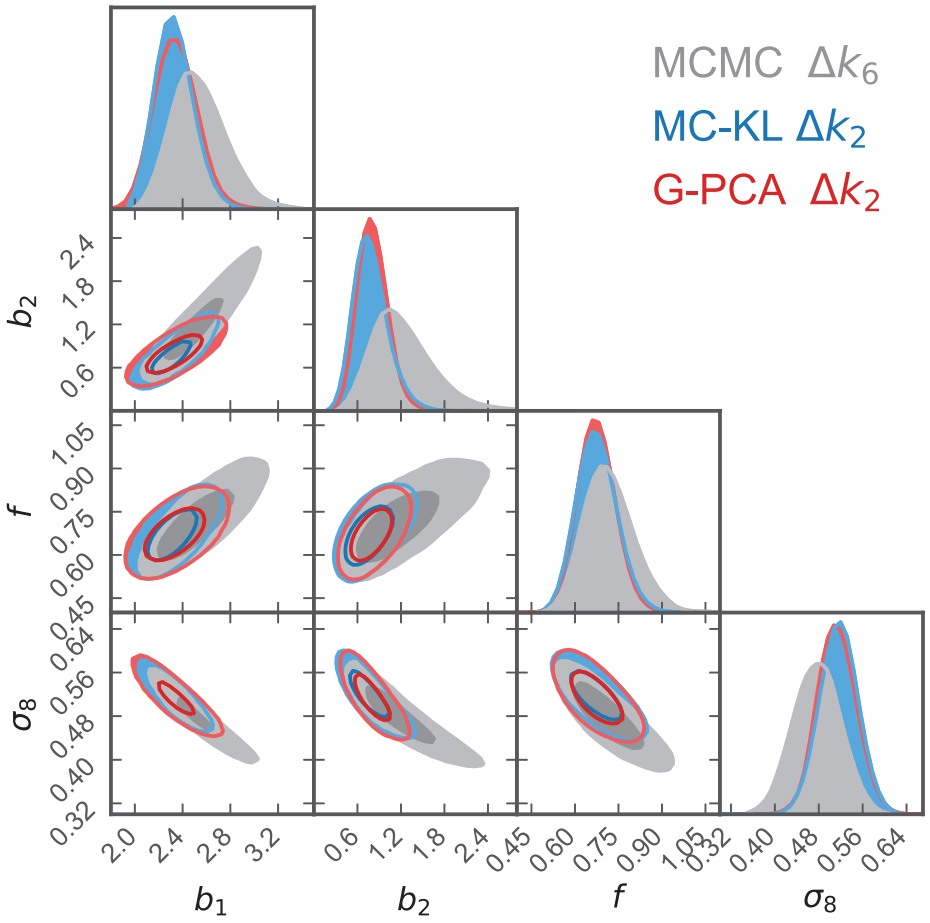


Figure 5. MCMC versus MC–KL versus G-PCA. 2D 68 per cent and 95 per cent credible contours are shown, respectively, for the Δk_6 MCMC (grey), Δk_2 MC–KL (blue), and Δk_2 G-PCA (red) cases. It is possible to observe the substantial improvement in parameter constraints through applying either compression method to a data vector containing approximately ∼23 times more triangle bins than the one used for the MCMC sampling case. The agreement between the MC–KL and G-PCA posterior distributions is remarkable. Using more triangle bins helps with lifting the strong degeneracy between the model parameters, as can be seen from the shrinkage of the 2D contours along the degeneracy directions.

Therefore, running a preliminary low-resolution MCMC sampling on the full data vector (which can be shorter than the one that will be later compressed, as we have done in our analysis) is an efficient solution to determine a reasonable fiducial model for deriving the compression.

7.1 Comparison with BOSS DR12 bias constraints

BOSS galaxy sample results from the bispectrum are reported by Gil-Marín et al. (2017; in table 3 at p. 18) from the same CMASS sample data set, at the same redshift, for the following parameter combinations: $b_1\sigma_8 = 1.2479 \pm 0.0072$, $b_2\sigma_8 = 0.641 \pm 0.066$, and $f\sigma_8 = 0.432 \pm 0.018$.[9] If we recast our results obtained using the MCMC for the Δk_6 case in terms of the same parameter combinations, these are $b_1\sigma_8 = 1.203 \pm 0.008$, $b_2\sigma_8 = 0.557 \pm 0.140$, and $f\sigma_8 = 0.339 \pm 0.019$.

In the BOSS analysis, a larger range of scales has been considered. In particular, BOSS analysis goes up to $k \sim 0.2h\,\mathrm{Mpc}^{-1}$ for both power spectrum monopole/quadrupole and bispectrum monopole while we stop at $k \sim 0.09h\,\mathrm{Mpc}^{-1}$ and $k \sim 0.12h\,\mathrm{Mpc}^{-1}$, respectively. This could explain the larger value we obtained for $b_2\sigma_8$. A more complex model for the power spectrum was used in the BOSS analysis, including loop corrections beyond the tree-level approximation. Moreover, the BOSS analysis also modelled the effect of the survey window function for both power spectrum and bispectrum.

As we saw from Fig. 4, the power spectrum monopole is the most constraining part of the full data vector, having error bars of less than 5 per cent. Therefore, it is possible that our simple tree-level approximation for the power spectrum, besides limiting the k-range analysed, could be the cause of the discrepancy between the BOSS results with respect to the relative lower values obtained for the combined parameters $b_1\sigma_8$, $b_2\sigma_8$, and $f\sigma_8$ in this work.

Moreover, in the BOSS analysis the FoG parameters $\sigma^{\mathrm{B}}_{\mathrm{FoG}}$ and $\sigma^{\mathrm{P}}_{\mathrm{FoG}}$ were left free to vary in order to better model the non-linear regime and were detected with high significance ($\sigma^{\mathrm{B}}_{\mathrm{FoG}} = 7.54 \pm 0.70$ and $\sigma^{\mathrm{P}}_{\mathrm{FoG}} = 3.50 \pm 0.14$). The BOSS model also included a noise amplitude parameter A_{noise}, which modelled divergence from Poissonian shot noise. In our model, we had included A_{noise} initially, however, we set it to zero after having checked that, if let free to vary, its posterior distribution was compatible with zero for the k-range considered. These differences in the modelling and scales considered could explain the discrepancy in the best-fitting parameters.

[9]We compare our results with the BOSS analysis standard deviation values obtained considering only the statistical contributions and not the systematics ones.

In Appendix B and in particular Fig. B3, we show how to test the limitation to the data vector constraining power, implied by our choice of k-range, by running an MCMC sampling for two-parameter sets: (b_1, b_2, f, σ_8) and (b_1, b_2, f). In the second case, by fixing σ_8 to its fiducial value, we recover maximum likelihood values for (b_1, b_2, f) very different from the ones corresponding to the four-parameter case reported in Table 1 (1.98 ± 0.01, $0.39 \pm 0.06, 0.53 \pm 0.03$). However, this discrepancy is not reflected in the reduced χ^2 values for the two different sets of best-fitting parameters: for the (b_1, b_2, f, σ_8) case $\chi^2_{\rm red} = 0.98$, while for the set (b_1, b_2, f) $\chi^2_{\rm red} = 1.05$. The fact that both best-fitting parameter sets very well fit the data implies that the constraining power of the data vector on the data in the four-parameter case is not sufficient to lift the degeneracies present in the parameter space. Therefore, in order to lift the degeneracies and to avoid a shift in the inferred parameters when σ_8 is also fitted, a more accurate model for the power spectrum monopole and quadrupole including loop correction is needed.

7.2 Difference in time and computer resources needed

There is no significant difference between MCMC and MC–KL in terms of time taken for the pipeline to run or computing resources needed. For the parameter set (b_1, b_2, f, σ_8), the running time varied between 20 min for 116 triangle bins to ∼10 h for 2734 triangle bins on 14 2.2 GHz Intel i7 cores. G-PCA proved to be faster when many triangle bins are used. Considering ∼30 min for the preliminary MCMC with 116 triangle bins and ∼ 2 h for the Gaussianization part, it took between ∼5 min (116 triangle bins) and ∼30 min (2734 triangle bins) using only one 2.2 GHz Intel i7 core for the compression plus posterior evaluation to run. Therefore, by running once the preliminary MCMC and Gaussianization algorithm, we were able to run the PCA part for all the binning cases considered in less than total ∼ 3 h wall clock time.

We used CAMB (Lewis, Challinor & Lasenby 2000) to compute the linear matter power spectrum. The time difference between MCMC/MC–KL and G-PCA would have been much more significant in the case of a parameter set for which the linear matter power spectrum needs to be recomputed for every model realization.

8 CONCLUSIONS

In this paper, we have shown the results of applying both compression methods for the galaxy redshift-space bispectrum, presented in Paper I, to the measurements from the SDSS-III BOSS DR12 CMASS sample (Gil-Marín et al. 2017). We considered as original data vector the combination of the power spectrum monopole and quadrupole with the bispectrum monopole, which are obtained by averaging over the angles describing the orientation with respect to the line of sight. The first method called MC–KL consists of running an MCMC sampling on the compressed data vector obtained by taking the scalar product between the original data vector and a set of weights derived as first shown by Tegmark et al. (1997). The second method, which we denoted as G-PCA, is the modification of the PCA + KL method presented in Paper I obtained by adding a Gaussianization transformation of the parameter set (Schuhmann et al. 2016) before rotating it using a PCA transformation followed by the KL compression. By transforming the physical parameter space into an orthogonal one, it is possible to just randomly sample 1D posterior distributions, avoid altogether the need of running an MCMC routine.

In order to derive the posterior distributions for the set of parameters considered, the galaxy bias parameters b_1 and b_2, the growth rate f, and the normalization of the dark matter perturbations amplitude σ_8, we numerically estimated the covariance matrix using 1400 and 700 galaxy mocks catalogues for the full data vector and compressed data vector cases, respectively.

The following points represent the main conclusions of our analysis:

(i) In order to obtain the weights for the compression methods, we derived an analytic approximation of the leading terms of the covariance matrix relative to the considered data vector. The final expressions of these computations are reported in Section 2, while the full derivations are shown in Appendix A.

(ii) In Section 5, we have shown that both compression methods recover the posterior distributions obtained via MCMC using the full data vector with little loss of information (∼4 per cent and ∼13 per cent larger 68 per cent credible intervals than the MCMC ones on average for MC–KL and G-PCA, respectively). More importantly, even if slightly broader, the posterior distributions recovered through compression have the same shape and modes as the MCMC counterparts.

(iii) Adding a pre-Gaussianization step removes the PCA + KL limitation linked to a strongly degenerate parameter space described in Paper I. It is, however, necessary to run a preliminary MCMC in order to derive the Gaussianization transformation parameters. Nevertheless, once these parameters have been derived for a number of triangle bins case for which it is possible to run an MCMC on the full data vector, they can then be used to compress a data vector with an arbitrary number of triangle bins. The decrease in the compression performances shown in Fig. 3 due to a far from optimal choice of fiducial model parameters is also solved by rerunning the compression using as fiducial model the parameters inferred in the first run.

(iv) In Section 6, we show the main result of this work, namely the substantial improvement in parameter constraints obtained by compressing a much larger number of triangle bins with respect to standard MCMC data vector. For the uncompressed data vector, the number of triangle bins is limited by the number of mock catalogues available to estimate the covariance matrix. For both compression methods and for any number of triangle configuration considered, the dimension of the compressed data vector is always equal to the number of model parameters constrained.
For the highest number of triangle bins considered, this leads to an improvement in terms of the 68 per cent 1D credible intervals by (−35 per cent, −45 per cent, −23 per cent, −23 per cent) and (−25 per cent, −53 per cent, −26 per cent, −21 per cent) for the MC–KL and G-PCA methods, respectively.

(v) By way of summary, in Fig. 5 we show the results for both MC–KL and G-PCA methods using 2734 triangle bins and for the MCMC on the uncompressed data vector containing 116 triangle bins. The two compression methods agree well and produce substantially tighter and less-degenerate constraints. Furthermore, the G-PCA approach allowed for a computational speed up, requiring only approximately a third of the time taken by the MCMC and MC–KL methods, including also the low-resolution MCMC necessary for the Gaussianization transformation. Considering only the PCA part, the speed up factor rises to ∼20−100 times depending on the parameter set considered.

(vi) Finally, we would like to point out that the compressing methods used in this work represent a straightforward approach to include higher order statistics such as the trispectrum or the tetraspectrum in the analysis of current and future data sets. This is due to the fact that the number of elements of the data vector, after the maximal compression, corresponds exactly to the number of

model parameters. Both MC–KL and G-PCA have the potential to fully exploit the constraining power of higher order statistics applied to data sets from future surveys such as DESI, EUCLID, and PFS.

ACKNOWLEDGEMENTS

DG is grateful to L. Whiteway for the useful discussions and to E. Edmondson for the technical help in using the UCL computer cluster. DG is supported by the Perren and the IMPACT studentships. HGM is supported by Labex ILP (reference ANR-10-LABX-63) part of the Idex SUPER, and received financial state aid managed by the Agence Nationale de la Recherche, as part of the programme Investissements d'avenir under the reference ANR-11-IDEX-0004-02.e. MM acknowledges funding from STFC Consolidated Grants RG84196 and RG70655 LEAG/506 and has received funding from the European Union's Horizon 2020 research and innovation programme under Marie Skłodowska–Curie grant agreement No 6655919. OL acknowledges support from a European Research Council Advanced Grant FP7/291329. C (Kernighan 1988) and PYTHON 2.7 (Rossum 1995) have been used together with many packages like IPYTHON (Perez & Granger 2007), Numpy (van der Walt, Colbert & Varoquaux 2011), Scipy (Jones et al. 2001), and Matplotlib (Hunter 2007). The corner plots have been realized using PYGTC developed by Bocquet & Carter (2016), while the violin plots have been created using SEABORN by Waskom et al. (2016).

REFERENCES

Ade P. A. R. et al., 2014, A&A, 571, A16
Alam S. et al., 2017, MNRAS, 470, 2617
Alsing J., Wandelt B., 2018, MNRAS, 476, L60
Alsing J., Wandelt B., Feeney S., 2018, MNRAS, 477, 2874
Bernardeau F., Brax P., 2011, J. Cosmol. Astropart. Phys., 1106, 019
Bocquet S., Carter F. W., 2016, J. Open Source Softw., 1, 46
Borisov A., Jain B., 2009, Phys. Rev. D, 79, 103506
Chan K. C., Blot L., 2017, Phys. Rev. D, 96, 023528
Dark Energy Survey Collaboration, 2016, MNRAS, 460, 1270
Dawson K. S. et al., 2013, AJ, 145, 10
de Jong J. T. A., Verdoes Kleijn G. A., Kuijken K. H., Valentijn E. A., 2013, Exp. Astron., 35, 25
Eisenstein D. J. et al., 2011, AJ, 142, 72
Feldman H. A., Kaiser N., Peacock J. A., 1994, ApJ, 426, 23
Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, PASP, 125, 306
Frieman J. A., Gaztanaga E., 1994, ApJ, 425, 392
Fry J. N., 1984, ApJ, 279, 499
Fry J. N., Gaztanaga E., 1993, ApJ, 413, 447
Fry J. N., Scherrer R. J., 1994, ApJ, 429, 36
Fu L. et al., 2014, MNRAS, 441, 2725
Gangui A., Lucchin F., Matarrese S., Mollerach S., 1994, ApJ, 430, 447
Gil-Marín H. et al., 2016a, MNRAS, 460, 4188
Gil-Marín H. et al., 2016b, MNRAS, 460, 4210
Gil-Marín H., Schmidt F., Hu W., Jimenez R., Verde L., 2011, J. Cosmol. Astropart. Phys., 11, 019
Gil-Marín H., Wagner C., Fragkoudi F., Jimenez R., Verde L., 2012, J. Cosmol. Astropart. Phys., 2, 047
Gil-Marín H., Wagner C., Noreña J., Verde L., Percival W., 2014, J. Cosmol. Astropart. Phys., 12, 029
Gil-Marín H., Noreña J., Verde L., Percival W. J., Wagner C., Manera M., Schneider D. P., 2015, MNRAS, 451, 539
Gil-Marín H., Percival W. J., Verde L., Brownstein J. R., Chuang C.-H., Kitaura F.-S., Rodríguez-Torres S. A., Olmstead M. D., 2017, MNRAS, 465, 1757
Groth E. J., Peebles P. J. E., 1977, ApJ, 217, 385
Gualdi D., Manera M., Joachimi B., Lahav O., 2018, MNRAS, 476, 4045
Hartlap J., Simon P., Schneider P., 2007, A&A, 464, 399
Heavens A. F., Matarrese S., Verde L., 1998, MNRAS, 301, 797
Heavens A. F., Jimenez R., Lahav O., 2000, MNRAS, 317, 965
Heavens A. F., Sellentin E., de Mijolla D., Vianello A., 2017, MNRAS, 472, 4244
Hoffmann K., Mao Y., Xu J., Mo H., Wandelt B. D., 2018, preprint (arXiv:1802.02578)
Howlett C., Percival W. J., 2017, MNRAS, 472, 4935
Hunter J. D., 2007, Comput. Sci. Eng., 9, 90
Jarvis M., Bernstein G., Jain B., 2004, MNRAS, 352, 338
Joachimi B., Shi X., Schneider P., 2009, A&A, 508, 1193
Jones E., Oliphant T., Peterson P. et al., 2001, SciPy: Open source scientific tools for Python. Available at: http://www.scipy.org/
Kayo I., Takada M., 2013, preprint (arXiv:1306.4684)
Kayo I., Takada M., Jain B., 2013, MNRAS, 429, 344
Kernighan B. W., 1988, The C Programming Language, 2nd edn. Prentice Hall, Upper Saddle River, NJ
Kilbinger M., Schneider P., 2005, A&A, 442, 69
Kitaura F.-S. et al., 2016, MNRAS, 456, 4156
Kitaura F.-S., Heß S., 2013, MNRAS, 435, L78
Laureijs R. et al., 2011, preprint (arXiv:1110.3193)
Levi M. et al., 2013, preprint (arXiv:1308.0847)
Lewis A., Challinor A., Lasenby A., 2000, ApJ, 538, 473
Liguori M., Sefusatti E., Fergusson J. R., Shellard E. P. S., 2010, Adv. Astron., 2010, 980523
LSST Science Collaboration, 2009, preprint (arXiv:0912.0201)
Manera M. et al., 2013, MNRAS, 428, 1036
Matarrese S., Verde L., Heavens A. F., 1997, MNRAS, 290, 651
Pearson D. W., Samushia L., 2017, MNRAS, 478, 4500
Percival W. J. et al., 2014, MNRAS, 439, 2531
Perez F., Granger B. E., 2007, Comput. Sci. Eng., 9, 21
Planck Collaboration XIII, 2016, A&A, 594, A13
Pyne S., Joachimi B., Peiris H. V., 2017, J. Cosmol. Astropart. Phys., 12, 043
Rossum G., 1995, Technical report, Python Reference Manual. Amsterdam, The Netherlands, The Netherlands
Schmidt F., Vallinotto A., Sefusatti E., Dodelson S., 2008, Phys. Rev. D, 78, 043513
Schneider P., Kilbinger M., Lombardi M., 2005, A&A, 431, 9
Schuhmann R. L., Joachimi B., Peiris H. V., 2016, MNRAS, 459, 1916
Scoccimarro R., 2000, ApJ, 544, 597
Scoccimarro R., Sheth R. K., 2002, MNRAS, 329, 629
Scoccimarro R., Colombi S., Fry J. N., Frieman J. A., Hivon E., Melott A., 1998a, Astrophys. J., 496, 586
Scoccimarro R., Colombi S., Fry J. N., Frieman J. A., Hivon E., Melott A., 1998b, ApJ, 496, 586
Sefusatti E., Crocce M., Pueblas S., Scoccimarro R., 2006, Phys. Rev. D, 74, 023522
Sellentin E., Heavens A. F., 2016, MNRAS, 456, L132
Simon P. et al., 2013, MNRAS, 430, 2476
Simon P. et al., 2015, MNRAS, 449, 1505
Slepian Z. et al., 2017a, MNRAS, 468, 1070
Slepian Z. et al., 2017b, MNRAS, 469, 1738
Takada M., Jain B., 2004, MNRAS, 348, 897
Takada M. et al., 2014, PASJ, 66, R1
Taruya A., Nishimichi T., Saito S., 2010, Phys. Rev. D, 82, 063522
Tegmark M., Taylor A. N., Heavens A. F., 1997, ApJ, 480, 22
Tellarini M., Ross A. J., Tasinato G., Wands D., 2016, J. Cosmol. Astropart. Phys., 6, 014
van der Walt S., Colbert S. C., Varoquaux G., 2011, Comput. Sci. Eng., 13, 22
Verde L., Heavens A. F., Matarrese S., Moscardini L., 1998, MNRAS, 300, 747
Verde L., Wang L., Heavens A. F., Kamionkowski M., 2000, MNRAS, 313, 141
Waskom M. et al., 2016, seaborn: v0.7.0 (January 2016). https://doi.org/10.5281/zenodo.45133
Zablocki A., Dodelson S., 2016, Phys. Rev. D, 93, 083525

APPENDIX A: ESTIMATORS AND COVARIANCE TERMS

A1 Power spectrum monopole/quadrupole and bispectrum monopole estimators

The computations for the power spectrum and bispectrum multipoles reported next are original of this work. Expressions for the matter power spectrum and bispectrum were derived also by Scoccimarro et al. (1998b) and Sefusatti et al. (2006), however, in this work we proceed similar to what is done by Kayo et al. (2013).

The analytical model for the redshift-space galaxy power spectrum monopole and quadrupole is given by equation (4).

It is therefore natural to define the estimator as

$$\hat{\mathrm{P}}_{\mathrm{g}}^{(\ell)}(k) = \left(\frac{2\ell+1}{2}\right)\frac{1}{(2\pi)^3 N_{\mathrm{p}}(k)}\int_{V_p}\int_{V_q} d^3\boldsymbol{p}d^3\boldsymbol{q}\, L_\ell(\mu)\,\delta_{\mathrm{D}}(\boldsymbol{q}+\boldsymbol{p})\,\delta_g^{\mathrm{s}}(\boldsymbol{q})\,\delta_g^{\mathrm{s}}(\boldsymbol{p}), \tag{A1}$$

where $V_{\boldsymbol{p},\boldsymbol{q}}$ are the spherical shell volumes characterized by $k-\Delta k/2 \leq q, p \leq k+\Delta k/2$. μ is the cosine of the angle with respect to the line of sight of the $\boldsymbol{q}$ wave vector, and $L_\ell(\mu)$ is the Legendre polynomial of order ℓ. δ_D is the 3D Dirac delta. N_{p} is the number of grid point pairs in the integration volume in Fourier space and can be computed as

$$N_{\mathrm{p}}(k) = \frac{V_k}{k_f^3} = k_f^{-3}\int_{V_p}\int_{V_q} d^3\boldsymbol{p}d^3\boldsymbol{q}\,\delta_{\mathrm{D}}(\boldsymbol{q}+\boldsymbol{p}) \simeq \frac{4\pi k^2\Delta k}{k_f^3}, \tag{A2}$$

where $V_k \simeq 4\pi k^2\Delta k$ is the spherical integration shell defined by $k-\Delta k/2 \leq q, p \leq k+\Delta k/2$ as defined in Scoccimarro et al. (1998b). k_f is the fundamental frequency defined in terms of the survey volume V_{e} as $k_f^3 = \dfrac{(2\pi)^3}{V_{\mathrm{e}}}$. We check that the estimator defined in equation (A1) is unbiased:

$$\begin{aligned}
\langle\hat{\mathrm{P}}_{\mathrm{g}}^{(\ell)}(k)\rangle &= \left(\frac{2\ell+1}{2}\right)\frac{1}{(2\pi)^3 N_{\mathrm{p}}(k)}\int_{V_p}\int_{V_q} d^3\boldsymbol{p}d^3\boldsymbol{q}\, L_\ell(\mu)\,\delta_{\mathrm{D}}(\boldsymbol{q}+\boldsymbol{p})\,\langle\delta_g^{\mathrm{s}}(\boldsymbol{q})\,\delta_g^{\mathrm{s}}(\boldsymbol{p})\rangle \\
&= \left(\frac{2\ell+1}{2}\right)\frac{1}{(2\pi)^3 N_{\mathrm{p}}(k)}\int_{V_p}\int_{V_q} d^3\boldsymbol{p}d^3\boldsymbol{q}\, L_\ell(\mu)\,\delta_{\mathrm{D}}(\boldsymbol{q}+\boldsymbol{p})^2\,(2\pi)^3\mathrm{P}_{\mathrm{g}}^{\mathrm{s}}(\boldsymbol{p}) \\
&= \left(\frac{2\ell+1}{2}\right)\frac{1}{(2\pi)^3 N_{\mathrm{p}}(k)}\int_{V_p}\int_{V_q} d^3\boldsymbol{p}d^3\boldsymbol{q}\, L_\ell(\mu)\,\delta_{\mathrm{D}}(\boldsymbol{q}+\boldsymbol{p})\,V_{\mathrm{e}}\,\mathrm{P}_{\mathrm{g}}^{\mathrm{s}}(\boldsymbol{p}) \\
&= \left(\frac{2\ell+1}{2}\right)\frac{1}{V_{\mathrm{e}}V_k}\int_{V_p}\int_{V_q} d^3\boldsymbol{p}d^3\boldsymbol{q}\, L_\ell(\mu)\,\delta_{\mathrm{D}}(\boldsymbol{q}+\boldsymbol{p})\,V_{\mathrm{e}}\,\mathrm{P}_{\mathrm{g}}^{\mathrm{s}}(\boldsymbol{p}) \\
&= \left(\frac{2\ell+1}{2}\right)\frac{1}{V_k}\int_{V_p}\int_{V_q} d^3\boldsymbol{p}d^3\boldsymbol{q}\, L_\ell(\mu)\,\delta_{\mathrm{D}}(\boldsymbol{q}+\boldsymbol{p})\,\mathrm{P}_{\mathrm{g}}^{\mathrm{s}}(\boldsymbol{p}) \\
&\approx \left(\frac{2\ell+1}{2}\right)\int_{-1}^{+1} d\mu\,\mathrm{P}_{\mathrm{g}}^{\mathrm{s}}(k,\mu)\,L_\ell(\mu),
\end{aligned} \tag{A3}$$

where we used the approximation made in Joachimi et al. (2009) that $\delta_{\mathrm{D}}^2 \approx \dfrac{V_{\mathrm{e}}}{(2\pi)^3}\delta_{\mathrm{D}} = k_f^{-3}\delta_{\mathrm{D}}$. In the last step, it has been made the common approximation that $\boldsymbol{p}$ and $\boldsymbol{q}$ are very close to k in module for thin enough shells (small Δk). The standard definition of the redshift galaxy power spectrum has also been used:

$$\langle\delta_g^{\mathrm{s}}(\boldsymbol{q})\,\delta_g^{\mathrm{s}}(\boldsymbol{p})\rangle = (2\pi)^3\delta_{\mathrm{D}}(\boldsymbol{q}+\boldsymbol{p})\,\mathrm{P}_{\mathrm{g}}^{\mathrm{s}}(\boldsymbol{p}). \tag{A4}$$

The redshift-space galaxy bispectrum is defined as

$$\langle\delta_g^{\mathrm{s}}(\boldsymbol{q}_1)\,\delta_g^{\mathrm{s}}(\boldsymbol{q}_2)\,\delta_g^{\mathrm{s}}(\boldsymbol{q}_3)\rangle = (2\pi)^3\delta_{\mathrm{D}}(\boldsymbol{q}_1+\boldsymbol{q}_2+\boldsymbol{q}_3)\,\mathrm{B}_{\mathrm{g}}^{\mathrm{s}}(\boldsymbol{q}_1,\boldsymbol{q}_2,\boldsymbol{q}_3). \tag{A5}$$

The analytical expression for the bispectrum monopole model was given in equation (7). Analogously to the power spectrum multipoles, the estimator for the bispectrum monopole can be defined as

$$\hat{\mathrm{B}}_{\mathrm{g}}^{(0)}(k_1,k_2,k_3) = \frac{1}{4\pi}\frac{V_{\mathrm{e}}(2\pi)^{-6}}{N_{\mathrm{t}}(k_1,k_2,k_3)}\prod_{i=1}^{3}\int_{V_{q_i}} d^3\boldsymbol{q}_i\,\delta_D(\boldsymbol{q}_1+\boldsymbol{q}_2+\boldsymbol{q}_3)\,\delta_g^{\mathrm{s}}(\boldsymbol{q}_1)\,\delta_g^{\mathrm{s}}(\boldsymbol{q}_2)\,\delta_g^{\mathrm{s}}(\boldsymbol{q}_3), \tag{A6}$$

where $N_{\mathrm{t}}(k_1, k_2, k_3)$ is the number of independent grid points triplets inside the integration volume in Fourier space. As shown in the weak-lensing 2D case by Kayo et al. (2013), this is computed as

$$N_{\mathrm{t}}(k_1,k_2,k_3) = \frac{V_{k_{123}}}{k_f^6} = k_f^{-6}\int_{V_{q_1}}\int_{V_{q_2}}\int_{V_{q_3}} d^3\boldsymbol{q}_1 d^3\boldsymbol{q}_2 d^3\boldsymbol{q}_3\,\delta_{\mathrm{D}}(\boldsymbol{q}_1+\boldsymbol{q}_2+\boldsymbol{q}_3) \simeq \frac{8\pi^2 k_1 k_2 k_3\Delta k_1\Delta k_2\Delta k_3}{k_f^6}. \tag{A7}$$

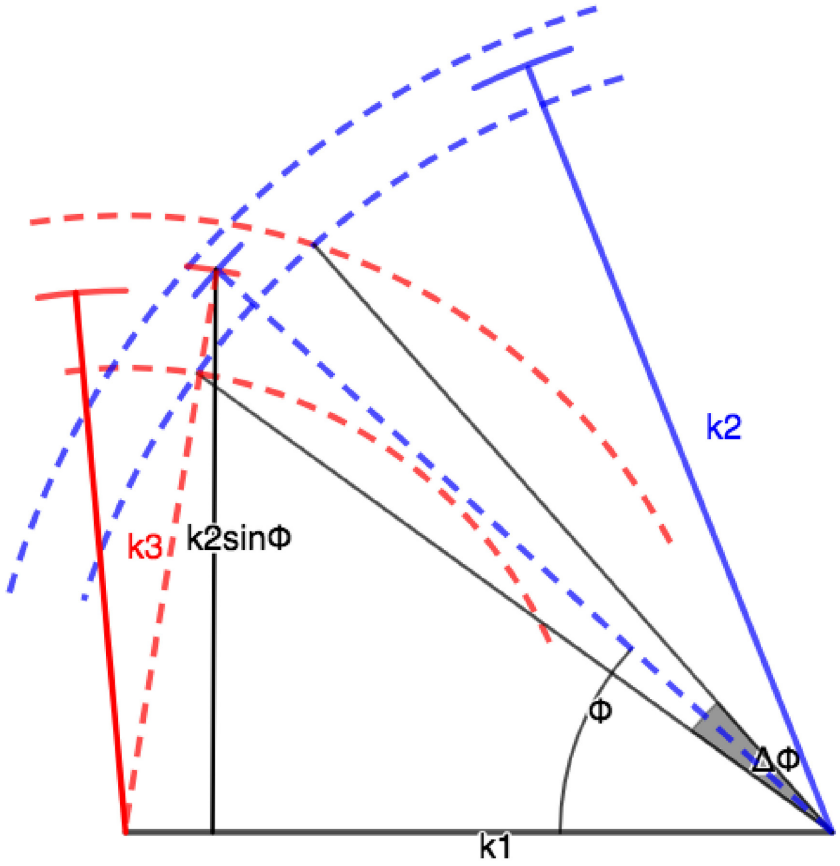


Figure A1. Computation of the integration volume in Fourier space in the case of the bispectrum monopole. Once the side k_1 of the triangle is fixed, the other two sides are free to vary in the intersection given by two sphere of radius $k_2 - \Delta k_2/2 \leq r_2 \leq k_2 + \Delta k_2/2$ and $k_3 - \Delta k_3/2 \leq r_3 \leq k_3 + \Delta k_3/2$, respectively. In the figure above, the 2D projection of the annuli of thickness Δk_2 (blue) and Δk_2 (red) is shown. The angle ϕ corresponds to the angle ϕ_{12} in the text.

It is important to notice that the result of the above integral must be symmetric in the k-vectors arguments. Therefore, the best way to derive the integral results is through geometrical considerations. Starting from q_1, this can be chosen in a spherical shell with volume $V_{k_1} \simeq 4\pi k_1^2 \Delta k$. Once q_1 is fixed, considering the plane in which both q_2 and q_3 lie, they must connect to each other inside the 2D intersection formed by the two annuli defined by $k_2 - \Delta k_2/2 \leq q_2 \leq k_2 + \Delta k_2/2$ and $k_3 - \Delta k_3/2 \leq q_3 \leq k_3 + \Delta k_3/2$. This has approximately an area equal to $A_{k_{23}} \simeq k_2 \Delta\phi_{12} \Delta k_2$. From Fig. A1, it is possible to see that $\Delta\phi_{12}$ is defined by varying k_3 by Δk_3. ϕ_{12} can be obtained from

$$\cos\phi_{12} = \frac{k_1^2 + k_2^2 - k_3^2}{2k_1k_2}, \tag{A8}$$

and therefore $\Delta\phi_{12}$ can be found differentiating with respect to k_3:

$$\frac{d\cos\phi_{12}}{dk_3} = -\frac{d\phi_{12}}{dk_3}\sin\phi_{12} = -\frac{k_3}{k_1k_2} \qquad \Longrightarrow \qquad \Delta\phi_{12} = \frac{\Delta k_3 k_3}{k_1k_2}(\sin\phi_{12})^{-1}. \tag{A9}$$

Finally, the volume of the intersection between k_2 and k_3 is obtained by rotating the area just found around the axis defined by k_1:

$$V_{k_{23}} = 2\pi A_{k_{23}}(k_2 \sin\phi_{12}), \tag{A10}$$

which allows to compute $V_{k_{123}} = V_{k_1} V_{k_{23}}$ in equation (A7).

A2 Power spectrum monopole and quadrupole covariance matrix: Gaussian term

Following the Appendix of Gualdi et al. (2018), we can check that also the bispectrum monopole estimator defined in equation (A6) is unbiased. Moreover, it is possible to compute the Gaussian term of the covariance for the power spectrum monopole and quadrupole as

follows:

$$
\begin{aligned}
&\mathrm{C_G^{P_g^{(\ell)}P_g^{(\ell)}}}(k_1;k_2) = \left(\frac{2\ell+1}{2}\right)^2 \frac{(2\pi)^{-6}}{N_\mathrm{p}(k_1)\,N_\mathrm{p}(k_2)} \int_{V_{q_1}}\int_{V_{q_2}}\int_{V_{p_1}}\int_{V_{p_2}} d^3\boldsymbol{q}_1 d^3\boldsymbol{q}_2 d^3\boldsymbol{p}_1 d^3\boldsymbol{p}_2 L_\ell(\mu_1)\,L_\ell(\mu_2)\,\delta_D(\boldsymbol{q}_1+\boldsymbol{p}_1)\,\delta_D(\boldsymbol{q}_2+\boldsymbol{p}_2)\\
&\times 2(2\pi)^6\delta_D(\boldsymbol{q}_1+\boldsymbol{q}_2)\,\delta_D(\boldsymbol{p}_1+\boldsymbol{p}_2)\,\mathrm{P_g^s}(\boldsymbol{q}_1)\,\mathrm{P_g^s}(\boldsymbol{p}_2)\\
&= \left(\frac{2\ell+1}{2}\right)^2 \frac{2}{N_\mathrm{p}(k_1)\,N_\mathrm{p}(k_2)} \int_{V_{q_1}}\int_{V_{q_2}} d^3\boldsymbol{q}_1 d^3\boldsymbol{q}_2 L_\ell(\mu_1)\,L_\ell(\mu_2)\,\delta_D(\boldsymbol{q}_1+\boldsymbol{q}_2)^2\,\mathrm{P_g^s}(\boldsymbol{q}_1)\,\mathrm{P_g^s}(\boldsymbol{q}_2)\\
&= \left(\frac{2\ell+1}{2}\right)^2 \frac{2k_f^{-3}}{N_\mathrm{p}(k_1)\,N_\mathrm{p}(k_2)} \int_{V_{q_1}}\int_{V_{q_2}} d^3\boldsymbol{q}_1 d^3\boldsymbol{q}_2 L_\ell(\mu_1)\,L_\ell(\mu_2)\,\delta_D(\boldsymbol{q}_1+\boldsymbol{q}_2)\,\mathrm{P_g^s}(\boldsymbol{q}_1)\,\mathrm{P_g^s}(\boldsymbol{q}_2)\\
&\approx \left(\frac{2\ell+1}{2}\right)^2 \frac{2k_f^{-3}}{N_\mathrm{p}(k_1)\,N_\mathrm{p}(k_2)}\quad \mathrm{P_g^{(\ell)}}(k_1)\,\mathrm{P_g^{(\ell)}}(k_2) \int_{V_{q_1}}\int_{V_{q_2}} d^3\boldsymbol{q}_1 d^3\boldsymbol{q}_2 \delta_D(\boldsymbol{q}_1+\boldsymbol{q}_2)\\
&= \left(\frac{2\ell+1}{2}\right)^2 \frac{2\delta_{12}^\mathrm{K}}{N_\mathrm{p}(k_1)}\quad \mathrm{P_g^{(\ell)}}(k_1)^2\,,
\end{aligned}
\tag{A11}
$$

where again we used the approximation made in Joachimi et al. (2009) that $\delta_\mathrm{D}^2 \approx \frac{V_\mathrm{e}}{(2\pi)^3}\delta_\mathrm{D} = k_f^{-3}\delta_\mathrm{D}$. δ_{12}^K is the Kronecker delta indicating that the vectors $\boldsymbol{q}_1$ and $\boldsymbol{q}_2$ are identical (in the second step, trivial δ_K has been omitted in order to avoid making the notation heavier by adding also the wave-vector letter). In the last steps, we made the approximation that the power spectrum monopole and quadrupoles do not vary significantly when integrated over the bin in Fourier space.

A3 Bispectrum monopole covariance matrix: Gaussian term

Analogously to the above, we now compute the diagonal term of the bispectrum monopole covariance matrix:

$$
\begin{aligned}
&\mathrm{C_G^{B_g^0B_g^0}}(k_1,k_2,k_3;k_4,k_5,k_6) =\\
&= \frac{1}{16\pi^2}\frac{(2\pi k_f)^{-6}}{N_\mathrm{t}(k_1,k_2,k_3)\,N_\mathrm{t}(k_4,k_5,k_6)} \prod_{i=1}^{6}\int_{V_{q_i}} d^3\boldsymbol{q}_i\,\delta_D(\boldsymbol{q}_1+\boldsymbol{q}_2+\boldsymbol{q}_3)\,\delta_D(\boldsymbol{q}_4+\boldsymbol{q}_5+\boldsymbol{q}_6)\\
&\times(2\pi)^9\delta_D(\boldsymbol{q}_1+\boldsymbol{q}_4)\,\delta_D(\boldsymbol{q}_2+\boldsymbol{q}_5)\,\delta_D(\boldsymbol{q}_3+\boldsymbol{q}_6)\,\mathrm{P_g^s}(\boldsymbol{q}_1)\,\mathrm{P_g^s}(\boldsymbol{p}_2)\,\mathrm{P_g^s}(\boldsymbol{q}_3) \quad +5 \quad \text{perm.}\\
&= \frac{\mathrm{D}_{123456}}{16\pi^2}\frac{(2\pi)^3 k_f^{-6}}{N_\mathrm{t}(k_1,k_2,k_3)^2} \prod_{i=1}^{3}\int_{V_{q_i}} d^3\boldsymbol{q}_i\,\delta_D(\boldsymbol{q}_1+\boldsymbol{q}_2+\boldsymbol{q}_3)^2\,\mathrm{P_g^s}(\boldsymbol{q}_1)\,\mathrm{P_g^s}(\boldsymbol{p}_2)\,\mathrm{P_g^s}(\boldsymbol{q}_3)\\
&= \frac{\mathrm{D}_{123456}}{16\pi^2}\frac{V_\mathrm{e} k_f^{-6}}{N_\mathrm{t}(k_1,k_2,k_3)^2} \prod_{i=1}^{3}\int_{V_{q_i}} d^3\boldsymbol{q}_i\,\delta_D(\boldsymbol{q}_1+\boldsymbol{q}_2+\boldsymbol{q}_3)\,\mathrm{P_g^s}(\boldsymbol{q}_1)\,\mathrm{P_g^s}(\boldsymbol{p}_2)\,\mathrm{P_g^s}(\boldsymbol{q}_3)\\
&\approx \frac{\mathrm{D}_{123456}}{16\pi^2}\frac{V_\mathrm{e} k_f^{-6}}{N_\mathrm{t}(k_1,k_2,k_3)^2}\mathrm{P_g^{(0)}}(k_1)\,\mathrm{P_g^{(0)}}(k_2)\,\mathrm{P_g^{(0)}}(k_3) \prod_{i=1}^{3}\int_{V_{q_i}} d^3\boldsymbol{q}_i\,\delta_D(\boldsymbol{q}_1+\boldsymbol{q}_2+\boldsymbol{q}_3)\\
&= \frac{\mathrm{D}_{123456}}{16\pi^2}\frac{V_\mathrm{e}}{N_\mathrm{t}(k_1,k_2,k_3)}\mathrm{P_g^{(0)}}(k_1)\,\mathrm{P_g^{(0)}}(k_2)\,\mathrm{P_g^{(0)}}(k_3)\,,
\end{aligned}
\tag{A12}
$$

where D_{123456} stands for all the possible permutations and has values 6, 2, and 1, respectively, for equilateral, isosceles, and scalene triangles. Again, it has been assumed that the power spectrum monopole does not vary significantly inside the integration volume.

A4 Bispectrum monopole covariance matrix: non-Gaussian term

In this work, we use only one of the non-Gaussian terms of the bispectrum monopole covariance matrix. This is because we just need to model the covariance matrix analytically in order to derive the weights for the compression. This additional term allows to better capture the

correlation between different triangle bins. We leave to future work the analytic computation of the remaining terms:

$$
\begin{aligned}
&\mathrm{C}_{\mathrm{NG}}^{\mathrm{B}_{\mathrm{g}}^{0}\mathrm{B}_{\mathrm{g}}^{0}}(k_1,k_2,k_3;k_4,k_5,k_6) = \\
&= \frac{1}{16\pi^2}\frac{(2\pi k_f)^{-6}}{N_{\mathrm{t}}(k_1,k_2,k_3)\,N_{\mathrm{t}}(k_4,k_5,k_6)}\prod_{i=1}^{6}\int_{V_{q_i}} d^3\boldsymbol{q}_i\,\delta_D(\boldsymbol{q}_1+\boldsymbol{q}_2+\boldsymbol{q}_3)\,\delta_D(\boldsymbol{q}_4+\boldsymbol{q}_5+\boldsymbol{q}_6) \\
&\times(2\pi)^6\delta_D(\boldsymbol{q}_1+\boldsymbol{q}_2+\boldsymbol{q}_4)\,\delta_D(\boldsymbol{q}_3+\boldsymbol{q}_5+\boldsymbol{q}_6)\,\mathrm{B}_{\mathrm{g}}^{\mathrm{s}}(\boldsymbol{q}_1,\boldsymbol{q}_2,\boldsymbol{q}_4)\,\mathrm{B}_{\mathrm{g}}^{\mathrm{s}}(\boldsymbol{q}_3,\boldsymbol{q}_5,\boldsymbol{q}_6) \quad +8 \quad \text{perm.} \\
&= \frac{1}{16\pi^2}\frac{k_f^{-6}\delta_{34}^{\mathrm{K}}}{N_{\mathrm{t}}(k_1,k_2,k_3)\,N_{\mathrm{t}}(k_3,k_5,k_6)}\int_{V_{q_1}}\int_{V_{q_2}}\int_{V_{q_3}}\int_{V_{q_5}}\int_{V_{q_6}} d^3\boldsymbol{q}_1 d^3\boldsymbol{q}_2 d^3\boldsymbol{q}_3 d^3\boldsymbol{q}_5 d^3\boldsymbol{q}_6\,\delta_D(\boldsymbol{q}_1+\boldsymbol{q}_2+\boldsymbol{q}_3) \\
&\times\delta_D(\boldsymbol{q}_3+\boldsymbol{q}_5+\boldsymbol{q}_6)^2\,\mathrm{B}_{\mathrm{g}}^{\mathrm{s}}(\boldsymbol{q}_1,\boldsymbol{q}_2,-\boldsymbol{q}_3)\,\mathrm{B}_{\mathrm{g}}^{\mathrm{s}}(\boldsymbol{q}_3,\boldsymbol{q}_5,\boldsymbol{q}_6) \quad +8 \quad \text{perm.} \\
&= \frac{1}{16\pi^2}\frac{k_f^{-9}\delta_{34}^{\mathrm{K}}}{N_{\mathrm{t}}(k_1,k_2,k_3)\,N_{\mathrm{t}}(k_3,k_5,k_6)}\int_{V_{q_1}}\int_{V_{q_2}}\int_{V_{q_3}}\int_{V_{q_5}}\int_{V_{q_6}} d^3\boldsymbol{q}_1 d^3\boldsymbol{q}_2 d^3\boldsymbol{q}_3 d^3\boldsymbol{q}_5 d^3\boldsymbol{q}_6\,\delta_D(\boldsymbol{q}_1+\boldsymbol{q}_2+\boldsymbol{q}_3) \\
&\times\delta_D(\boldsymbol{q}_3+\boldsymbol{q}_5+\boldsymbol{q}_6)\,\mathrm{B}_{\mathrm{g}}^{\mathrm{s}}(\boldsymbol{q}_1,\boldsymbol{q}_2,-\boldsymbol{q}_3)\,\mathrm{B}_{\mathrm{g}}^{\mathrm{s}}(\boldsymbol{q}_3,\boldsymbol{q}_5,\boldsymbol{q}_6) \quad +8 \quad \text{perm.} \\
&\approx \frac{1}{16\pi^2}\frac{k_f^{-3}\delta_{34}^{\mathrm{K}}}{N_{\mathrm{t}}(k_3,k_5,k_6)}\mathrm{B}_{\mathrm{g}}^{(0)}(k_1,k_2,k_3)\,\mathrm{B}_{\mathrm{g}}^{(0)}(k_3,k_5,k_6)\int_{V_{q_i}} d^3\boldsymbol{q}_5 d^3\boldsymbol{q}_6\,\delta_D(\boldsymbol{q}_3+\boldsymbol{q}_5+\boldsymbol{q}_6) \quad +8 \quad \text{perm.} \\
&= \frac{\delta_{34}^{\mathrm{K}}}{16\pi^2}\frac{k_f^3}{4\pi k_3^2\Delta k_3}\mathrm{B}_{\mathrm{g}}^{(0)}(k_1,k_2,k_3)\,\mathrm{B}_{\mathrm{g}}^{(0)}(k_3,k_5,k_6) \quad +8 \quad \text{perm.},
\end{aligned}
\tag{A13}
$$

where the usual approximations have been used together with equation (A10) that in the last step has been used to simplify the integration over the volume in Fourier space once one of the $\boldsymbol{k}$-vectors is fixed.

A5 cross-covariance term

For what concerns the cross-covariance term between power spectrum (monopole/quadrupole) and bispectrum monopole, we use only the first leading term in our model:

$$
\begin{aligned}
&\mathrm{C}^{\mathrm{P}_{\mathrm{g}}^{(\ell)}\mathrm{B}_{\mathrm{g}}^{0}}(k_1;k_2,k_3,k_4) = \\
&= \frac{1}{4\pi}\left(\frac{2\ell+1}{2}\right)\frac{(2\pi)^{-6}k_f^{-3}}{N_{\mathrm{p}}(k_1)\,N_{\mathrm{t}}(k_2,k_3,k_4)}\int_{V_{q_1}}\int_{V_{p_1}} d^3\boldsymbol{q}_1 d^3\boldsymbol{p}_1\prod_{i=2}^{4}\int_{V_{q_i}} d^3\boldsymbol{q}_i\,\delta_D(\boldsymbol{q}_1+\boldsymbol{p}_1)\,\delta_D(\boldsymbol{q}_2+\boldsymbol{q}_3+\boldsymbol{q}_4)\;L_\ell(\mu_1) \\
&\times 2(2\pi)^6\delta_D(\boldsymbol{q}_1+\boldsymbol{q}_2)\,\delta_D(\boldsymbol{p}_1+\boldsymbol{q}_3+\boldsymbol{q}_4)\,\mathrm{P}_{\mathrm{g}}^{\mathrm{s}}(\boldsymbol{q}_2)\,\mathrm{B}_{\mathrm{g}}^{\mathrm{s}}(\boldsymbol{q}_2,\boldsymbol{q}_3,\boldsymbol{q}_4) \quad +2 \quad \text{perm.} \\
&= \frac{1}{2\pi}\left(\frac{2\ell+1}{2}\right)\frac{k_f^{-3}}{N_{\mathrm{p}}(k_1)\,N_{\mathrm{t}}(k_2,k_3,k_4)}\prod_{i=1}^{4}\int_{V_{q_i}} d^3\boldsymbol{q}_i\;L_\ell(\mu_1)\,\delta_D(\boldsymbol{q}_1+\boldsymbol{q}_2)\,\delta_D(\boldsymbol{q}_2+\boldsymbol{q}_3+\boldsymbol{q}_4)^2\,\mathrm{P}_{\mathrm{g}}^{\mathrm{s}}(\boldsymbol{q}_2)\,\mathrm{B}_{\mathrm{g}}^{\mathrm{s}}(\boldsymbol{q}_2,\boldsymbol{q}_3,\boldsymbol{q}_4) \quad +2 \quad \text{perm.} \\
&= \frac{1}{2\pi}\left(\frac{2\ell+1}{2}\right)\frac{k_f^{-6}\delta_{12}^{\mathrm{K}}}{N_{\mathrm{p}}(k_2)\,N_{\mathrm{t}}(k_2,k_3,k_4)}\prod_{i=2}^{4}\int_{V_{q_i}} d^3\boldsymbol{q}_i\;L_\ell(\mu_2)\,\delta_D(\boldsymbol{q}_2+\boldsymbol{q}_3+\boldsymbol{q}_4)\,\mathrm{P}_{\mathrm{g}}^{\mathrm{s}}(\boldsymbol{q}_2)\,\mathrm{B}_{\mathrm{g}}^{\mathrm{s}}(\boldsymbol{q}_2,\boldsymbol{q}_3,\boldsymbol{q}_4) \quad +2 \quad \text{perm.} \\
&\approx \frac{1}{2\pi}\left(\frac{2\ell+1}{2}\right)\frac{\delta_{12}^{\mathrm{K}}}{N_{\mathrm{p}}(k_2)}\mathrm{P}_{\mathrm{g}}^{(\ell)}(k_2)\,\mathrm{B}_{\mathrm{g}}^{(0)}(k_2,k_3,k_4) \quad +2 \quad \text{perm.},
\end{aligned}
\tag{A14}
$$

where once more we have used the same approximation of the power spectrum multipoles and bispectrum monopole not varying significantly inside the integration volume.

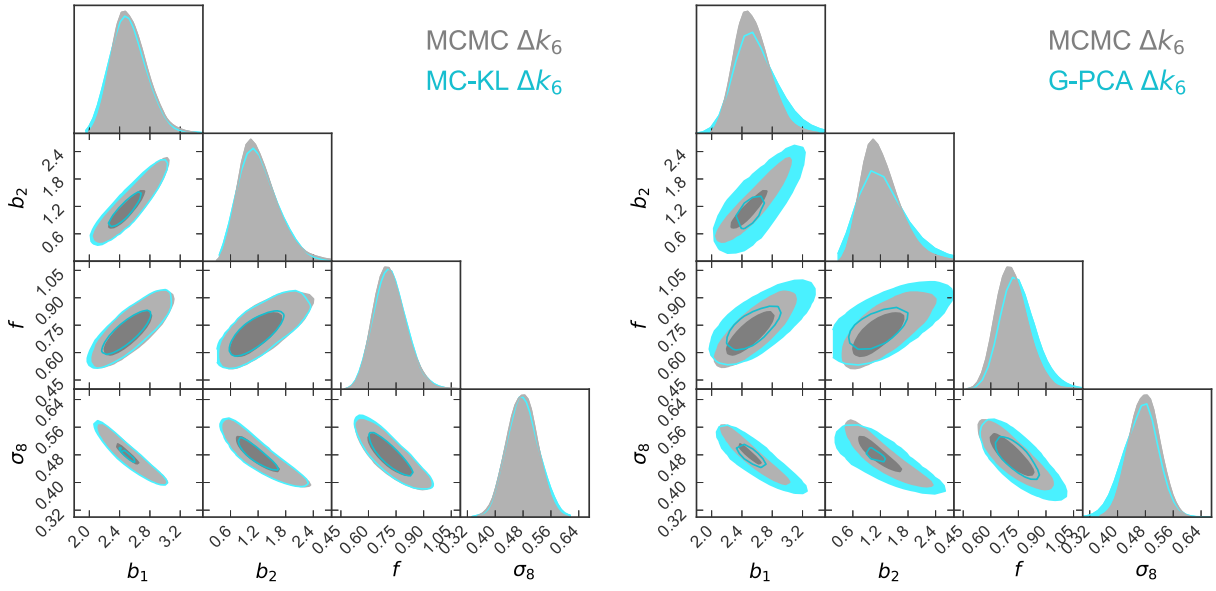
APPENDIX B: VALIDATION TESTS

In Table B1, we report the results obtained compressing the bispectrum with respect to the shifted fiducial parameter sets. This is to test whether the performance of the compression is affected by the choice of fiducial set of parameter values. In particular, we consider two cases by varying the fiducial cosmology by adding/subtracting 1σ 1D credible intervals (derived from the MCMC) to all the parameters. The table quantifies that the shifts in the means of the 1D posterior distributions produced by considering a non-optimal fiducial cosmology are small compared to the 1σ 1D credible intervals of the MCMC results.

In Figs B1 and B2, the 1D and 2D posterior distributions obtained via MCMC/MC–KL/G-PCA for the test cases relative to the Δk_6 and Δk_5 binning cases are shown. MC–KL recovers with very good approximation the 1D and 2D posterior distributions derived by the MCMC. G-PCA shows a slightly greater loss of information for the Δk_6 case. However, this is noticeably closer to the MCMC/MC–KL result when the number of triangle bins used is increased (Δk_5 case).

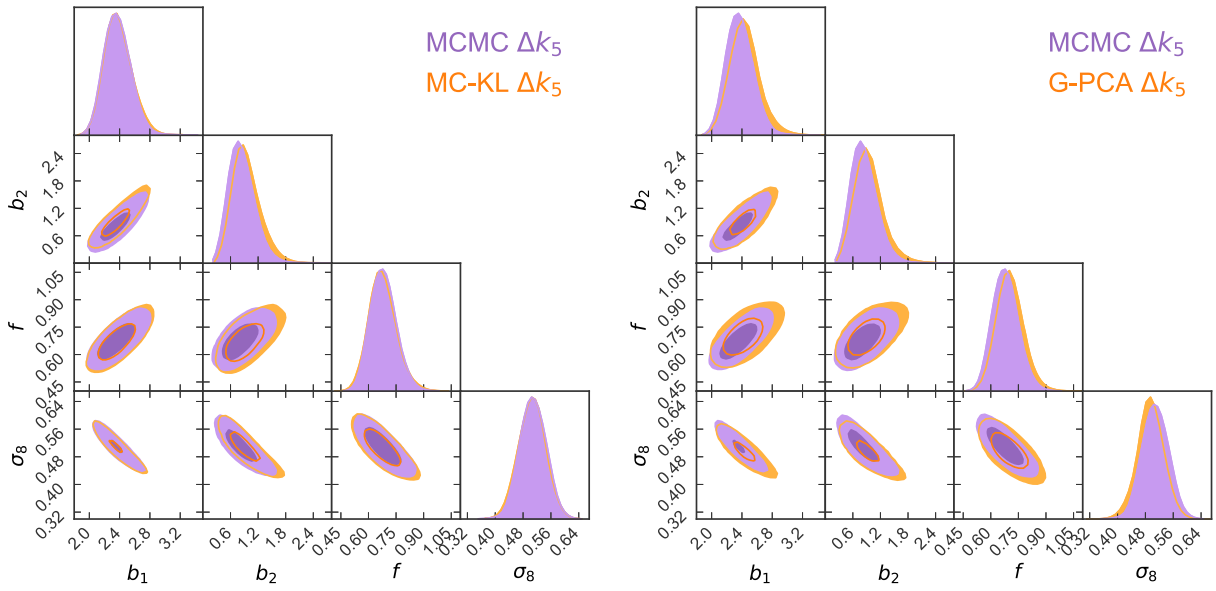
In Fig. B3, we compare the best-fitting model obtained by varying four parameters (b_1, b_2, f, and σ_8) with the best-fitting model corresponding to a fit done via standard MCMC sampling with only three parameters varied, (b_1, b_2, and f), with $\sigma_8 = \sigma_8^{\mathrm{fid.}}$. For the three-parameter case, we find running the MCMC: $b_1 = 1.98 \pm 0.01$, $b_2 = 0.39 \pm 0.06$, and $f(z_{\mathrm{CMASS}}) = 0.52 \pm 0.03$ with $\sigma_8^{\mathrm{fid.}}(z_{\mathrm{CMASS}}) = 0.61$.

Thereby, we show that the discrepancy between the results of this paper and the ones presented in the BOSS collaboration analysis Gil-Marín et al. (2017) is, together with the different model used for the power spectrum monopole and quadrupole, probably due to the



(a) MC-KL (b) G-PCA

Figure B1. Joint data vector $\left[P_g^{(0)}, P_g^{(2)}, B_g^{(0)}\right]$ posteriors: MC–KL and G-PCA four-parameter Δk_6 case. (a) 2D 68 per cent and 95 per cent credible regions are shown in order to compare the MC–KL (cyan) performance to one of the standard MCMC (grey) for the full data vector. The difference between MC–KL and MCMC contours is quantified in Table 1. (b) The same as (a) but for the G-PCA method.



(a) MC-KL (b) G-PCA

Figure B2. Joint data vector $\left[P_g^{(0)}, P_g^{(2)}, B_g^{(0)}\right]$ posteriors: MC–KL and G-PCA four-parameter Δk_5 case. Both (a) and (b) are the same as for Fig. B1 for the Δk_5 case.

different range of scales considered. Indeed, by limiting our analysis to a smaller range of scales in k-space, the degeneracy between the amplitude-like parameters b_1 and σ_8 is much stronger. That is visible in Fig. B3, where the models given by sets of parameters with very different b_1, b_2, and σ_8 parameters produce very similar predictions of the signals all with good $\chi^2_{\rm red.}$ and p-values.

Table B1. Four-parameter case, checking consistency for shifted fiducial cosmology. *Upper half:* Mean values of the posterior distributions and 68 per cent credible intervals for the MCMC and the MC–KL / G-PCA compression methods. We report the values for the Δk_6 binning case for both compression methods in three cases consisting in using for the compression: the fiducial cosmology, the fiducial cosmology shifted by $+1\,\sigma$, and the fiducial cosmology shifted by $-1\,\sigma$. *Lower half:* In the compression columns, we report the relative difference between the posterior modes obtained via MCMC and the ones obtained via compression (MC–KL or G-PCA). In the MCMC columns, the relative size of the 68 per cent credible intervals obtained via MCMC sampling is shown. By comparing the MCMC columns to the compression ones, it is clear that the difference between the mean parameter values obtained via MCMC and the ones obtained via compression (MC–KL or G-PCA) is evidently within the 68 per cent credible intervals given by the MCMC on the full data vector.

	Δk_6			$\Delta k_6 \quad + 1\sigma$		$\Delta k_6 \quad - 1\sigma$	
	MCMC	MC–KL	G-PCA	MC–KL	G-PCA	MC–KL	G-PCA
b_1	2.41 ± 0.22	2.41 ± 0.23	2.49 ± 0.27	2.47 ± 0.23	2.41 ± 0.12	2.54 ± 0.24	2.34 ± 0.37
b_2	1.00 ± 0.40	1.04 ± 0.42	1.08 ± 0.47	1.04 ± 0.40	1.29 ± 0.25	1.03 ± 0.44	0.93 ± 0.67
f	0.69 ± 0.08	0.72 ± 0.09	0.72 ± 0.09	0.70 ± 0.08	0.69 ± 0.05	0.72 ± 0.09	0.68 ± 0.12
σ_8	0.50 ± 0.04	0.48 ± 0.05	0.48 ± 0.05	0.49 ± 0.04	0.49 ± 0.03	0.46 ± 0.05	0.50 ± 0.07
	$\frac{\Delta\theta^{\rm mc}}{\theta^{\rm mc}}$ [%]	$\frac{\theta^{\rm comp.}-\theta^{\rm mc}}{\theta^{\rm mc}}$ [%]		$\frac{\theta^{\rm comp.}-\theta^{\rm mc}}{\theta^{\rm mc}}$ [%]		$\frac{\theta^{\rm comp.}-\theta^{\rm mc}}{\theta^{\rm mc}}$ [%]	
b_1	9.2	−0.3	3.3	2.15	−0.26	8.57	0.31
b_2	40.3	3.5	7.5	3.47	28.68	25.29	13.26
f	12.1	4.4	4.4	0.84	0.51	6.96	0.26
σ_8	8.5	−5.1	−5.5	−3.25	−2.91	−8.94	−1.39

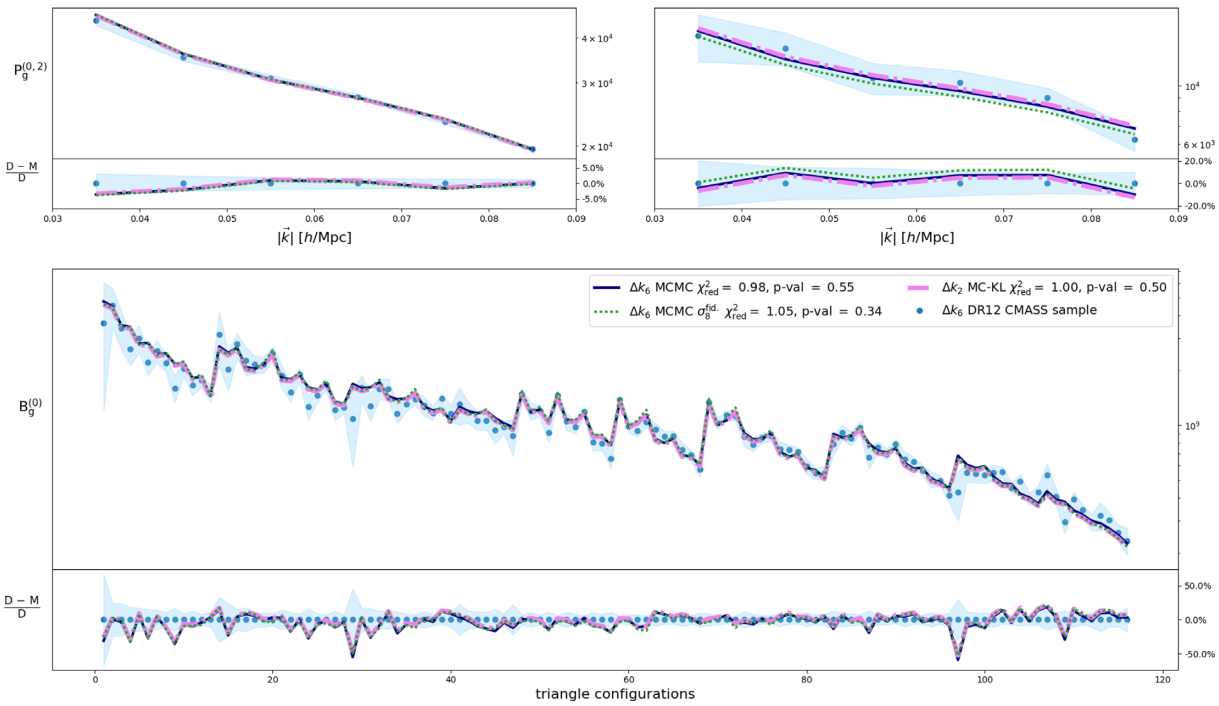


Figure B3. Reduced χ^2 and p-values for the best-fitting parameters obtained using the MCMC/MC–KL methods with varying σ_8 and for the MCMC leaving $\sigma_8 = \sigma_8^{\rm fid.}$ fixed. The k-binnings shown for the four-parameter case (b_1, b_2, f, and σ_8) are, respectively, the standard Δk_6 (navy) for the MCMC and the Δk_2 (pink) for the MC + KL. The line corresponding to the fit obtained by letting free to vary only the parameters (b_1, b_2, f) is shown in green. The two upper panels are for the power spectrum monopole (left) and quadrupole (right), while the bottom panel refers to the bispectrum monopole. The lower part of each panel shows the relative difference between the data measurements and the different models. Even if, for example, b_1 and σ_8 values are shifted in the cases of Δk_6 and Δk_2, this is due to the strong degeneracy between them, and both models are practically identical to the one given by the three-parameters fit (b_1, b_2, and f) with $\sigma_8 = \sigma_8^{\rm fid.}$. The only way to converge to the results obtained by the BOSS collaboration is to consider a larger range of scales (as they have done) for both power spectrum and bispectrum that, however, involves a more complex modelling of the data vector.

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