Minkowski Functionals of SDSS-III BOSS: Hints of Possible Anisotropy in the Density Field?

Stephen Appleby, Changbom Park, Pratyush Pranav, Sungwook E. Hong, Ho Seong Hwang, Juhan Kim, Thomas Buchert

▶ **To cite this version:**

Stephen Appleby, Changbom Park, Pratyush Pranav, Sungwook E. Hong, Ho Seong Hwang, et al.. Minkowski Functionals of SDSS-III BOSS: Hints of Possible Anisotropy in the Density Field?. Astrophys.J., 2022, 928 (2), pp.108. 10.3847/1538-4357/ac562a . hal-03406871

HAL Id: hal-03406871
https://hal.science/hal-03406871

Submitted on 20 Jun 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

The Astrophysical Journal, 928:108 (23pp), 2022 April 1 https://doi.org/10.3847/1538-4357/ac562a
© 2022. The Author(s). Published by the American Astronomical Society.
OPEN ACCESS

Minkowski Functionals of SDSS-III BOSS: Hints of Possible Anisotropy in the Density Field?

Stephen Appleby[1,2], Changbom Park[3], Pratyush Pranav[4], Sungwook E. Hong (홍성 욱)[5,6], Ho Seong Hwang[7,8], Juhan Kim[9], and Thomas Buchert[4]
[1] Asia Pacific Center for Theoretical Physics, Pohang, 37673, Republic of Korea; stephen.appleby@apctp.org
[2] Department of Physics, POSTECH, Pohang 37673, Republic of Korea
[3] School of Physics, Korea Institute for Advanced Study, 85 Hoegiro, Dongdaemun-gu, Seoul, 02455, Republic of Korea
[4] Univ Lyon, ENS de Lyon, Univ Lyon1, CNRS, Centre de Recherche Astrophysique de Lyon UMR5574, F69007, Lyon, France
[5] Korea Astronomy and Space Science Institute, 776 Daedeokdae-ro, Yuseong-gu, Daejeon 34055, Republic of Korea
[6] Astronomy Campus, University of Science and Technology, 776 Daedeok-daero, Yuseong-gu, Daejeon 34055, Republic of Korea
[7] Astronomy Program, Department of Physics and Astronomy, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 08826, Republic of Republic of Korea
[8] SNU Astronomy Research Center, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 08826, Republic of Korea
[9] Center for Advanced Computation, Korea Institute for Advanced Study, 85 Hoegiro, Dongdaemun-gu, Seoul, 02455, Republic of Korea
Received 2021 October 22; revised 2022 February 15; accepted 2022 February 15; published 2022 March 30

Abstract

We present measurements of the Minkowski functionals extracted from the SDSS-III BOSS catalogs. After defining the Minkowski functionals, we describe how an unbiased reconstruction of these statistics can be obtained from a field with masked regions and survey boundaries, validating our methodology with Gaussian random fields and mock galaxy snapshot data. From the BOSS galaxy data, we generate a set of four density fields in three dimensions corresponding to the northern and southern skies of LOWZ and CMASS catalogs, smoothing over large scales (Gaussian smoothing scale of 35 Mpc) such that the field is perturbatively non-Gaussian. We extract the Minkowski functionals from each data set separately, and measure their shapes and amplitudes by fitting a Hermite polynomial expansion. For the shape parameter of the Minkowski functional curves a_0, that is related to the bispectrum of the field, we find that the LOWZ-South data presents a systematically lower value of $a_0 = -0.080 \pm 0.040$ than its northern sky counterpart $a_0 = 0.032 \pm 0.024$. Although the significance of this discrepancy is not very high, it potentially indicates some systematics in the data or that the matter density field exhibits anisotropy at low redshift. By assuming a standard isotropic flat ΛCDM cosmology, the amplitudes of the Minkowski functionals from the combination of northern and southern sky data give the constraints $\Omega_c h^2 n_s = 0.110 \pm 0.006$ and 0.111 ± 0.008 for CMASS and LOWZ, respectively, which is in agreement with the Planck ΛCDM best-fit $\Omega_c h^2 n_s = 0.116 \pm 0.001$.

Unified Astronomy Thesaurus concepts: Cosmology (1146); Large-scale structure of the universe (902); Cosmic web (330)

1. Introduction

The Minkowski functionals (MFs) describe the morphological, i.e., geometrical and topological, characteristics of a field. They have a long and venerable history within cosmology, both theoretical and computational, as well as observational, starting with the genus of isodensity excursion sets (Melott et al. 1989; Gott et al. 1990; Park & Gott 1991; Park et al. 1992; Mecke et al. 1994; Schmalzing & Buchert 1997; Schmalzing & Gorski 1998; Park et al. 2001; Park & Kim 2010; Zunckel et al. 2011). Often referred to merely as summary statistics in cosmological literature, they emerge from topo-geometrical description of space and manifolds, and together with the homology characteristics encoded in the Betti numbers (van de Weygaert et al. 2011; Park et al. 2013; Shivshankar et al. 2016; Pranav et al. 2017, 2019a, 2019b; Feldbrugge et al. 2019), and its hierarchical extension persistent homology (Edelsbrunner & Harer 2010; Pranav et al. 2017), as well as the Minkowski tensors (Beisbart et al. 2001a, 2001b; Chingangbam et al. 2017a, 2017b; Ganesan & Chingangbam 2017; Appleby et al. 2018a, 2018b; Kapahtia et al. 2018; Joby et al. 2019; Kapahtia et al. 2019; Wilding et al. 2020; Chingangbam et al. 2021; Goyal & Chingangbam 2021), present a high-level description of the topo-geometrical properties of the fields of interest in cosmology. Extracting information from these statistics remains an open and challenging program (Park et al. 2001; Hikage et al. 2001, 2002, 2003; Sheth et al. 2003; Park et al. 2005; Sheth & Sahni 2005; Gott et al. 2008, 2009; James et al. 2009; Choi et al. 2010; Zhang et al. 2010; Petri et al. 2013; Blake et al. 2014; Parihar et al. 2014; Wiegand et al. 2014; Wang et al. 2015; Buchert et al. 2017; Wiegand & Eisenstein 2017; Sullivan et al. 2019; Lippich & Sánchez 2021; Matsubara & Kuriki 2021; Appleby et al. 2021; Pranav 2021; Shim et al. 2021).

For D-dimensional sets $\delta(x_1,\ldots,x_D)$, there are $(D+1)$ MFs, which correspond to topo-geometrical properties of the set. Considering a manifold $\mathbb{M}$, in three dimensions, the MFs correspond to the enclosed volume (W_0) of $\mathbb{M}$, as well as the surface area (W_1), mean (W_2) and Gaussian curvatures (W_3) of the boundary $\partial\mathbb{M}$. In our particular setting, the manifolds of interest are the excursion sets of three-dimensional cosmological density fields.

In a recent series of works, we have measured the genus of two-dimensional slices of the Sloan Digital Sky Survey (SDSS) III Baryon Oscillation Spectroscopic Survey (BOSS) data, extracting cosmological information from the genus amplitude,

Original content from this work may be used under the terms of the Creative Commons Attribution 4.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

and placing constraints on the scalar spectral index n_s and dark matter fraction $\Omega_c h^2$, assuming a flat ΛCDM cosmology (Appleby et al. 2020). Given that the BOSS data is spectroscopic, we have precise redshift information. Hence it is possible to extract the MFs from the full three-dimensional data, which contains more information than two-dimensional shells.[10]

With this motivation in mind, in this work, we extract the MFs from the full, three-dimensional galaxy distribution. Specifically, we use the SDSS-III BOSS DR12 catalog to reconstruct four distinct, three-dimensional density fields corresponding to the LOWZ/CMASS data in the northern and southern Galactic Planes. After generating smoothed number density fields from the point galaxy distribution, we measure the statistics, ensuring that the complex survey geometry does not impact our results. We then extract information from the amplitudes of these functions, using similar methodology to Appleby et al. (2020). We also measure the bispectrum components from the MFs, although we do not convert this information into cosmological parameter constraints in this work.

The paper will proceed as follows. We define the MFs, and review their theoretical expectation values for a compact and boundaryless field in Section 2. In Section 3, we review our methodology, including the generation of a density field from a point distribution, and the mock data used to estimate statistical uncertainties. Time-poor readers can proceed directly to Section 4, which contains our principle results—MF measurements extracted from the BOSS galaxy samples and cosmological parameter estimation using their amplitudes. We discuss our results in Section 5.

Some of the details of our methodology and analysis are contained in appendices. In Appendix A, we define the MFs for a generic manifold. In Appendix B, the numerical algorithm for MF extraction is elucidated, and we confirm that our method is unbiased by survey boundaries, using masked Gaussian random fields (GRFs) and mock galaxy snapshot data. The effect of redshift space distortion (RSD) is briefly reviewed in Appendix C. Potential systematics that could impact our analysis are described in Appendix D.

2. Integral-geometry of Manifolds: Minkowski Functionals

In a cosmological setting, we are typically interested in the properties of a scalar random field u, defined on a manifold $\mathbb{M}$, which will be Euclidean and three-dimensional in this work. We take u to be mean subtracted and root mean square normalized $\langle u^2\rangle = 1$, and Gaussian in this subsection. The MFs $W_0,\ldots,W_3$ are topo-geometric quantifiers that encode the geometry and topology induced by the fluctuations of the field, in combination with the geometric characteristics of the manifold itself. The usual practice is to examine the properties of the *excursion set* of the manifold, defined by

$$E_\nu = \{x \in \mathbb{M}\colon u(x) \geqslant \nu\}, \tag{1}$$

where ν is an isofield threshold value. When dealing with compact boundaryless manifolds, such as the 2-sphere or the periodically tiled Euclidean grid (found for example in cosmological simulation snapshot boxes), the expressions for the MFs simplify substantially, and the volume-normalized expressions may be written in terms of curvature integrals

$$W_0 = \frac{1}{V}\int_{E_\nu} dV, \tag{2}$$

$$W_1 = \frac{1}{6V}\int_{\partial E_\nu} dA, \tag{3}$$

$$W_2 = \frac{1}{3\pi V}\int_{\partial E_\nu} G_2\, dA, \tag{4}$$

$$W_3 = \frac{1}{4\pi^2 V}\int_{\partial E_\nu} G_3\, dA, \tag{5}$$

where $G_2 = (\kappa_1 + \kappa_2)/2$ and $G_3 = \kappa_1.\,\kappa_2$ are the mean and Gaussian curvatures of the boundary of the excursion set ∂E_ν. Here, κ_1, κ_2 are the principal curvatures of the surface ∂E_ν at a point, and V is the total volume of the space under consideration. The equations for the 2-sphere were derived by Doroshkevich (1970), while the generic case including boundary effects was developed by Adler (1981). The usage of the Euler characteristic was introduced to cosmology by Gott and collaborators (Gott et al. 1990; Park & Gott 1991), while the full set of volume-normalized expressions in three dimensions was introduced in cosmology by Mecke et al. (1994), Schmalzing & Buchert (1997).

For a GRF, the ensemble average of these curvature integrals can be written as (Doroshkevich 1970; Adler 1981; Gott et al. 1986; Hamilton et al. 1986; Tomita 1986; Gott et al. 1987; Weinberg et al. 1987; Ryden et al. 1989)

$$\langle W_k\rangle = \frac{1}{(2\pi)^{(k+1)/2}}\frac{\omega_3}{\omega_{3-k}\omega_k}\left(\frac{\sigma_1^2}{3\sigma_0^2}\right)^{k/2} e^{-\nu^2/2} H_{k-1}(\nu), \tag{6}$$

where $H_k(\nu)$ is kth Hermite polynomial, and σ_i are the two-point cumulants

$$\sigma_i^2 = \frac{1}{(2\pi)^3}\int d^3k k^{2i} P(k) W^2(kR_G). \tag{7}$$

$P(k)$ is the power spectrum of the field. $W(kR_G)$ is a smoothing kernel with corresponding scale R_G, which we take to be Gaussian throughout this work: $W(kR_G) \propto e^{-k^2R_G^2/2}$. The amplitude of $\langle W_k\rangle$ is defined as

$$A_{k,G} \equiv \frac{1}{(2\pi)^{(k+1)/2}}\frac{\omega_3}{\omega_{3-k}\omega_k}\left(\frac{\sigma_1^2}{3\sigma_0^2}\right)^{k/2}. \tag{8}$$

The constants ω_m are the volume of the unit ball in $\mathbb{R}^m$:

$$\omega_m = \frac{\pi^{m/2}}{\Gamma\left(\frac{n}{2}+1\right)}, \tag{9}$$

i.e., $\omega_0 = 1$, $\omega_1 = 2$, $\omega_2 = \pi$, and $\omega_3 = 4\pi/3$.

The curvature integrals (Equations (2)–(5)) are equal to the MFs only when the manifold possesses no boundary. For a field defined on a generic manifold, the Gaussian Kinematic Formula (GKF) describes the MFs instead. The GKF is defined in Appendix A.

Practically, cosmological data sets are always defined on domains with boundaries. For example the Cosmic Microwave Background temperature field is measurable over the entire 2-sphere, but realistically foreground masks ensure that only part of the all-sky data is used. Similarly, galaxy catalogs possess both angular and radial survey boundaries, and the

[10] Shells were analyzed in Appleby et al. (2017, 2018c, 2020, 2021) with the long-term goal of comparing the results with photometric galaxy catalogs.

resulting matter density field is measured only over a finite volume with a boundary. The MFs of these cosmological data sets are therefore *not* described by the curvature integrals defined in Equations (2)–(5). However, it is possible to generate an unbiased estimate of the curvature integrals from a field with a boundary, and directly compare the results to the ensemble expectation values (Equation (6)).

To see this, we take one of the curvature integrals (3) as an example. By using an integral transform, we can write

$$W_1 = \frac{1}{6V}\int_{\partial E_\nu} \mathrm{dA} = \frac{1}{6V}\int_{\mathbb{M}} \mathrm{dV}|\nabla u|\,\delta_D(u-\nu). \tag{10}$$

We recognise the final term in this equation as the volume average of the scalar quantity $\lambda \equiv \delta_D(\nu-u)\sqrt{u_i u^i}/6$, where u is the field, u_i are the gradients with respect to some arbitrary coordinate system ($i=1,\,2,\,3$), and δ_D is the delta function, which is defined in a distributional sense as we are taking a volume average. If we generate a pixelated GRF, W_1 can be estimated as

$$W_1 \simeq \frac{1}{6N_{\rm pix}}\sum_{\ell=1}^{N_{\rm pix}} \tilde{\delta}_D(u_\ell-\nu)\sqrt{u_{\ell,i}u_\ell^i}, \tag{11}$$

where $N_{\rm pix}$ is the number of pixels in the discretized field, and we discretize the delta function as

$$\tilde{\delta}_D(u_\ell-\nu) = \begin{cases} \frac{1}{\Delta\nu}, & \text{for } \nu-\frac{\Delta\nu}{2} \leqslant u_\ell < \nu+\frac{\Delta\nu}{2} \\ 0, & \text{otherwise} \end{cases}. \tag{12}$$

The important point is that the numerical approximation of W_1 defined in Equation (11) can be estimated from a finite subset of a field. That is, we do not require an entire field defined on a boundaryless manifold to estimate the curvature integral (Equation (3)); any unbiased sampling of pixels can be used to generate an unbiased estimator of this quantity.

The property of ergodicity is then used to equate the volume and ensemble averages of λ. The ensemble average of λ can be written as

$$\langle W_1\rangle = \frac{1}{6}\int P(u,\,u_k)\,du\,du_k\,\delta_D(u-\nu)\sqrt{u_i u^i}, \tag{13}$$

where $P(u,\,u_k)$ is the joint probability density function of the field and its derivatives. If we take u, u_k to be uncorrelated and Gaussian distributed, this ensemble average yields the standard result in Equation (6).

Again, we stress that the volume average in Equation (11) does not require the field to be complete and boundaryless, and can be approximated by a subset of pixels of masked data. Hence we can estimate the curvature integrals from cosmological data, but we should understand that these quantities do not represent the MFs for a field defined on a manifold with a boundary. This is intuitively obvious, as the curvature integrals are intrinsically *local* quantities in the sense that they can be estimated from the average properties of the field and its derivatives at points on the manifold. In contrast, the topology of the manifold is an intrinsically global quantity.

Throughout this work, we will use the terms *curvature integrals* and *Minkowski functionals* interchangeably, but the reader should understand the distinction made above, and this work is concerned solely with the curvature integrals.

2.1. Weakly Non-Gaussian Fields

For a weakly non-Gaussian field, the amplitude and shape of the MFs is modified. To linear order in σ_0, the following *Edgeworth* expansion has been constructed (Melott et al. 1988; Matsubara 1994a, 1994b; Matsubara & Yokoyama 1996; Matsubara & Suto 1996; Matsubara 2000; Hikage et al. 2008; Pogosyan et al. 2009; Gay et al. 2012; Codis et al. 2013):

$$\langle W_k\rangle = A_{k,\rm G}e^{-\nu^2/2}\Bigg\{H_{k-1}(\nu) + \Bigg[\frac{1}{6}S^{(0)}H_{k+2}(\nu) + \frac{k}{3}S^{(1)}H_k(\nu) + \frac{k(k-1)}{6}S^{(2)}H_{k-2}(\nu)\Bigg]\sigma_0 + \mathcal{O}(\sigma_0^2)\Bigg\}. \tag{14}$$

The quantities $S^{(0)}$, $S^{(1)}$, and $S^{(2)}$ are proportional to the three-point cumulants of the field:

$$S^{(0)} = \frac{\langle\delta^3\rangle}{\sigma_0^4}, \tag{15}$$

$$S^{(1)} = -\frac{3}{4}\frac{\langle\delta^2(\nabla^2\delta)\rangle}{\sigma_0^2\sigma_1^2}, \tag{16}$$

$$S^{(2)} = -\frac{9}{4}\frac{\langle(\nabla\delta\cdot\nabla\delta)(\nabla^2\delta)\rangle}{\sigma_1^4}. \tag{17}$$

The amplitude and shape of the MFs contain information pertaining to the N-point cumulants of the field, which in turn are sensitive to cosmological parameters.[11]

Throughout this work, we rescale the isodensity threshold ν to $\nu_{\rm A}$, which is the threshold defined such that the excursion set has the same volume fraction as a corresponding Gaussian field:

$$f_{\rm A} = \frac{1}{\sqrt{2\pi}}\int_{\nu_A}^{\infty} e^{-t^2/2}\,dt, \tag{18}$$

where $f_{\rm A}$ is the fractional volume of the field above $\nu_{\rm A}$. Expressing the MFs as a function of $\nu_{\rm A}$ as opposed to ν mitigates the non-Gaussianity in the MFs (Gott et al. 1987; Weinberg et al. 1987; Melott et al. 1988), although obviously does not completely remove it. Additional non-Gaussian information is retained in the mapping $\mathcal{F}_{\rm A}$: $\nu\to\nu_{\rm A}$ but is not used in this work. The non-Gaussian expansion of MFs as a function of $\nu_{\rm A}$ is (Matsubara 2000; Hikage et al. 2008)

$$\langle W_k\rangle = A_{k,\rm G}e^{-\nu_{\rm A}^2/2}\{H_{k-1}(\nu_{\rm A}) + \Bigg[\frac{k}{3}(S^{(1)}-S^{(0)})H_k(\nu_{\rm A}) + \frac{k(k-1)}{6}(S^{(2)}-S^{(0)})H_{k-2}(\nu_{\rm A})\Bigg]\sigma_0 + \mathcal{O}(\sigma_0^2)\Bigg\}. \tag{19}$$

The amplitude of the MF W_k, which is the coefficient of Hermite polynomial $H_{k-1}(\nu_{\rm A})$ in the perturbative non-Gaussian expansion, predominantly contains the Gaussian information of the field (with second-order corrections $\sim\mathcal{O}(\sigma_0^2)$). All other

[11] See Matsubara et al. (2020) for a recent expansion of $\langle W_k\rangle$ at $\mathcal{O}(\sigma_0^2)$ and Gay et al. (2012) for a general expansion. For a model-independent approach applying MFs to the CMB and using general Hermite expansions of the discrepancy functions with respect to the analytical Gaussian predictions, together with a generalization of Matsubara's second-order expansion, see Buchert et al. (2017).

Hermite polynomial coefficients contain only higher-point cumulant information, which are induced by the non-Gaussianity of the field.

As we have seen in Equations (6), (14), (19), the Hermite polynomial expansion is useful to extract information from the MF curves. In particular, one can extract their coefficient by using the following orthogonality relation:

$$\int_{-\infty}^{\infty} H_{\rm m}(\nu) H_n(\nu) e^{-\nu^2/2}\, d\nu = \sqrt{2\pi}\, n!\, \delta_{mn}, \tag{20}$$

where δ_{mn} is the Kronecker delta. Provided we are smoothing over scales $R_{\rm G}$ such that the σ_0 expansion of the MFs is applicable, we can multiply the measured MF by a Hermite polynomial and integrate over ν to obtain the coefficient of the polynomial. Performing the integral in Equation (20) over $\nu_{\rm A}$ rather than ν is recommended, as the MFs as a function of ν are more strongly asymmetric around $\nu = 0$. To reliably utilize the orthogonality property of the functions, one must integrate over large $|\nu|$ ranges. The extraction of polynomial coefficients using this method is only formally valid if we have measured the MF over a sufficiently large threshold range. Any truncation in the integral will correlate the Hermite polynomial coefficients obtained using this method.

The intention of this paper is to provide a method of numerically reconstructing the curvature integrals as in Equations (2)–(5) from a discretized and masked density field, and then extracting cosmological information from $A_{k,{\rm G}}$, the Gaussian amplitude of these functions. We are not measuring the *true* MFs of the bounded manifold on which the matter density field is defined, as explained in the previous section. We intend to pursue the difference between the curvature integrals and MFs in detail in future work, as there is additional information contained within the boundary. The presence of a mask can profoundly modify the global properties of the manifold.

3. Methodology

We now review and subsequently validate our methodology. First, we describe our reconstruction of a smoothed density field from the point distribution, and introduce the mock galaxy catalogs used to reconstruct the statistical error of our measurements. In Appendix B.1, we provide a detailed explanation for our numerical algorithm for measuring the curvature integrals, and validate our analysis with mock data.

3.1. Density Field Reconstruction

We measure the MFs of the SDSS-III BOSS (York et al. 2000). The twelfth data release of the SDSS (Alam et al. 2015) imaged 9,376 deg^2 in the *ugriz* bands (Fukugita et al. 1996). The survey was executed with the 2.5 m Sloan telescope (Gunn et al. 2006) at the Apache Point Observatory in New Mexico. The extra-galactic catalog contains 1,372,737 unique galaxies, with redshifts extracted using an automated pipeline described in Bolton et al. (2012).

The SDSS-III BOSS data is decomposed into two catalogs. The LOWZ sample is composed of galaxies predominantly at redshift $z < 0.4$, and are selected using numerous color–magnitude cuts that are intended to match the evolution of a passively evolving stellar population. The purpose is to extend the bright and red *low-redshift* galaxy population measured in the SDSS-II Luminous Red Galaxies to relatively higher redshift. The CMASS galaxies, on the other hand, are selected using a set of color–magnitude cuts to identify *high-redshift* galaxies at $0.4 < z < 0.7$. In contrast to the LOWZ data, the sample is not biased toward red galaxies as some of the color limits imposed on the SDSS-II sample have been removed. The color–magnitude cut is varied with redshift to ensure massive objects are sampled as uniformly as possible over the survey volume. We direct the reader to Reid et al. (2016) for further details of the galaxy samples.

Table 1
Fiducial Cosmological Parameters Used to Reconstruct Distances in This Work, and in the Creation of the BOSS Patchy Mocks

Parameter	Fiducial Value
$\Omega_{\rm m}$	0.307
h	0.6777
$w_{\rm de}$	−1
$n_{\rm s}$	0.9611
σ_8	0.8288
$R_{\rm G}$	35 Mpc

Note. $R_{\rm G}$ is the smoothing scale used throughout this work, using a Gaussian kernel.

Throughout this work, we treat the LOWZ and CMASS catalogs separately, and also treat the north and south sky data as independent. Hence we have four practically independent data sets—CMASS and LOWZ, north/south—from which we extract the MF statistics. All steps below are repeated individually for each subset of the data.

The preparation of the data requires some care, so we now list the steps taken in this work to convert the galaxy data to a smooth density field.

1. First each galaxy is weighted to account for observational systematics. Specifically, the following combined weight was applied to each galaxy in the LOWZ and CMASS sample:
$$w_{\rm tot} = w_{\rm systot}(w_{\rm cp} + w_{\rm noz} - 1), \tag{21}$$
where $w_{\rm cp}$ is the correction factor to account for the subsample of galaxies that are not assigned a spectroscopic fiber (cp stands for "close pair"), $w_{\rm noz}$ is for the failure in the pipeline to assign redshifts for certain galaxies, and $w_{\rm systot}$ represents noncosmological fluctuations in the CMASS target density due to stellar density and seeing (Reid et al. 2016).
2. We bin the galaxies into redshift shells of thickness $\Delta z = 0.02$ over the range $0.10 < z < 0.48$ and $0.40 < z < 0.68$ for the LOWZ/CMASS data, respectively. The redshift bin thickness is ∼85 to ∼60 Mpc in comoving distance over this redshift range. Each galaxy contributes $w_{\rm tot}$ to its redshift bin. We then select galaxies in each shell to match the number density as $\bar{n} = 6.25 \times 10^{-5}\,{\rm Mpc}^{-3}$. The selection is made using a lower mass cut based on the predicted stellar mass of the galaxies from the *Portsmouth* model found in Tinker et al. (2017). If the total number density in a given redshift shell is below $\bar{n}$, it is not used in our analysis. To generate a number density, which is a dimension-full quantity, we use the fiducial cosmology presented in Table 1 to define the volume of the shells. For practical purposes, the variation of this cosmology will not affect

Table 2
Size of Box That We Enclose the Four Data Sets in, and the Resolution of Each Box

Data Set	$L_{\rm box}$ (Mpc)	Δ_g (Mpc)
CMASS N	4200	8.2
CMASS S	3330	6.5
LOWZ N	3050	6.0
LOWZ S	2400	4.7

Note. We use Gaussian smoothing with scale $R_{\rm G} = 35$ Mpc to mitigate any discrete pixel effects.

our results as the shells are used only to generate mass cuts. The number density cut restricts our analysis to the redshift ranges $0.20 < z < 0.40$ and $0.45 < z < 0.60$ for LOWZ/CMASS, respectively. We could use lower-redshift data, but the volume over the range $0.1 < z < 0.2$ is insufficient to affect our results.

The construction of a uniform number density subset always requires some form of redshift-dependent selection. However, our tests indicate that the sampling is robust under reasonable variations of $0.015 < \Delta z < 0.03$,[12] and in Appendix D, we repeat our analysis for a randomly selected subset of the galaxies and find no significant change in our results.

3. The selected galaxies are aggregated into a uniform cubic grid of size $L_{\rm box}$ and resolution Δ_g along each dimension, where these values are provided in Table 2 for the four different data sets. The three-dimensional positions of the galaxies are generated from their angular positions and redshifts using the fiducial cosmology in Table 1. Each galaxy contributes $w_{\rm tot}$ to its nearest pixel, and this procedure generates a discrete number field n_{ijk}, where i, j, k subscripts run over the lattice in $x_{1,2,3}$ directions, respectively.
4. We also generate a mask lattice M_{ijk} by projecting the angular selection function ω_ℓ into the same cubic grid and applying the radial boundaries. The angular selection function ω_ℓ is a Healpix[13] (Gorski et al. 2005) $N_{\rm side} = 512$ map that takes value $\omega_\ell = 0$ if the ℓ pixel lies outside the survey boundary, and $0 \leqslant \omega_\ell \leqslant 1$ if ℓ is within the mask ($1 \leqslant \ell \leqslant 12 \times N_{\rm side}^2$ is the pixel identifier). The angular mask is projected into a three-dimensional cube such that $M_{ijk} = \omega_\ell$ if the (i, j, k) pixel lies inside the survey geometry, and $M_{ijk} = 0$ otherwise. We then apply a binary mask to n_{ijk} by setting $n_{ijk} = 0$ if $M_{ijk} < M_{\rm cut}$, where we set $M_{\rm cut} = 0.9$. Note that $w_{\rm tot}$ is a weight applied directly to the galaxies, so it is not used for the mask lattice.
5. We define $\bar{n}$ as the average number of galaxies within the unmasked pixels, and define a mean-subtracted number density as $n_{ijk} \to (n_{ijk}/\bar{n} - 1)$. The remaining masked pixels have an arbitrary *bad pixel* value δ_b and are not used.
6. We smooth both the field n_{ijk} and mask M_{ijk} with a Gaussian kernel $W(kR_{\rm G}) \propto \exp[-k^2 R_{\rm G}^2/2]$, where $R_{\rm G} = 35$ Mpc is selected such that the field is expected to be in the weakly non-Gaussian regime (Matsubara et al. 2020). We denote the smoothed field and mask as $\tilde{\delta}_{ijk}$ and $\tilde{M}_{ijk}$, respectively. We make a second mask cut and set $\tilde{\delta}_{ijk} = \delta_{\rm b}$ if $\tilde{M}_{ijk} \leqslant M_{\rm cut}$, where $M_{\rm cut} = 0.9$. This second masking procedure cuts regions of the density field close to the survey boundary.
7. Finally, we calculate the average μ and root mean square σ of all unmasked pixels, and define a mean-subtracted, unit variance field $\delta_{ijk} = (\tilde{\delta}_{ijk} - \mu)/\sigma$.

Having constructed a smoothed, discretized density field δ_{ijk}, we next extract the MFs from the unmasked pixels in the following section using the method described in Appleby et al. (2018b) but accounting for the presence of a mask. A discussion of how we adjust our algorithm to account for the mask can be found in Appendix B.1.

In this work, we smooth on a relatively large scale $R_{\rm G} = 35$ Mpc. Such a large smoothing is chosen because we intend to compare our measurements to the perturbative Edgeworth expansion derived in Matsubara (1994a, 1994b); Matsubara & Suto (1996). If we smooth on smaller scales, the amplitude of the MFs contains information on higher-point cumulant contributions and nonlinear RSD. An alternative approach is to smooth on small scales and correct for nonlinear gravitational interactions using simulations (Li et al. 2016; Appleby et al. 2021). This method has the advantage of yielding much stronger constraining power, but requires more careful analyses to remove nonlinear systematics and to take into account their dependence on cosmological models. Throughout this work, we follow Appleby et al. (2020) and avoid (as far as possible) correcting the measured MFs using simulations. We do correct for RSDs using simulations, but this is a $\sim\mathcal{O}(1\%)$ effect.

In what follows, we extract the $W_{0,1,2,3}$ MFs from the masked, smoothed galaxy density fields at $N = 41$, $\nu_{\rm A}$ threshold values, equi-spaced over the range $-3 < \nu_{\rm A} < 3$. To determine the magnitude of the statistical fluctuations on these measurements, we repeat our analysis on a set of mock galaxies with similar properties to the data. The mock analysis is described next.

3.2. Mock Galaxy Catalogs

To estimate the statistical uncertainty of the $W_k(\nu_A)$ measurements, we use $N_{\rm r} = 250$ Multidark-Patchy mocks (Kitaura et al. 2016; Rodríguez-Torres et al. 2016). A detailed explanation of their construction can be found in Kitaura et al. (2016). Briefly, the mocks were generated using an iterative procedure to mimic a reference galaxy catalog using gravity solvers and statistical biasing models (Kitaura et al. 2014). The reference catalog is the Big-MultiDark N-body simulation, which used Gadget-2 (Springel 2005) to gravitationally evolve 3840^3 particles in a $(2.5h^{-1}{\rm Gpc})^3$ volume. Halo abundance matching was utilized to reproduce the clustering of galaxy data. The Patchy code (Kitaura et al. 2014, 2015) matches the two- and three-point clustering statistics with the reference simulation in multiple redshift bins. Stellar masses are estimated and mock lightcones are generated, including masks and other selection effects. The mock catalogs accurately reproduce the number density, two-point correlation function, selection function, and survey geometry of the SDSS-III BOSS DR12 observational data. The simulations adopted a Planck standard ΛCDM cosmology with $\Omega_{\rm m} = 0.307$, $\Omega_{\rm b} = 0.048$, $n_{\rm s} = 0.961$, $H_0 = 67.77\,{\rm km\,s^{-1}Mpc^{-1}}$, the same as the fiducial cosmology adopted in this study.

[12] We repeated our analysis using $\Delta z = 0.015$ and $\Delta z = 0.03$, and found no significant change to our results.
[13] http://healpix.sourceforge.net

For each mock catalog, we repeat our methodology. We begin by sorting the galaxies into redshift shells, applying a mass cut, then projecting the surviving galaxies into a uniform lattice. We then apply our masking and smoothing procedure and use the marching tetrahedra to extract the MFs from the density field $W_k^p(\nu_A)$, where p represents the pth mock realization. As for the actual data, we measure the MFs at 41 values of ν_A over the range $-3 < \nu_A < 3$. The method is repeated separately for mock CMASS/LOWZ north ("N") and south ("S") data. From these measurements, a set of covariance matrices $\Sigma_k^{m,n}$ can be generated as

$$\Sigma_k^{m,n} = \frac{1}{N_r - 1}\sum_{p=1}^{N_r}(W_{k,p}^n - \langle W_k^n\rangle)(W_{k,p}^m - \langle W_k^m\rangle), \quad (22)$$

where $W_{k,p}^n$ is the value of the kth MF at the nth threshold value $\nu_{A,n}$ in the pth mock realization, and $\langle W_k^n\rangle$ is its average. Since we use a transformation to write the MFs as a function of ν_A, there is no information in the W_0 volume fraction,[14] and we restrict our analysis to $k = 1, 2, 3$. Hence there are a total of three covariance matrices $\Sigma_{1,2,3}^{m,n}$, for each of our four data sets: LOWZ/CMASS N/S. We present these covariance matrices in Figure 1; these are $W_{m,n}^k$ for the mock CMASS N catalogs. The other three sets are very similar and not exhibited. It is clear that the values of the MFs are strongly correlated between the different threshold bins, and can also be anticorrelated. The correlation is largest for $k = 1$ and lowest for $k = 3$. When we utilize the covariance matrices for the parameter estimation, we follow Hartlap et al. (2007) and correct their inverses with a factor of $(N_r - N_b - 2)/(N_r - 1)$, where $N_b = 41$ is the number of ν_A bins.

In Figure 2 (left panels), we present the MFs obtained from the patchy mock data, treating the LOWZ/CMASS N/S data separately. We exhibit them as a function of ν_A, and present $W_{1,2,3}$. The solid gold/green/blue/red lines are the mean values extracted from $N_{real} = 250$ mock realizations from the four different data sets, and the solid gray region is the standard deviation of the LOWZ S realizations. The mean values of all statistics are consistent within the statistical error of the measurements, which confirms the insensitivity of our analysis to the data mask. Each of the four subsets of data have very different survey geometries and volumes.

4. Results

We now present the $W_{1,2,3}$ statistics extracted from the BOSS galaxy data. We first present the results of our numerical analysis, and in Appendix D, we test their robustness under variation of the assumptions implicit within our methodology.

4.1. Minkowski Functionals of BOSS Data

In Figure 2 (right panels), we present the MFs of the four BOSS data sets—CMASS/LOWZ N/S—as a function of ν_A. We again plot the 1σ standard deviation of the LOWZ-South patchy mock realizations as gray filled regions, to provide a visual guide of the statistical errors on the measurements. The LOWZ S data have the largest uncertainties of the four, as it encompasses the smallest volume. The gold/green/blue/red solid lines represent the CMASS N/S and LOWZ N/S data, respectively. The right-hand panels show much larger statistical fluctuations than the left, because the left panels present the mean of $N = 250$ mock realizations whereas the right constitute a single data realization. In contrast to the mock data, in the right panels, one can observe some discrepancy between the MF curves in the northern and southern skies, with the LOWZ S (red lines) in particular presenting anomalously low values. The difference is most clearly observed in W_1 but is present in all three panels. The CMASS N and LOWZ N data occupy the largest volumes, and are consistent with the patchy mock realizations (blue/yellow lines).

From the covariance matrices presented in Figure 1, it is clear that $W_{1,2,3}$ are correlated between ν_A threshold bins. For this reason, care should be taken not to perform the statistical analysis *by eye*, using Figure 2. To proceed, we fit the following functions to each curve,

$$\tilde{W}_k = A_{k,G}e^{-\nu_A^2/2}\{H_{k-1}(\nu_A) + \frac{k}{3}a_0H_k(\nu_A) + \frac{k(k-1)}{6}a_2H_{k-2}(\nu_A)\}, \quad (23)$$

by minimizing the χ_k^2 functions

$$\chi_k^2 = \sum_{n,m}(W_{k,n} - \tilde{W}_{k,n})(\Sigma^{-1})_k^{n,m}(W_{k,m} - \tilde{W}_{k,m}), \quad (24)$$

assuming a Gaussian likelihood. In Equation (24), W_n^k is the measured value of the kth MF at the nth, ν_A threshold, and $\Sigma_k^{m,n}$ is defined in Equation (22). The parameters varied are $A_{k,G}$, a_0, and a_2. The quantities a_0, a_2 contain information pertaining to the bispectrum, which can be explicitly written as (Matsubara 2003)

$$a_0 = (S^{(1)} - S^{(0)})\sigma_0, \quad (25)$$

$$a_2 = (S^{(2)} - S^{(0)})\sigma_0, \quad (26)$$

where $S^{(0)}$, $S^{(1)}$, $S^{(2)}$ are given in Equations (15)–(17). We do not use a_0 and a_2 for the cosmological parameter estimation in this work, but each MF $W_{1,2,3}$ should measure consistent a_0, a_2 values when extracted from the same data set. This provides a consistency check of our methodology, assuming that Equation (23) is a viable fitting function. Furthermore, if the galaxy distribution is isotropic, then we should expect that the north and south data in each catalog will yield consistent $A_{k,G}$ and $a_{0,2}$ values. However, CMASS and LOWZ will not necessarily yield consistent results, as they constitute two distinct galaxy samples with different selection criteria and redshifts.

We minimize the function (Equation (24)) for each of the four data sets and each W_k function separately, to obtain a set of twelve measurements of $A_{k,G}$, a_0, and a_2. The prior ranges used are $-1 < a_{0,2} < 1$ and $-18.5 < \log[A_{k,G}] < -10$, and a variation of these limits does not affect our results. When fitting Equation (23) to W_3, we multiply the functions extracted from the data (see Figure 2, bottom panels) by -1, as the convention in cosmology is to present W_3 in terms of $-H_2$.

In Figures 3 and 4, we present the best fit and marginalized 1σ limits of the parameters $A_{k,G}$ and $a_{0,2}$, respectively, for each of the four data sets. The gold/green/blue/red diamonds and solid error bars are the best fit and 1σ uncertainties obtained by minimizing the χ_k^2 function (Equation (24)) for CMASS N/S and LOWZ N/S data, respectively. In Figure 4, the light-to-

[14] The information in W_0 has been transferred to the $\mathcal{F}_A$: $\nu \to \nu_A$ mapping, which we do not use here.

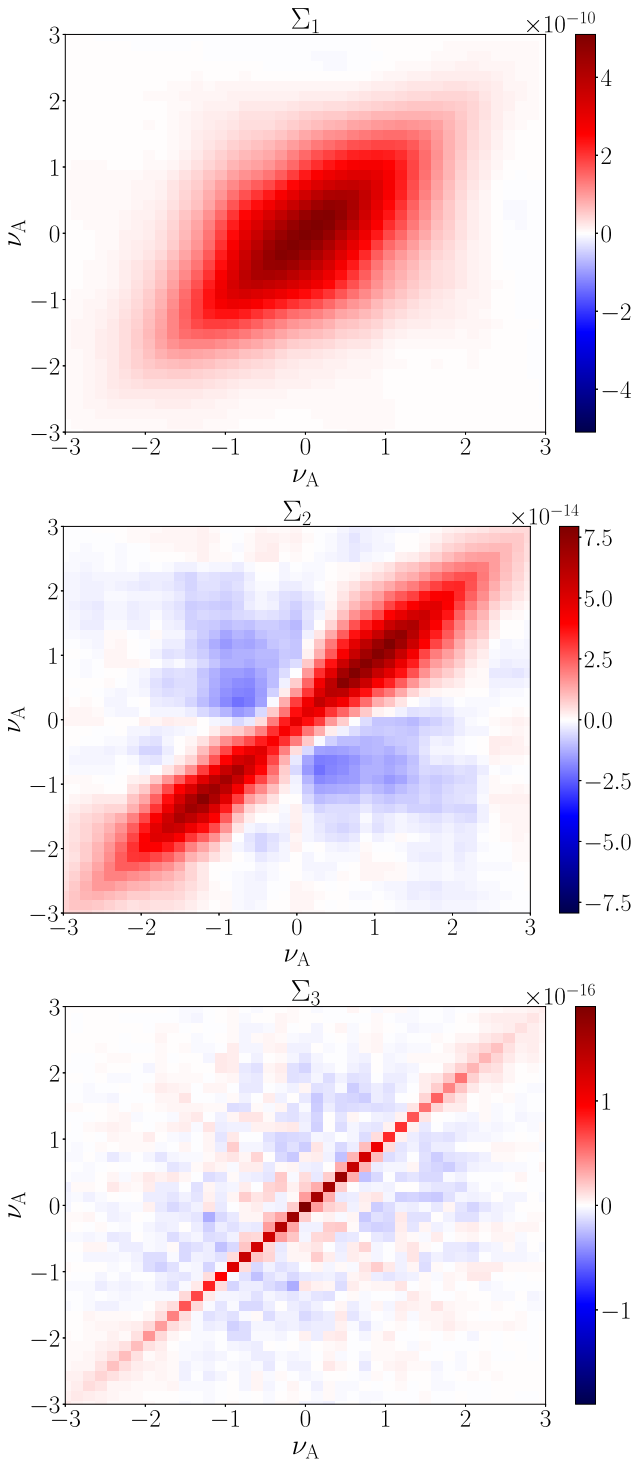

Figure 1. The covariance of the W_1, W_2, W_3 MFs (top–bottom) as a function of ν_A threshold. There are large correlations between threshold values for W_1 (red), and some anticorrelation in W_2, W_3 (blue regions). W_3 (bottom panel) exhibits the lowest level of cross-correlation between threshold values.

dark points/error bars are the values of a_0 and a_2 obtained from each MF $W_{1,2,3}$ within the same subset of data. Note that W_1 is independent of a_2 due to the $(k-1)$ factor in the expansion (Equation (23)), so it is not included in the lower panel of Figure 4.

For comparison, the small points and dashed error bars in the figures are the mean and rms values of $A_{k,G}$ and $a_{0,2}$ obtained from the patchy mock catalogs, obtained using the expressions

$$A_{k,G} \simeq \frac{1}{\sqrt{2\pi}\,(k-1)!}\int_{-4}^{4} W_k(\nu_A) H_{k-1}(\nu_A)\, d\nu_A, \tag{27}$$

$$a_0 \simeq \frac{3}{k\sqrt{2\pi}\,k!A_{k,G}}\int_{-4}^{4} W_k(\nu_A) H_k(\nu_A)\, d\nu_A, \tag{28}$$

$$a_2|_{k>1} \simeq \frac{6}{\sqrt{2\pi}\,k!A_{k,G}}\int_{-4}^{4} W_k(\nu_A) H_{k-2}(\nu_A)\, d\nu_A, \tag{29}$$

where W_k are the MFs extracted from the mocks. Finally, the dark/light gray horizontal dashed lines in Figure 3 are the Gaussian expectation values (Equation (8)) for the MF amplitudes, assuming the cosmological parameters in Table 1 and $\bar{n} = 6.25 \times 10^{-5}\,\mathrm{Mpc}^{-3}$, and the linear galaxy bias $b = 2$. The power spectrum adopted is $P(k, z) = b^2 P_m(k, z) + 1/\bar{n}$, where $P_m(k, z)$ is the linear matter power spectrum at redshift $z = 0.3$ and 0.5 (light/dark dashed lines). We correct Equation (8) from real- to redshift-space by applying a constant factor $A^{rsd}_{k,G} = \alpha_k A_{k,G}$, where $\alpha_k = 0.99$, 0.98, 0.97 for $k = 1$, 2, 3, respectively. This correction factor is derived from mock catalogs and is discussed further in Appendix C.

The patchy mock results (small points and dashed error bars) are entirely self-consistent, in the sense that the four data sets yield values of $A_{k,G}$ and $a_{0,2}$ that are in agreement within 1σ. Furthermore, $W_{1,2,3}$ as measured within each data set yield consistent values of $a_{0,2}$, which serves as a check that the three-point cumulants are being correctly measured. The amplitudes $A_{k,G}$ are in close agreement with the Gaussian expectation values (Equation (8)), except the systematically high values of $A_{2,G}$ from the patchy mocks (middle panel in Figure 3). We can provide no compelling explanation for this discrepancy, other than our estimator for W_2 might be marginally biased by the presence of the mask. Although the mean values are practically consistent with the Gaussian prediction at 1σ, the reconstructed values are systematically high.

The BOSS data results (diamond points, solid error bars) present some peculiarities. The amplitudes $A_{k,G}$ of $W_{1,2,3}$ extracted from the LOWZ S data are systematically lower than the other three data sets. The statistical significance of this discrepancy is low, due to the large smoothing scales adopted in this work. The bispectrum term a_0 is also large and negative in LOWZ S, which suggests that the discrepancy in the data is not restricted to the two-point cumulants. In addition, the data reconstruction of a_2 presents a mild discrepancy in the CMASS S data (see Figure 4, bottom panel). The reconstructed value of a_2 obtained from W_2 is high compared to the same quantity extracted from W_3. Some weak systematic offset is also observed in the mock reconstruction, which could again indicate some effect of the mask on the W_2 estimation. Extracting cosmological information from the bispectrum terms a_0 and a_2 will be considered in future work, and here we simply report the anomalous behavior of the southern sky data. The marginalized best fit and 1σ uncertainties on the parameters $a_{0,2}$ for each of the measurements in Figure 4 are presented in Table 3.

So far, we have proceeded under the assumption that the perturbative expansion (Equation (19)) can be applied to the data, and we truncated the expansion at order $\mathcal{O}(\sigma_0)$. At order σ_0^2, multiple new terms are introduced that are related to the four-point cumulants $\sim\langle\delta^4\rangle/\sigma_0^4$, and the amplitude $A_{k,G}$ also

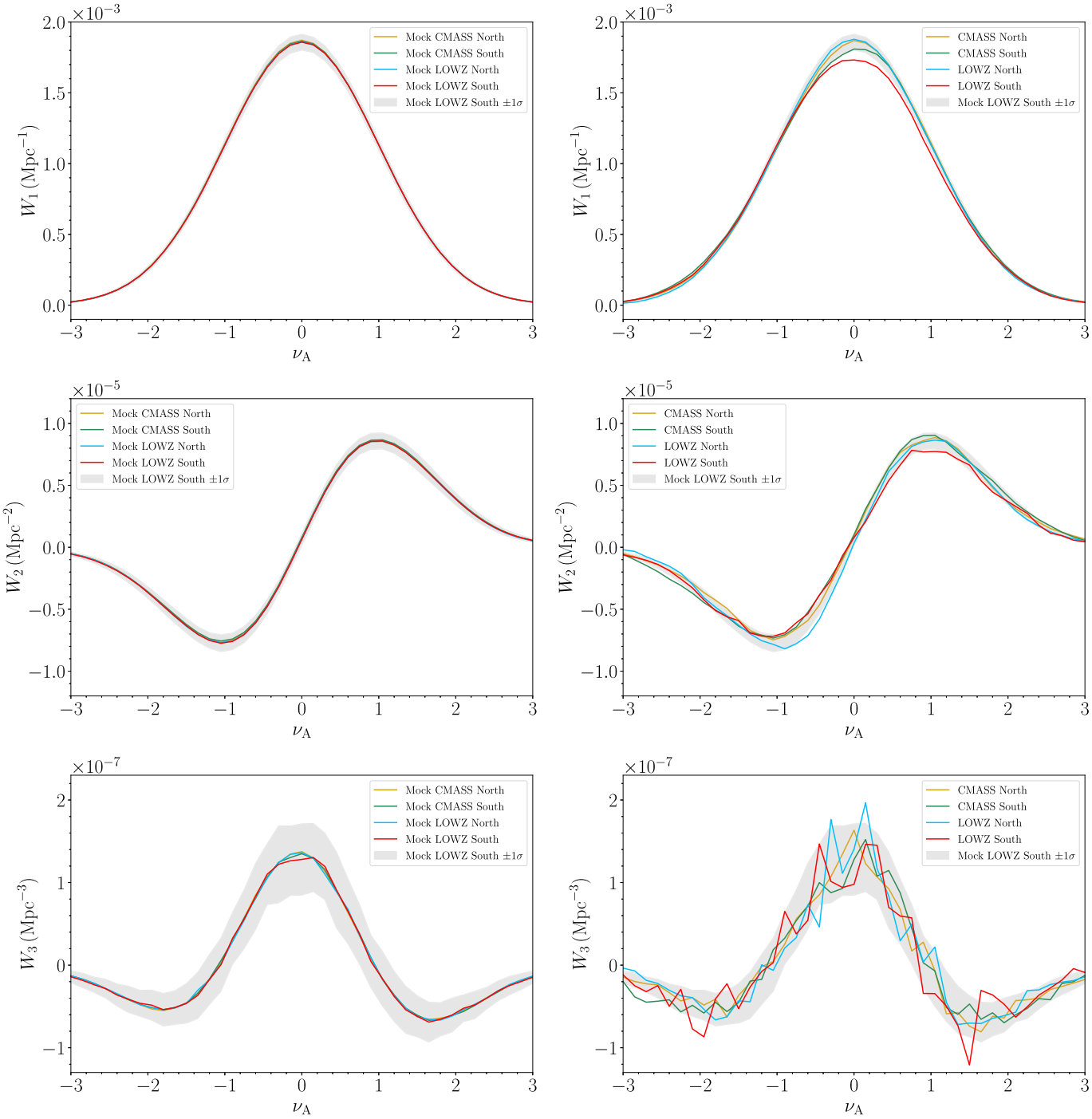

Figure 2. Left panels: the MFs extracted from $N_{\rm rea} = 250$ Patchy mock catalogs as a function of volume fraction threshold $\nu_{\rm A}$. The gray filled region is the $\pm 1\sigma$ variation of $W_{1,2,3}$ (top/middle/bottom panels) from the LOWZ S realizations. The gold/green/blue/red solid lines are the mean values of the MFs extracted from the mock CMASS N/S and LOWZ N/S realizations, respectively. All data sets exhibit close agreement. Right panels: the same statistics, extracted from the BOSS data. The color schemes are the same as in the left panel. The LOWZ S data (red solid line) exhibits lower values than the other three data sets. All data sets have been smoothed with smoothing scale $R_{\rm G} = 35$ Mpc.

receives a correction (Matsubara et al. 2020). Although we do not pursue higher-order terms in this work, it is instructive to introduce a single additional Hermite polynomial coefficient to the fitting procedure, and check if it does not significantly alter our conclusions. We end this section by fitting the following functions to the data:

$$\tilde{W}_1 = A_{1,\rm G} e^{-\nu_{\rm A}^2/2}\left(H_0 + \frac{a_0}{3}H_1 + h_2 H_2\right), \tag{30}$$

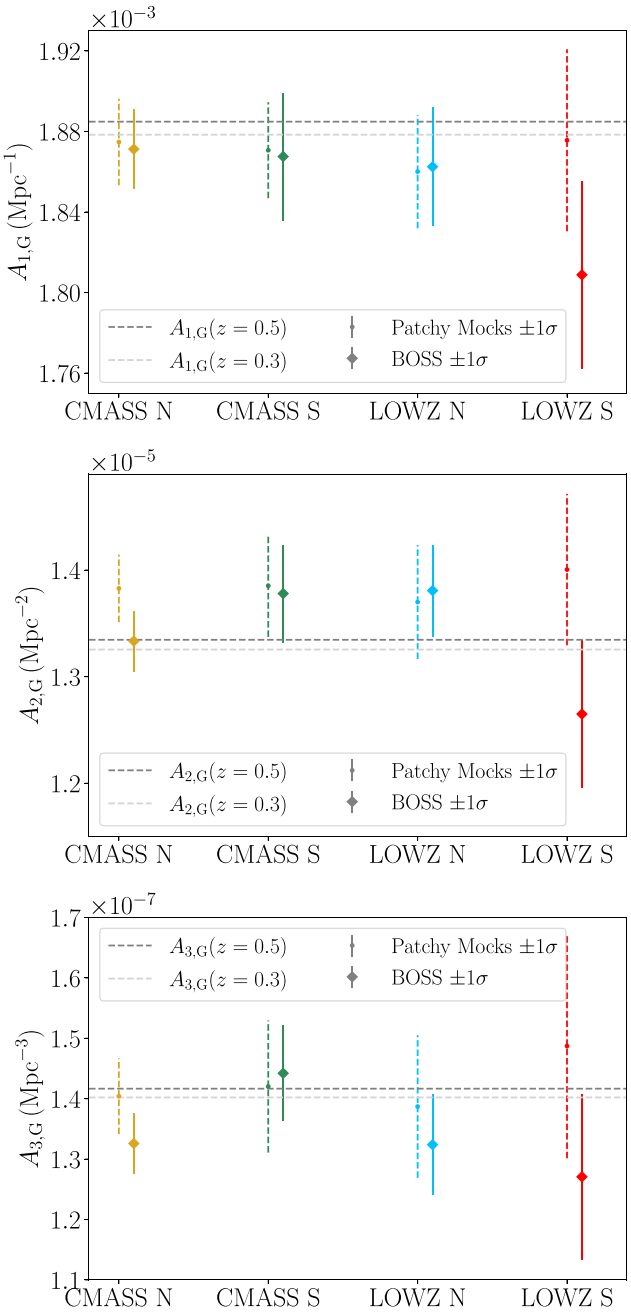

Figure 3. The best fit and 1σ uncertainties on the amplitudes $A_{k,G}$ of the MFs $W_{1,2,3}$ (top/middle/bottom panels) for the four distinct BOSS data sets CMASS N/S and LOWZ N/S (gold/green/blue/red diamonds and solid error bars). The small points and dashed error bars are the mean and rms values of $A_{k,G}$ extracted from the patchy mock catalogs. The dark/light gray dashed lines are the Gaussian expectation values (Equation (8)) based on the patchy mock cosmological parameters and power spectrum at $z = 0.5$ and 0.3, respectively.

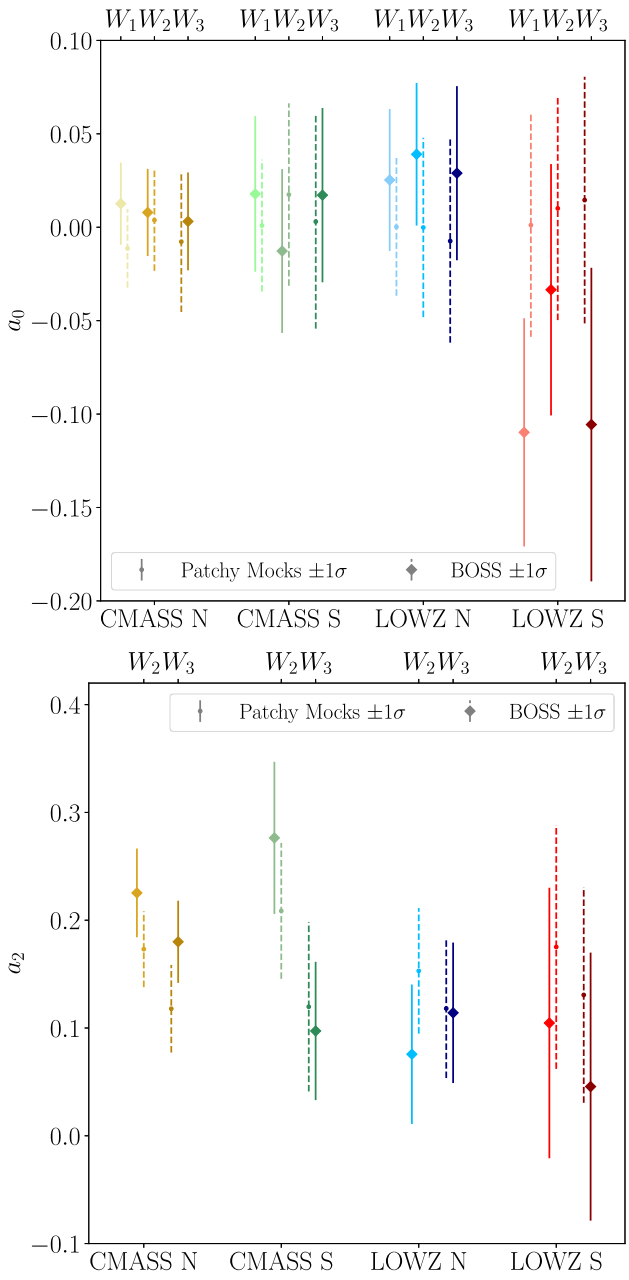

Figure 4. The best fit and 1σ uncertainties on a_0 (top panel) and a_2 (bottom panel) extracted from the BOSS data (diamonds and solid error bars), and the mean and rms value of these quantities from the patchy mocks. Gold/green/blue/red points correspond to CMASS N/S and LOWZ N/S, respectively. The light-to-dark range of colors represent the values of a_0, a_2 extracted from W_1, W_2, W_3, respectively, from the same data set.

$$\tilde{W}_2 = A_{2,G} e^{-\nu_A^2/2}\left(H_1 + \frac{2a_0}{3}H_2 + \frac{a_2}{3}H_0 + h_3 H_3\right), \quad (31)$$

$$\tilde{W}_3 = A_{3,G} e^{-\nu_A^2/2}(H_2 + a_0 H_3 + a_2 H_1 + h_0 H_0), \quad (32)$$

where $h_{0,2,3}$ are additional free parameters. We select these terms as they correspond to coefficients of the lowest-order Hermite polynomials introduced at order σ_0^2 for each MF. Additional higher-order polynomials should also be included, but they require increasing the information from the large $|\nu_A|$ tails to accurately measure.

We again minimize the χ_k^2 functions (Equation (24)), but with the additional parameters free to vary over the range $-1 < h_{0,2,3} < 1$. In Figure 5, we present the marginalized one-dimensional probability distribution functions for h_2, h_3, h_0 (top/middle/bottom panels). The vertical filled bars are the 1σ ranges of these parameters obtained directly from the patchy

Table 3
Marginalized Best Fit and 1σ Uncertainties on the Parameters Fit by Minimizing the χ^2 Function (Equation (24)) for the Data Sets

Data/MF	$\Omega_c h^2\, n_s$	a_0	a_2	χ^2_r
CMASS	0.110 ± 0.006	0.006 ± 0.012	0.183 ± 0.021	1.32
LOWZ	0.111 ± 0.008	0.007 ± 0.020	0.082 ± 0.036	1.07
CMASS N	0.104 ± 0.008	0.008 ± 0.014	0.191 ± 0.023	1.35
CMASS S	0.119 ± 0.010	0.002 ± 0.025	0.158 ± 0.040	1.27
LOWZ N	0.117 ± 0.009	0.032 ± 0.024	0.098 ± 0.042	1.10
LOWZ S	0.095 ± 0.012	-0.080 ± 0.040	0.066 ± 0.073	0.95
CMASS N W_1	0.109 ± 0.010	0.012 ± 0.022	⋯	1.25
CMASS S W_1	0.109 ± 0.015	0.008 ± 0.040	⋯	1.20
LOWZ N W_1	0.109 ± 0.014	0.025 ± 0.037	⋯	1.35
LOWZ S W_1	0.091 ± 0.017	-0.116 ± 0.063	⋯	0.93
CMASS N W_2	0.116 ± 0.010	0.008 ± 0.024	$0.224 \pm 0.042.$	1.31
CMASS S W_2	0.129 ± 0.017	-0.007 ± 0.042	0.284 ± 0.066	1.02
LOWZ N W_2	0.133 ± 0.016	0.038 ± 0.039	0.070 ± 0.065	0.99
LOWZ S W_2	0.100 ± 0.020	-0.025 ± 0.067	0.123 ± 0.121	1.25
CMASS N W_3	0.099 ± 0.010	0.001 ± 0.026	0.177 ± 0.037	1.42
CMASS S W_3	0.122 ± 0.016	0.028 ± 0.046	0.113 ± 0.062	1.18
LOWZ N W_3	0.103 ± 0.016	0.030 ± 0.046	0.114 ± 0.064	0.86
LOWZ S W_3	0.098 ± 0.023	-0.102 ± 0.079	0.043 ± 0.116	0.68
Planck	0.116 ± 0.001	⋯	⋯	⋯

Note. CMASS and LOWZ correspond to an overall fit of the functions (Equation (34)) to the north and south data combined $\chi^2 = \chi^2_1 + \chi^2_2 + \chi^2_3$. CMASS N/S and LOWZ N/S are fits of Equation (34) separately to each of the four data sets. The following twelve rows are fitting (Equation (34)) separately to each data set and each MF. The final row is the Planck best fit and 1σ uncertainty on the parameter combination $\Omega_c h^2\, n_s$, assuming ΛCDM.

mocks, using

$$h_m \simeq \frac{1}{\sqrt{2\pi}\, m! A_{k,\mathrm{G}}} \int_{-4}^{4} H_m(\nu_\mathrm{A})\, W_k(\nu_\mathrm{A})\, d\nu_\mathrm{A}. \tag{33}$$

There is slight evidence that the northern and southern sky data exhibits a dichotomy in h_2 [top panel], although, again, the statistical significance is low. The introduction of $h_{0,2,3}$ does not move the best-fit values of the other parameters $A_{k,G}$, $a_{0,2}$ outside of their 1σ ranges, indicating that our results are stable under the addition of the higher-point cumulants. The parameter values $A_{k,\mathrm{G}}$, $a_{0,2}$ with and without the $h_{0,2,3}$ terms are provided in Table 6 of Appendix D.

4.2. Cosmological Parameter Estimation from the Minkowski Functional Amplitudes

Finally, we repeat our χ^2_k minimization procedure of the previous section, but now fit the function

$$\tilde{W}_k = \frac{\alpha_k}{(2\pi)^{(k+1)/2}} \frac{\omega_3}{\omega_{3-k}\omega_k} \left(\frac{\sigma_1^2}{3\sigma_0^2}\right)^{k/2} e^{-\nu_\mathrm{A}^2/2} \{H_{k-1}(\nu_\mathrm{A}) + \frac{k}{3} a_0 H_k(\nu_\mathrm{A}) + \frac{k(k-1)}{6} a_2 H_{k-2}(\nu_\mathrm{A})\}, \tag{34}$$

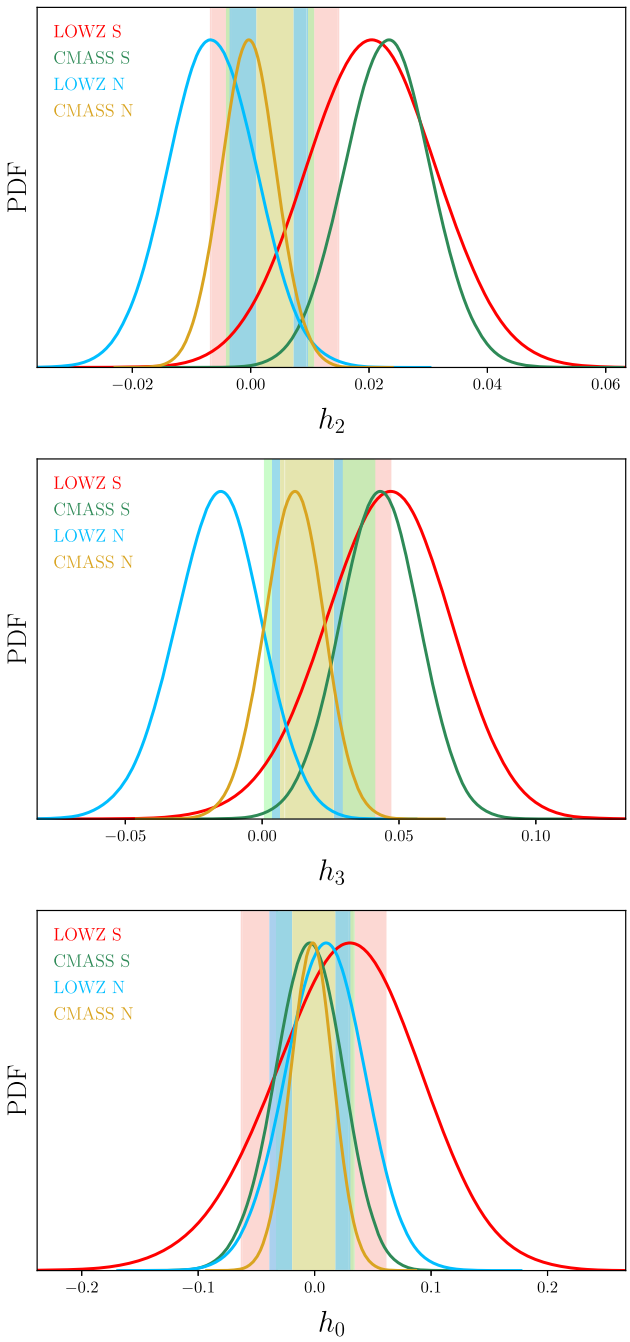

Figure 5. One-dimensional marginalized probability distribution functions for the parameters h_2, h_3,and h_0 (top–bottom panels) for the CMASS N/S and LOWZ N/S data (gold/green/blue/red solid lines, respectively). The vertical matching color bars are the rms values of these parameters extracted from the patchy mock data.

to the MF curves extracted from the BOSS data. This is the same function as Equation (23), but now we fit a cosmological model to the amplitudes rather than treating $A_{k,\mathrm{G}}$ as arbitrary constants. Cosmology enters via the ratio of two-point cumulants σ_1 and σ_0, which is given by

$$\frac{\sigma_1^2}{\sigma_0^2} = \frac{\int k^2 P_g(z,\, k) W^2(kR_\mathrm{G})\, d^3k}{\int P_g(z,\, k) W^2(kR_\mathrm{G})\, d^3k}. \tag{35}$$

Table 4
Prior Parameter Ranges Used in Section 4.2

Parameter	Range
$\Omega_c h^2$	[0.05, 0.30]
n_s	[0.6, 1.2]
a_0	[−1, 1]
a_2	[−1, 1]

We approximate the galaxy power spectrum in real space as $P_g(z, k) = b^2 P_m(z, k) + 1/\bar{n}$, where b is the linear galaxy bias, P_m is the underlying linear matter power spectrum, and $\bar{n}$ is the number density of the galaxy sample being utilized. We fix $b = 2$ based on the mock catalogs, but our results will be practically insensitive to variation of this parameter. This is a valid assumption provided we restrict our analysis to scales at which shot noise is negligible compared to the signal. We fix $z_{\rm LOWZ} = 0.3$ and $z_{\rm CMASS} = 0.5$ in the power spectrum $P_g(z, k)$, but these values will not affect our conclusions. We fix the baryon fraction and σ_8 to their Planck values $\Omega_b h^2 = 0.0224$, $\sigma_8 = 0.8288$, as our statistics are only very weakly sensitive to these parameters.

The quantities in Equations (23), (35) have been defined in real space, but the measured MFs are in redshift space. To account for this discrepancy, we correct the measured MF curves by a constant factor $\alpha_k = [0.99, 0.98, 0.97]$ for W_1, W_2, and W_3, respectively. These correction factors were obtained by measuring the MF statistics in a mock galaxy snapshot box in real and redshift space, and calculating the ratio of their amplitudes. Because the redshift space correction is so small, we do not expect any model dependence in this effect to be significant. This point is discussed further in Appendix C.

In total, for each MF curve in each data set, we vary four parameters $\Omega_c h^2$, n_s, a_0, and a_2 over the prior ranges given in Table 4. We again perform the χ^2_k minimization for each of the four data sets separately, and each MF separately (for a total of twelve sets of parameter constraints). We also combine the information from W_1, W_2, W_3 for each data set, by summing their χ^2_k values, to obtain four distinct measurements labeled CMASS N, CMASS S, LOWZ N, and LOWZ S. Finally, we combine the north and south results, by simply summing their chi-squared values, to obtain overall CMASS and LOWZ results.

In Figure 6, we present the two-dimensional contours for the parameters $\Omega_c h^2$, n_s, a_0, and a_2 for CMASS N/S and LOWZ N/S, (gold, green, blue, red filled contours), and the combined CMASS and LOWZ results (dashed/solid black empty contours). The $a_{0,2}$ parameters are orthogonal to $\Omega_c h^2$ and n_s, due to the orthogonal nature of the Hermite polynomials. Therefore, we do not present a_0,a_2–$\Omega_c h^2$,n_s contours, as they are not informative.

There exists a strong degeneracy between $\Omega_c h^2$ and n_s; this was also observed in two-dimensional slices of the BOSS data in Appleby et al. (2020). Since we cannot simultaneously constrain these two parameters, we rotate the parameter plane and obtain an effective one-dimensional constraint on the $\Omega_c h^2\, n_s$ combination.[15] The LOWZ S data presents a lower value of both a_0 and $\Omega_c h^2\, n_s$, but the uncertainties are large due to this data set occupying the smallest volume. In Table 3, we present the one-dimensional marginalized best fit and 1σ uncertainties on $\Omega_c h^2\, n_s$, a_0, and a_2 for each data set, and the corresponding reduced chi-squared values. We also present the Planck best fit of the combination of parameters $\Omega_c h^2\, n_s$ (Aghanim et al. 2020). Note that the combined CMASS N/S and LOWZ N/S data are consistent with the Planck cosmology, despite the LOWZ S data being systematically low. This is because the north data simply possesses more constraining power and is closer to the Planck cosmology. The difference in a_0 between LOWZ N and S ($a_0 = 0.032 \pm 0.024$ and -0.080 ± 0.040) is the most significant discrepancy observed in this work, and brings into question the suitability of the expansion (Equation (23)) for the low-redshift galaxy data. The cosmological parameters inferred from LOWZ N and S are practically consistent ($\Omega_c h^2\, n_s = 0.117 \pm 0.009$ and 0.095 ± 0.012).

5. Discussion

The MFs provide a complementary approach to extracting information from cosmological data sets. In this work, we have measured the MFs from the SDSS-III DR12 BOSS galaxy data. To do so, we binned the point distribution onto a uniform lattice and Gaussian smoothed the discrete field with comoving scale $R_G = 35$ Mpc. At these large scales, the perturbative non-Gaussian expansion (Equation (14)) can be used in principle.

After validating our analysis with GRFs, and mock galaxy snapshot data, we measured the amplitude and shape of the MF curves obtained from the BOSS data. The resulting analysis yielded some quirks. Specifically, the LOWZ-South data possesses systematically low MF amplitudes compared to LOWZ north, and also low values of the shape parameter a_0, which contains information from the three-point cumulants. The exact values can be found in Table 3. The significance of these discrepancies is not high, because we must smooth over relatively large scales to reconcile late universe measurements of the large-scale structure with the perturbative non-Gaussian expansion typically used in cosmology (Matsubara et al. 2020). However, the presence of such anomalies could indicate either some unknown systematics in the data or some physical anisotropy in the low-redshift large-scale structure. The higher-redshift CMASS data does not present any discrepancy between northern and southern sky data. Similarly, the patchy mock data is remarkably consistent between CMASS/LOWZ north/south subsamples, which suggests that the problem does not lie with our analysis pipeline. The LOWZ data considered in this work lies at cosmological distances $d_c \sim 850$–1600 Mpc relative to the observer, scales at which we expect the data to be close to isotropic and homogeneous within the standard cosmological model. Measurements of the MFs from earlier large-scale structure catalogs support our findings (Kerscher et al. 1997, 1998, 2001); the north and south sky data consistently present different morphological properties.

If we overlook the north/south discrepancy and simply combine the data sets into two overall catalogs (CMASS and LOWZ), we find that the amplitude of the MFs are consistent with the Planck ΛCDM best fit. This is primarily due to the northern sky data simply occupying a larger volume, and the curious southern sky results are mitigated. Also, we are only using the two-point information contained within the

[15] This is a different parameter combination than the two-dimensional results of Appleby et al. (2020); this is due to the different parameter sensitivity in the two- and three-dimensional statistics.

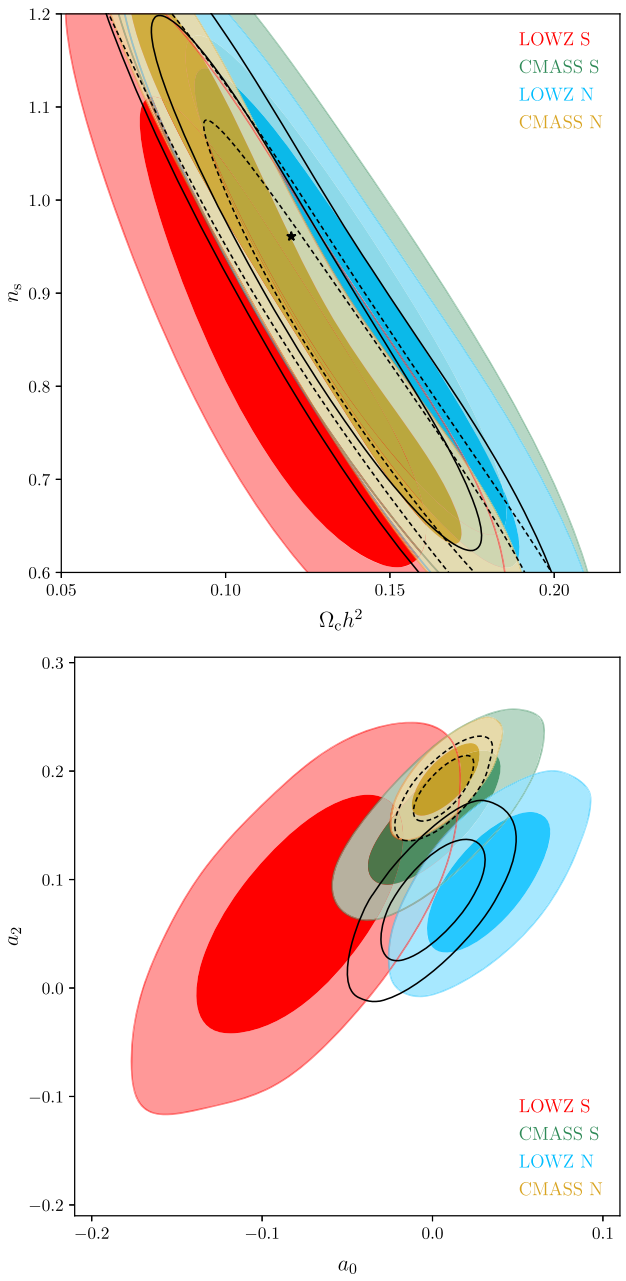

Figure 6. Marginalized two-dimensional contours in the n_s–$\Omega_c h^2$ and a_2–a_0 planes. The gold/green/blue/red contours are the 68%–95% limits obtained from the CMASS N/S and LOWZ N/S data, respectively. The unfilled black dashed/solid lines are the contours obtained by combining north and south data into two catalogs CMASS and LOWZ, respectively. The black star in the top panel is the Planck best fit of these parameters, assuming ΛCDM.

amplitudes, and the two-point function is not sensitive to the structure of the cosmic web.

The BOSS data has been exhaustively studied in the literature (Manera et al. 2015; Ross et al. 2017; Slepian et al. 2017; Zhai et al. 2017; Colas et al. 2020; D'Amico et al. 2020; Hamaus et al. 2020; Ivanov et al. 2020). A direct comparison to our work and two-point correlation function and power spectrum analyses is difficult, because different parameters are varied, sampling choices are made, and typically north and south data are not separately analyzed. However, the literature consensus is that the BOSS data is consistent with the best-fit Planck ΛCDM cosmology, when the two-point information is extracted.[16] We are in agreement with this conclusion. A recent, comprehensive analysis of the BOSS power spectrum in Ivanov et al. (2020) generated a constraint of $\Omega_c h^2 = 0.1127 \pm 0.0046$. This is consistent (both the best fit and approximately the statistical uncertainty) with our measurements if we fix $n_s = 0.965 \pm 0.004$ using a Planck prior. Similarly, our results are consistent with a previous analysis of the two-dimensional genus extracted from shells of the BOSS data in Appleby et al. (2020). We provide a comparison of these results in Table 5.

Table 5
A Review of the Constraints on $\Omega_c h^2$ in This Work (After Applying a Planck Prior on n_s), and Corresponding Measurements of the Same Parameter in Other Works in the Literature

Measurement	$\Omega_c h^2$
CMASS (This work)	0.114 ± 0.005
LOWZ (This work)	0.114 ± 0.007
Ivanov et al. (2020)	0.113 ± 0.005
CMASS (Appleby et al. 2020)	0.121 ± 0.006
LOWZ (Appleby et al. 2020)	0.116 ± 0.008

Regarding the north/south comparison, Tojeiro et al. (2014) noted some tension between north and south sky data in previous versions of the BOSS data, but the difference was not deemed significant. In this work, the discrepancy between the north and south sky is present predominantly in the non-Gaussian, higher-point cumulants to which the MFs are sensitive. The effect is modest, and future galaxy catalogs will provide more information on the non-Gaussian nature of the late-time gravitational field. In Sullivan et al. (2019), the MFs of the BOSS data were extracted using a germ-grain method, and the non-Gaussian properties were rigorously studied. The authors of Sullivan et al. (2019) concluded that there is no strong statistical evidence of any north/south discrepancy in the DR12 data. In Figure 6 and Table 3 of that work, the LOWZ data presents a small offset between the north and south sky, which is most evident in the W_1 MF. The significance is low—the quoted p-value associated with the hypothesis that the difference is consistent with random fluctuations is $p \simeq 0.02$, which is approximately in agreement with our $\gtrsim 2\sigma$ discrepancy in a_0 from the LOWZ data. In contrast, the CMASS data is fully consistent. The result is suggestive rather than conclusive—given that other large-scale structure data sets have presented north/south discrepancies (Kerscher et al. 1997, 1998, 2001), it would be interesting to study the low-redshift galaxy density field in more detail.

There is increasing discussion in the literature on the potential existence of a dipole in various data sets (Colin et al. 2019; Mohayaee et al. 2020; Luongo et al. 2021; Secrest et al. 2021), beyond the kinematic dipole observed in the CMB (Aghanim et al. 2014). A related observational framework to measure multipoles in low-redshift data can be found in Heinesen (2021). The BOSS data is not a magnitude-limited sample, and its complex selection criteria and incomplete sky coverage make it difficult to relate our findings to other claims

[16] With the caveat that the MF amplitudes are not sensitive to σ_8, only the shape of the power spectrum.

in the literature. A study of the topology of all-sky density fields is an interesting direction of future study.

At fixed comoving smoothing scales $R_G = 35$ Mpc, the majority of information contained in the late universe density field is washed out. Also, the MFs themselves are *summary statistics* and do not contain all topological information. To proceed further, we should unmoor ourselves from the model-dependent non-Gaussian perturbative expansion in σ_0 cumulants (see footnote 2), and also consider the more complex class of topological statistics that can be applied to a point distribution. For point processes, a direct MF analysis using the decoration of galaxies with Boolean grains without constructing a density field, hence without extra smoothing, provides an alternative strategy (Mecke et al. 1994; Kerscher et al. 1997, 1998, 2001; Wiegand et al. 2014). This methodology naturally contains boundary corrections according to the GKF and is model-independent by construction. Such an analysis is currently being pursued by the authors to further determine the properties of the observed large-scale structure.

S.A. is supported by an appointment to the Junior Research Groups (JRG) Program at the APCTP through the Science and Technology Promotion Fund and Lottery Fund of the Korean Government, and was also supported by the Korean Local Governments in Gyeongsangbuk-do Province and Pohang City. This work is also part of a project that has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement ERC adG No. 740021ARThUs, PI: T.B.). S.E.H. was supported by the project 우주거대구조를이용한암흑우주연구 ("Understanding Dark Universe Using Large Scale Structure of the Universe"), funded by the Ministry of Science. H.S.H. acknowledges the support by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT; No. 2021R1A2C1094577).

The authors would like to acknowledge the support of the Korea Institute for Advanced Study (KIAS) grant funded by the government of Korea. Computing resources were supplied by the KIAS Center for Advanced Computation Linux Cluster System.

Funding for SDSS-III has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, and the U.S. Department of Energy Office of Science. The SDSS-III website is http://www.sdss3.org/. SDSS-III is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS-III Collaboration including the University of Arizona, the Brazilian Participation Group, Brookhaven National Laboratory, Carnegie Mellon University, University of Florida, the French Participation Group, the German Participation Group, Harvard University, the Instituto de Astrofisica de Canarias, the Michigan State/Notre Dame/JINA Participation Group, Johns Hopkins University, Lawrence Berkeley National Laboratory, Max Planck Institute for Astrophysics, Max Planck Institute for Extraterrestrial Physics, New Mexico State University, New York University, Ohio State University, Pennsylvania State University, University of Portsmouth, Princeton University, the Spanish Participation Group, University of Tokyo, University of Utah, Vanderbilt University, University of Virginia, University of Washington, and Yale University.

The massive production of all MultiDark-Patchy mocks for the BOSS Final Data Release has been performed at the Barcelona Supercomputing Center Marenostrum supercomputer, the Hydra cluster at the Instituto de Fsica Teorica UAM/CSIC, and NERSC at the Lawrence Berkeley National Laboratory. We acknowledge support from the Spanish MICINNs Consolider-Ingenio 2010 Programme under grant MultiDark CSD2009-00064, MINECO Centro de Excelencia Severo Ochoa Programme under grant SEV-2012-0249, and grant AYA2014-60641-C2-1-P. The MultiDark-Patchy mocks was an effort led from the Institute of Theoretical Physics (IFT) UAM-CSIC by F. Pradas group (C.-H. Chuang, S. Rodriguez-Torres, and C. Scoccola) in collaboration with C. Zhao (Tsinghua U.), F.-S. Kitaura (AIP), A. Klypin (NMSU), G. Yepes (UAM), and the BOSS galaxy clustering working group.

Some of the results in this paper have been derived using the healpy and HEALPix package.

Appendix A
Geometry of Random Fields on Manifolds

Going under various names and orderings in different settings, such as MFs, curvature integrals, intrinsic volumes, and Lipschitz-Killing curvatures, there are $(D+1)$ quantifiers associated with the geometry of a D-dimensional manifold $\mathbb{M}$. Taking V_D to be the D-dimensional Lebesgue measure, which quantifies the D-dimensional volume, and for convex $\mathbb{M}$, there exist a set of numbers $W_0,\ldots,W_D$ known as *Minkowski functionals*, which are associated with the volume of the *tube* of radius ϵ around $\mathbb{M}$, where ϵ is small, through the *tube formula*:

$$V_D(Tube(\mathbb{M}, \epsilon)) = \sum_{j=0}^{D} \frac{\epsilon^j}{j!} W_j(\mathbb{M}). \tag{A1}$$

Restricting to three dimensions, W_0 measures the volume, W_1 measures the surface area, and W_2 is associated with contour length and measures the caliper diameter of $\mathbb{M}$. W_3, or equivalently W_D of a D-dimensional manifold, is associated with a purely topological quantity called the *Euler characteristic*.

The MFs of the excursion sets of *stochastic* fields on manifolds are defined in the usual sense of the tube formula in Equation (A1), with the exception that the Lebesgue measure is replaced by the probability measure, such that all measures of size are weighted with respect to probability content, giving the probabilistic version of the tube formula

$$\Pr\left\{X \in \left\{x \in \mathbb{R}^d : \min_{y \in E_\nu} \|y - x\|\right\} \leqslant \epsilon\right\} = \sum_{j=0}^{\infty} \frac{\epsilon^j}{j!} \mathcal{M}_j^d(E_\nu). \tag{A2}$$

The above equation is a Taylor series expansion, in which the coefficients $\mathcal{M}_j^d$ are known as the *Gaussian Minkowski functionals*; they play the role of the usual MFs and encode the geometric properties of the manifold induced by the random field u.

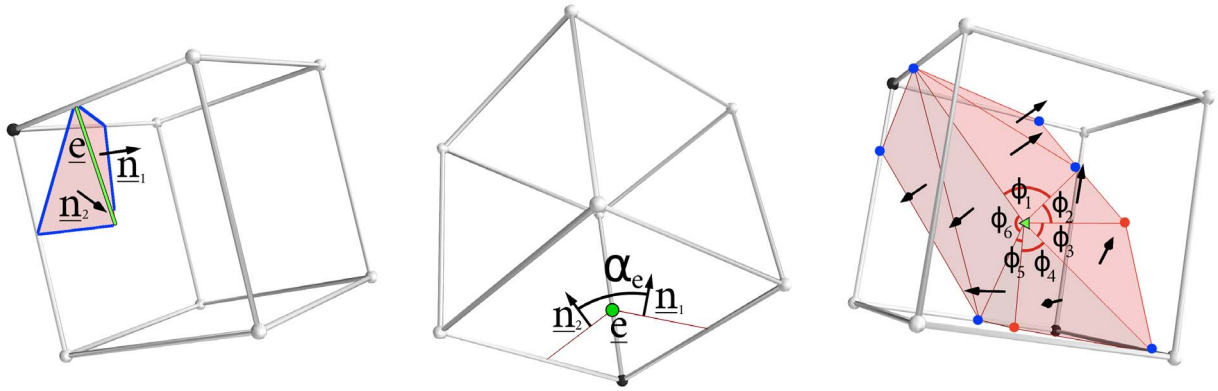

Figure 7. Two examples of pixel boxes that highlight our numerical algorithm. The black/white points are pixels in the discrete field lattice δ_{ijk} that are inside/outside the excursion set. The red triangles are the triangulated mesh of constant field value $\delta = \nu$ that our algorithm generates. The normal vectors, edges, and vertices of triangles are presented in the left and right panels. The quantity α_e, used in the extraction of W_2, is presented in the middle panel. The middle panel is the same box as the left panel, rotated such that the line of sight is parallel to the green edge. The images presented here are modified versions of Figures 8, 9 in Appleby et al. (2018b). © AAS. Reproduced with permission.

Restricting to cubical Euclidean grids, the MFs of the excursion sets are given via the *Gaussian Kinematic Formula*:

$$\langle W_i(E_\nu)\rangle = e^{-\nu^2/2}\sum_{j=0}^{i}\begin{bmatrix} D+j-i \\ j \end{bmatrix}\begin{bmatrix} i \\ j \end{bmatrix} \times \frac{\omega_j j!\lambda^j}{(2\pi)^{(j+1)/2}} H_{j-1}(\nu) W_{i-j}(\mathbb{M}). \tag{A3}$$

In the above equation, λ is proportional to the second spectral moment of the power spectrum, or equivalently, proportional to the second-order gradient of the correlation function. The combinatorial *flag coefficients* are defined by

$$\begin{bmatrix} n \\ j \end{bmatrix} = \binom{n}{j}\frac{\omega_n}{\omega_{n-j}\,\omega_j}, \tag{A4}$$

where H_n is the nth Hermite polynomial, defined for $n \geqslant 0$,

$$H_n(x) = n!\sum_{j=0}^{\lfloor n/2\rfloor}\frac{(-1)^j x^{n-2j}}{j!\,(n-2j)!\,2^j};$$

while, for $n = -1$, we fix

$$H_{-1}(x) = \sqrt{2\pi}\,e^{x^2/2}\Psi(x), \tag{A5}$$

where

$$\Psi(x) = \frac{1}{\sqrt{2\pi}}\int_u^\infty e^{-x^2/2}\,dx, \tag{A6}$$

is the Gaussian tail probability.

The GKF describes the MFs for a field defined on a generic manifold, in the presence of boundaries or otherwise. It is equal to the curvature integrals (Equations (2)–(5)) only when the field is boundaryless. Many works in the cosmological literature (including this one) actually extract the curvature integrals from data sets, such as Appleby et al. (2020), Schmalzing & Gorski (1998), rather than the MFs Pranav (2021).

Appendix B
Unbiased Estimators of the Curvature Integrals

For large-scale structure catalogs, one must account for radial and angular selection functions, masks, and complex survey geometries. Given that the ensemble averages quoted in Section 2 apply only to the unbounded fields, we must carefully construct unbiased estimators for these statistics when the data is masked. In this section, we review our numerical algorithm and then test our method by applying it to GRFs and mock galaxy snapshot data.

B.1. Numerical Reconstruction of Minkowski Functionals

In Appleby et al. (2018b), we provided a detailed description on an algorithm to extract the MFs from a discretized field on a uniform lattice. To briefly review, the method requires a set of field values u_{ijk} on a uniform lattice, where i, j, k subscripts denote pixel identifiers in the $x_{1,2,3}$ directions, respectively. We then form *pixel boxes* from eight adjacent pixels (u_{ijk}, $u_{i,j,k+1}$, $u_{i,j+1,k}$, $u_{i,j+1,k+1}$, $u_{i+1,j,k}$, $u_{i+1,j,k+1}$, $u_{i+1,j+1,k}$, $u_{i+1,j+1,k+1}$). Decomposing each individual pixel box into six nonoverlapping tetrahedra, we linearly interpolate along edges of the tetrahedra to find points at which $u = \nu$, where ν is some constant field value that we select. We then generate a triangulated surface mesh of constant $u = \nu$ from these points. This defines the excursion set boundary $\partial\mathbb{M}$ as a triangulated mesh. Finally, we extract the MFs from the triangulated boundary according to

$$W_0 = \frac{1}{V}\sum_{i\in N} V_{t,i}, \tag{B1}$$

$$W_1 = \frac{1}{6V}\sum_{i\in N} A_{t,i}, \tag{B2}$$

$$W_2 = \frac{1}{6\pi V}\sum_e |\boldsymbol{e}|\alpha_e, \tag{B3}$$

$$W_3 = \frac{1}{4\pi^2 V}\sum_v\left(1 - \frac{1}{2\pi}\sum_{T\in v}\phi_T^v\right), \tag{B4}$$

where V is the total volume occupied by the field; i, e, and v in the summation denote the pixel box's *unique* edge, and *unique* triangle vertex in the triangulated surface mesh, respectively. $V_{t,i}$ is the volume contained within the ith pixel box that is enclosed by the triangulated mesh, and $A_{t,i}$ is the total area of the triangulated mesh within the ith pixel box. $|\boldsymbol{e}|$ is the length of the edge $\boldsymbol{e}$, and α_e is the angle subtended by the normal's of the two triangles that share the edge $\boldsymbol{e}$. Finally, $\sum_{T\in v}\phi_T^v$ is the

sum of internal angles of all triangles T that share the common vertex v. The quantities α_e and ϕ_T^v are presented pictorially in Figure 7, which is reproduced from Appleby et al. (2018b).

This methodology was applied to "complete" fields with no masked regions and periodic boundary conditions in Appleby et al. (2018b). We now highlight the modifications required to reconstruct the same statistics from a restricted field. As before, we define a field on some regular three-dimensional lattice u_{ijk}, but now some of the domain is masked. We assign all masked pixels a particular *bad* value u_b ; $u_{ijk} = u_b$. We restrict our analysis only to pixel boxes for which all eight vertices (u_{ijk}, $u_{i,j,k+1}$, $u_{i,j+1,k}$, $u_{i,j+1,k+1}$, $u_{i+1,j,k}$, $u_{i+1,j,k+1}$, $u_{i+1,j+1,k}$, $u_{i+1,j+1,k+1}$) are not masked (that is, not assigned value u_b). We call such pixels boxes as unmasked. We denote the total number of pixel boxes in the entire volume V as N and the total number of unmasked pixel boxes used in our analysis as $\ell \leqslant N$. The total volume of the domain is $V = N\Delta_g^3$, and the corresponding masked volume is $V_\ell = \ell\Delta_g^3$, where Δ_g^3 is the volume of a single pixel box (the resolution along each $x_{1,2,3}$ dimension is Δ_g).

For each unmasked pixel box, we perform the standard marching tetrahedron algorithm, generate six tetrahedra, interpolate along their edges to points at which $u = \nu$, and construct a triangulated mesh internal to this particular pixel box. From this, we can calculate the triangulated surface area of isofield value $u = \nu$, and also the fractional volume enclosed by this triangulated surface. Hence the volume and surface area of the excursion set can be estimated locally within each pixel box. Our estimates of volume of the excursion set W_0 and surface area of its boundary W_1, *per unit volume*, are therefore given by

$$W_0 = \frac{1}{V_\ell}\sum_{i\in\ell} V_{t,i}, \tag{B5}$$

$$W_1 = \frac{1}{6V_\ell}\sum_{i\in\ell} A_{t,i}, \tag{B6}$$

where the sums are over all unmasked pixel boxes $i \in \ell$, $A_{t,i}$ is the total area of all triangles constructed within the ith pixel box, and $V_{t,i}$ is the volume enclosed by the triangulated mesh in the ith box. $A_{t,i}$ and $V_{t,i}$ can be calculated using trigonometry from the tetrahedral decomposition.

The remaining MFs—W_2, W_3—are also local quantities and can be estimated from a masked subset of data. However, unlike W_0 and W_1, they require information from adjacent boxes as they are determined by triangles in the surface mesh that share common edges and vertices. Each triangle edge can be shared by a maximum of two adjacent pixel boxes, and triangle vertices can be shared by a maximum of four adjacent pixel boxes. To estimate W_2, we only consider pixel boxes that are at least two pixels away from any mask or boundary, to ensure that all edges counted in the W_2 reconstruction have two matching triangles. This is necessary to construct α_e in Equation (B3). The estimator is simply

$$W_2 = \frac{1}{6\pi V_{\ell'}}\sum_{e'} |\boldsymbol{e}'|\alpha_{e'}, \tag{B7}$$

where ℓ' identifies the set of all pixel boxes at least two pixels from the boundary ($\ell \neq \ell'$), $\sum_{e'}$ represents the sum of all triangle edges within this subset of pixel boxes, and $V_{\ell'} = \ell'\Delta_g^3$.

For W_3, we adopt the following modified estimator

$$W_3 = \frac{1}{4\pi^2 V_\ell}\sum_{\tilde{v}}\left(w_{\tilde{v}} - \frac{1}{2\pi}\sum_{T\in\tilde{v}}\phi_T^v\right), \tag{B8}$$

where now the sum is over $\tilde{v}$, which is all triangle vertices extracted using the marching tetrahedral algorithm, which counts vertices multiple times. For example, if a triangle vertex is generated on the surface/edge of a pixel box, then it will be counted two/four times in the $\tilde{v}$ sum, respectively, because the marching tetrahedron algorithm will extract it from two/four pixel boxes. If a triangle vertex is generated internally to a pixel box, then it will be counted once. To remove the multiple counting, we weight each triangle vertex in the reconstruction by $w_{\tilde{v}}$, where $w_{\tilde{v}} = 1$, $1/2$, and $1/4$ depending on the vertex being internal, on the surface, or on the edge of a pixel box, respectively. If we are considering a field without any boundary, then Equation (B8) is equivalent to Equation (B4). In the presence of a mask, a triangle vertex may not contribute a total of unity to the first term in the sum (Equation (B8)), because the algorithm now skips masked boxes. However, we still obtain an unbiased reconstruction of W_3, because the sum over $\phi_T^{\tilde{v}}$ in Equation (B8) only includes the triangle angles in the unmasked pixel boxes.

We present an example of two pixel boxes used in our analysis in Figure 7. The solid black/white points denote pixels in the lattice that are in/out of the excursion set (that is, they have values such that $\delta_{ijk} > \nu$ and $\delta_{ijk} < \nu$, respectively). The pixel boxes are decomposed into six nonoverlapping tetrahedra as described in Appleby et al. (2018b), and a triangulated mesh of $\delta = \nu$ is generated (red triangles in the figure). α_e is presented in the middle panel, which is the pixel box in the left panel rotated to align with the green triangle edge. In the left panel, the green triangle edge is completely internal to the pixel box, and hence both triangles incident to it are internal. This means that α_e for the green edge can be obtained within the box, from the normal vectors n_1 and n_2. On the contrary, the blue triangle edges lie on the surfaces of the pixel box, and each require a triangle in adjacent boxes to define their corresponding α_e. For this reason, this pixel box will only be used to calculate W_2 if all adjacent pixel boxes also contain no bad pixels.

In the left panel, the volume enclosed is the volume occupied between the red triangles and the black pixel and the surface area is the total area of the triangles. These are the contributions to W_0 and W_1 from this particular pixel box.

In the right panel of Figure 7, we present a different pixel box. The triangulated surface $\delta = \nu$ is presented as a set of red triangles, and the triangle vertices are colored green/blue/red. The green vertex in the center is completely internal to the pixel box, and hence will contribute $w = 1$ to Equation (B8), and all incident triangles are present. The red dots are triangle vertices on the surface of the pixel box, and will contribute $w = 1/2$ when this particular box is encountered in the algorithm. The blue dots lie on the edges of the box and will contribute $w = 1/4$, as they are potentially shared by four other boxes. All triangle internal angles in the Figure are counted in the $\sum\phi$ term in Equation (B8).

Our methodology can be used to extract the local properties of a surface per unit volume. To perform this numerical calculation, we do not need to sample the entire data domain;

hence the presence of a mask is practically irrelevant. Due to the local nature of the curvature integrals, they can be extracted by sampling a subset of the surface, and hence our numerical algorithm can provide an unbiased estimate of the Edgeworth expansion of W_k. Only local statistics can be extracted in an unbiased manner using our methodology—the average curvature per unit volume, for example. Global properties such as topology cannot be extracted using this approach.

Next, we verify that our estimators provide an unbiased estimate of the curvature integrals on an unbounded domain, by applying them to masked GRFs and mock galaxy snapshot boxes. We also address the separate issue of smoothing masked fields.

B.2. Gaussian Random Fields

To confirm that the estimators described above can be used to reconstruct the underlying curvature integrals, we generate mock data. Initially, we take realizations of a GRF in a periodic box. We draw the random fields from a linear ΛCDM matter power spectrum with parameters given in Table 1, in a periodic box of volume $V = (3.15\,\mathrm{Gpc})^3$. We adopt a resolution of $\Delta = 6\,\mathrm{Mpc}$ and smooth the field with a Gaussian kernel of scale $R_\mathrm{G} = 35\,\mathrm{Mpc}$. We denote the unmasked, smoothed field δ_{ijk}. We then mask the field. First, we set all vertices within distance R_G of the boundary of the (periodic) volume to $\delta_{ijk} = \delta_\mathrm{b}$, where δ_b is some arbitrary *bad pixel* value. We then generate 200 cylinders through the box in the x_3 direction, of radius r_n and center $x_{1,n}$, $x_{2,n}$, where $1 \leqslant n \leqslant 200$. The values of $0 < r_n < 200\,\mathrm{Mpc}$ and $0 < x_{1,n} < 3150\,\mathrm{Mpc}$, $0 < x_{2,n} < 3150\,\mathrm{Mpc}$ are randomly generated from a uniform distribution. Any δ_{ijk} vertex within the cylinders is also assigned a *bad pixel* value $\delta_{ijk} = \delta_\mathrm{b}$. This simple mask is representative of an angular mask on the sky, which generates cylinders through data in the distant observer limit (more precisely, cones but we do not pursue this distinction here). We measure the MFs of the unmasked GRFs and the masked equivalents.

In Figure 8, we present the mean and standard deviation of the MFs for $N_\mathrm{real} = 50$ GRF realizations. The dark gray dashed lines (labeled "Mask I") represent the mean of the masked fields extracted using the algorithms described in Section B, and the solid black lines are the corresponding MFs extracted from the full, unmasked data. The gray solid region is the rms fluctuations of the statistics from the masked realizations. We observe no systematic deviation between the bounded and unbounded domains, and the solid and dashed lines practically overlap.

Following the main body of the text, we then extract the amplitudes $A_{k,\mathrm{G}}$ and $a_{0,2}$ from each realization using Equations (27)–(29). In Figure 9, we present the mean and rms values obtained from the unmasked/masked data (black/dark gray diamonds and error bars, respectively, labeled "No Mask" and "Mask I"). In the top panel, we also present the Edgeworth expansion, Gaussian expectation values of $A_{k,\mathrm{G}}$ from the expression (Equation (8)) as brown dashed horizontal lines. In all instances, the masked data is consistent with the unmasked equivalents, and consistent with the ensemble expectation value. The $a_{0,2}$ parameters should be consistent with zero for a Gaussian field, and this expectation is recovered in our analysis. There is a $\lesssim 1\%$ systematic discrepancy in a_0 extracted from the W_3 curve (lower left panel, right-hand side); this is due to our method of extracting this parameter. For W_3, a_0 is the coefficient of the cubic Hermite polynomial H_3, which has a relatively large tail in the high $|\nu_A|$ regime; whereas we truncate the integrals in Equations (27)–(29) at $|\nu_A| = 4$.

B.3. Mock Galaxy Catalogs

A GRF is a special example in the sense that all information is contained in the underlying power spectrum. In terms of the MFs, all information in W_k is contained within the H_{k-1} Hermite polynomial coefficient. The matter density field in the late universe is not well described by a GRF, even when smoothing on large scales $R_\mathrm{G} \sim 35\,\mathrm{Mpc}$. We now check that our estimators are also unbiased for gravitationally evolved, nonlinear matter fields.

To this end, we repeat our test on a mock galaxy snapshot box, gravitationally evolved to $z = 0$. Specifically, we use Horizon Run 4 (HR4)—a cosmological N-body simulation containing $N = 6300^3$ particles in a volume of $V = (3150\,\mathrm{Mpc}/h)^3$. The simulation uses a modified GOTPM code.[17] The cosmological parameters used are $h = 0.72$, $n_\mathrm{s} = 0.96$, $\Omega_\mathrm{m} = 0.26$, $\Omega_\mathrm{b} = 0.048$. Details of the numerical implementation and the method by which mock galaxies are constructed can be found in Hong et al. (2016). The mock galaxies are defined using the most bound halo particle galaxy correspondence scheme, and the survival time of satellite galaxies post merger is estimated via a modification of the merger timescale model described in Jiang et al. (2008).

We use the mock galaxy snapshot box at $z = 0$, making no redshift space corrections to the galaxy positions. We bin the galaxies into a regular lattice and generate a mean-subtracted number density field $(n_{ijk} - \bar{n})/\bar{n}$, where n_{ijk} is the number of galaxies in (i, j, k)-pixel and $\bar{n}$ is the mean number of galaxies in all pixel boxes. We smooth the field with Gaussian kernel, and then mask the data using the same procedure as for the GRFs. Defining the smoothed field as $\tilde{\delta}_{ijk}$, we finally define a mean-subtracted, unit variance field $\delta_{ijk} = (\tilde{\delta}_{ijk} - \mu)/\sigma$, where μ, σ are the mean/rms of $\tilde{\delta}_{ijk}$. Despite only having a single realization of the data, we generate subsamples by repeating the masking procedure over $N_\mathrm{real} = 50$ realizations, each time drawing a different set of masked cylinder positions and radii. Each time, we measure the MF curves, and resulting Hermite polynomial coefficients. Although the resulting error bars will underrepresent the *true* statistical uncertainty because the samples are not independent, we can use the mean values to confirm that the masked MFs are unbiased with respect to the MFs of the full unbounded field. We present the results in Figure 10. The dark solid curves are the MFs obtained from the full, unmasked box. The dark violet solid region is the 1σ uncertainty of the statistics over the $N_\mathrm{real} = 50$ resamplings. We observe no statistically significant bias in our reconstruction of the full data from the masked subsets, indicating that our methodology can be used to extract the *Edgeworth expansion* theoretical prediction from the masked data. We also present the parameters $A_{k,\mathrm{G}}$ and $a_{0,2}$ in Figure 11 (top panels, bottom left, and right, respectively). The darkest diamonds are the results from the unmasked, periodic field, and the dark pink diamonds/error bars (labeled "Mask I") are the mean and rms from the masked field. All parameters are consistent between masked and unmasked data. The dashed horizontal lines in the

[17] For a description of the original GOTPM code, please see Dubinski et al. (2004). A description of the modifications introduced in the Horizon Run project at https://astro.kias.re.kr/~kjhan/GOTPM/index.html.

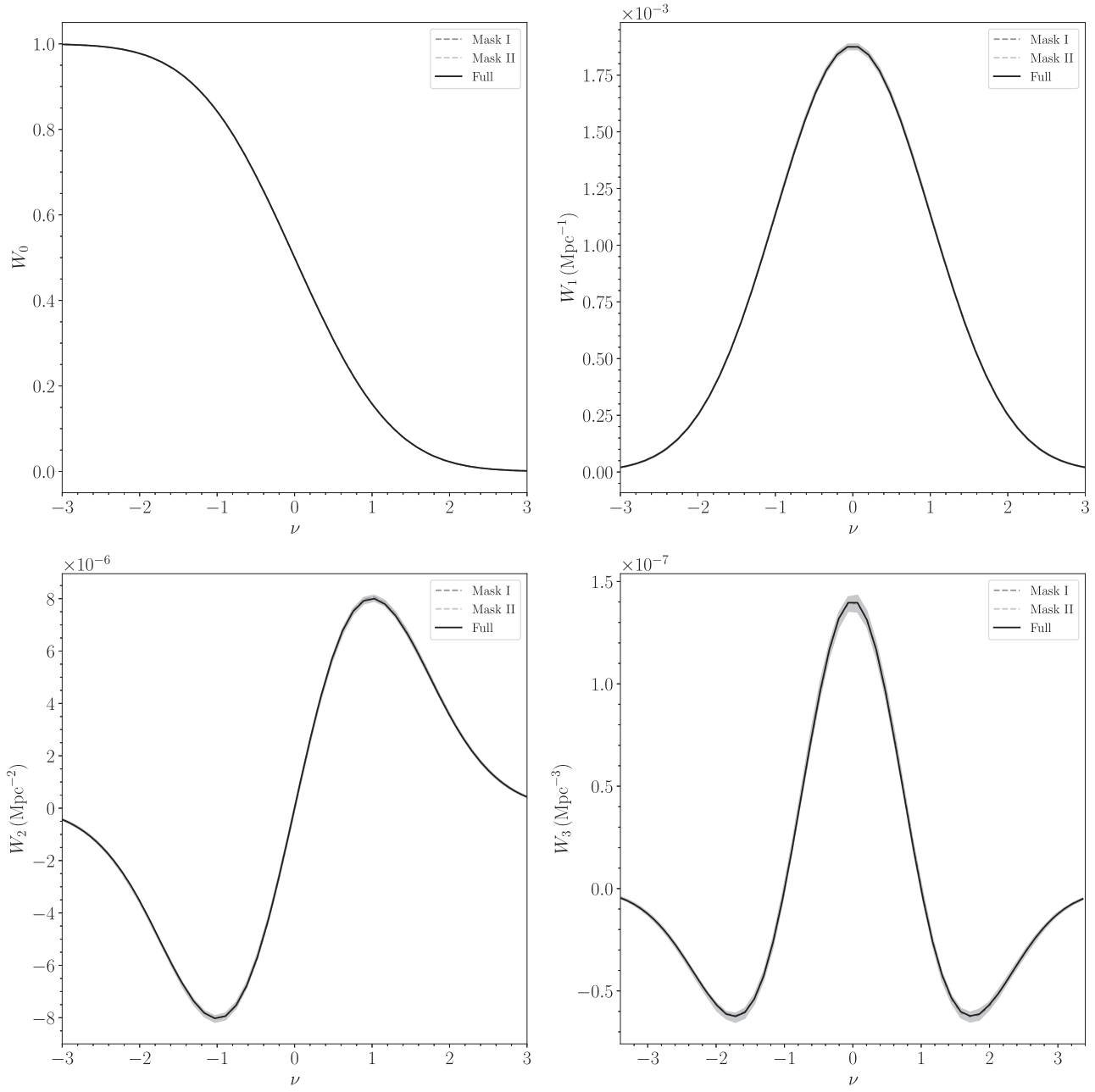

Figure 8. The MFs obtained from $N_\mathrm{r} = 50$ realizations of a Gaussian random field (GRF) with ΛCDM linear matter power spectrum. The solid gray region is the $\pm 1\sigma$ statistical uncertainty from the "Mask I" realizations. The black solid lines/dark-gray dashed lines/light-gray dashed lines are the mean values of the statistics obtained from the full unmasked field, and the result of two masking procedures explained in Sections B.2 and B.4.

top panels are the Gaussian expectation values of the amplitudes, which are consistent with the galaxy data.

B.4. Smoothing a Masked Field

In the previous section, we took a *complete* data set (that is, unmasked and with periodic boundary conditions), smoothed the field, then applied a mask. The goal was to confirm that the MF reconstruction used in this work is unbiased. However, in realistic situations, we cannot mask the data after smoothing, but rather must smooth the masked data. This introduces an additional complication, as the *true* field is not well reconstructed in regions close to the boundary after smoothing. We must therefore cut a region in the vicinity of the boundary from our analysis. We now quantify this statement.

Initially returning to GRFs, we perform the following steps. We generate a discrete field δ_{ijk} within a periodic box, and then apply the same masking procedure as in the previous subsection. We define the field as δ_{ijk}. We also define a discrete mask field M_{ijk}, where $M_{ijk} = 0$ if $\delta_{ijk} = \delta_\mathrm{b}$ is masked and $M_{ijk} = 1$ otherwise.

We then smooth both the masked field δ_{ijk} and the mask M_{ijk}, both with the same Gaussian smoothing kernel and scale $R_\mathrm{G} = 35$ Mpc. We denote the smoothed fields as $\tilde{\delta}_{ijk}$ and $\tilde{M}_{ijk}$. We then apply the smoothed mask to the smoothed data, taking $\tilde{\delta}_{ijk} = \delta_\mathrm{b}$ if $\tilde{M}_{ijk} < M_\mathrm{cut}$, where M_cut is some arbitrary value

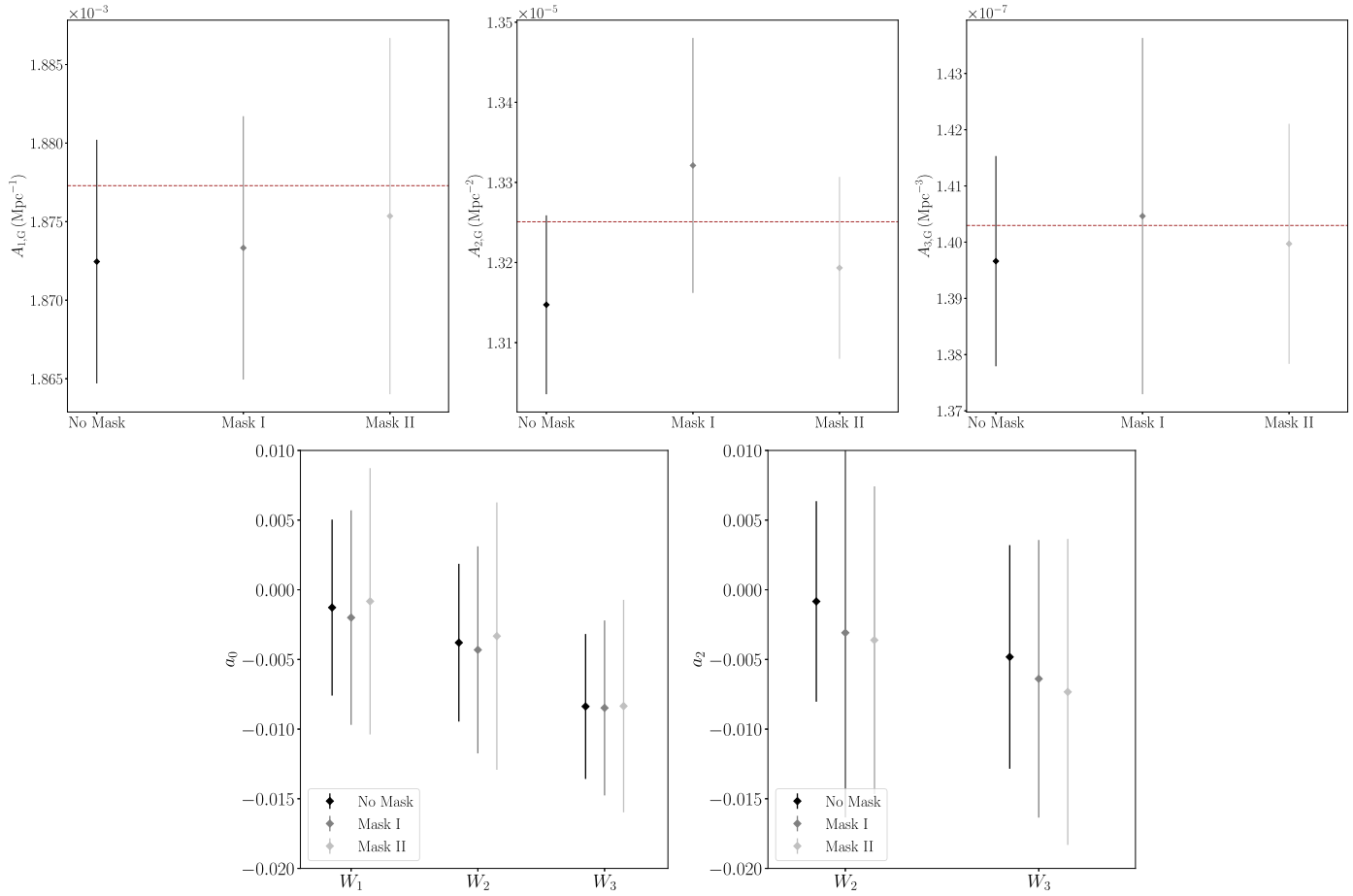

Figure 9. The amplitudes of the MFs (top three panels) and a_0, a_2 extracted from the $N_r = 50$ realizations of a GRF. The brown dashed lines in the top panels are the Gaussian expectation value from the Edgeworth expansion, and the black/dark gray/light gray diamonds and error bars are the mean and rms fluctuations of the quantities numerically extracted from the unmasked and masked data, respectively. For a GRF, $a_0 = a_2 = 0$.

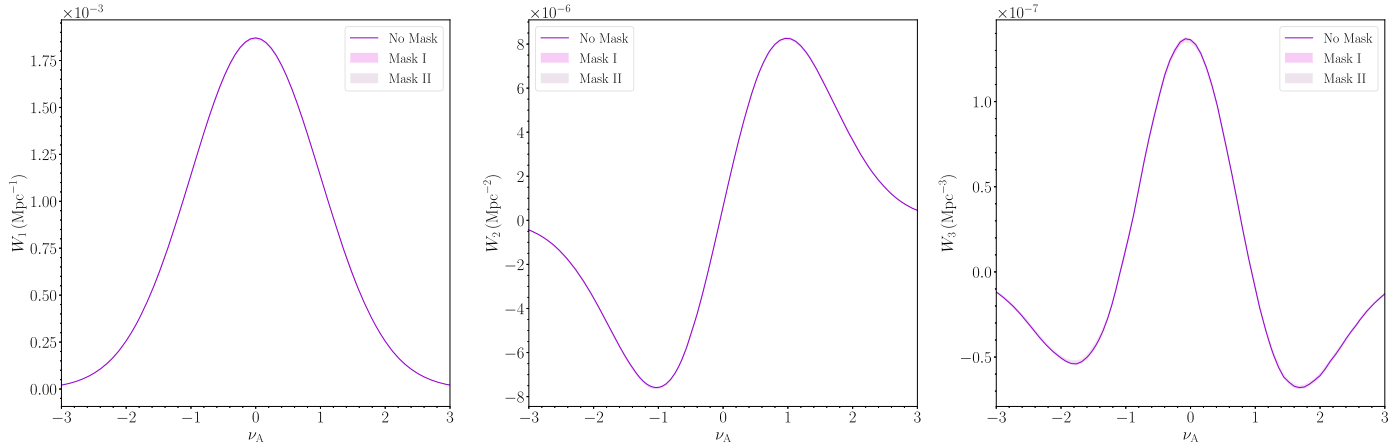

Figure 10. The MFs W_1, W_2, W_3 extracted from the Horizon Run 4 (HR4) snapshot box at $z = 0$. The purple line is the value obtained from the unmasked box, and the lighter solid regions are the 1σ rms fluctuations obtained from $N_{real} = 50$ random realizations of the masking procedures outlined in Sections B.2 and B.4.

selected to be $M_{cut} = 0.9$. Increasing M_{cut} increases the volume of data in the vicinity of the boundary that is cut during this second masking procedure. We then measure the MFs of the double-masked field $\tilde{\delta}_{ijk}$. The results are presented in Figures 8 and 10 for GRFs and the HR4 mock galaxy snapshot box at $z = 0$, respectively (light gray dashed line, pale pink filled region, labeled "Mask II"). We again observe an unbiased reconstruction of the MFs of the unmasked field. The parameters $A_{k,G}$ and $a_{0,2}$ are also presented in Figures 9 and 11, labeled "Mask II." The parameter reconstruction is unbiased within the uncertainties of the measurements.

We stress that the issues described in the previous subsection and here are not related. The former relates to the ability to reconstruct the MF statistics from a masked field, and the latter is the ability to faithfully reconstruct a field from a masked domain. We have shown that both issues can be ameliorated,

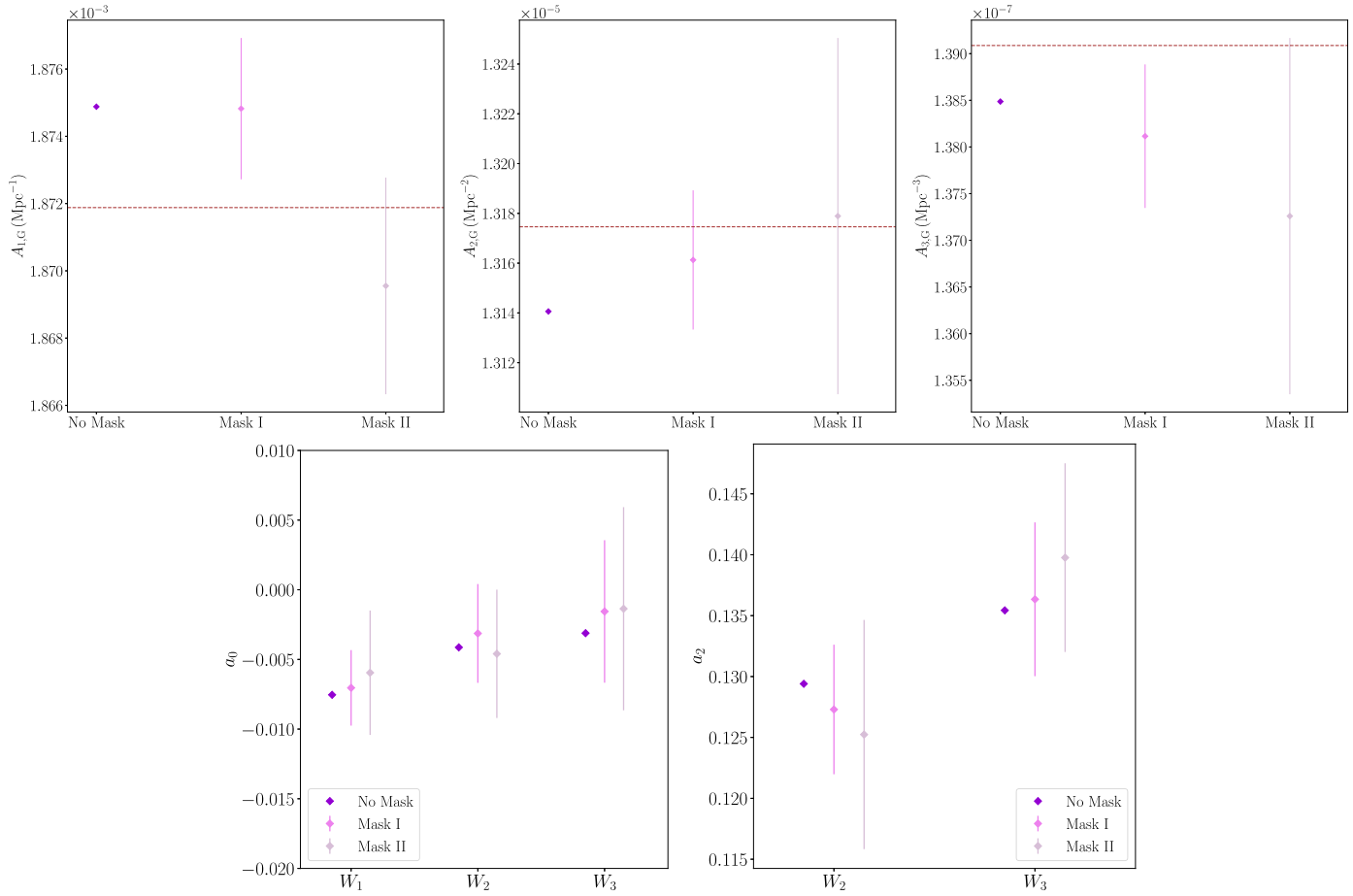

Figure 11. The amplitudes of the MFs and a_0, a_2 parameters extracted from the HR4 snapshot box. The deep purple diamond is the value obtained from the unmasked box (a single realization; no error bar) and the lighter purple diamonds/error bars are the values of the statistics from the $N_{\rm real} = 50$ realizations of the mask. The error bars will not represent the true statistical uncertainty, as we are not accounting for cosmic variance by using a single data realization. The horizontal dashed lines in the top panel are the Gaussian expectation values (Equation (8)) assuming HR4 cosmological parameters.

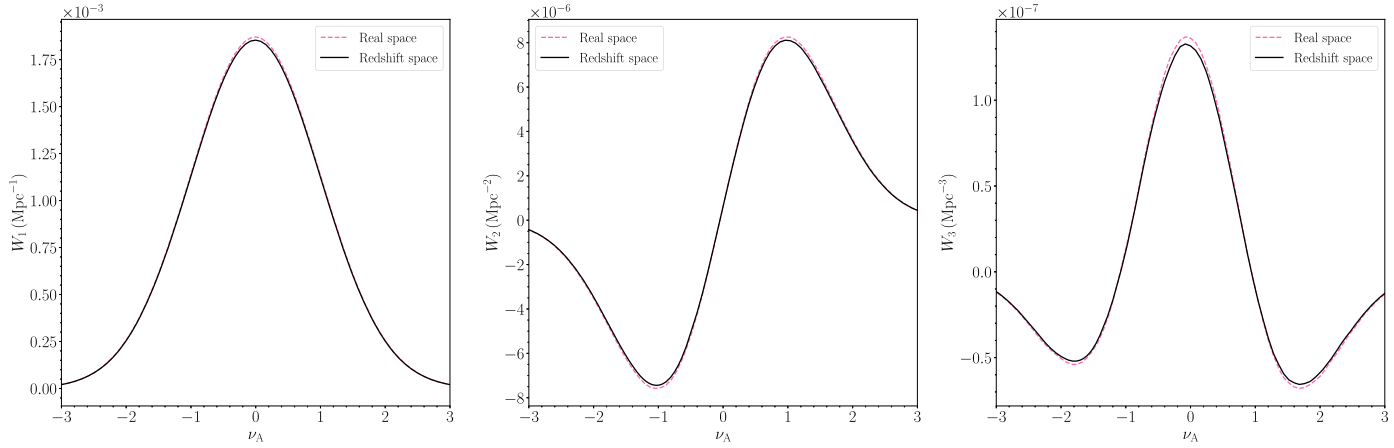

Figure 12. The MFs obtained from the HR4 $z = 0$ snapshot box in real (pink dashed line) and redshift (black solid line) space. The effect of redshift space distortion is to decrease the amplitude of the MFs.

with a judicious choice of statistical estimator and smoothing algorithm, respectively.

Appendix C
Redshift Space Distortion

RSD presents some difficulty with MF analysis. In Matsubara (1996), the effect was first theoretically predicted in the Kaiser limit and under the plane-parallel approximation.

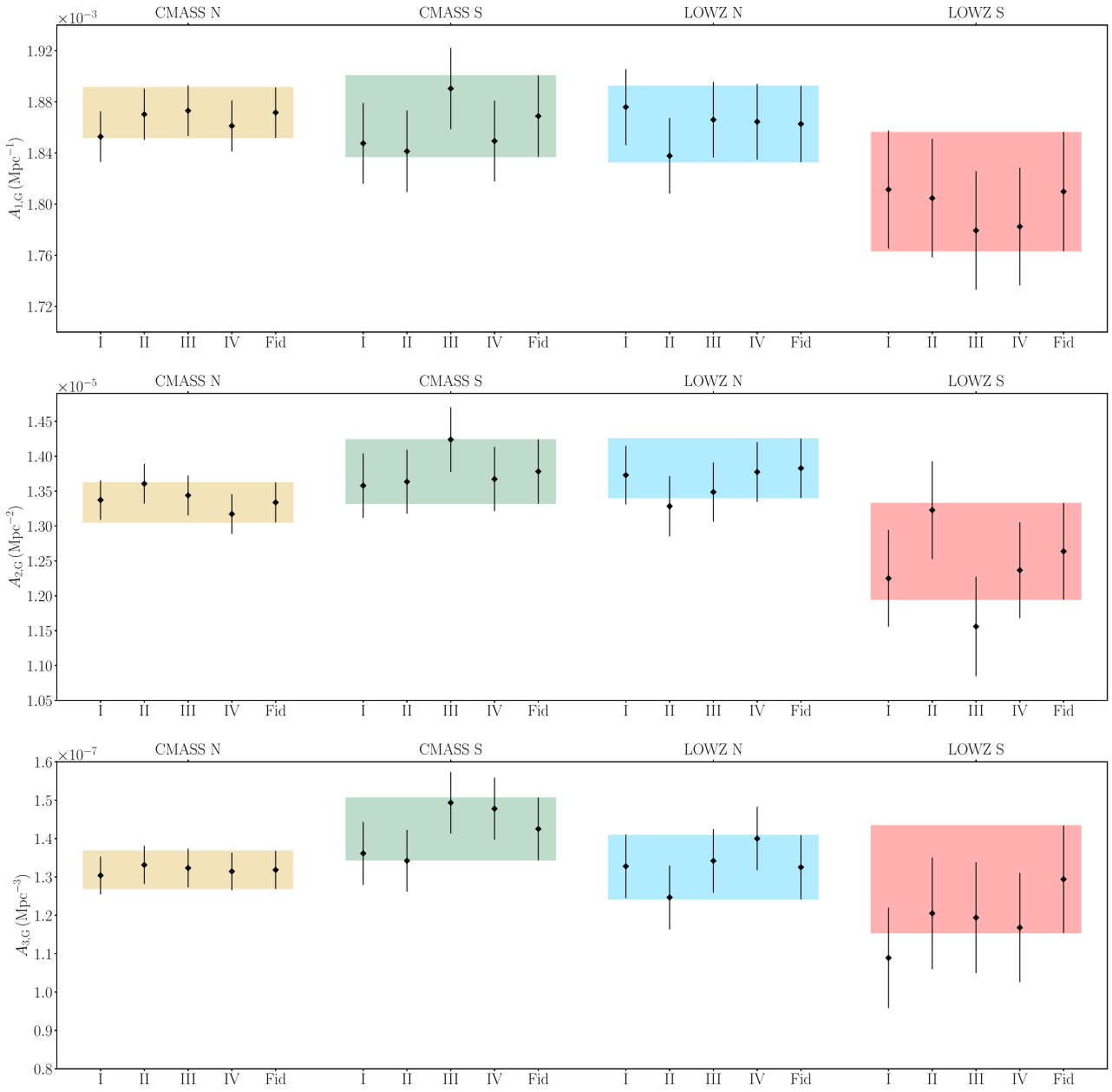

Figure 13. The best fit and 1σ uncertainties of the amplitudes of the MFs W_1, W_2, W_3 (top–bottom panels) obtained by varying the assumptions made in the main body of the paper and repeating our analysis. "Fid" represents the fiducial analysis in the paper, and I-IV are the four variations of our analysis discussed in Appendix D. The solid gold/green/blue/red regions are the 1σ uncertainties from the fiducial analysis of the CMASS N/S and LOWZ N/S data, respectively.

The analysis was continued in Codis et al. (2013), in which the Kaiser approximation was applied to higher-order cumulants. This represents the current state-of-the-art development in MF analysis. However, two outstanding issues remain. One is that the Finger-of-God effect could also impact the statistics, even on relatively large scales, as the cumulants σ_1 and σ_0 are integrated quantities. Second, large area sky surveys generically violate the plane-parallel, distant observer approximation, which breaks the homogeneity of the RSD signal.

Fortunately, the effect of RSD on the three-dimensional MFs is small. In Figure 12, we present the MFs of the HR4 mock galaxy snapshot box at $z = 0$ in real and plane-parallel redshift space. To generate the redshift space box, we perturb the galaxies along an arbitrary x_3 direction according to

$$x_3 \to x_3 + \frac{(1 + z_b)v_3}{H(z_b)}, \tag{C1}$$

where $z_b = 0$, and v_3 is the galaxy velocity in the x_3 direction. In Figure 12, the pink dashed/black solid lines are the MF statistics extracted from the real/redshift space boxes, respectively. We extract the amplitudes of these functions using the integral in Equation (27). We define the ratio $\alpha_k = A^{(s)}_{k,G}/A^{(r)}_{k,G}$, where r/s superscripts denote the amplitudes in real/redshift space, respectively. We find $\alpha_1 = 0.991$, $\alpha_2 = 0.981$, and $\alpha_3 = 0.971$. Our results indicate that the effect of RSD constitutes an approximately 1%, 2%, 3% decrease in the amplitude of $W_{1,2,3}$, respectively. In the main body of the

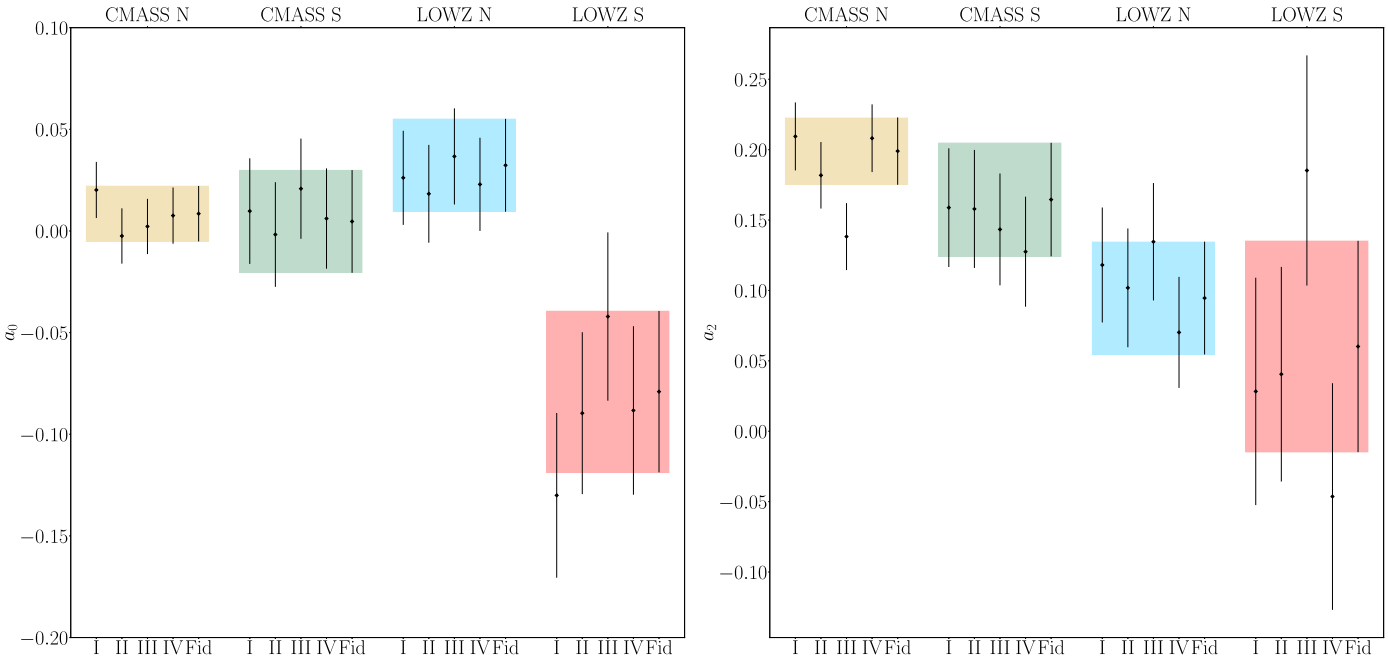

Figure 14. Same as Figure 13, but with parameters a_0 and a_2 (left/right panels).

Table 6
Marginalized Best Fit and 1σ Uncertainties on the Parameters $\log[A_{k,\mathrm{G}}]$, $a_{0,2}$, and $h_{0,2,3}$ Obtained by Fitting the Functions in Equations (30)–(32) to the Data by Minimizing the χ^2_k Function (Equation (24))

Data/MF	$\log[A_{k,\mathrm{G}}]$	a_0	a_2	h_0	h_2	h_3
CMASS N W_1	$-6.281(-6.281) \pm 0.010$	$0.013(0.013) \pm 0.022$	⋯	⋯	0.000 ± 0.005	⋯
CMASS S W_1	$-6.284(-6.283) \pm 0.017$	$0.010(0.018) \pm 0.041$	⋯	⋯	0.023 ± 0.007	⋯
LOWZ N W_1	$-6.287(-6.286) \pm 0.016$	$0.024(0.025) \pm 0.038$	⋯	⋯	-0.007 ± 0.007	⋯
LOWZ S W_1	$-6.317(-6.315) \pm 0.026$	$-0.110(-0.110) \pm 0.062$	⋯	⋯	0.020 ± 0.011	⋯
CMASS N W_2	$-11.211(-11.225) \pm 0.025$	$0.006(0.008) \pm 0.025$	$0.216(0.225) \pm 0.042$	⋯	⋯	0.012 ± 0.011
CMASS S W_2	$-11.145(-11.192) \pm 0.036$	$-0.021(-0.013) \pm 0.041$	$0.244(0.276) \pm 0.066$	⋯	⋯	0.043 ± 0.015
LOWZ N W_2	$-11.215(-11.190) \pm 0.038$	$0.039(0.035) \pm 0.039$	$0.075(0.076) \pm 0.066$	⋯	⋯	-0.017 ± 0.016
LOWZ S W_2	$-11.227(-11.288) \pm 0.058$	$-0.029(-0.033) \pm 0.064$	$0.100(0.104) \pm 0.116$	⋯	⋯	0.050 ± 0.023
CMASS N W_3	$-15.838(-15.836) \pm 0.038$	$0.003(0.003) \pm 0.026$	$0.180(0.0175) \pm 0.038$	-0.002 ± 0.019	⋯	⋯
CMASS S W_3	$-15.757(-15.752) \pm 0.058$	$0.019(0.017) \pm 0.047$	$0.099(0.097) \pm 0.065$	-0.006 ± 0.029	⋯	⋯
LOWZ N W_3	$-15.837(-15.838) \pm 0.062$	$0.030(0.029) \pm 0.046$	$0.116(0.114) \pm 0.065$	0.008 ± 0.034	⋯	⋯
LOWZ S W_3	$-15.873(-15.879) \pm 0.108$	$-0.097(-0.106) \pm 0.086$	$0.062(0.046) \pm 0.129$	0.025 ± 0.064	⋯	⋯

Note. The values inside the brackets are the corresponding *fiducial* values with $h_{0,2,3} = 0$.

text, we correct the theoretical expectation $A_{k,\mathrm{G}}$ MF amplitudes by α_k to convert them into redshift space.

Predicting the consequences of the breakdown of the plane-parallel approximation is beyond the scope of this work, and will be considered elsewhere. However, if the overall effect on the MFs due to RSD is $\sim\mathcal{O}$(few%) on quasi-linear scales, we can expect this subtlety to not significantly affect our conclusions. Similarly, we expect the cosmological parameter dependence of the RSD effect to be a small correction to α_k, which is itself a small correction. We therefore treat α_k as constant correction factors in the main body of the paper.

Appendix D
Testing for Systematics

Finally, we consider the various systematics that could impact our measurements of the MFs, to test the robustness of our measurements. We perform a number of stress-tests on our results by significantly varying the assumptions made. Each time, we vary a single assumption used in the main body of the paper, and run the new data through our analysis pipeline. We fit the functional form in Equation (23) to the data sets, and present the best fit and 1σ marginalized uncertainties on $A_{k,\mathrm{G}}$.

We identify the following potential issues:

I. The galaxy weights significantly affect the MF reconstruction.
II. Our estimator of the MFs is affected by the mask/boundaries of the data.
III. Changing the sampling method will yield subsamples of galaxies with different properties.
IV. Finite pixel resolution effects.

We consider each point in turn.

The galaxy weights are one possible issue. We have accounted for systematic variations in the galaxy number

density by weighting each galaxy, as detailed in Reid et al. (2016). We have also removed areas of the data in which the completeness is below 90%. Due to the high completeness of the sample, and the large smoothing scales adopted, we do not anticipate that our conclusions will be affected by the galaxy weighting scheme. However, to test this, we repeat our analysis on each of the data sets without weighting the galaxies with *any* observational systematic correction, assuming that each galaxy contributes a total of unity to the number density. This is obviously an inappropriate procedure, but will inform us of how significant the effect of the galaxy weights are on our conclusions.

Given that the northern/southern sky data possess very different survey geometries, it is possible that our methodology is affected by the boundaries of the data. However, this seems unlikely given that our treatment of the boundary is repeated exactly for the patchy mock data, and all subsets of the mock catalogs present excellent consistency. To test the impact of the mask further, we decrease the mask cut $M_{\rm cut}$ applied to the data to be $M_{\rm cut} = 0.8$, to test the sensitivity of our results to this parameter. As we lower $M_{\rm cut}$, we are including the regions closer to the survey boundary, where the field may not be well reconstructed.

Regarding the sampling, in principle the MFs are insensitive to galaxy sampling and linear bias. This implies that on large scales any dependence on linear galaxy bias drops out. However, in practise this is not always precisely true because the presence of shot noise in a field reconstructed from a point distribution makes the measured power spectrum the sum of two distinct quantities—the cosmological and noise power spectra. The ratio between these two contributions will impact the MF amplitudes, and is a function of the matter power spectrum amplitude and galaxy bias. To test the sensitivity of the statistics to sample selection, we randomly select galaxies in the CMASS and LOWZ data to match the fiducial number density $\bar{n} = 6.25 \times 10^{-5}\,{\rm Mpc}^{-3}$, rather than mass selecting as in the main body of the paper. It was shown in Kim et al. (2014) that the difference between randomly and mass selecting galaxies can significantly impact the genus statistic. However, for the BOSS data, we do not expect such a pronounced difference because the entire data set constitutes mostly massive, highly biased galaxies, and with our number density cut, we are using the majority of the sample over the range $0.2 < z < 0.4$ and $0.45 < z < 0.6$. Hence randomly and mass selecting the galaxies will yield a similarly biased sample.

Finally, we consider finite resolution pixel effects. This is also an unlikely source of contamination because we are Gaussian smoothing over at least four pixels in each data set, and also the patchy mock catalogs are subject to the same pixel resolution and are self-consistent. However, we repeat our analysis, using the same box sizes as in Table 2 but with resolutions $\Delta = 5.5, 4.3, 4.0, 3.1$ Mpc for the CMASS N/S and LOWZ N/S data, respectively. This corresponds to a change from 512^3 to 768^3 pixels within the uniform lattices onto which we aggregate the data.

The results of these tests are presented in Figure 13 for the amplitudes $A_{k,{\rm G}}$ and Figure 14 for s_0 and s_2, respectively (left/right panels). We present a combined fit of five parameters $A_{k,{\rm G}}$, s_0, s_2 to $W_{1,2,3}$ extracted separately from the CMASS/LOWZ N/S data. That is, we fit the functional form in Equation (23) to $W_{1,2,3}$, by minimizing $\chi^2 = \chi^2_1 + \chi^2_2 + \chi^2_3$, with χ^2_k defined in Equation (24). The black diamonds and error bars are the result of repeating our analysis in the main body of the paper, relaxing each assumption made in points I–IV above. The "Fid" points in the figures are the fiducial values obtained in the main text body, and the solid gold/green/blue/red filled areas are the $\pm 1\sigma$ ranges of the parameters from the fiducial analysis for CMASS N/S and LOWZ N/S, respectively, included as a visual guide.

The CMASS N/S and LOWZ N/S data present self-consistent results for $A_{k,{\rm G}}$ and $a_{0,2}$ for practically all tests performed in this section. The only seemingly significant peculiarity is the low value of a_2 in the CMASS N data for test III (randomly sampling the galaxies; right panel in Figure 14). However, it is clear that a_2 is the least well measured quantity that we extract from the data, and it is not clear if we are obtaining a systematically high value of this quantity in the CMASS N data. This could also simply be a statistical fluctuation, as the discrepancy between this point and the fiducial measurement is not high (less than 2σ). The amplitudes that we use for the cosmological parameter estimation present excellent stability when we modify our analysis, which indicates that they are robust cosmological measurements.

Finally in this appendix, we include the results of including $h_{0,2,3}$ when fitting a Hermite polynomial expansion to the BOSS MF curves. In Table 6, we present the best fit and 1σ uncertainties in all parameters with $h_{0,2,3}$ included in the fits. Also included are the best-fit values with $h_{0,2,3} = 0$ (shown in brackets).

ORCID iDs

Changbom Park https://orcid.org/0000-0001-9521-6397
Sungwook E. Hong (홍성 욱) https://orcid.org/0000-0003-4923-8485
Ho Seong Hwang https://orcid.org/0000-0003-3428-7612
Juhan Kim https://orcid.org/0000-0002-4391-2275
Thomas Buchert https://orcid.org/0000-0002-0828-3901

References

Adler, R. 1981, The Geometry of Random Fields (New York: Wiley)
Aghanim, N., Armitage-Caplan, C., Arnaud, M., et al. 2014, A&A, 571, A27
Aghanim, N., Akrami, Y., Ashdown, M., et al. 2020, A&A, 641, A6
Alam, S., Albareti, F. D., Allende Prieto, C., et al. 2015, ApJS, 219, 12
Appleby, S., Chingangbam, P., Park, C., et al. 2018a, ApJ, 858, 87
Appleby, S., Chingangbam, P., Park, C., Yogendran, K. P., & Joby, P. K. 2018b, ApJ, 863, 200
Appleby, S., Park, C., Hong, S., & Kim, J. 2018c, ApJ, 853, 17
Appleby, S., Park, C., Hong, S. E., et al. 2021, ApJ, 907, 75
Appleby, S., Park, C., Hong, S. E., & Kim, J. 2017, ApJ, 836, 45
Appleby, S. A., Park, C., Hong, S. E., Hwang, H. S., & Kim, J. 2020, ApJ, 896, 145
Beisbart, C., Buchert, T., & Wagner, H. 2001a, Phy, A293, 592
Beisbart, C., Valdarnini, R., & Buchert, T. 2001b, A&A, 379, 412
Blake, C., James, J. B., & Poole, G. B. 2014, MNRAS, 437, 2488
Bolton, A. S., Schlegel, D. J., Aubourg, É., et al. 2012, AJ, 144, 144
Buchert, T., France, M. J., & Steiner, F. 2017, CQGra, 34, 094002
Chingangbam, P., Ganesan, V., Yogendran, K. P., & Park, C. 2017a, PhL, B771, 67
Chingangbam, P., Goyal, P., Yogendran, K. P., & Appleby, S. 2021, PhRvD, 104, 123516
Chingangbam, P., Yogendran, K. P., K, J. P., et al. 2017b, JCAP, 12, 023
Choi, Y.-Y., Park, C., Kim, J., et al. 2010, ApJS, 190, 181
Codis, S., Pichon, C., Pogosyan, D., Bernardeau, F., & Matsubara, T. 2013, MNRAS, 435, 531
Colas, T., D'amico, G., Senatore, L., Zhang, P., & Beutler, F. 2020, JCAP, 06, 001
Colin, J., Mohayaee, R., Rameez, M., & Sarkar, S. 2019, A&A, 631, L13
D'Amico, G., Gleyzes, J., Kokron, N., et al. 2020, JCAP, 05, 005
Doroshkevich, A. G. 1970, Ap, 6, 320

Dubinski, J., Kim, J., Park, C., & Humble, R. 2004, NewA, 9, 111
Edelsbrunner, H., & Harer, J. 2010, Computational Topology - an Introduction, I–XII (Providence, RI: American Mathematical Society), 1
Feldbrugge, J., van Engelen, M., van de Weygaert, R., Pranav, P., & Vegter, G. 2019, JCAP, 1909, 052
Fukugita, M., Ichikawa, T., Gunn, J. E., et al. 1996, AJ, 111, 1748
Ganesan, V., & Chingangbam, P. 2017, JCAP, 1706, 023
Gay, C., Pichon, C., & Pogosyan, D. 2012, PhRv, D85, 023011
Gorski, K. M., Hivon, E., Banday, A. J., et al. 2005, ApJ, 622, 759
Gott, J. R., Choi, Y.-Y., Park, C., & Kim, J. 2009, ApJL, 695, L45
Gott, J. R., Dickinson, M., & Melott, A. L. 1986, ApJ, 306, 341
Gott, J. R., Weinberg, D. H., & Melott, A. L. 1987, ApJ, 319, 1
Gott, J. R. I., Hambrick, D. C., Vogeley, M. S., et al. 2008, ApJ, 675, 16
Gott, J. R., III, Park, C., Juszkiewicz, R., et al. 1990, ApJ, 352, 1
Goyal, P., & Chingangbam, P. 2021, JCAP, 2021, 006
Gunn, J. E., Siegmund, W. A., Mannery, E. J., et al. 2006, AJ, 131, 2332
Hamaus, N., Pisani, A., Choi, J.-A., et al. 2020, JCAP, 12, 023
Hamilton, J. S. A., Gott, J. R., & Weinberg, D. 1986, ApJ, 309, 1
Hartlap, J., Simon, P., & Schneider, P. 2007, A&A, 464, 399
Heinesen, A. 2021, JCAP, 2021, 008
Hikage, C., Coles, P., Grossi, M., et al. 2008, MNRAS, 385, 1613
Hikage, C., Schmalzing, J., Buchert, T., et al. 2003, PASJ, 55, 911
Hikage, C., Suto, Y., Kayo, I., et al. 2002, PASJ, 54, 707
Hikage, C., Taruya, A., & Suto, Y. 2001, ApJ, 556, 641
Hong, S. E., Park, C., & Kim, J. 2016, ApJ, 823, 103
Ivanov, M. M., Simonović, M., & Zaldarriaga, M. 2020, JCAP, 05, 042
James, J. B., Colless, M., Lewis, G. F., & Peacock, J. A. 2009, MNRAS, 394, 454
Jiang, C. Y., Jing, Y. P., Faltenbacher, A., Lin, W. P., & Li, C. 2008, ApJ, 675, 1095
Joby, P. K., Chingangbam, P., Ghosh, T., Ganesan, V., & Ravikumar, C. D. 2019, JCAP, 1901, 009
Kapahtia, A., Chingangbam, P., & Appleby, S. 2019, JCAP, 2019, 053
Kapahtia, A., Chingangbam, P., Appleby, S., & Park, C. 2018, JCAP, 1810, 011
Kerscher, M., Mecke, K., Schmalzing, J., et al. 2001, A&A, 373, 1
Kerscher, M., Schmalzing, J., Buchert, T., & Wagner, H. 1997, in Research in Particle-Astrophysics, Proceedings of a Workshop , ed. R. Bender et al. (Munich: Max-Planck-Institut)
Kerscher, M., Schmalzing, J., Buchert, T., & Wagner, H. 1998, A&A, 333, 1
Kim, Y.-R., Choi, Y.-Y., Kim, S. S., et al. 2014, ApJS, 212, 22
Kitaura, F.-S., Gil-Marín, H., Scóccola, C. G., et al. 2015, MNRAS, 450, 1836
Kitaura, F.-S., Rodríguez-Torres, S., Chuang, C.-H., et al. 2016, MNRAS, 456, 4156
Kitaura, F.-S., Yepes, G., & Prada, F. 2014, MNRAS, 439, L21
Li, X.-D., Park, C., Sabiu, C. G., et al. 2016, ApJ, 832, 103
Lippich, M., & Sánchez, A. G. 2021, MNRAS, 508, 3771
Luongo, O., Muccino, M., Colgáin, E. O., Sheikh-Jabbari, M. M., & Yin, L. 2021, arXiv:2108.13228
Manera, M., Samushia, L., Tojeiro, R., et al. 2015, MNRAS, 447, 437
Matsubara, T. 1994a, ApJL, 434, L43
Matsubara, T. 1994b, in Large Scale Structure in the Universe, Proc. of an International Workshop, ed. J. P. Mücket et al. (Singapore: World Scientific) arXiv:9501076
Matsubara, T. 1996, ApJ, 457, 13
Matsubara, T. 2000, arXiv:astro-ph/0006269
Matsubara, T. 2003, ApJ, 584, 1
Matsubara, T., Hikage, C., & Kuriki, S. 2020, arXiv:2012.00203
Matsubara, T., & Kuriki, S. 2021, PhRvD, 104, 103522
Matsubara, T., & Suto, Y. 1996, ApJ, 460, 51
Matsubara, T., & Yokoyama, J. 1996, ApJ, 463, 409
Mecke, K. R., Buchert, T., & Wagner, H. 1994, A&A, 288, 697
Melott, A. L., Cohen, A. P., Hamilton, A. J. S., Gott, J. R., & Weinberg, D. H. 1989, ApJ, 345, 618
Melott, A. L., Weinberg, D. H., & Gott, J. R. 1988, ApJ, 328, 50
Mohayaee, R., Rameez, M., & Sarkar, S. 2020, arXiv:2003.10420
Parihar, P., Vogeley, M. S., Gott, J. R., III, et al. 2014, ApJ, 796, 86
Park, C., Choi, Y.-Y., Vogeley, M., et al. 2005, ApJ, 633, 11
Park, C., & Gott, J. R. 1991, ApJ, 378, 457
Park, C., Gott, J. R., & Choi, Y. J. 2001, ApJ, 553, 33
Park, C., Gott, J. R., Melott, A. L., & Karachentsev, I. D. 1992, ApJ, 387, 1
Park, C., & Kim, Y.-R. 2010, ApJL, 715, L185
Park, C., Pranav, P., Chingangbam, P., et al. 2013, JKAS, 46, 125
Petri, A., Haiman, Z., Hui, L., May, M., & Kratochvil, J. M. 2013, PhRvD, 88, 123002
Pogosyan, D., Gay, C., & Pichon, C. 2009, PhRvD, 80, 081301
Pranav, P. 2021, arXiv:2101.02237
Pranav, P., Adler, R. J., Buchert, T., et al. 2019a, A&A, 627, A163
Pranav, P., Edelsbrunner, H., van de Weygaert, R., et al. 2017, MNRAS, 465, 4281
Pranav, P., van de Weygaert, R., Vegter, G., et al. 2019b, MNRAS, 485, 4167
Reid, B., Ho, S., Padmanabhan, N., et al. 2016, MNRAS, 455, 1553
Rodríguez-Torres, S. A., Chuang, C.-H., Prada, F., et al. 2016, MNRAS, 460, 1173
Ross, A. J., et al. 2017, MNRAS, 464, 1168
Ryden, B. S., Melott, A. L., Craig, D. A., et al. 1989, ApJ, 340, 647
Schmalzing, J., & Buchert, T. 1997, ApJL, 482, L1
Schmalzing, J., & Gorski, K. M. 1998, MNRAS, 297, 355
Secrest, N. J., von Hausegger, S., Rameez, M., et al. 2021, ApJL, 908, L51
Sheth, J. V., & Sahni, V. 2005, arXiv:astro-ph/0502105
Sheth, J. V., Sahni, V., Shandarin, S. F., & Sathyaprakash, B. S. 2003, MNRAS, 343, 22
Shim, J., Codis, S., Pichon, C., Pogosyan, D., & Cadiou, C. 2021, MNRAS, 502, 3885
Shivshankar, N., Pranav, P., Natarajan, V., et al. 2016, ITVCG, 22, 1745
Slepian, Z., et al. 2017, MNRAS, 468, 1070
Springel, V. 2005, MNRAS, 364, 1105
Sullivan, J. M., Wiegand, A., & Eisenstein, D. J. 2019, MNRAS, 485, 1708
Tinker, J. L., et al. 2017, ApJ, 839, 121
Tojeiro, R., et al. 2014, MNRAS, 440, 2222
Tomita, H. 1986, PThPh, 76, 952
van de Weygaert, R., et al. 2011, arXiv:1110.5528
Wang, Y., Xu, Y., Wu, F., et al. 2015, in Proc. of Advancing Astrophysics with the Square Kilometre Array (AASKA14) (Trieste: PoS), 033
Weinberg, D. H., Gott, J. R., & Melott, A. L. 1987, ApJ, 321, 2
Wiegand, A., Buchert, T., & Ostermann, M. 2014, MNRAS, 443, 241
Wiegand, A., & Eisenstein, D. J. 2017, MNRAS, 467, 3361
Wilding, G., Nevenzeel, K., van de Weygaert, R., et al. 2020, arXiv:2011.12851
York, D. G., Adelman, J., Anderson, J. E., Jr., et al. 2000, AJ, 120, 1579
Zhai, Z., et al. 2017, ApJ, 848, 76
Zhang, Y., Springel, V., & Yang, X. 2010, AJ, 722, 812
Zunckel, C., Gott, J. R., III, & Lunnan, R. 2011, MNRAS, 412, 1401