



Universiteit
Leiden
The Netherlands

Towards testing the theory of gravity with DESI: summary statistics, model predictions and future simulation requirements

Alam, S.; Arnold, C.; Aviles, A.; Bean, R.; Cai, Y.-C.; Cautun, M.; ... ; Zheng, Y.

Citation

Alam, S., Arnold, C., Aviles, A., Bean, R., Cai, Y. -C., Cautun, M., ... Zheng, Y. (2021). Towards testing the theory of gravity with DESI: summary statistics, model predictions and future simulation requirements. *Journal Of Cosmology And Astroparticle Physics*, 2021(11). doi:10.1088/1475-7516/2021/11/050

Version: Accepted Manuscript
License: [Creative Commons CC BY 4.0 license](https://creativecommons.org/licenses/by/4.0/)
Downloaded from: <https://hdl.handle.net/1887/3254584>

Note: To cite this publication please use the final published version (if applicable).

Testing the theory of gravity with DESI: estimators, predictions and simulation requirements

Shadab Alam,¹ Alejandro Aviles,^{2,3} Rachel Bean,⁴ Yan-Chuan Cai,¹ Marius Cautun,^{5,6} Jorge L. Cervantes-Cota,² Carolina Cuesta-Lazaro,⁵ N. Chandrachani Devi,⁷ Alexander Eggemeier,⁵ Sebastien Fromenteau,⁸ Alma X. Gonzalez-Morales,^{3,9} Vitali Halenka,¹⁰ Jian-hua He,^{11,5} Wojciech A. Hellwing,¹² César Hernández-Aguayo,^{13,14,5} Mustapha Ishak,¹⁵ Kazuya Koyama,¹⁶ Baojiu Li,⁵ Axel de la Macorra,¹⁷ Jennifer Meneses Rizo,¹⁷ Christopher Miller,^{18,10} Eva-Maria Mueller,¹⁹ Gustavo Niz,⁹ Pierros Ntelis,²⁰ Matias Rodríguez Otero,¹⁷ Cristiano G. Sabiu,²¹ Zachary Slepian,^{22,23} Alejo Stark,¹⁸ Octavio Valenzuela,⁷ Georgios Valogiannis,^{4,24} Mariana Vargas-Magaña,¹⁷ Hans A. Winther,^{16,25} Pauline Zarrouk,^{26,5} Gong-Bo Zhao,^{27,28,16} and Yi Zheng^{29,30}

¹*Institute for Astronomy, University of Edinburgh, Edinburgh, UK*

²*Departamento de Física, Instituto Nacional de Investigaciones Nucleares, Apartado Postal 18-1027, Col. Escandón, Ciudad de México, 11801, México*

³*Consejo Nacional de Ciencia y Tecnología, Av. Insurgentes Sur 1582. Colonia*

Credito Constructor, Del. Benito Jurez C.P. 03940, México D.F. México

⁴*Department of Astronomy, Cornell University, Ithaca, NY 14853, USA*

⁵*Institute for Computational Cosmology, Department of Physics, Durham University, South Road, Durham DH1 3LE, UK*

⁶*Leiden Observatory, Leiden University, PO Box 9513, NL-2300 RA Leiden, the Netherlands*

⁷*Instituto de Astronomía, Universidad Nacional Autónoma de México, Apdo. Postal 20-364, México*

⁸*Instituto de Ciencias Físicas, Universidad Nacional Autónoma de México, Av. Universidad s/n, 62210 Cuernavaca, Mor., México*

⁹*Departamento de Física, División de Ciencias e Ingenierías, Campus León, Universidad de Guanajuato, León 37150, México*

¹⁰*University of Michigan, Department of Physics, 1085 South University, Ann Arbor, 48109, USA*

¹¹*School of Astronomy and Space Science, Nanjing University, Nanjing 210093, China*

¹²*Center for Theoretical Physics, Polish Academy of Sciences, Al. Lotników 32/46, 02-668 Warsaw, Poland*

¹³*Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Str. 1, D-85748, Garching, Germany*

¹⁴*Excellence Cluster ORIGINS, Boltzmannstrasse 2, D-85748 Garching, Germany*

¹⁵*Department of Physics, The University of Texas at Dallas, Richardson, Texas 75080, USA*

¹⁶*Institute of Cosmology and Gravitation, University of Portsmouth, Portsmouth, Hampshire, PO 3FX, UK*

¹⁷*Instituto de Física, Universidad Nacional Autónoma de México, Apdo. Postal 20-364, Ciudad de México, México*

¹⁸*University of Michigan, Department of Astronomy, Department of Physics, 1085 South University, Ann Arbor, 48109, USA*

¹⁹*Department of Physics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH, UK*

²⁰*Aix Marseille Univ, CNRS/IN2P3, CPPM, Marseille, France*

²¹*Department of Astronomy, Yonsei University, Seoul, Korea*

²²*Einstein Fellow, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA*

²³*Department of Astronomy, University of Florida, 211 Bryant Space Science Center, Gainesville, FL 32611, USA*

²⁴*Department of Physics, Harvard University, Cambridge, MA 02138, USA*

²⁵*Institute of Theoretical Astrophysics, University of Oslo, 0315 Oslo, Norway*

²⁶*Sorbonne Université, Université Paris Diderot, CNRS/IN2P3,*

Laboratoire de Physique Nucléaire et de Hautes Energies, LPNHE, 4 Place Jussieu, F-75252 Paris, France

²⁷*National Astronomy Observatories, Chinese Academy of Science, Beijing, 100101, China*

²⁸*School of Astronomy and Space Science, University of Chinese Academy of Sciences, Beijing 100049, China*

²⁹*School of Physics and Astronomy, Sun Yat-sen University, 2 Daxue Road, Tangjia, Zhuhai, 519082, China*

³⁰*School of Physics, Korea Institute for Advanced Study, Hoegiro 85, Seoul 02455, Korea*

Shortly after its discovery, General Relativity (GR) was applied to predict the behavior of our Universe on the largest scales, and later became the foundation of modern cosmology. Its validity has been verified on a range of scales and environments from the Solar system to merging black holes. However, experimental confirmations of GR on cosmological scales have so far lacked the accuracy one would hope for – its applications on those scales being largely based on extrapolation and its validity there sometimes questioned in the shadow of the discovery of the unexpected cosmic acceleration. Future astronomical instruments surveying the distribution and evolution of galaxies over substantial portions of the observable Universe, such as the Dark Energy Spectroscopic Instrument (DESI), will be able to measure the fingerprints of gravity and their statistical power will allow strong constraints on alternatives to GR.

In this paper, based on a set of N-body simulations and mock galaxy catalogs, we study the predictions of a number of traditional and novel estimators beyond linear redshift distortions in two well-studied modified gravity models – chameleon $f(R)$ gravity and a braneworld model – and the potential of testing these deviations from GR using DESI. These estimators employ a wide array of statistical properties of the galaxy and the underlying dark matter field, including two-point and higher-order statistics, environmental dependence, redshift space distortions and weak lensing. We find that they hold promising power for testing GR to unprecedented precision. The major future challenge is to make realistic, simulation-based mock galaxy catalogs for both GR

and alternative models to fully exploit the statistic power of the DESI survey (by matching the volumes and galaxy number densities of the mocks to those in the real survey) and to better understand the impact of key systematic effects. Using these, we identify future simulation and analysis needs for gravity tests using DESI.

CONTENTS

I. Introduction	4
II. Models, simulations and galaxy catalogs	5
A. Modified gravity models	6
B. N-body simulations and halo/galaxy catalogs	8
III. Estimators	14
A. Redshift-space galaxy clustering	14
B. Density-dependent environmental effects on galaxy clustering	22
C. Beyond two-point statistics	31
D. Gravitational lensing statistics	47
IV. Conclusions	52
Acknowledgments	55
References	55

I. INTRODUCTION

Since its discovery over a century ago, General Relativity (GR) has been established as our standard theory of gravity, thanks to the many experimental and observational tests of its various predictions. Most of these tests have been laboratory and Solar System tests in the weak-field limit, and its predicted effects from strong gravitational fields have also been detected, through the orbital decay of binary pulsars [1] and gravitational waves from merging binary compact objects [2, 3]. It has passed yet another test recently by the further detection of gravitational waves from a binary neutron star merger (GW170817) associated with a short Gamma-ray burst [3] and various other electromagnetic counterparts [4], which confirmed the equivalence between the speeds of gravity and light with high accuracy. Its application to cosmology involves a huge extrapolation from length scales at which it has been tested accurately to the Universe as a whole, but this was rarely questioned given the success of GR in predicting a diverse set of observations: the Hubble expansion, Big Bang Nucleosynthesis (BBN) and the Cosmic Microwave Background (CMB). The observations of an accelerated rate of the Hubble expansion based on the dimming of distant type Ia supernovae (SNIe) in 1998 [5, 6], which have since then been supported by various independent probes, however, raised the possibility that GR might have to be replaced by an alternative model on cosmological scales. This is not the only possibility, because the cosmic acceleration may well be driven by a positive cosmological constant Λ (as in the standard Λ CDM paradigm), or some dynamical dark energy component motivated by beyond-standard-model particle physics [7]. However, given that there is currently not a widely-accepted theoretical explanation, these different possibilities have all been kept open, and there has been a great body of research on modified gravity (MG) models in recent years (see, e.g., [8–11] for some recent reviews). From a practical point of view, the studies of modified gravity and dark energy models can be considered as different faces of a common theme: *testing GR-based Λ CDM in cosmology*. This field has now matured enough, thanks to developments of both theoretical frameworks, which tells us what happens in cosmology, if there is a deviation from GR, and observational techniques, which tells us whether such a deviation is supported by the data.

DESI (Dark Energy Spectroscopic Instrument) will perform one of these advanced wide-area galaxy and quasar redshift surveys. It is a stage-IV ground-based experiment designed to have a sky coverage of $14,000 \text{ deg}^2$, using four different classes of spectroscopic targets – luminous red galaxies (LRGs) up to $z \simeq 1$, emission line galaxies (ELGs) up to $z \simeq 1.7$, quasars and Ly- α features up to $z \simeq 3.5$, and bright galaxies at low redshifts ($z_{\text{median}} \simeq 0.2$) – as tracers of the underlying dark matter field. It will measure ~ 30 million galaxy and quasar redshifts to obtain precise measurements of the Baryon Acoustic Oscillation (BAO) features, Redshift-Space Distortions (RSD), as well as the full galaxy power spectrum, increasing the Dark Energy Task Force (DETF) [12] Figure of Merit (FoM) to above 700 when all these measurements are used [13]. Such unprecedented measurement will make DESI an ideal experiment to study fundamental sciences, such as constraints on neutrino mass, parameters of inflation, understanding of the cosmic acceleration, and finally, cosmological tests of gravity theories.

In this paper, we consider two representative examples of modified gravity models: chameleon $f(R)$ gravity [14, 15] and the 5D brane-world Dvali-Gabadadze-Porrati (DGP) model [16] (we look at the normal branch of the DGP model, nDGP). Those models can be considered as a subclass of a general class of modified gravity models, named as *Actions of Effective Field Theories* [17]. However, the chosen models serve as ideal testbeds of gravity on cosmological scales for a few reasons: (i) they can potentially realise the screening mechanisms [18–20] to pass local gravity tests and leave observable signatures on much larger, cosmological, scales – indeed, each of them represents a class of screening mechanisms, the chameleon and Vainshtein mechanism, (ii) the cosmological behavior of these models – with varying parameters – qualitatively represents that of various other classes of models, (iii) these models have been studied in great details to date, and simulations for them, which are essential to test them using galaxy surveys, have been developed and matured (see below). As a result, by studying them in detail, we hope to understand and to quantify the constraints that future astronomical observations would place on deviations from GR at astrophysical and cosmological scales, and therefore test the validity of GR in a completely different regime from traditional tests.

In the standard cosmological scenario, the formation and growth of structure is driven by hierarchical gravitational collapse: small structures collapse first, which over time merge and grow larger and more nonlinear. The galaxies that we observe today mostly reside in regions with highly nonlinear density, which cannot be adequately described using linear perturbation theory. Numerical simulations are a more reliable tool for predicting the clustering of matter in such regimes. In the MG models studied here, the enhanced gravity means that structures become more nonlinear. However, as far as simulations are concerned, the main effect these models introduce is screening, which is an inherently nonlinear phenomenon. This is reflected in the highly nonlinear equations governing the behavior of the fifth force, that need to be solved accurately in simulations. For example, linearizing the equations can cause non-negligible errors on length scales as large as $k \sim \mathcal{O}(10^{-2}) \text{ hMpc}^{-1}$ [21] in an otherwise fully consistent simulation. Other approximations that allow to simplify simulations significantly, such as the quasi-static approximation, were shown to be valid in the cosmological regime [22]. Such considerations have motivated the developments of a wide array of numerical tools to simulate the structure formation in modified gravity models, including complete simulation codes [23–38] (see [39] for a detailed comparison of different modified codes) and approximate methods [40–43].

Recent progresses in optimizing the simulation codes [44, 45] and in developing faster substitutes [43] have largely been motivated by the desire to keep up with the rapidly increasing demand to match the size, resolution and variety of observables of ongoing and future galaxy surveys, including DESI. However, to compare theoretical studies with the real observations of galaxy

catalogs, there are various intermediate steps, all of which could affect the reliability of the final tests of gravity by introducing sources of uncertainty.

The main objective of this paper is to identify the key issues in using DESI observations to constrain gravity models, which can help us move to the next stage of preparation for DESI science cases. Redshift-space distortions provide a promising way to constrain modified gravity models through the measurement of the growth rate. The two MG models considered in this paper have been constrained by RSD measurements from Baryon Oscillation Spectroscopic Survey (BOSS) [46–48]. We consider model parameters that are consistent on the 2σ level with these measurements. In this paper, we will focus on novel estimators to constrain modified gravity models exploiting information on non-linear scales. A main objective is to quantitatively assess the potential constraining power of these various estimators in testing the two classes of modified gravity models mentioned above. To this end we consider three main categories of estimators: (i) two-point and N-point statistics and marked statistics of galaxy catalogs, (ii) statistics that employ velocity information, including redshift space distortions (RSD) and escape velocity, and (iii) statistics relying on synergy with weak gravitational lensing (WL) data. This enables us to make theoretical predictions of various estimators based on the same set of mock galaxy catalogs, and hence allows a like-for-like comparison. Where possible, we consider different designs in the same estimator class; an example is the marked galaxy 2-point correlation function (2PCF), which quantifies the correlations of galaxy internal and external properties (marks) [49], where we will test different definitions of marks. Following this philosophy, we have made public the mock galaxy data used in this paper and encourage the community to use these data in their analyses so that their results can be compared with those from this work. The data can be found at this webpage¹.

The second main objective is to provide a guidance on future simulations and mock requirements for modified gravity tests using galaxy clustering data from the likes of DESI and Euclid. In this paper we will mainly use two sets of simulations: (i) the ELEPHANT simulations described in [50] with a simulation box size $L_{\text{box}} = 1024h^{-1}\text{Mpc}$ and particle number $N_p = 1024^3$, and (ii) the LIMINALITY simulation described in [51] with $L_{\text{box}} = 64h^{-1}\text{Mpc}$ and $N_p = 512^3$. The former have been used to construct halo occupation distribution (HOD) galaxy mocks which match the number densities of luminous red galaxies (LRGs) in current galaxy surveys, while the latter is used to build mock galaxies with similar number densities as DESI bright galaxy survey (BGS), using subhalo abundance matching (SHAM). The simulations considered in this paper do not cover the volume of the full DESI survey, however they do have number densities that are comparable to the full BOSS LRG sample². Recognizing the high cost of large volume, high resolution simulations, this paper is a preliminary study to determine the statistical potential to test gravity for data with the richness of DESI, to help frame simulation specifications and model choices, before embarking on future, heavy computational efforts for survey-tailored full scale simulations.

The final objective is to make an assessment of the statistical and, where possible, systematical uncertainties associated with the individual estimators. For most estimators, we use five independent realizations of simulations and mock galaxy catalogs to estimate the sample variance, while theoretical covariance matrices are used in other cases (e.g., galaxy galaxy lensing). This analysis is intended to set the stage for more comprehensive, quantitative estimate of the impact of systematic errors, beyond the scope of this paper.

The plan of this paper is as follows. In Section II we briefly review the modified gravity models studied in this work, focusing on their key differences from ΛCDM , the simulations for these models, and the dark matter halo and galaxy catalogs used for the analyses. In Section III we have a number of subsections, each of which focuses on a particular estimator and studies the potential of using such estimators from DESI galaxy survey data to constrain the theoretical models. In Section IV, the conclusions of the analyses and implications for future work are summarized.

II. MODELS, SIMULATIONS AND GALAXY CATALOGS

In this section we introduce the modified gravity models that are exemplified in this paper – $f(R)$ gravity and the nDGP model. These are two of the most well-studied modified gravity models, and they represent two main classes of screening mechanisms, thin-shell chameleon [18, 19] and Vainshtein [20] screening. Cosmological simulations for these models have been carried out by various groups, and we will briefly review the simulations to be used in this paper. Then we will introduce the mock galaxy catalogs constructed using these simulations, which will be used in the calculation of the various estimators in the next section. More details of these models, along with the simulation techniques for them, can be found in [52], and more comprehensive reviews of these models and their cosmological tests can be found in the recent review papers, e.g., [10, 53].

¹ <http://icc.dur.ac.uk/data/>

² Fig. 3.8 DESI Technical Design Report Part I: Science, Targeting, and Survey Design, July 27, 2015

A. Modified gravity models

Ever since the discovery of the accelerated expansion of the Universe, many dark energy and modified gravity models have been proposed, [8–11]. The recent detection of a binary neutron star merger by gravitational waves (GW170817) associated with various electromagnetic counterparts [3, 4] has had a substantial impact on this field [54–57]. In the four-dimensional scalar tensor theory, the high-precision measurement of the equivalence between the speeds of photons and gravitational waves leaves very limited scope for a modified gravity model that naturally produces self-acceleration without a cosmological constant [54] while being compatible with other observations such as the integrated Sachs Wolfe effect (ISW; e.g., [58–61]). The two models that we consider in this paper – $f(R)$ gravity and nDGP model – are compatible with this measurement. The former is a leading example of the class of models in which gravity is modified by a scalar field that has a self-interaction described by a nonlinear potential and interactions with other matter species, while the latter belongs to the category of models where the self-interaction of the scalar field is caused by derivative couplings.

Before moving on to the details of these models, it is worthwhile to make a couple of comments:

1. in both models, the accelerated expansion of the Universe is not driven by the modification of gravity itself, but has to be explained by an extra cosmological constant or dark energy species. The models may lose some of their original appeals as a consequence. However, they provide useful testbeds for cosmological tests of gravity, which is the prevailing topic of this paper.
2. while representative, the two models cannot be expected to cover all behaviors of potential MG models. As a few examples, in both models gravity can only be enhanced rather than weakened, and in $f(R)$ gravity the maximum of this enhancement (which is $1/3$) cannot be tuned freely; the expansion history of $f(R)$ models has to be practically indistinguishable from that of Λ CDM for the model to pass stringent Solar System tests of gravity [62–64]. Our choices of models are therefore inflexible in some aspects, and such inflexibility is the price to pay for having a manageable number of models to study in greater depth. We believe that an analysis of various estimators predicted by these two models could offer insight into the cosmological tests of more general models and therefore serve as a starting point to prepare for the tests of gravity using galaxy surveys such as DESI. One alternative to the model-by-model approach here is to use a general parameterization of MG models (see, e.g., [65], for a review), but this has the disadvantages of having a significantly larger parameter space to explore and – more importantly – some of the popular parameterization schemes only mimic the full models in certain (e.g., linear perturbation) regimes and therefore can not be relied on for a fully nonlinear study. Nevertheless, when describing the models below, it is useful to know their links to and position in certain parameterization schemes which are widely used in the literature, for qualitative comparisons.

A popular parameterization scheme for modified gravity is the μ - γ parameterization, which we briefly introduce here as we shall try to make connections to the two MG models studied in this paper. In the Newtonian gauge, the line element is given by

$$ds^2 = -[1 + 2\Psi(\mathbf{x}, t)] dt^2 + a^2 [1 - 2\Phi(\mathbf{x}, t)] \delta_{ij} dx^i dx^j, \quad (1)$$

where $a = a(t)$ is the scale factor, δ_{ij} is the Kronecker δ and Ψ, Φ , which are functions of space and time, denote the gravitational potentials. We introduce parameterization functions, $\mu(k, t)$ and $\gamma(k, t)$, in the following linearized Einstein equations, now written in k space,

$$k^2 \Psi = -4\pi \mu(k, t) G \bar{\rho}_m a^2 \delta_m, \quad (2)$$

$$\Psi = \gamma(k, t)^{-1} \Phi \quad (3)$$

where δ_m is the matter density contrast and $\bar{\rho}_m$ is the mean matter density ($\delta \rho_m = \bar{\rho}_m \delta_m$); these equations are written in Fourier space so that Φ, Ψ, δ_m should be understood as the Fourier transforms of the corresponding fields. Notice that $\mu = \gamma = 1$ for GR and deviations of these functions from 1 could be signatures of modified gravity. This model has been constructed under the assumption of statistical cosmic homogeneity and isotropy. This is a well motivated assumption, since even though we cannot prove it, there are good observational evidences [66, 67].

1. $f(R)$ gravity

In this model, the standard Einstein-Hilbert action is extended to include an additional function of the Ricci curvature $f(R)$ (see [68] for a review). This theory is equivalent to the scalar-tensor theory with a potential for the scalar field that is determined by the function $f(R)$, and which realizes the chameleon screening mechanism [18, 19] (see [53] for a recent review). Deviations from standard GR can be suppressed in regions with deep gravitational potential by the chameleon mechanism. In regions where the potential is shallow, the theory is in the unscreened regime, in which massive particles experience an additional fifth force mediated by the scalar field $f_R \equiv df(R)/dR$, whose strength can be comparable to that of standard gravity. In $f(R)$ gravity,

the maximum strength of the fifth force is $1/3$ of Newtonian gravity, while in general chameleon models this can be a free model parameter.

In the quasi-static³ and weak-field limits, the equations for the Newtonian potential Φ and the scalar field f_R are given by

$$\nabla^2 \Phi = 4\pi G a^2 \delta \rho_m - \frac{1}{2} \nabla^2 f_R, \quad (4)$$

$$\nabla^2 f_R = -\frac{a^2}{3} \delta R - \frac{8\pi G a^2}{3} \delta \rho_m, \quad (5)$$

where $\delta \rho_m$ and δR are the matter density perturbations and the perturbation of the Ricci curvature respectively. δR can be written as a function of f_R , which plays the role of the potential for f_R . As a specific example, we consider one of the models proposed by Hu & Sawicki (HS) [69], in which the functional form of $f(R)$ is given by

$$f(R) = -6\Omega_\Lambda H_0^2 + |f_{R0}| \frac{\bar{R}^2}{R}, \quad (6)$$

where H_0 is the present-day Hubble parameter and $f_{R0} < 0$ is the value of f_R today. Note that the HS $f(R)$ model has another free model parameter, an integer n , which has been set to 1 in this paper; having different values of n could lead to quantitative differences in the model behaviour, but we expect the case of $n = 1$ to be representative for the qualitative properties of the fifth force and screening. f_{R0} is conventionally used as a parameter of this model to describe the deviation from Λ CDM, with smaller $|f_{R0}|$ values denoting weaker deviation from GR, as can be seen from Eqs. (4,6). For small $|f_{R0}|$ the background expansion history can be approximated by that in Λ CDM and $\bar{R}$ can be identified with the background Ricci curvature today in Λ CDM,

$$\bar{R} = 3\Omega_m H_0^2 \left(1 + 4 \frac{\Omega_\Lambda}{\Omega_m} \right). \quad (7)$$

Hereafter we will use this approximation so that there is no difference between $f(R)$ and Λ CDM in the background. As mentioned above, $f(R)$ gravity models, and chameleon models in general, have to closely resemble Λ CDM background expansion history in order to have a sufficiently efficient chameleon screening mechanism to pass the solar system tests of gravity.

The chameleon screening mechanism works because the scalar field has a position-dependent mass m , give by

$$m^2 \simeq \frac{1}{3f_{RR}} \equiv \frac{1}{3} \left| \frac{d^2 f(R)}{dR^2} \right|^{-1}, \quad (8)$$

so that the fifth force it mediates is of Yukawa type with its potential of the form $\sim \exp(-mr)/r$, where r is the distance between two bodies. In other words, unlike the standard Newtonian force, the fifth force has a finite range characterized by the Compton wavelength m^{-1} , beyond which it decays exponentially. In high-density regions, m is large and the fifth force is exponentially suppressed, causing the screening. An important property of $f(R)$ gravity is its prediction of scale-dependent linear growth rate: this is because even in the linear regime, where m can be replaced by its background value $\bar{m}(a)$, the fifth force still has a finite range within which the growth of matter density perturbations is suppressed.

In the μ - γ parameterization framework described in Eqs. (2,3), $f(R)$ gravity can be described as

$$\mu = \frac{4 + 2\omega}{3 + 2\omega}, \quad (9)$$

$$\gamma = \frac{1 + \omega}{2 + \omega}. \quad (10)$$

with

$$\omega = \omega(k, a) = \frac{3a^2 \bar{m}^2}{2k^2}, \quad (11)$$

where $\bar{m} = \bar{m}(a)$ is the background scalar field mass mentioned above. These equations indicate that, in the large-scale limit where $a\bar{m} \gg k$ or $\omega \gg 1$, $\mu, \gamma \rightarrow 1$ and GR is recovered, while $\mu \rightarrow 4/3$ in the opposite limit (therefore the scale dependence). As mentioned earlier, this parametrization only works for the linear perturbations regime. For models with larger $|f_{R0}|$, such as $|f_{R0}| = 10^{-4}$ used in this paper, chameleon screening is inefficient and this parameterization can be a good description, but for smaller values of $|f_{R0}|$ we expect it to miss some important effects of screening [28].

³ This means that the time evolution of the gravitational potentials is assumed to be small compared to the Hubble time so one can assume the derivatives of the potentials to be zero for sub-Hubble-horizon scales. For scalar-tensor theories, this approximation also means that one neglects the time derivatives of the fluctuations in the scalar field at scales below the scalar perturbation sound horizon.

2. DGP model

The DGP model [16] is a braneworld model where standard model particles are confined to a four-dimensional brane in a five-dimensional spacetime. This model has one parameter r_c of length dimension, below which gravity becomes four dimensional. On scales smaller than r_c and in the quasi-static and weak-field limits, the equations for the Newtonian potential and the scalar field φ that represents the brane-bending mode is given by

$$\nabla^2\Phi = 4\pi G a^2 \delta\rho_m + \frac{1}{2}\nabla^2\varphi, \quad (12)$$

$$\nabla^2\varphi + \frac{r_c^2}{3\beta a^2} \left[(\nabla^2\varphi)^2 - (\nabla_i\nabla_j\varphi)^2 \right] = \frac{8\pi G a^2}{3\beta} \delta\rho_m, \quad (13)$$

where

$$\beta = \beta(a) = 1 + 2Hr_c \left(1 + \frac{\dot{H}}{3H^2} \right). \quad (14)$$

Note that we assumed the normal branch of the DGP model. This branch requires an additional dark energy to explain the cosmic acceleration, but does not suffer from the instabilities of the self-accelerating branch, see, e.g., [70–72]. In order to make the comparison between Λ CDM and $f(R)$ models easier, we tune the dark energy equation of state so that the background expansion history is identical with that of Λ CDM [73].

In the DGP model, massive particles also feel a fifth force – as can be seen from Eq. (12) – whose potential is governed by the scalar field φ . The model realizes the so-called Vainshtein screening mechanism [20], by which the fifth force can be suppressed in regions where the second derivatives of the scalar field φ ($\nabla^2\varphi$) is large. This can be seen from Eq. (13): in regions where $\nabla^2\varphi$ is small, nonlinear terms such as $(\nabla^2\varphi)^2$ and $(\nabla_i\nabla_j\varphi)^2$ are subdominant so that $\nabla^2\varphi \sim \nabla^2\Phi$ for $\beta \sim \mathcal{O}(1)$; while in regions where $\nabla^2\varphi$ is large, the nonlinear terms are dominant and so $|\nabla^2\varphi| \ll |\nabla^2\Phi|$.

Unlike in $f(R)$ gravity, in the DGP model the linear growth rate is scale-independent as the scalar field is massless. Detailed comparisons between nDGP and $f(R)$ gravity, in particular Vainshtein and chameleon screening mechanisms can be found in [74, 75]. This feature can again be seen if we map this model to the μ - γ parameterization above, which corresponds to Eqs. (2,3,9,10) with

$$\omega = \omega(a) = \frac{3}{2} [\beta(a) - 1]. \quad (15)$$

B. N-body simulations and halo/galaxy catalogs

In this subsection we outline the simulations used in the analysis and present some of the statistics derived from the simulated dark matter halo and mock galaxy catalogs.

1. N-body simulations

The $f(R)$ simulations were performed using an optimized version of the ECOSMOG code [30, 45], and the nDGP simulations were done using an optimized version of the ECOSMOG-V code [32, 44]. Both codes are extensions to the publicly-available N-body and hydrodynamical simulation code RAMSES [76], with new routines added to solve the scalar field and modified Einstein equations in the MG models. These codes are parallelized using MPI and use the adaptive-mesh-refinement (AMR) technical to achieve high resolution in overdense regions where the requirement for the force resolution is high and the screening effect is strong. The simulations start with a uniform (domain) grid with $N_{\text{dc}}^{1/3}$ cells a side which covers a cubic box of size L_{box} . The cells are refined (split into eight daughter cells) if the number of particles contained in them grows over some pre-set threshold (N_{ref}), in such a way as to hierarchically refine the domain grid by adding higher-resolution meshes.

The cosmological and technical parameters of the simulations are given in Table I. The former are chosen as the best-fit Λ CDM parameters of the WMAP9 cosmology [77]. The simulations were started at $z_{\text{ini}} = 49$, from initial conditions generated using Zel’dovich approximation with the publicly available MPGRAFIC code [78]. Because the $f(R)$ and nDGP model parameters are chosen such that they only deviate from Λ CDM non-negligibly at late times, at z_{ini} the modified gravity effect can be neglected, and so all our simulations started from exactly the same initial condition. In order to estimate the effect of sample variance, we have used five independent realizations of boxes, whose initial conditions differ only in their random phases of the density field.

parameter	physical meaning	value
Ω_m	present fractional matter density	0.281
Ω_Λ	$1 - \Omega_m$	0.719
h	$H_0/(100 \text{ km s}^{-1} \text{ Mpc}^{-1})$	0.697
n_s	primordial power spectral index	0.971
σ_8	r.m.s. linear density fluctuation	0.820
$ f_{R0} $	Hu & Sawicki $f(R)$ parameter	0 (GR) 10^{-6} (F6), 10^{-5} (F5), 10^{-4} (F4)
$H_0 r_c$	nDGP parameter	5.0 (N5), 1.0 (N1)
L_{box}	simulation box size	$1024 h^{-1} \text{ Mpc}$
N_p	simulation particle number	1024^3
m_p	simulation particle mass	$7.78 \times 10^{10} h^{-1} M_\odot$
N_{dc}	domain grid cell number	1024^3
N_{ref}	refinement criterion	8

TABLE I. The parameters and technical specifications of the N-body simulations of this work. Note that the refinement criterion N_{ref} is the same for refinement levels, and that σ_8 is for the Λ CDM model and only used to generate the initial conditions – its value for $f(R)$ gravity is different but is irrelevant here.

We shall refer to these different realizations as ‘Box 1’ to ‘Box 5’. For $f(R)$ gravity we ran three variants of the HS model, with $\log(|f_{R0}|) = -6, -5, -4$ (with increasing deviation from GR), which we shall refer to as F6, F5 and F4 respectively in what follows; note that GR, or Λ CDM, is a special subcase of $f(R)$ gravity with $f_{R0} = 0$. For nDGP we consider two variants with $H_0 r_c = 5.0, 1.0$ (again with increasing deviation from GR) and refer to them as N5 and N1 respectively; Λ CDM is a special case of nDGP with $H_0 r_c = \infty$.

To gain some quick insight into the qualitative behavior of the different models, we show the predictions of some cosmological quantities here.

Figure 1 shows the matter power spectra of the MG models at two redshifts, $z = 0$ (left) and $z = 0.5$ (right), as well as their relative difference with respect to Λ CDM (bottom subpanels). For $z = 0$, we show the results for Box 1 only, while for $z = 0.5$ we show the results for all boxes (the different line styles are for individual realizations, to highlight the good agreement between them). The line styles and color scheme in Fig. 1 (see the legend and caption for more details) will be used in other plots across this paper.

Figure 1 confirms that in $f(R)$ gravity the linear growth rate is scale dependent while in nDGP it is scale independent, as can be seen from the bottom subpanels at $k \lesssim 0.1 h \text{ Mpc}^{-1}$, where linear theory works relatively well for both models. The amount of deviation from Λ CDM in the MG models follows the expected order, and in all models it increases with time as the effect of enhanced gravity accumulates. In both F4 and N1, the enhancement of $P(k)$ starts to decrease at $k \sim 0.8 h \text{ Mpc}^{-1}$. For F4 this is not a signature of chameleon screening — but is related to the internal structures of halos [21] — as can be realized from the facts that F5 and F6, which both have stronger screening effect, actually do not show a similar decrease of $\Delta P/P_{\text{GR}}$ at that scale. For N1, in contrast, the decrease of $\Delta P/P_{\text{GR}}$ is a real signal of Vainshtein screening, which very efficiently suppresses the fifth force near and inside halos.

2. Halo catalogs and halo mass functions

Dark matter halos and the self-bound substructures associated with them are identified using the publicly-available ROCKSTAR halo finder⁴[80]. ROCKSTAR uses the six-dimensional phase-space information from the dark matter particles to identify halos. Note that, in principle, the presence of the fifth force in $f(R)$ gravity⁵ would require a modification to the unbinding procedure in ROCKSTAR, but the effect is expected to be small [81] and so we use identical versions of ROCKSTAR for GR and MG simulations, and we also use the same halo mass definition, M_{200c} , which is the mass enclosed in R_{200c} , the radius from halo center within which the mean mass density is 200 times the critical density $\rho_{\text{crit}}(z)$. In this paper, we make use of only independent (‘main’) halos, and not their substructures, partly because of the relatively low resolution of our simulations.

Figure 2 shows the cumulative halo mass function (HMF) of all models at $z = 0$ (left panel) and $z = 0.5$ (right panel). The bottom subpanels show the relative differences between the MG models and GR. In the case of $z = 0.5$, there is good agreement between the five realizations (the different line styles) again.

As the halo catalogs are the starting point of the mock galaxy catalogs to be described below, it is useful to note some main features and their physical origins. Although $\Delta n/n_{\text{GR}}$ is smaller at higher redshift, the qualitative features are the same in both

⁴ <https://bitbucket.org/gfcanford/rockstar>.

⁵ In nDGP, the fifth force is strongly suppressed in halos and we can safely neglect its effect.

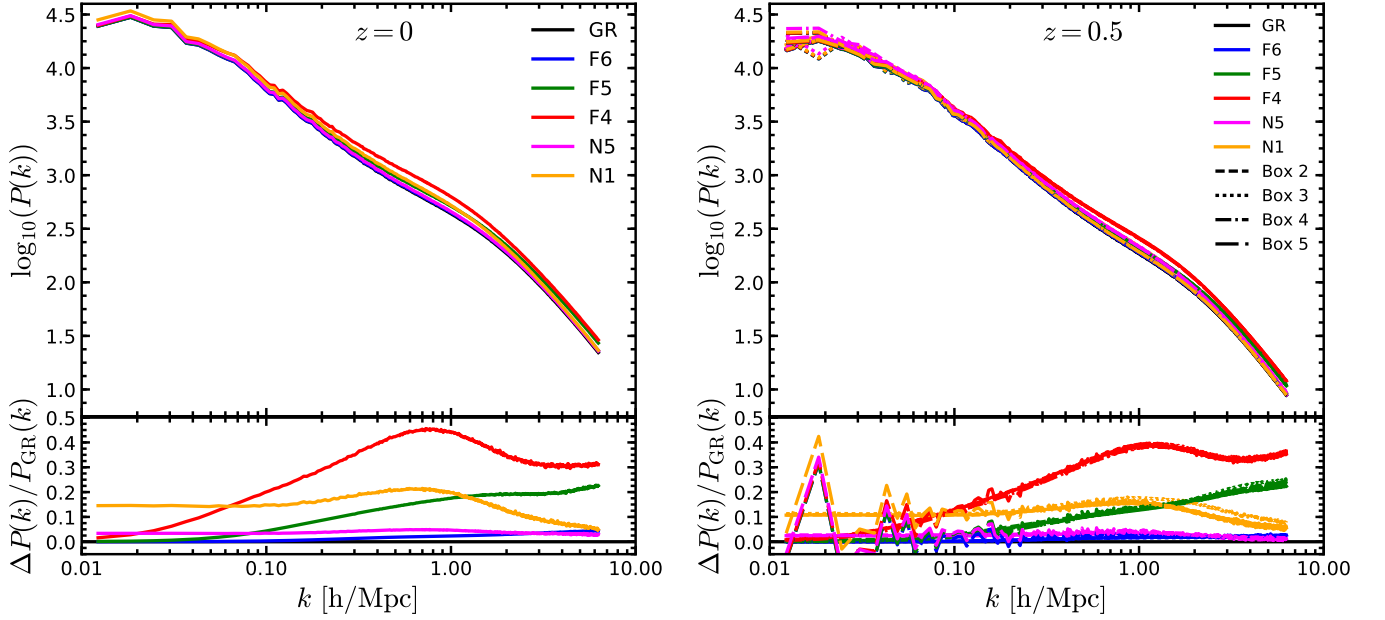


FIG. 1. (Color Online) The matter power spectra of the different models compared in this work (upper panels) and their relative differences of the modified gravity models with respect to Λ CDM (lower panels). *Left panels*: results at $z = 0$. *Right panels*: results at $z = 0.5$, for all five boxes. This plot sets the convention of line styles that will be used in the rest of this paper: black, blue, green, red, magenta and orange respectively represent GR, F6, F5, F4, N5 and N1, while solid, dashed, dotted, dash-dotted and long-dash-dotted lines represent results from Boxes 1, 2, 3, 4 and 5. This convention will be used in other plots unless otherwise stated. Note that for $z = 0$ we only show the results for Box 1, while for $z = 0.5$ we show all five boxes – the upper panel shows the average $P(k)$ from these boxes, while in the lower panel we plot individual curves of $\Delta P/P_{\text{GR}}$ for the five boxes, in order to show the agreement between the different realizations. All power spectra have been measured using the publicly-available code POWMES [79]

redshifts. Of the $f(R)$ variants, in F6 the difference is strongly suppressed by the chameleon mechanism except for the smallest halos for which the screening is weak; this feature remains in F5, though the deviation from GR now starts at higher halo mass; for F4, the screening is essentially non-existent, leading to a significant increase in the number density of the most massive halos resolved in the simulations ($M_{200c} > 10^{14.5} h^{-1} M_{\odot}$). Due to the faster mergers of small halos to form larger ones, F4 actually produces fewer halos in the mass range $10^{13} \sim 10^{13.5} h^{-1} M_{\odot}$ than F5. The nDGP models are qualitatively similar to F4, but with smaller differences from GR. The Vainshtein mechanism does not prevent more massive halos from forming in N1 and N5 as compared with GR, because the growth of halos is largely determined by how much matter the halos can accrete from their surroundings: while the Vainshtein mechanism is efficient in suppressing the fifth force close to and inside the halos, gravity can still be stronger than in GR within regions of size $\mathcal{O}(10)h^{-1}\text{Mpc}$ from halos, which means that the largest structures end up growing more by accreting more matter from further away.

As will be discussed next, the differences in the HMFs of the different models means that we have to slightly tune the galaxy populating scheme to obtain galaxy catalogs with the same desired clustering properties.

3. Mock galaxy catalogs

To map the halo catalogs to a corresponding galaxy distribution, we populate halos with galaxies using the Halo Occupation Distribution (HOD) method [82, 83], in which it is assumed that the probability for a halo to host a certain number of galaxies can be computed through a simple functional dependence on the mass of the host halo. We use the form of the HOD suggested by [84], in which the mean number of central galaxies, $\langle N_{\text{cen}}(M) \rangle$, and the mean number of satellite galaxies, $\langle N_{\text{sat}}(M) \rangle$, in a halo of mass M , are given respectively by:

$$\begin{aligned} \langle N_{\text{cen}}(M) \rangle &= \frac{1}{2} \left[1 + \text{erf} \left(\frac{\log M - \log M_{\text{min}}}{\sigma_{\log M}} \right) \right], \\ \langle N_{\text{sat}}(M) \rangle &= \langle N_{\text{cen}} \rangle \left(\frac{M - M_0}{M_1} \right)^{\alpha}, \end{aligned} \quad (16)$$

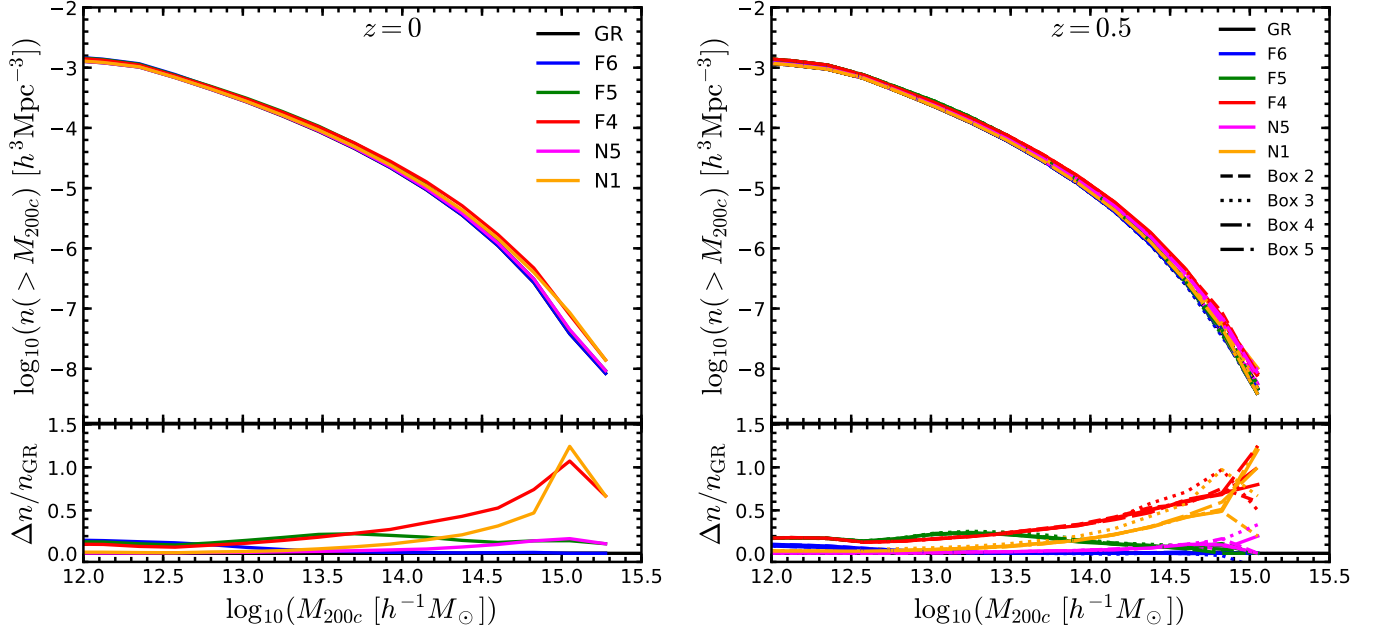


FIG. 2. (Color Online) The cumulative halo mass function for the various models simulated here, at $z = 0.0$ (left panel) and $z = 0.5$ (right panel). The bottom panels show the relative difference between the various modified gravity models and Λ CDM. For the $z = 0.5$ plot we have used all 5 realizations, and it shows that the model differences are not sensitive to the specific realization. For the $z = 0$ plot we only used Box 1. The line colors and styles are the same as in Fig. 1.

where $M_{\min}$, M_0 , M_1 , $\sigma_{\log M}$ and α are free parameters of the HOD model. Once their values have been specified, the mean number of galaxies in a halo of mass M is then given by $\langle N(M) \rangle = \langle N_{\text{cen}}(M) \rangle + \langle N_{\text{sat}}(M) \rangle$. From Eq. (16), it can be seen that $M_{\min}$ and M_0 , respectively, denote the threshold halo mass required to host at least one central or one satellite galaxy. When placing HOD galaxies in halos, central galaxies are assumed to reside at the center of potential of their host halo. Satellites, on the other hand, are distributed between $[0, R_{200c}]$ of the host halo center, according to a Naranro-Frenk-White (NFW, [85, 86]) profile with the concentration of the host halo as computed by ROCKSTAR. Furthermore, central galaxies are assigned the center-of-mass velocity of the host halo, V_{CM} ; the velocity of a satellite galaxy is V_{CM} plus a perturbation along the x, y and z axes sampled from a 3D Gaussian distribution with a dispersion equal to the root-mean-squared (RMS) velocity dispersion of the host halo.

If the HOD catalogs in the MG models had been constructed using the same HOD parameters as in the Λ CDM model, there would generally be a difference of order 10-20% in the resulting number density and two-point correlation function (2PCF) of HOD galaxies, reflecting the MG effects on the halo abundance and clustering. Since there is only one observed Universe, if we do not know which cosmological model is the correct one, a more conservative way is to demand that all models make predictions that are consistent with observations. For this consideration, we have tuned the HOD parameters for the MG models to ensure that their resulting galaxy catalogs have roughly the same number densities and clustering properties as the corresponding Λ CDM catalogs. The assumption that MG models have different HOD parameters from GR is reasonable, because the evolution of the matter field and the assembly histories of galaxies are generally different in these models. This tuning of HOD parameters to fix galaxy clustering actually can help to remove one source of contamination when it comes to the model differences predicted by the various estimators to be studied below. As a result of this tuning, some estimators, such as the projected two point correlation functions, by construction cannot be used to discriminate the MG models from GR, and we need to find other ways to use the galaxy catalogs.

In practice, our tuning of MG HOD parameters was carried out using a Nelder-Mead simplex search through the 5-dimensional HOD parameter space. From a MG and a Λ CDM HOD catalogs, the projected galaxy 2PCFs, $w_p(r_p)$, were measured⁶ with the comoving projected separation range of $0.1 < r_p < 50 h^{-1} \text{Mpc}$. The RMS difference between two models was calculated with an identical weight of 1.0 for all $\log(r_p)$ bins. To ensure that the two models have similar galaxy number density $n(z)$, the fractional difference in the respective $n(z)$ values was also used in the calculation of the RMS difference, with a (somewhat

⁶ This was obtained by projecting the 3D redshift-space 2PCF $\xi(r_p, r_\pi)$, with r_p, r_π being respectively the galaxy pair separations transverse and parallel to the line of sight, using a projection depth of $90 h^{-1} \text{Mpc}$. $\xi(r_p, r_\pi)$ was measured using the Correlation Utilities and Two-point Estimation (CUTE) code [87], with the distant-observer approximation. For the main results in this paper we have used the z axis of the simulation box as the line-of-sight direction, but we have checked the HOD tuning when using the x and y axes as the lines of sight, and found very similar values of the best-fit HOD parameters.

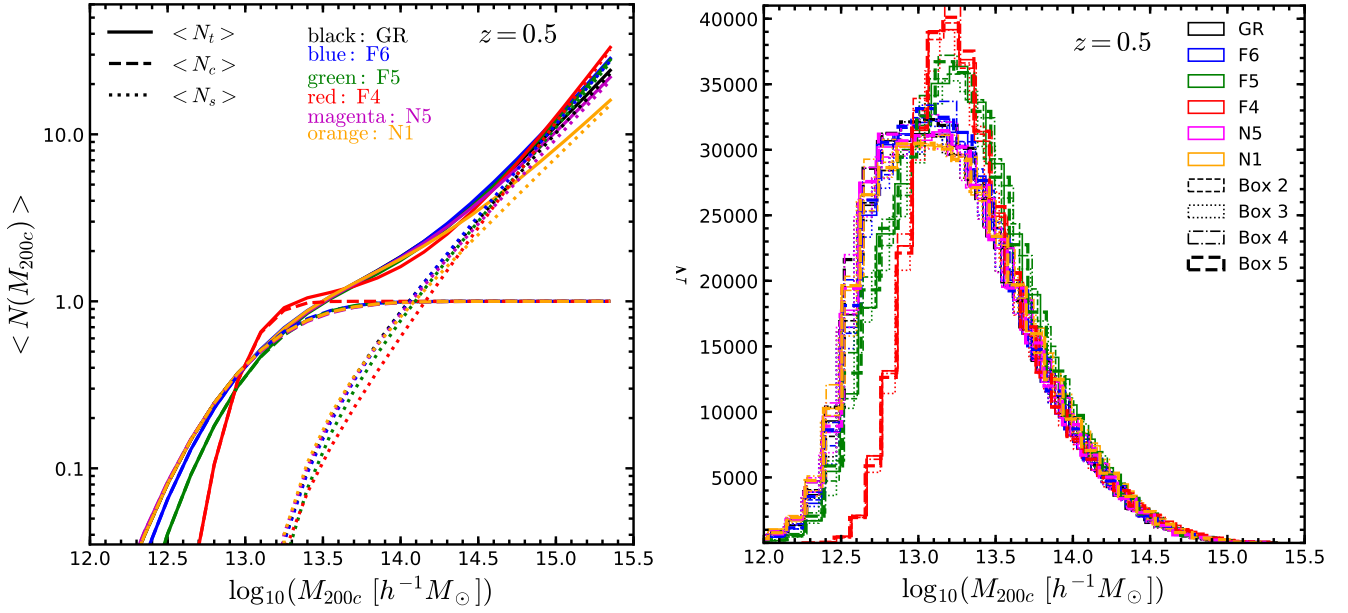


FIG. 3. (Color Online) *Left panel*: the mean galaxy number as a function of host halo mass for central (dashed), satellite (dotted) and all (solid) galaxies for Λ CDM and the five MG models. The figure is made using Eq. (16) with the tuned HOD parameters (cf. Table II) for the individual models. *Right panel*: histograms of galaxy numbers as a function of the host halo mass for all models and all realizations; this is essentially the product of the mean occupancy number (as shown in the left panel) and the halo mass function. The line colors and styles in the right panel are the same as in Fig. 1. All results are at $z = 0.5$.

arbitrary) weight of 8.0. The code then walked through the 5D HOD parameter space to look for the smallest RMS difference, and the search stopped if the value dropped below 1.5% (with the exception of N1 for which the minimum value the code found was 2.2%). As the estimators to be studied later in this paper are all evaluated at $z = 0.5$, we only did this tuning to produce $z = 0.5$ HOD galaxy catalogs.

This is a simplified and less rigorous approach in several ways. First, unlike a Markov chain Monte Carlo approach, the search only led to the ‘best-fit’ HOD parameters rather than their posterior distributions. Second, the fitting of HOD parameters did not involve real data; instead, we adopted the best-fit HOD parameter values taken from [88] for the CMASS data in the GR halo catalogs (all five realizations) to produce the Λ CDM HOD catalogs, and then tuned the HOD parameters for the MG models to match the Λ CDM results. Finally, often the HOD parameters are constrained simultaneously with cosmological parameters using a combination of different probes, while here we used a single probe (the projected galaxy 2PCF) to fix the HOD parameters and study other probes afterwards, as we wanted to use the same HOD catalogs to study a variety of probes. Note that for each MG model a single set of ‘best-fit’ HOD parameters were tuned and used in producing the HOD catalogs for all five realizations. To check the sensitivity of the physical results presented below to the way in which the best-fit HOD parameters were obtained, we made two additional checks by: (1) tuning the parameters so that the MG models match the Λ CDM prediction of the real-space 3D galaxy 2PCF $\xi_{gg}(r)$ instead of the projected 2PCF, and (2) tuning the parameters individually for each simulation realization of each MG model. In both cases, we found little difference from the default case in terms of the halo occupancy properties and the physical results of the various cosmological probes.

The left panel of Fig. 3 shows the mean halo occupancy numbers for the different models in Box 1 (with other boxes being in good agreement), where a complicated pattern can be observed. For example, in F4 N_{cen} is substantially higher than in other all models at $10^{13} \lesssim M/(h^{-1} M_{\odot}) \lesssim 10^{13.5}$ but decays much faster at $M \lesssim 10^{13} h^{-1} M_{\odot}$. N_{cen} in the two nDGP variants both agree very well with that in Λ CDM, because for massive halos N_{cen} is equal to 1 anyway, while for smaller halos these models have very similar HMFs, cf. Fig. 2. The right panel of Fig. 3 shows the histograms of the numbers of HOD galaxies in host halos of different masses, which is essentially the product of the mean occupancy number multiplied by the host halo mass function. We have plotted the results from all five boxes using different line styles, and a good agreement among them (for a given model) is visible. As expected, more galaxies reside in more massive halos in F4 and F5 than in the other models.

The right panel of Fig. 4 shows the projected galaxy 2PCFs $w_p(r_p)$ of the HOD galaxies for all models (top subpanel; average over five boxes), and the relative differences of the MG models from GR (bottom subpanel; individual boxes). This verifies that the HOD parameter tuning for the different MG models has served its purpose of making $w_p(r_p)$ in the different models agree

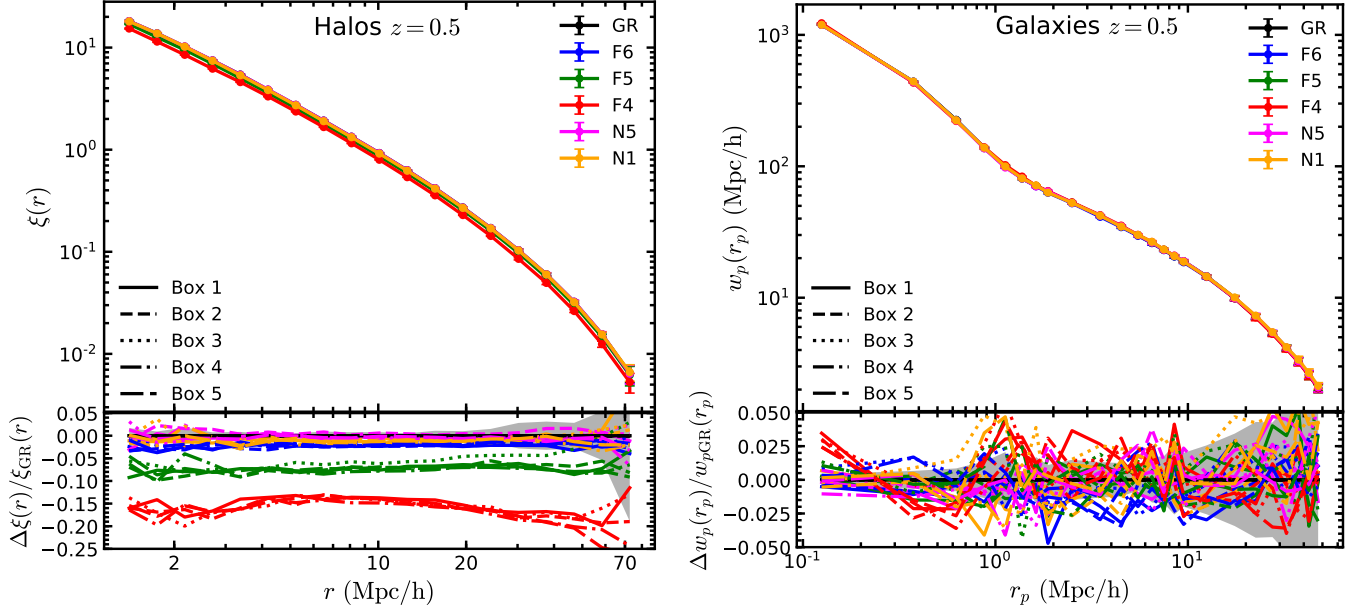


FIG. 4. (Color Online) *Left panels*: the 3D halo two-point correlation functions, $\xi_{hh}(r)$, for the six models, averaged over all five realizations, as a function of halo separation r (upper panel) and the relative differences with respect to Λ CDM (lower panel; the five simulation realizations are shown individually using different line styles as in Fig. 1). Note that only halos more massive than $10^{13}h^{-1}M_{\odot}$ are used, so that the halo number densities are different in the different models. *Right panels*: the same as the left panels, but for the projected correlation functions of HOD galaxies, $w_p(r_p)$. All results are at $z = 0.5$. The error bars in the top panels and the shaded regions in the bottom panels are the standard deviation of 5 Λ CDM realizations.

	$\log(M_{\min}/[h^{-1}M_{\odot}])$	$\log(M_1/[h^{-1}M_{\odot}])$	$\log(M_0/[h^{-1}M_{\odot}])$	$\sigma_{\log M}$	α
GR	13.09	14.00	13.077	0.596	1.0127
F6	13.090	14.011	13.035	0.552	1.0766
F5	13.132	14.049	13.061	0.512	1.1008
F4	13.030	14.132	12.939	0.260	1.2393
N5	13.098	14.017	13.054	0.610	0.9930
N1	13.126	14.036	13.111	0.633	0.8967

TABLE II. The tuned HOD parameter values (columns 2-6) for the GR, $f(R)$ (F6, F5, F4) and nDGP (N5, N1) models at $z = 0.5$ and different realizations (Boxes 1-5). Note that all GR simulations use the same set of HOD parameters, which are taken from the best-fit parameters using CMASS data [88]. The galaxy number density for all HOD catalogs is around $n_g = 3.2 \times 10^{-4}[h^{-1}\text{Mpc}]^{-3}$.

within ~ 1.5 - 2.5% , and that the same HOD parameters applied to different simulation realizations do give convergent results of $w_p(r_p)$.

As a comparison, we also show, in the left panel of Fig. 4, the same results but the 2PCFs of dark matter halos with $M_{200c} \geq 10^{13}h^{-1}M_{\odot}$. Here we see that $f(R)$ models generally have weaker clustering than GR, since for the same $M_{200c,\min}$ the model with stronger gravity would have a higher halo number density. This means that some of its halos correspond to initial density peaks too small to form halos with $M \geq M_{200c,\min}$ in a weaker gravity model. Note again the good agreement of halo 2PCFs in the different realizations.

It is interesting to see that very different gravity models can give nearly identical HOD galaxy number densities and clustering, showing the flexibility of the 5-parameter HOD model used here. This highlights the fact that galaxy-halo connection can bring a main theoretical uncertainty in using galaxy clustering to test models. Note also that in screened MG models one may expect an additional assembly bias because the strength of gravity can vary from region to region, and that might lead to corresponding variations of N_{cen} and N_{sat} ; we shall not pursue this extra complication here.

4. Fast simulations and mocks

A rigorous analysis of the estimators and their capability to constrain models requires to generate a substantial number of realizations of simulations to explore the statistical properties of the estimators and characterize the significance of the differences observed. Furthermore, for comparing models one also needs the covariance matrices for the different estimators. Since full MG simulations introduced above are expensive, it is not practical to use them for that purpose. Thus, it is helpful to use approximate methods to generate a large number of simulations.

One of the existing fast simulation codes is MG-COLA (COMoving Lagrangian Acceleration) [43], which is an approximation method based on second order Lagrangian Perturbation Theory (2LPT), to generate quick mock catalogs for the different MG models considered. MG-COLA has flexibility to allow a simulation to be run with varying numbers of time steps: the use of very few time steps gives the predictions of 2LPT, while the use of more time steps can improve the accuracy towards that of a full simulation. This flexibility, however, means that *a priori* we do not know the smallest number of time steps to be used in order to meet our target accuracy. Therefore, in order to use MG-COLA simulations, we need to calibrate them using full simulations to make sure that the setup can reproduce the results of the latter on the scales of interest, and this step should be performed separately for each estimator of interest to us. Apparently, this will be a substantial effort that goes beyond the scope of this paper.

As a quick check, we have run 20 simulations for each of the six models considered in this paper, using the same specifications as given in Table I except that the particle number is 8 times higher (2048^3). With 50 code time steps (30 steps for $30 \leq z \leq 0.5$ and 20 up to $z = 0$), each MG-COLA simulation takes about 300 core hours, which is a factor of $\mathcal{O}(100)$ faster than full simulations which take about 350 coarse and $\sim 10,000$ fine time steps. The huge saving of computing time partly comes from the fact that full simulations spend a substantial fraction of their running time solving the nonlinear field equations for the scalar fields on refinements. The HMFs of these fast simulations agree with those from the full simulations within $\sim 20\%$ between $10^{12.5} \sim 10^{14.5} h^{-1} M_\odot$, but the agreement worsens rapidly at $> 10^{14.5} h^{-1} M_\odot$. The two-point correlation functions of halos from the fast simulations agree with those of the full simulations within 15% between $3 \sim 30 h^{-1} \text{Mpc}$, and the discrepancy increases quickly for $r > 30 h^{-1} \text{Mpc}$. The disagreements between $3 \sim 30 h^{-1} \text{Mpc}$ are at a similar level to the model differences that we are interested in tests, and so are not suitable for estimating the observables or their covariances. This suggests that the fast simulations may need to be run at higher time, space or mass resolutions in order to agree with full simulations, especially for probes that employ small-scale information such as within dark matter halos. Future developments in this area, e.g., in optimising existing codes, developing possible new fast codes, or proposing (semi)analytical methods for covariance estimation, will therefore be of great importance and much welcomed effort.

III. ESTIMATORS

Having introduced the theoretical models, simulations, and halo and mock galaxy catalogs in the previous section, we can now move on to showing how the various estimators studied in this paper predict different signals for the different gravity models. Because of the restrictions in the number densities and volumes of our mock galaxies, and due to the various systematic effects which are not accounted for, we shall not conduct a rigorous statistical analysis of the future constraints from DESI, but instead focus on understanding the model behaviors and signatures, quantifying roughly the significance at which different models can be statistically distinguished, and identifying future simulation and analysis needs.

In this section we consider a diverse range of estimators. In III A, we consider the use of redshift-space clustering on large and small scales. In III B, we investigate how clustering in under-dense and over-dense regions can be contrasted through both the explicit consideration of statistics in dense and void environments and the application of environment-dependent, or “marked” statistics. In III C, the applications of a range of real- and Fourier-space statistics beyond the two-point correlation function are studied, including the three-point correlation function, the bispectrum, hierarchical clustering, Minkowski functionals and phase space statistics. In III D, the use of lensing information in addition to spectroscopic clustering information is considered both in clustered and void environments.

A. Redshift-space galaxy clustering

Peculiar velocities of galaxies induce anisotropies in redshift space and leave distinctive imprints on the clustering pattern at different regimes. On large and linear scales, galaxies infall into high-density regions such as clusters, producing a squashing effect of these regions along the line-of-sight: this is the Kaiser effect [89]. On smaller (nonlinear) scales, the random motions of galaxies in virialized objects produce the Fingers-of-God (FoG) effect where the density field becomes stretched and structures appear elongated along the line of sight [90].

In this section we consider the potential constraints from redshift space distortions (RSD) on large scales, section III A 1, and small scales, in section III A 2.

1. Large-scale RSD

In linear theory, the amplitude of the RSD is related to the distortion parameter β , defined as

$$\beta(z) \equiv \frac{f(z)}{b(z)}, \quad (17)$$

where f is the linear growth rate and b is the linear galaxy bias, as functions of redshift.

The linear growth for the matter fluctuations in different gravity models can be obtained by solving the equation of the linear growth factor, D_+ ,

$$D_+'' + \left[2 - \frac{3}{2}\Omega_m(a)\right] D_+' - \frac{3}{2}\mu(k, a)\Omega_m(a)D_+ = 0, \quad (18)$$

where $'$ denotes a derivative with respect to $\ln a$ and μ , introduced in Eq. (9), is the ratio between the effective Newton constant G_{eff} and the true one G :

$$\mu(k, a) = \frac{G_{\text{eff}}}{G} = \begin{cases} 1 & \text{for GR,} \\ 1 + \frac{k^2}{3(k^2 + a^2 m_{f_R}^2)} & \text{for } f(R) \text{ gravity,} \\ 1 + \frac{1}{3\beta_{\text{DGP}}(a)} & \text{for nDGP,} \end{cases} \quad (19)$$

where k is the wavenumber of a perturbation mode, m_{f_R} is the mass of the scalaron field defined by $m_{f_R}^2 \simeq [3|f_{RR}|]^{-1}$ and $\beta_{\text{DGP}}(a)$ is given by Eq. (14). Note that $G_{\text{eff}}^{f(R)}$ is a function of time and scale, which means that the linear growth of structure for $f(R)$ gravity is scale dependent, while for GR and nDGP is scale independent. The linear growth rate, f , is defined as

$$f \equiv \frac{d \ln D_+}{d \ln a}. \quad (20)$$

Large-scale redshift distortions have been studied with a large variety of tracers, including luminous red galaxies, e.g., [91, 92], cosmic voids [93–95] and quasi-stellar-objects (QSOs) [96–98], and been successfully used to extract cosmological information by assuming a standard cosmological model, Λ CDM, based on GR. Current studies of modified gravity have used redshift-space distortions to put constraints on the β parameter, Eq. (17), see, e.g., Ref. [99]. In Ref. [100], the authors studied a coupled model of $f(R)$ gravity and massive neutrinos to break the degeneracy between the enhancement of the growth of large-scale structure produced by MG models and the suppression due to the free-streaming of massive neutrinos at late times.

The effects of redshift-space distortions can be measured using the redshift-space correlation function of galaxies, $\xi(r_p, r_\pi)$, which is the excess probability of finding a pair of galaxies at separations transverse (r_p) and parallel (r_π) to the line of sight (LOS). The left panel of Fig. 5 shows $\xi(r_p, r_\pi)$ as a function of separation (r_p, r_π) at $z = 0.5$, using the HOD catalogs described in II B 3 for the different gravity models. For comparison, the black dashed curve corresponds to the spherical two-dimensional correlation function in real-space of the Λ CDM (GR) model. We clearly see that along the LOS at $r_p \lesssim 2 h^{-1} \text{Mpc}$ the clustering is enhanced by the FoG effect, while at $r_p > 2 h^{-1} \text{Mpc}$ the clustering pattern is squashed by the Kaiser effect.

Given the symmetry along the line of sight, the transverse and parallel separations, (r_p, r_π), can be expressed as a distance in redshift space, s , and the cosine of the angle between s and the LOS direction,

$$s = \sqrt{r_\pi^2 + r_p^2}, \quad \mu = \frac{r_\pi}{s}. \quad (21)$$

The resulting anisotropic correlation function, $\xi(s, \mu)$, can then be decomposed into multipole moments,

$$\xi_l(s) = (2l + 1) \int_0^1 \xi(s, \mu) P_l(\mu) d\mu, \quad (22)$$

where $P_l(\mu)$ are the Legendre polynomials. In linear theory, the $l = 0, 2$ and 4 moments are non-zero with $P_0(\mu) = 1$, $P_2(\mu) = (3\mu^2 + 1)/2$, $P_4(\mu) = (35\mu^4 - 30\mu^2 + 3)/8$, corresponding to the monopole, quadrupole and hexadecapole moments.

The right panel of Fig. 5 shows the multipole moments, $\xi_l(s)$, of the correlation functions measured from the same galaxy catalogues. From the monopole, $\xi_0(s)$, we observe that the position of the baryon acoustic oscillations (BAO) peak, at $s_{\text{BAO}} \simeq 105 h^{-1} \text{Mpc}$ or 150Mpc , is not affected by modified gravity. Higher-order multipole moments such as the quadrupole $\xi_2(s)$ and the hexadecapole $\xi_4(s)$ encode the anisotropies induced by redshift distortions. In the case of the quadrupole, $\xi_2(s)$, N1 shows the strongest deviation with respect to GR, especially on scales $s > 20 h^{-1} \text{Mpc}$, followed by F4 and N5. Our measurements of the hexadecapole are almost indistinguishable among all models studied here. Comparing the plot with the left panel of Fig. 4

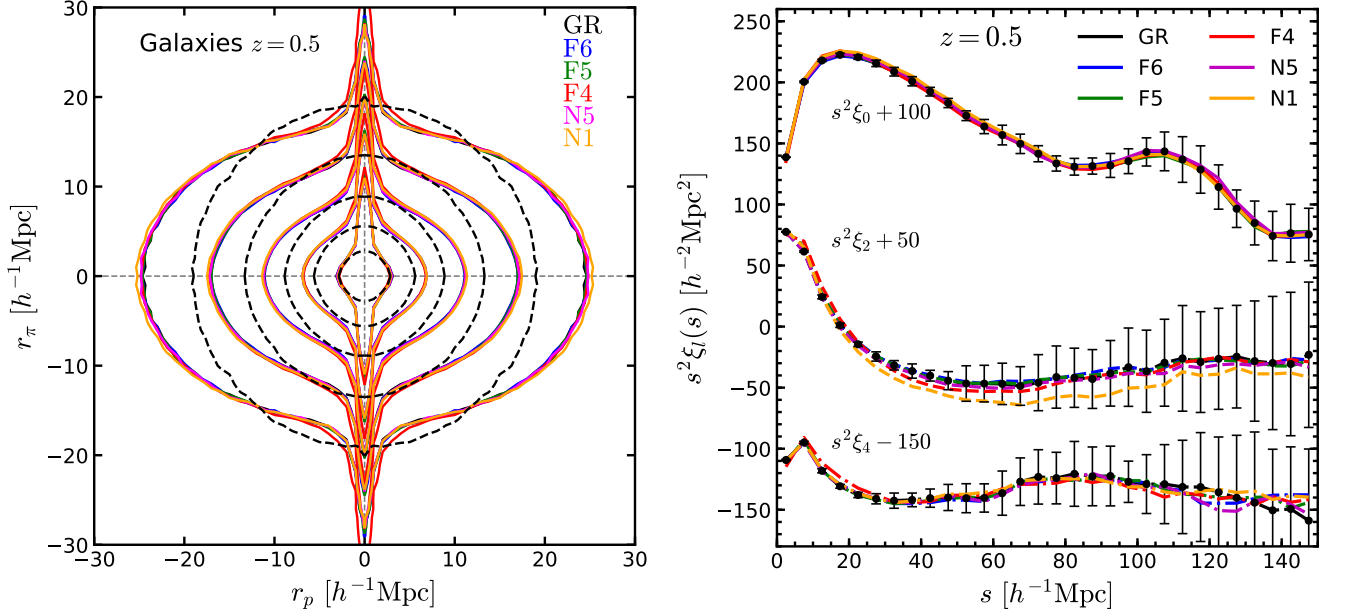


FIG. 5. *Left panel:* The two dimensional galaxy correlation function $\xi(r_p, r_\pi)$ measured from our mock catalogues at $z = 0.5$ as a function of separation across (r_p) and along (r_π) the line-of-sight. Contours show lines of constant $\xi(r_p, r_\pi)$ at $\xi(r_p, r_\pi) = 5, 2, 1, 0.5, 0.25$. The correlation functions correspond to the average of fifteen measurements obtained by projecting five realizations over the three LOS directions. The black dashed contours correspond to the real-space measurements of the correlation function at the same values of its counterpart in redshift-space, for simplicity we show the GR case only for these contours. *Right panel:* Monopole, quadrupole and hexadecapole moments of the correlation function, Eq. (22), for our six gravity models at $z = 0.5$. The moments have been shifted by a factor of 100, 50 and -150 for better visualization. The error bars correspond to the standard deviation over fifteen GR measurements. Different colour line correspond to different gravity model as specified in the legend.

of [99], in which the HOD parameters were tuned to match the real-space galaxy two point correlation functions in MG and GR, we can see that the results are almost identical.

To investigate the impact of MG on RSD more quantitatively, we discuss two methods to constrain the distortion parameter, β . The first is based on the Kaiser linear model [89], considering an estimator, $Q(s)$, to obtain the distortion parameter β [108]:

$$Q(s) \equiv \frac{\xi_2(s)}{\xi_0(s) - \xi_0(s)} = \frac{(4/3)\beta + (4/7)\beta^2}{1 + (2/3)\beta + (1/5)\beta^2} \quad (23)$$

The second method is based on an extension of the Galilean-invariant renormalized perturbation theory [gRPT; 109], where the anisotropic correlation function is obtained as the inverse Fourier transform of the power spectrum.

To estimate $\beta(z)$ from $Q(s)$ using the linear theory model, we use a χ^2 -test by minimizing the χ^2 defined as

$$\chi^2(\beta) = \sum_i \left[\frac{Q(s_i) - Q^{\text{th}}(s_i; \beta)}{\sigma_{Q_i}} \right]^2, \quad (24)$$

where $Q(s)$ is the average of the linear estimator given by Eq. (23) from the 15 redshift-space HOD catalogues constructed from the 5 real-space HOD catalogs, σ_Q is the standard deviation in the same catalogs, and $Q^{\text{th}}(s; \beta)$ is the theoretical prediction of $Q(s)$ given by the third expression of Eq. (23). We searched in a grid of values in $\beta \in [0, 1]$, with a step size of $\Delta\beta = 10^{-4}$, for the theoretical estimator and identified the value of β that minimizes χ^2 as $\chi^2_{\text{min}} = \chi^2(\beta_{\text{fit}})$. As we vary only one parameter, the 1σ error on β corresponds to $\Delta\chi^2 \equiv \chi^2 - \chi^2_{\text{min}} = 1$. The fit was done using the range of scales $s = 42.5 - 147.5 \, h^{-1} \text{Mpc}$.

To obtain the constraints of β using the gRPT model, we used Bayesian statistics and maximize the likelihood,

$$\mathcal{L}(\xi|\lambda) \propto \exp \left[-\frac{1}{2} (\xi - \xi_{\text{model}}(\lambda))^T \Psi (\xi - \xi_{\text{model}}(\lambda)) \right], \quad (25)$$

where the $\Psi = \text{C}^{-1}$ is the inverse of the covariance matrix. We applied the Gaussian recipe [110] to estimate the covariance matrix, which is then rescaled by the number of simulations. The parameters that enter the default fitting are $\{f, b_1, b_2, \gamma_3^-, a_{\text{vir}}\}$,

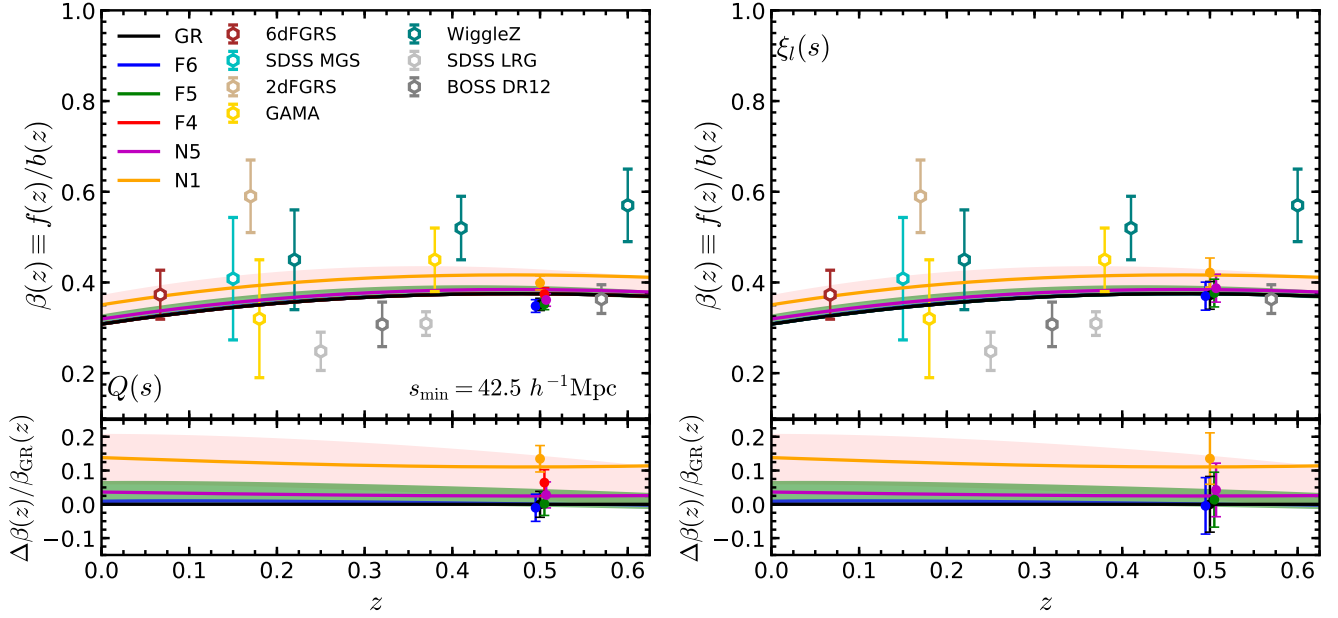


FIG. 6. Evolution of β as a function of redshift. The curves show the theoretical predictions for the gravity models as shown in the legend, for $f(R)$ gravity models the theoretical predictions are shown as a shaded region for wavenumbers $0.01 < k/[h \text{ Mpc}^{-1}] \leq 0.1$. Open coloured symbols correspond to current constraints from observational galaxy survey data: 6dFGRS [101], SDSS MGS [102], 2dFGRS [103], GAMA [104], Wiggle Z [105], SDSS LRG [106] and BOSS DR12 [107]. Filled symbols, at $z = 0.5$, denote the β estimates obtained from the simulations using the estimators: [Left panel] $Q(s)$ Eq. (23) using linear theory and [Right panel] $\xi_l(s)$ Eq. (22) using non-linear theory with $s_{\min} = 42.5 h^{-1} \text{Mpc}$ at $z = 0.5$. The lower subpanels show the relative difference between the modified gravity models and GR. Error bars correspond to the 1σ confidence level.

where b_1 and b_2 are local galaxy biases to linear and second order; γ_3^- is a non-local bias coefficient, see, e.g. [111]; and a_{vir} is a free parameter that describes the kurtosis of the velocity distribution on small scales. Two additional parameters, $\{q_{\parallel}, q_{\perp}\}$, relating fiducial and real distances, are needed when applying the Alcock–Paczynski test [112]. We marginalize over the nuisance parameters to find the probability distribution of the distortion parameter $\beta = f/b_1$. For more details of the method, we refer the readers to [99, 113].

In the upper panels of Fig. 6, we compare the theoretical predictions for $\beta(z)$ to current observational measurements of the distortion parameter from galaxy surveys, including the 6dFGRS at $z \simeq 0.067$ [101], the SDSS MGS at $z \simeq 0.15$ [102], the 2dFGRS at $z \simeq 0.17$ [103], GAMA at $z \simeq 0.18$ and 0.38 [104], Wiggle Z at $z \simeq 0.22, 0.41$ and 0.6 [105], the SDSS LRG at $z \simeq 0.25$ and 0.37 [106] and BOSS DR12 at $z \simeq 0.32$ and 0.57 [107].

In the left-hand panel, we also include the best-fit β values for all gravity models extracted from the simulations at $z = 0.5$ using the linear Kaiser method, with 1σ error bars. The extracted estimates consistently underestimate the β value for all gravity models compared with the theoretical prediction. This is due non-linearities that produce smaller values of $Q(s)$. In general, the linear Kaiser model fails to model RSD in configuration space even on large scales ($s_{\min} = 42.5 h^{-1} \text{Mpc}$). In the right-hand panel of Fig. 6, the constraints on β using the gRPT model are shown, where we observe a good agreement between the best-fit values and the theoretical predictions for all models, and the additional details of the RSD model corrects the inaccuracy of the Kaiser model on linear scales.

In the lower subpanels of Fig. 6, we plot the relative differences of the MG models with respect to GR. Despite the different best-fit β values in the linear and non-linear methods, the relative differences between models predicted by them are almost the same. However, the difference of the F6 and F5 models with respect to GR is $\lesssim 1\%$, making these three models statistically indistinguishable from each other, while F4 keeps a difference of $\sim 5\%$ with respect to GR; this reflects the fact that the growth of structure in $f(R)$ gravity is not enhanced on large scales, beyond the Compton wavelength of the scalaron field. On the other hand, N5 and N1 models show a difference of respectively $\sim 2.5\%$ and $\sim 12\%$ with respect to GR, due to the long-range nature of the fifth force.

We conclude that with current data it is difficult to distinguish between the various gravity models simply by using constraints on β . Future data from surveys like DESI will likely improve on this situation, though tests of models like F5, N5 and F6 may still remain a challenge [see e.g. 114]. As we discuss next, the use of RSD to distinguish between gravity models may require the inclusion of smaller scale information.

2. Small-scale redshift-space galaxy clustering

We have seen in the previous subsection that large-scale RSD can be a useful tool to test gravity theories which strongly affect structure formation on large scales, such as the nDGP model, while for models such as $f(R)$ gravity, in which gravity is modified on small, nonlinear, scales, the constraints are weaker. However, this conclusion should be read with the particular context of the analysis in mind. Neither the perturbative theoretical templates for RSD nor the numerical results from our HOD catalogs are accurate enough for the highly nonlinear regime, where the FoG effect due to the virial motions of small galaxies dominates the anisotropies in galaxy clustering and can potentially be affected by an enhanced gravity force. For this reason, it is worthwhile to explore this regime in greater details using different techniques. In this subsection, we describe the subhalo abundance matching (SHAM) technique [115–117] to predict galaxy clustering, in particular, the small-scale RSDs in $f(R)$ gravity. This subsection is largely a review of the results presented in Ref. [118], to make this paper complete.

The SHAM technique assumes that galaxies reside in subhalos, and that through a monotonic relation between a property of a subhalo and an observed property of a galaxy, subhalos selected in a simulation correspond to galaxies observed in a galaxy survey. The clustering of subhalos from the simulation therefore can be compared to the observed clustering of galaxies directly. An advantage of SHAM is that there is no ambiguity of galaxy bias. Moreover, the method can fully explore nonlinear effects, such as the FoG, on small scales as it utilizes N-body simulations directly. However, straightforward as the SHAM method may seem to be, it is indeed non-trivial to practically implement it and effectively control systematics.

On the observational side, instead of photometrically selected samples in optical bands, stellar-mass-selected samples should be used. State-of-the-art hydrodynamic simulations (e.g., [119, 120]) suggest that it is the stellar mass of a galaxy, rather than its r -band luminosity, that has a tighter correlation with v_{peak} (the peak value of the maximum circular velocity over a subhalo’s merger history) of a subhalo. However, unlike its luminosity, a galaxy’s stellar mass cannot be directly measured but has to be derived from a stellar population synthesis model. A challenge we face here is the uncertainty in the estimation of stellar mass. There are two main sources of the systematics: one is in theory, especially the stellar initial mass function (IMF); the other is the uncertainty in determining the total flux of a galaxy. In order to mitigate these systematics, we have constructed volume-limited samples that are complete in stellar mass, with galaxies being selected in terms of number densities rather than stellar mass cut. The idea here is to keep the ranking orders of galaxies, and as demonstrated in [118] this method can effectively minimize the impact of systematics in the estimation of stellar mass on the measured RSD multipoles, particularly for samples with higher number densities, $n_g \gtrsim 1 \times 10^{-2} [h^{-1}\text{Mpc}]^{-3}$.

On the theoretical side, a high mass resolution of the simulation is crucial for SHAM. A satellite subhalo close to the host halo center may substantially lose mass due to tidal stripping. Sometimes even a massive halo at an earlier time in a simulation can be completely disrupted by tidal stripping and can not be resolved at a later time, leading to a phenomenon called “orphan galaxies” [117], which can result in an under-estimation of the galaxy clustering on small scales [117]. A test of the completeness of the subhalo catalog, therefore, is important for the SHAM method [118].

Figure 7 shows a comparison between the observed redshift space galaxy clustering and the predictions in ΛCDM for galaxy samples with $n_g = 1 \times 10^{-2} [h^{-1}\text{Mpc}]^{-3}$. The SHAM predictions in ΛCDM agree very well with the observation especially for the Fingers-of-God on small scales. To demonstrate the robustness of SHAM in constraining MG models, we also present its predictions for the F6 model in Fig. 7. The predictions are based on set of high-resolution $f(R)$ and GR simulations presented in Ref. [51], using the “effective halo” technique developed in Refs. [121–123]. As we can observe from this figure, the $f(R)$ result deviates substantially from both the ΛCDM predictions and observational data for all three RSD multipoles.

To further understand the substantial differences between F6 and GR in Fig. 7, we have plotted in Fig. 8 the real space 2PCFs of the SHAM mock galaxies, for two sample number densities: $n_g = 1 \times 10^{-2} [h^{-1}\text{Mpc}]^{-3}$ (left) and $2 \times 10^{-2} [h^{-1}\text{Mpc}]^{-3}$ (right). In both cases there is a 20 ~ 40% difference at $r \lesssim 6 h^{-1}\text{Mpc}$, which partially explains the behavior of Fig. 7 (note the difference in the galaxy velocities between the two models also contributes to the model difference in redshift space in Fig. 7). The clustering is actually weaker in F6, which is because on small scales the enhanced gravitational force makes structures grow faster, which means that lower initial density peaks, which would not have become massive enough to host halos for the chosen n_g cutoff in GR, have indeed become galaxy bearing. These halos are intrinsically less clustered than the ones which form from higher initial density peaks. This is clearly a feature that can only be probed in the nonlinear regime due to the short-range nature of the fifth force in this model. In nDGP models, in contrast, we expect the effect to be weaker due to the Vainshtein screening of the fifth force even for small halos.

Small-scale RSD is therefore a promising tool when using future galaxy surveys such as DESI, in particular its Bright Galaxy Survey (BGS), to constrain gravity models. The BGS can obtain redshifts of “bright” galaxies up to two magnitudes fainter than the limit of the Sloan Digital Sky Survey (SDSS) main galaxy redshift survey [125]. In its current design, the BGS will observe approximately 17 million galaxies over 14,000 deg^2 , in two completeness tiers: $r < 19.5$ galaxies with completeness of $\approx 90\%$ and $19.5 < r < 20.0$ galaxies with completeness of $\approx 75\%$. The exceptionally high sampling density allows the best achievable measurements of RSD with unparalleled accuracy. Using SHAM, the BGS can offer an unprecedented test of gravity.

Although SHAM is a simple model to assign galaxies to halos, the physics motivation and robustness of this method has been tested by modern state-of-the-art hydrodynamic simulations. Compared with other properties of dark matter halos, such as the maximum circular velocity of a subhalo at today, due to the prevalence of disruption in dark matter halos, quantities before this

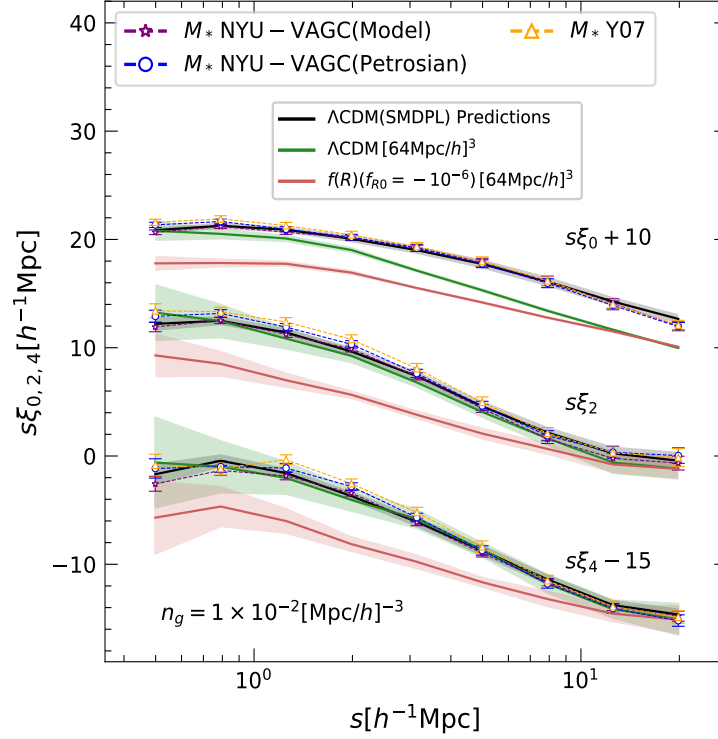


FIG. 7. A comparison of RSD measurements using three different stellar mass models: a template-fit method adopted in the NYU catalog with the SDSS model magnitude (stars with dashed line), the same template-fit method but with the SDSS Petrosian magnitude (circles with dashed lines), and a single-color method (triangles with dashed lines). Different stellar mass models give rise to convergent results. The error-bars in the above plot are estimated using a jackknife re-sampling technique with 133 realizations. Here the error-bars represent 3σ statistical error. The predicted RSD multipoles (monopole ξ_0 , quadrupole ξ_2 and hexadecapole ξ_4) in Λ CDM from SHAM mock are in good agreement with observations. Note that the SHAM mock is based on the SMDPL simulation [124] and has the same geometry as the real data. The plot is reproduced from [118]. The green and red lines show the predicted RSD multipoles for Λ CDM and $f(R)$ gravity using a suite of simulations with $64[h^{-1}\text{Mpc}]^3$ box size. The shaded regions give the 1σ uncertainty around the expectation value, derived from 500 realizations with line-of-sight along different directions of the simulation box. Despite the deficit of power in the monopole due to the small box size of the simulations, the higher order multipoles, such as quadrupole ξ_2 and hexadecapole ξ_4 , show significant differences, indicating that even a model like F6 can be testable given the accuracy of the clustering measurement of DESI.

disruption such as v_{peak} can correlate more tightly to galaxy stellar mass[126]. However, despite its success in Λ CDM, in MG models it still needs to be tested using high-resolution hydrodynamical galaxy formation simulations, such that we can more accurately assess its constraining power with DESI BGS.

3. Redshift-space distortions around voids

Cosmic voids are regions in our Universe that are underdense in terms of tracer numbers and matter. Redshift space distortions around voids can be used to probe the growth rate of structures around these regions [127, 128]. The void-galaxy correlation function is distorted in redshift space because of the peculiar motions of galaxies. While such peculiar motions respond only to the Newtonian potential in GR, in MG models they can also be affected by the fifth force, causing the distortion patterns to change – a diagnostic that can then be used to distinguish the model from GR. Since the fifth force is expected to be unscreened in voids [129–132], the effect should be larger around voids.

We conduct an analysis for the redshift-space distortions around voids using the redshift-space void-galaxy correlation function ξ_{vg}^s , using the mock galaxy catalogs from the GR and modified gravity simulations described in previous section. Voids are identified in the redshift-space galaxy fields using the ZOBOV void finder, which makes use of Voronoi tessellation of the galaxy field [133] (details for the definitions of ZOBOV voids including void center and radius are described in Section 3.3 of [50]). We then measure the void-galaxy correlation function ξ_{vg}^s in redshift space. An example for the GR model is shown in the left panel of Fig. 9. The extracted monopole ξ_0^s and quadrupole ξ_2^s moments are shown in the top of the right panel of Fig. 9. All voids with radius r_v greater than $20h^{-1}\text{Mpc}$ are used for the analysis.

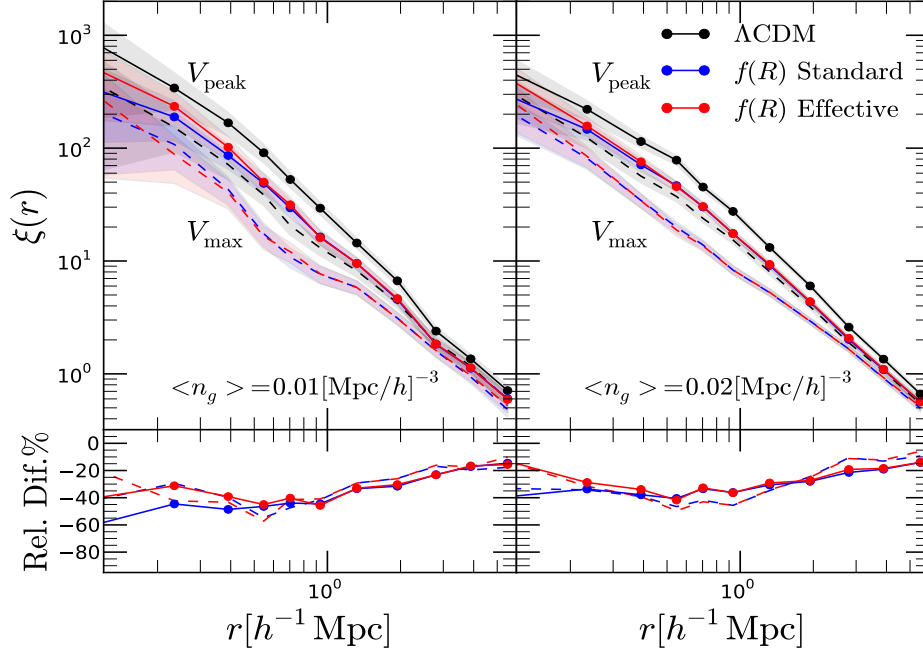


FIG. 8. The predicted three-dimensional two-point galaxy correlation functions from the SHAM model (upper panels). The shaded regions represent the 1σ Poisson errors. The lower panels show the fractional differences between the $f(R)$ model and Λ CDM. The left panels show the results for a galaxy density $\langle n_g \rangle = 0.01 [h^{-1} \text{Mpc}/h]^{-3}$ and the right panels are for $\langle n_g \rangle = 0.02 [h^{-1} \text{Mpc}/h]^{-3}$. For comparison, the dashed lines show the results obtained using the current maximum circular velocity v_{max} . The plot is reproduced from Ref. [123]

In GR, Following the linear model of [127], the ratio between the quadrupole and monopole is a constant $G = 2\beta/(3 + \beta)$, where $\beta = f/b$ is the distortion parameter introduced above. We can estimate β using the estimator

$$\tilde{G}(\beta) = \frac{\xi_2^s(r)}{\xi_0^s(r) - \frac{3}{r^3} \int_0^r \xi_0^s(r') r'^2 dr'} = \frac{2\beta}{3 + \beta}. \quad (26)$$

The multipoles of the redshift-space correlation function can be obtained by

$$\xi_\ell^s(r) = \int_0^1 \xi_{\text{vg}}^s(r, \mu) (1 + 2\ell) P_\ell(\mu) d\mu, \quad (27)$$

where $P_0(\mu) = 1$ and $P_2(\mu) = \frac{1}{2}(3\mu^2 - 1)$, and $\mu = \cos \theta$ where θ is the angle between the line connecting a galaxy-void pair and the LOS. In linear theory, the model has only one free parameter, β .

We follow the same procedure as in [127] for constructing the covariance matrix and for the parameter fitting. The correlation functions from all the 5 boxes of simulation at $z = 0.5$ are treated as independent and used for the fitting. We have also viewed the simulation box along three different major axes of the box to further increase the data size, although we do not expect the volume to increase by three times since they are not independent.

We appreciate that the growth rate in non-GR models may be scale dependent even in the linear regime. In this study, we do not explore this subtlety and simply recover the effective growth rate parameters $\beta_{\text{effective}}$ for all models. We have also treated the linear bias as scale independent and taken the measurement from the galaxy 2PCF versus the dark matter correlation function between 20 to 50 $h^{-1} \text{Mpc}$. The linear galaxy biases for all different models are very close to each other, $b \sim 1.9$ at $z = 0.5$.

We find that the best-fit growth rate parameter for the GR model agrees with the true answer within the 1σ range. However, with the same fitting procedure, we recover similar growth rates for F6 and N5, not distinguishable from GR. Both the F5 and N1 models may be distinguished from GR at the $2-3\sigma$ level, with the F5 model having a slightly lower best-fit value of β and the N1 model having a higher value of β than GR, which is somewhat unexpected as the linear growth rate in both these two models should be higher than in the GR model with the same expansion history. This may have been complicated by non-linearity and scale-dependent galaxy bias being different among different models. To fully exploit the information from these measurements, more accurate models of RSD which work in the non-linear regime are needed, e.g., [134].

We have also found that error budget for β does not goes down with the square root of the volume, which suggests that viewing the simulation along three different axes does not give us more independent data. Having a higher number density of

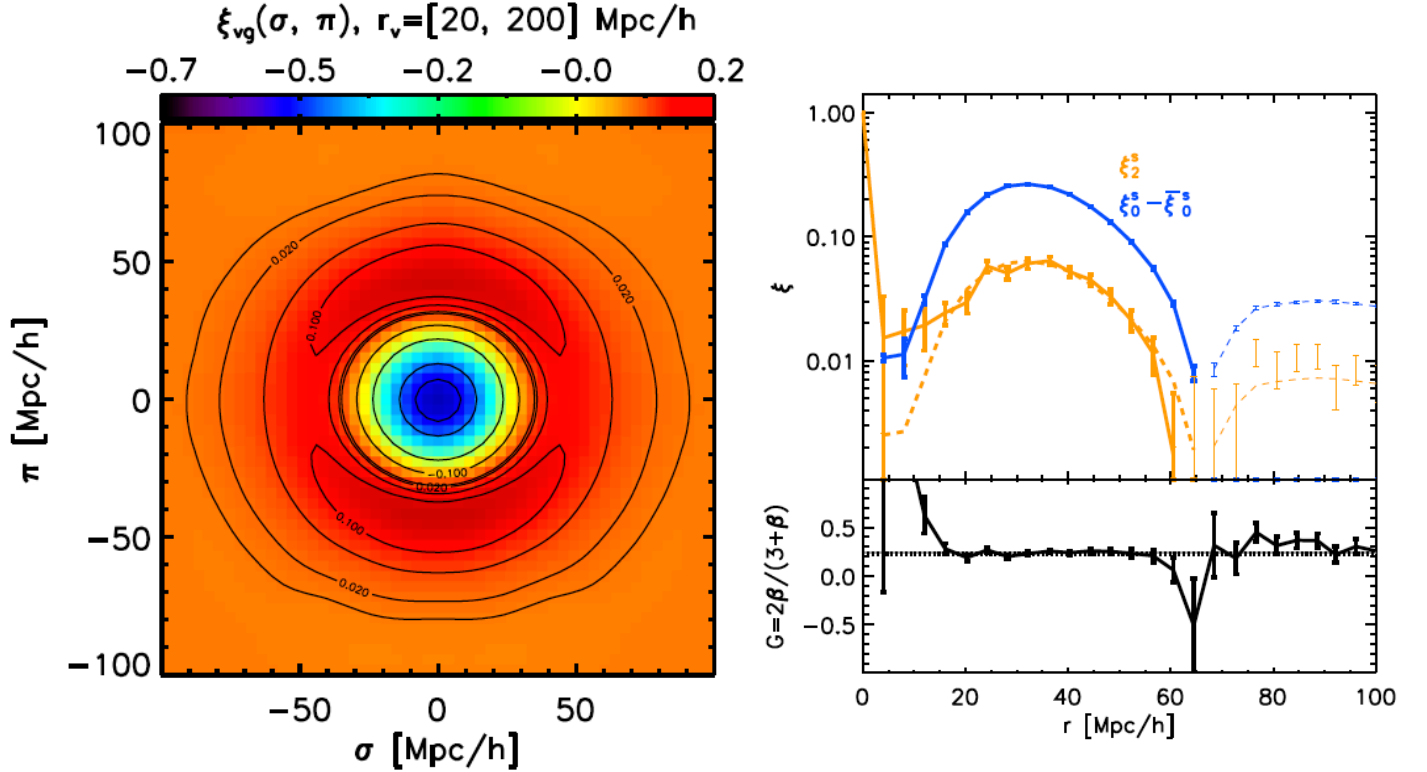


FIG. 9. Left: the void-galaxy correlation function in redshift space for the GR model for voids identified by ZOBOV with the HOD galaxies in 5 boxes of the $1024h^{-1}\text{Mpc}$ -aside simulation. Voids with effective radius $r_v > 20h^{-1}\text{Mpc}$ are used. Right: the upper panel shows the monopole (blue) and quadrupole (orange) moments of the redshift-space correlation function. Thin lines indicate the absolute values of negative parts. The lower panel shows the ratio of the quadrupole versus monopole. The black curve is from taking all measurements of voids along three different major axes of the simulation box, with the error bars showing the error on the mean. The dotted line is the true answer for the growth rate parameter G given by the cosmology and galaxy bias. The black shaded region shows the 1σ range around the best-fit value.

galaxies will help to resolve smaller voids and increase their number for a fixed volume. This will be guaranteed for the DESI BGS survey, where the number density of galaxies will be at least an order of magnitude higher than the current CMASS mocks. This may help to further reduce the errors and increase the distinguishing power between the MG models and GR. Void RSD is a relatively unexplored probe, and its potential in testing gravity will require further investigations in the future.

4. Discussion

Redshift-space distortions, through combining information from both large and small scales, can be a potentially powerful tool to discriminate between different gravity scenarios. Numerical, semi-analytic and analytic techniques may each be valuable to accurately predict RSD in gravity models that deviate from GR. In the ΛCDM model, various RSD models have been proposed. Substantial progress has been made based on perturbation theory such as the one given by [135, 136], and the Taruya-Nishimichi-Saito model [137] and the effective theory of large-scale structure [138] (see [139–141] for comparisons of these models). These need to be combined with perturbative galaxy bias models such as the one studied in [111] (see [142] for a review).

In configuration space, the redshift-space correlation function is often modelled by the Gaussian streaming model (GSM) [143] combined with Lagrangian perturbation theory [144]. GSM models assume the pairwise velocity distribution of galaxies can be modelled as a Gaussian. However, N-body simulations have demonstrated that gravitational clustering introduces non-negligible skewness and kurtosis [136], that play an important role in shaping the redshift-space galaxy clustering multipoles on small scales. Various models aimed at incorporating these properties have been proposed in recent years, e.g., [145–149], which can improve the accuracy of redshift-space clustering predictions on smaller scales (e.g., $\lesssim 20h^{-1}\text{Mpc}$) than the Gaussian case, but their accuracy is often limited by the ability of perturbation theory to reproduce the pairwise velocity moments accurately. In the near future, this limitation might be overcome by calibrating non-Gaussian models to N-body simulations.

More generally, N-body simulations are being used extensively not only to test the accuracy of perturbation theory approaches,

but also to directly constrain gravity by means of either hybrid techniques, such as the one presented in [150] where the redshift-space power spectrum is computed on large scales using perturbation theory and a template approach is taken for the small scales, or by taking simulations a step further to create emulators for clustering estimators [151, 152]. The main challenges for emulator approaches are their expensive computational requirements, and their reliability on accurate models of the galaxy-halo connection to produce unbiased constraints of gravity from data.

The advantage of the perturbation-theory-based RSD models is that they can be extended to different gravity models. This has been done for several MG models including $f(R)$ gravity and nDGP models [153–158]. This enables one to directly constrain modified gravity parameters such as $|f_{R0}|$ by taking into account consistently the scale dependence of the growth rate and non-linear interactions due to screening [46]. The disadvantage is that perturbative approaches are not able to model the RSD effect accurately on small scales and a careful analysis is required to determine the scales that can be used using galaxy mocks (as we shall discuss in the following subsection). On the other hand, the emulator approaches can exploit the information on much smaller scales but they require a large number of high-resolution N-body simulations in a given gravity model, although COLA approaches have brought some alleviation to such a requirement [159, 160]. It is important to develop more efficient and optimal RSD models to distinguish different gravity models for surveys like DESI.

B. Density-dependent environmental effects on galaxy clustering

We have seen above that many MG models have environmentally-dependent behaviors, such as stronger deviation from GR in under dense regions. The mock galaxy catalogs used for this paper have been produced using HOD parameters tuned for galaxies from all environments and so, while the different gravity models match each other in their overall projected 2PCFs, it is natural to ask whether stronger differences can be found in subsets of galaxies from specific environments. In this section, we investigate this possibility through a variety of statistics in contrasting environments. We consider statistics that utilize density transformations, or “marked tracers”, to test modified gravity with simulated data.

A general marked correlation function can be defined as [49]:

$$M(r) \equiv \frac{1}{n(r)\bar{m}^2} \sum_{ij} m_i m_j = \frac{1 + W(r)}{1 + \xi(r)}, \quad (28)$$

where the sum is over all tracer pairs with a given separation, r , $n(r)$ is the number of such pairs, m_i the mark for the i th tracer and $\bar{m}$ the mean mark for the entire sample. In the second equality, $\xi(r)$ denotes the standard two-point correlation function and $W(r)$ is the ‘weighted’ correlation function at a separation r .

In the following discussion we consider estimators based on number counts of halos and galaxies in III B 1 and the gravitational potential that accentuate under-dense regions in III B 2. In addition to the estimators derived from the simulations, we also discuss how marked correlation functions can be predicted from analytical estimates in section III B 3. Finally we consider correlation functions that highlight over-dense regions, in III B 4.

1. Marks defined using tracer number densities

Screening in modified gravity models, which occurs in regions of high density, has led to the consideration of density transformations that up-weight lower density, unscreened, regimes as a mechanism to enhance the modified gravity signatures. Such transformations include: the logarithmic transform of the density field [161–163], for which the transformed field becomes more Gaussian, facilitating easier information extraction from the 2-point function; a clipped density field [164–166] in which density peaks are all allocated a common value δ_0 ,

$$\delta' = \delta_c = \begin{cases} \delta & \text{if } \delta < \delta_0 \\ \delta_0 & \text{if } \delta > \delta_0. \end{cases} \quad (29)$$

and more general “marked” density transformations that up-weight low densities. An example of such a mark is the one proposed in [167],

$$\delta' = m(\delta) = \left(\frac{\rho_* + 1}{\rho_* + \rho} \right)^p = \left(\frac{\rho_* + 1}{\rho_* + (\delta + 1)} \right)^p, \quad (30)$$

where ρ_* and p are free parameters and ρ the dark matter mass or tracer number density field in a grid cell, in units of the mean density $\bar{\rho}$. Marked statistics have been previously explored in the context of MG in [168], using CDM simulations produced in [42], and in [169, 170]. Tested on the $f(R)$ and symmetron [171, 172] MG models, the density-marked statistic was found to

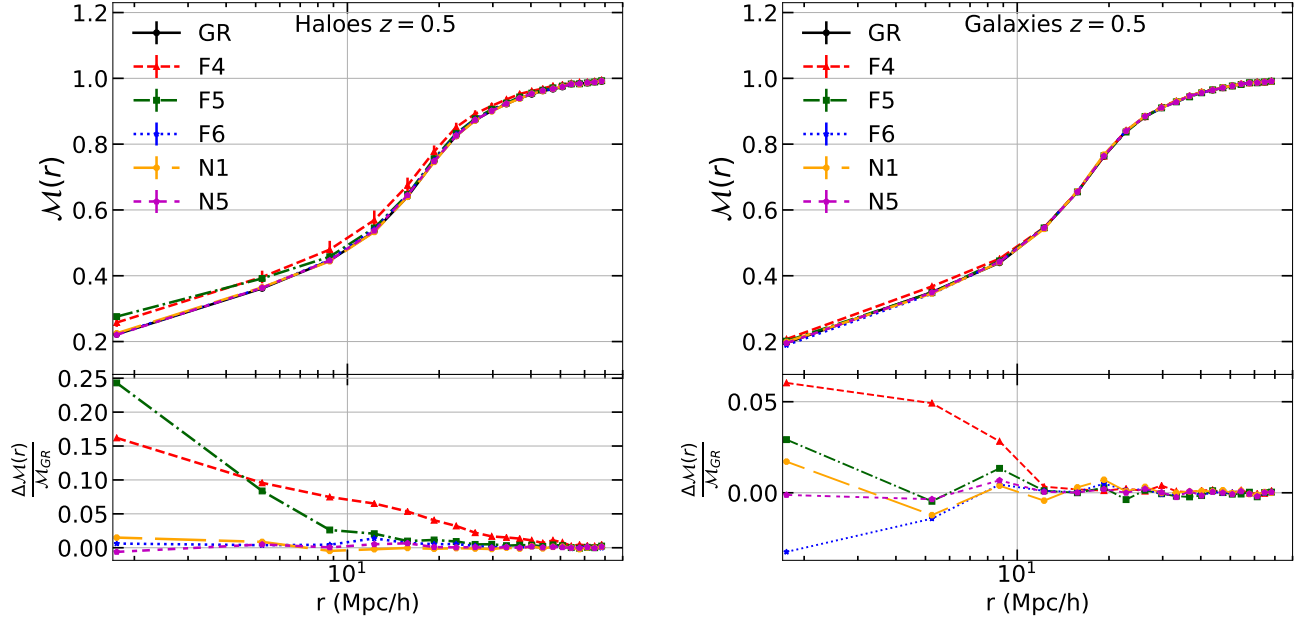


FIG. 10. (Color Online) The galaxy-number-density marked correlation statistic for *halos* (left) and *galaxies* (right) with $\rho_* = 4, p = 10$, for GR (black), $f(R)$ with $|f_{R0}| = 10^{-4}$ (red), $|f_{R0}| = 10^{-5}$ (green), $|f_{R0}| = 10^{-6}$ (blue) and nDGP with $r_c H_0 = 1$ (orange), $r_c H_0 = 5$ (magenta). The curves correspond to the averages over the 5 realizations of HOD catalogs, while the bottom panels are the fractional deviation $\Delta\mathcal{M}/\mathcal{M} = \mathcal{M}_{\text{MG}}/\mathcal{M}_{\text{GR}} - 1$ for each MG model. The error bars correspond to the standard deviations over the 5 boxes.

boost the signal-to-noise ratio encoded in the two-point statistics, providing thus additional discriminatory power with respect to Λ CDM.

To calculate the marked correlation function for our models, we measure the galaxy number density using the nearest-grid-point scheme, by dividing the simulation box into cells of the same size, and then counting the number of galaxies inside each cell and using this to assign a density to the cell. Hence, we can compute the overdensity, δ , as

$$1 + \delta \equiv \frac{n_g}{\bar{n}_g}, \quad (31)$$

where n_g is the number of galaxies in each cell and $\bar{n}_g$ is the mean number of galaxies in cells of a given size over the simulation volume. To compute the galaxy density, we have used 60^3 cells of size $\sim 17 h^{-1}$ Mpc.

In the left-hand panel of Fig. 10, we show the marked correlation statistic Eq. (28) for halos of mass $M_{200c} > 10^{13} h^{-1} M_\odot$, evaluated using galaxy-number-weighted density estimates in Eq. (30), with $\rho_* = 4, p = 10$, for GR and all the modified gravity models considered. In addition, the right-hand panel of Fig. 10 shows the same marked statistic, for the same models using the same parameters for the mark (ρ_*, p), but calculated using the simulated galaxy catalogs as described in § II B. The correlation function calculations were performed using the publicly available code Super W Of Theta (SWOT) [173].

In the case of the halo marked correlation functions, the fractional deviations from GR are more pronounced for the F4 and F5 models, as expected, reaching values of $\sim 20\%$ in the lowest r bin. The deviations in the F6 case are at the sub-percent level in both cases and at all scales, while the nDGP models predict maximum deviations of $\sim 5\%$ in the mass weighted estimator. The predicted differences are overall less pronounced in the marked correlations of galaxies, with the various models predicting deviations not larger than $\sim 5\%$ in the smallest scales for all models. This is likely a consequence of the HOD parameters used to populate galaxies in each halo being tuned to match the unmarked correlation functions. An explanation of this behavior is presented in § III B 5.

For completeness, we also studied the case with negative p . Figure 11 shows the marked correlation using the galaxy-number-weighted density estimate for halos and galaxies with $\rho_* = 1, p = -1$. For dark matter halos, the deviations from GR are much smaller compared with the case with $\rho_* = 4, p = 10$, confirming our expectation that up-weighting low-density regions enhances MG signals. On the other hand, the galaxy marked correlation function shows similar small deviations to the case with $\rho_* = 4, p = 10$, with no clear trend among the MG models in their relative difference to GR. This, together with Fig. 10, implies that if we only use the galaxy number density as a mark, it is difficult to design a mark that can substantially enhance the model differences. This difficulty may be alleviated if one uses external information such as the gravitational potential to define marks

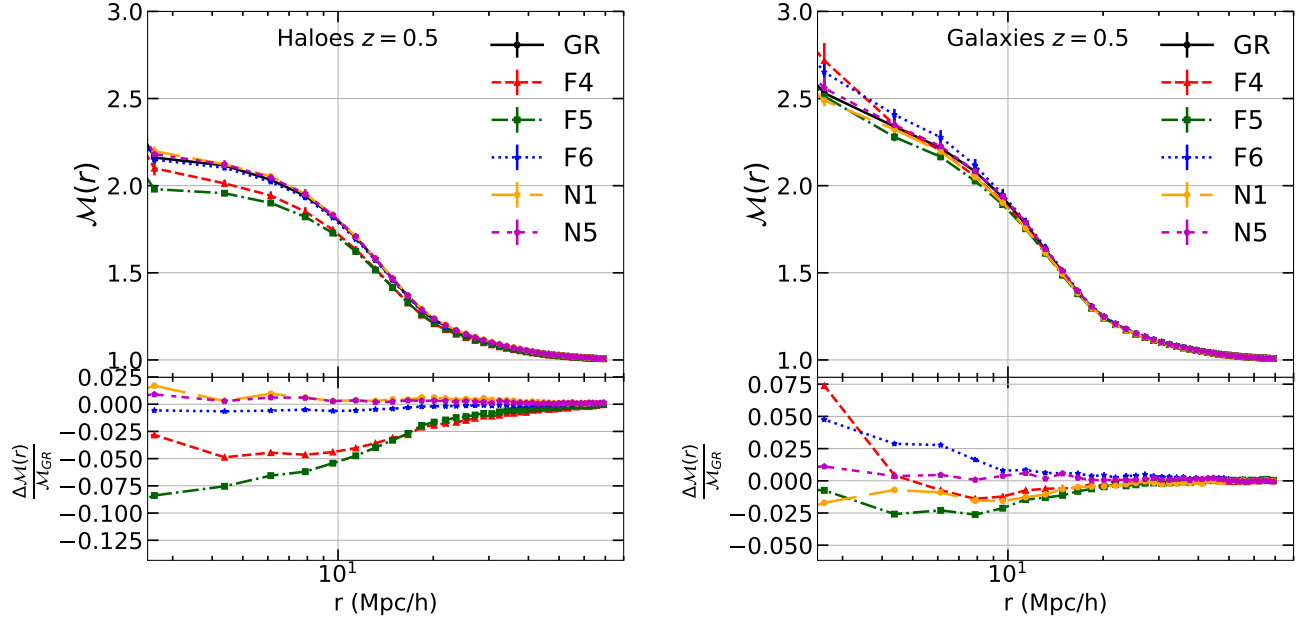


FIG. 11. (Color Online) Comparison of the galaxy-number-density marked correlation statistic for *dark matter halos* (left) and *galaxies* (right) for the case of up-weighting high density regions ($\rho_* = 1, p = -1$), for GR (black), $f(R)$ with $|f_{R0}| = 10^{-4}$ (red), $|f_{R0}| = 10^{-5}$ (green), $|f_{R0}| = 10^{-6}$ (blue) and nDGP with $r_c H_0 = 1$ (orange), $r_c H_0 = 5$ (magenta). The curves correspond to the averages over the 5 realizations. The bottom panels show the fractional deviation $\Delta \mathcal{M}/\mathcal{M} = \mathcal{M}_{\text{MG}}/\mathcal{M}_{\text{GR}} - 1$ for each MG model. The error bars correspond to the standard deviations over the 5 boxes.

as discussed in § III B 2

In the discussion so far, we have used the NGP assignment scheme to calculate the galaxy density field. We have also checked that the conclusions for Figs. 10 and 11 also apply if we use the matter density to define the marks in Eq. (30). For completeness, we further checked if the results are sensitive to the details of the density assignment method used. To this end we have tested two other ways to estimate the density, smoothed particle hydrodynamics (SPH) and space tessellation, respectively. Since their results are again consistent with each other, here we only discuss the results from the former case.

In Fig. 12, we estimate the local galaxy number density n_g at each galaxy's position by the spline kernel interpolation [174],

$$\rho_R = \sum_{i=1}^{20} W(r_i, h),$$

where $W(r_i, h)$ is the SPH spline kernel, r_i is the distance to i th galaxy and the smoothing length h being the distance to the 20th nearest galaxy [175],

$$W(r, h) = \frac{8}{\pi h^3} \begin{cases} 1 - 6 \left(\frac{r}{h}\right)^2 + 6 \left(\frac{r}{h}\right)^3, & 0 \leq \frac{r}{h} \leq \frac{1}{2}, \\ 2 \left(1 - \frac{r}{h}\right)^3, & \frac{1}{2} < \frac{r}{h} \leq 1, \\ 0, & \frac{r}{h} > 1. \end{cases}$$

The smoothing scale h for density estimation in this method is varying across the simulation box.

The left panel of Fig. 12 shows the galaxy-number-density-weighted marked correlation functions for HOD galaxies in real space at $z = 0.5$, using the same parameters (ρ_*, p) as in Fig. 10. The deviations from GR show the same trend as in Fig. 10, but the amplitudes are slightly larger, now reaching $\sim 10\%$ for F4 on small scales; for the two DGP variants the deviation from GR is negligible. The right panel of Fig. 12 represents the monopole of galaxy marked correlation functions in redshift space. Rather than averaging the nominator and denominator of Eq. (28) over five boxes and taking the ratio, we directly measure the marked correlation function by Eq. (28) and then take the average. There are some quantitative differences from the real-space case, with the deviations from GR generally slightly larger, but the overall behaviors are similar to the left panel. The only MG model that is significantly different from GR is F4. This shows that the result of marked correlation functions could be sensitive to the scheme of the mark evaluation, adding to the complexity in making theoretical predictions.

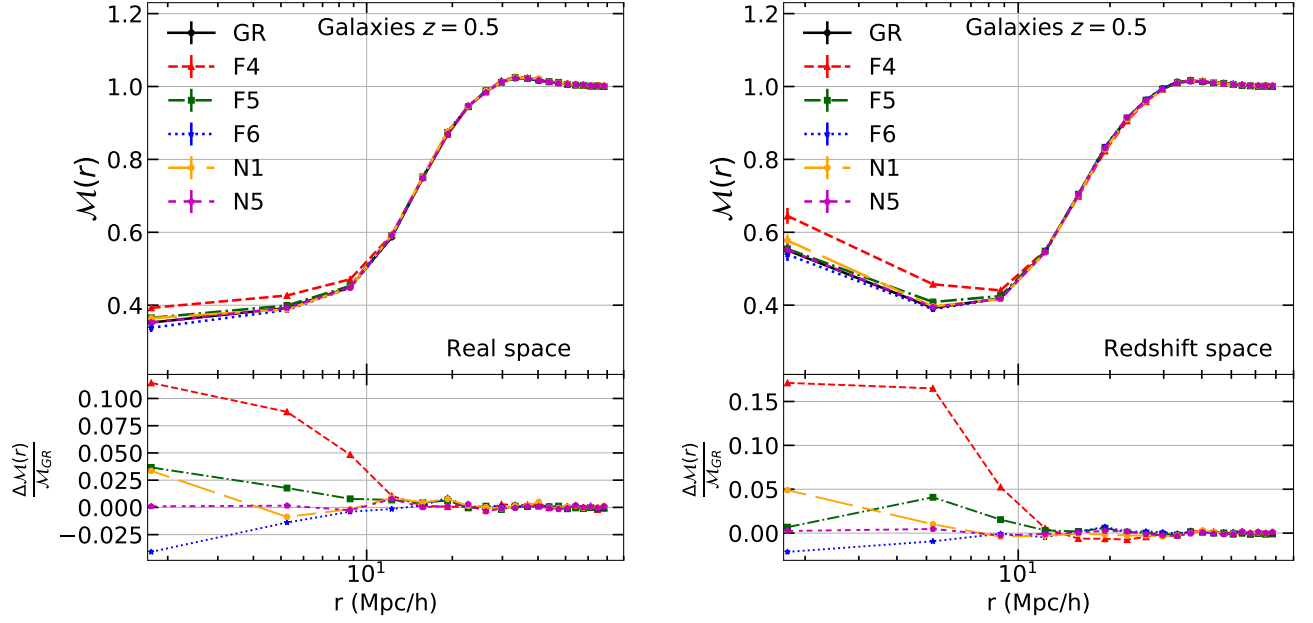


FIG. 12. (Color Online) Comparison of the galaxy-number-density marked correlation statistic for galaxies with $\rho_* = 4, p = 10$, for GR (black), $f(R)$ with $|f_{R0}| = 10^{-4}$ (red), $|f_{R0}| = 10^{-5}$ (green), $|f_{R0}| = 10^{-6}$ (blue) and nDGP with $r_c H_0 = 1$ (orange), $r_c H_0 = 5$ (magenta) when using SPH spline kernel interpolation for density estimate in real space (left) and redshift space (right). The curves correspond to the averages over the 5 realizations (in redshift space we consider 3 directions of each realization as independent samples), while the bottom panels show the fractional deviation $\Delta\mathcal{M}/\mathcal{M} = \mathcal{M}_{\text{MG}}/\mathcal{M}_{\text{GR}} - 1$ for each MG model. The error bars correspond to the standard deviations over the 5 boxes.

2. Marked correlation function with gravitational potential estimators

In the above we have used the galaxy density field, calculated in various ways, to define the mark. Due to the projected galaxy 2PCFs in the different models all matching each other, we have seen that the resulting marked 2PCFs show very mild difference which is generally of the same order as the residual mismatch in the projected galaxy 2PCFs, making it hard to be tied to any MG effect. This implies that a better definition of the mark may be obtained using quantities other than the density field. In the mean time, we know that in the MG models studied here, whether strong deviations from GR happen for a galaxy depends on certain properties of its environment. Here we try a mark defined using the Newtonian gravitational potential, Φ_N , of the galaxy's host dark matter halo, which is one measure of this environment (e.g., [176–178]). As Eq. (33) below shows, Φ_N is closely related to the halo mass M_{200c} ; the abundance and clustering of the latter are strongly affected by MG (cf. Figs. 2, 4), and the way galaxies populate in halos of different masses is also sensitive to gravity (cf. Fig. 3).

In many cosmological models, dark matter halos follow a NFW density profile [85, 86]:

$$\rho_{\text{NFW}} = \frac{\rho_s}{(r/r_s)(1 + r/r_s)^2}, \quad (32)$$

where r_s is the characteristic radius where the profile has a slope of -2 and ρ_s is the density at r_s . The NFW profile also works well for many MG models, including those studied here (e.g., [51, 179–182]). The gravitational potential is accordingly given by [86, 183, 184]:

$$\Phi_N = -\frac{GM_{200c}}{r_{200c}} \frac{\ln(1+c)}{\ln(1+c) - c/(1+c)}, \quad (33)$$

where G is Newton's gravitational constant and c is the concentration parameter defined as $c \equiv r_{200c}/r_s$.

The new mark we define here is a Gaussian function of Φ_N ,

$$m = \frac{1}{\sqrt{2\pi}\sigma_\Phi} \exp \left[-\frac{(\log_{10}(|\Phi_N|) - \Phi_*)^2}{2\sigma_\Phi^2} \right], \quad (34)$$

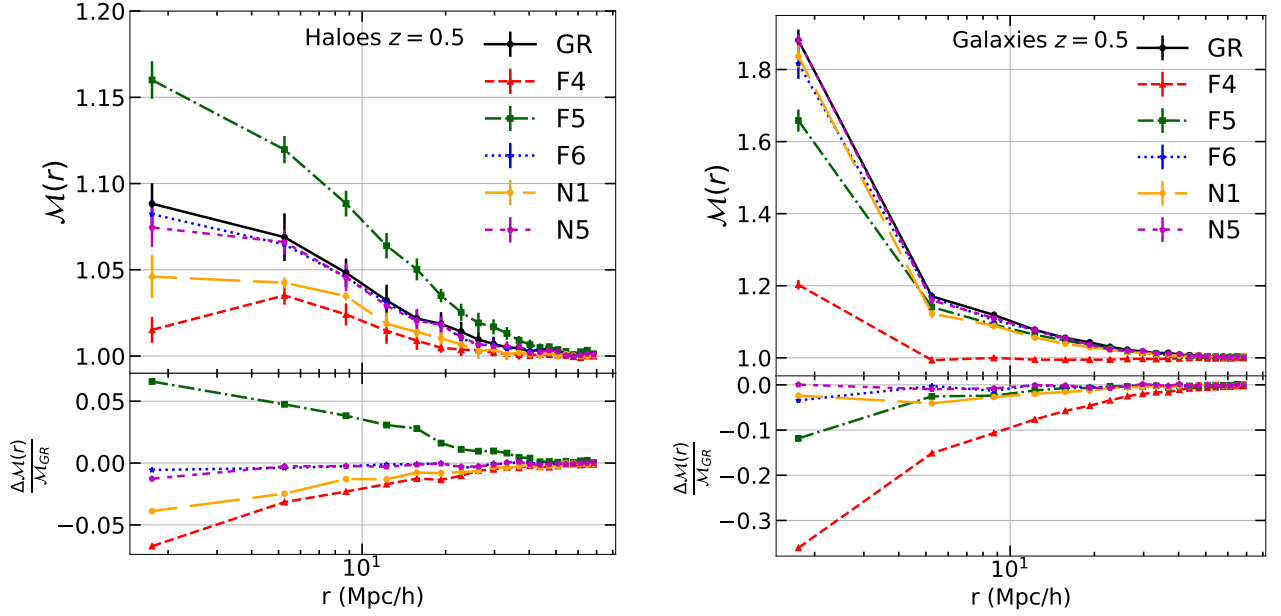


FIG. 13. Gravitational potential marked correlation function with $\Phi_* = -5.295$ and $\sigma_\Phi = 0.1$ at $z = 0.5$ for halos (left panel) and galaxies (right panel). The curves correspond to the averages over the 5 realizations. Different colors and line styles correspond to different models as labeled. The lower panels show the relative difference between the MG models and GR, and the error bars correspond to the standard deviation over the 5 realizations.

where Φ_* and σ_Φ are free parameters which control the amplitude and width of the Gaussian. This mark allows one to up-weight galaxies hosted by halos of certain mass, and as an illustration here we choose host halos with $M_{200c}/[h^{-1}M_\odot] \in [10^{13}, 10^{14}]$, which is where most of the galaxies in our HOD catalogues reside and where the different gravity models differ most substantially (cf. the right panel of Fig. 3). As above, we consider the marked 2PCFs for halos and galaxies separately, using parameter values $\Phi_* = -5.295$ and $\sigma_\Phi = 0.1$ for both tracers: halos and galaxies.

The results for halos are shown in the left-hand panel of Fig. 13 where we find that F5 differs most strongly from GR, with a maximum positive difference of $\sim 7\%$, whereas the other MG models all produce a negative difference with respect to GR. The distinct behavior of F5 is likely a consequence of the MG effect on the halo mass function (cf. right panel of Fig. 2) — the halo catalogs for all models shown in this plot are cut at $10^{13}h^{-1}M_\odot$, so that the numbers of halos in the different models are quite different. The result is qualitatively different if one cuts the halo catalogs to have the same halo number in all models (not shown here).

The right panel of Fig. 13 shows the gravitational potential marked correlation function at $z = 0.5$ for galaxies. We note that F4 differs most from GR, giving a maximum difference of $\sim 36\%$, while F6 and F5 give a monotonically increasing difference of $\sim 3\%$ and $\sim 12\%$, respectively. The nDGP models (N1 and N5) are both very close to GR with a relative difference of $< 2\%$. We have measured the marked 2PCFs in redshift space, and found very similar result. We have also checked the HOD catalogs for which the HOD parameters were tuned so that the 3D galaxy 2PCFs $\xi_{gg}(r)$ match between the MG and GR models, and they again showed very similar features, with slightly smaller (larger) differences of $f(R)$ gravity (nDGP) from GR. While accurately measure the Newtonian potentials of galaxy host halos is nontrivial, methods to estimate halo masses from observations do exist and have been constantly improved (e.g., [185–189]). The result here suggests that using information other than the density field itself to define the mark can be a potential way to increase the constraining power of the marked CF, and it will be worthwhile to pursue this direction further. In particular, it will be interesting to investigate how the results presented in Fig. 13 are affected by uncertainties in the halo mass or potential estimation.

3. Analytical predictions for the marked correlation function

In the discussion of marked correlation functions so far, the results have been obtained numerically from mock galaxy catalogs. It would be useful to have a theoretical template to make analytical predictions, for example based on perturbation theory, which can be used to check consistency at large scales and to more efficiently explore the model parameter space.

In this subsection we compute the White marked correlation function [eq. (30)] using the MG Lagrangian Perturbation Theory (LPT) of [190–192]. Since we are up-weighting low-density regions it is expected that higher than linear order corrections will be highly suppressed, and therefore we focus on the Zel’dovich approximation. The LPT considers the mapping $\mathbf{x}(t) = \mathbf{q} + \Psi(\mathbf{q}, t)$ between initial (Lagrangian) and final (Eulerian) coordinates, and performs a formal expansion on the Lagrangian displacement Ψ . To linear order

$$\Psi^{(1)}(\mathbf{k}, t) = i \frac{\mathbf{k}}{k^2} D_+(k, t) \delta_L(\mathbf{k}, t_0), \quad (35)$$

with $\delta_L(\mathbf{k}, t_0)$ the linear overdensity field, evaluated at t_0 , and D_+ the linear growth factor introduced in Eq. (18). Remember that while D_+ is scale-independent in Λ CDM, it can have a scale dependence in general MG models.

In the majority of MG models which are viable to explain the accelerated expansion of the Universe, including those considered in this work, the evolution of perturbations at a sufficiently early time is indistinguishable from GR, hence the linear power spectrum in MG can be obtained by the relation $P_L^{\text{MG}}(k, t) = [D_+(k, t)/D_+(k, t_0)]^2 P_L^{\text{GR}}(k, t_0)$. Otherwise, it can be obtained from an Einstein-Boltzmann code.

To model halos, we use the biasing prescription of Ref. [144, 193] that evolves initially Lagrangian biased tracers, and consider linear b_1 and second order b_2 local biases. However, the k dependence of the μ function in Eqs. (9,18) implies that even linear local bias becomes scale dependent. In order to model this, one can perform an expansion of the function μ in powers of k^2 and substitute the linear local bias by $b_1 \rightarrow b_1 + b_{\nabla^2 \delta} k^2 + \dots$ with $\nabla^2 \delta$ the curvature bias operator [142]. In addition to the biasing expansion, we expand the mark as $m(\delta_R) \simeq 1 + B_1 \delta_R + \frac{1}{2} B_2 \delta_R^2$, with $\delta_R(\mathbf{q})$ obtained by convolving the matter density field with a Gaussian kernel $W_R \propto e^{-|\mathbf{q}|^2/2R^2}$, and we choose the smoothing scale as $R = 6 h^{-1} \text{Mpc}$.

The analytical prediction is obtained following the standard methods of Convolution Lagrangian Perturbation Theory [144, 167, 194, 195]

$$\begin{aligned} 1 + W(r) = \int \frac{d^3 q}{(2\pi)^{3/2} (\det A)^{1/2}} e^{-\frac{1}{2}(r_i - q_i)(r_j - q_j) A_{ij}^{-1}} & \left\{ 1 + b_1^2 \xi_L - 2b_1 U_i g_i - (b_2 + b_1^2) U_i U_j G_{ij} \right. \\ & - 2b_1 b_2 \xi_L U_i g_i + B_1^2 \xi_{RR} - 2B_1 U_i^R g_i - (B_2 + B_1^2) U_i^R U_j^R G_{ij} - 2B_1 B_2 \xi_{RR} U_i^R g_i + 2b_1 B_1 \xi_R \\ & - 4b_1 B_1 U_i U_j^R G_{ij} - 2b_1^2 B_1 \xi_L U_i^R g_i - 2B_1^2 b_1 \xi_{RR} U_i g_i - 2(b_2 + b_1^2) B_1 \xi_R U_i g_i - 2(B_2 + B_1^2) b_1 \xi_R U_i^R g_i \\ & \left. + 2(1 + b_1 + B_1) b_{\nabla^2 \delta} \nabla^2 \xi_L(q) + b_{\nabla^2 \delta}^2 \nabla^4 \xi_L(q) \right\}. \end{aligned} \quad (36)$$

Here, $\Delta_i = \Psi_i(\mathbf{q}) - \Psi_i(0)$, and

$$\begin{aligned} A_{ij}(\mathbf{q}) &= \langle \Delta_i \Delta_j \rangle, \quad U_i(\mathbf{q}) = \langle \delta(\mathbf{q}) \Delta_i \rangle, \quad U_i^R(\mathbf{q}) = \langle \delta_R(\mathbf{q}) \Delta_i \rangle \\ \xi_L(q) &= \langle \delta(\mathbf{q}) \delta(0) \rangle, \quad \xi_R(q) = \langle \delta_R(\mathbf{q}) \delta(0) \rangle, \quad \xi_{RR}(q) = \langle \delta_R(\mathbf{q}) \delta_R(0) \rangle, \end{aligned} \quad (37)$$

and the tensors are defined as $g_i = A_{ij}^{-1}(q_j - r_j)$, $G_{ij} = A_{ij}^{-1} - g_i g_j$. The (unmarked) correlation function for tracer type X is given by $\xi_X(r) = W(r; B_i = 0)$. Eqs. (28,36) provide a good approximation for large scales as long as the RMS of matter Lagrangian displacements is smaller than the smoothing scale R . In other words, as larger the smoothing scale is, or as higher the evaluation redshift is, more accurate predictions are expected [195].

By applying the above formalism to Λ CDM and the three variants of $f(R)$ gravity considered here, we find that the trends of the marked correlation function for matter and biased tracers are qualitatively different, as shown in Fig. 14. In the matter case, the F4, F5, and F6 marked correlation functions fall below that of GR, a behavior that has been observed recently in simulations [168]. This can be interpreted by considering the mean mark $\bar{m} = (1 + b_1) B_1 \sigma_R^2$, which shows that for the unbiased case $\bar{m}^{\text{MG}} < \bar{m}^{\text{GR}}$, simply because $\sigma_R^{\text{MG}} > \sigma_R^{\text{GR}}$ and $B_1 < 0$. For the biased tracer case, we have used halos instead of HOD galaxies. A key factor here is to identify the halo bias, because its effect is to reduce the marked correlations [167] – the larger the bias the larger the reduction. A local bias is typically controlled by two parameters, the density threshold for collapse, $\delta_c(M)$, and the variance $\sigma(M)$ (roughly through their ratio). The former (latter) is smaller (bigger) in $f(R)$ gravity, resulting in smaller bias values for our MG models than for GR [191]. As a result, the halo bias may bring the marked correlation above that of GR, as is the case for the halo masses interval presented here. These results can be understood on physical grounds as even though low density regions in GR correspond to even more underdense regions in $f(R)$, the halos are more efficiently formed due to a larger gravitational strength.

To compare with simulations, we let free the bias parameters and fit them on scales $r > 20 h^{-1} \text{Mpc}$, finding a good agreement within the error bars of the data. Our best fit is given by linear local Lagrangian biases $b_1^{\text{GR}} = 1.15$, $b_1^{\text{F6}} = 1.12$, $b_1^{\text{F5}} = 1.07$, and $b_1^{\text{F4}} = 0.90$, and second order biases $b_2^{\text{GR}} = 0.8$, $b_2^{\text{F6}} = 0.7$, $b_2^{\text{F5}} = 0.5$, and $b_2^{\text{F4}} = 0.3$, while the curvature bias does not contribute significantly over this region. In Fig. 14, we also show the number weighted density estimates, as presented in the left panel of Fig. 10. Therefore, at least on scales $r \gtrsim 20 h^{-1} \text{Mpc}$ the Lagrangian perturbation approach is a promising tool to produce theoretical templates for future gravity tests against survey data.

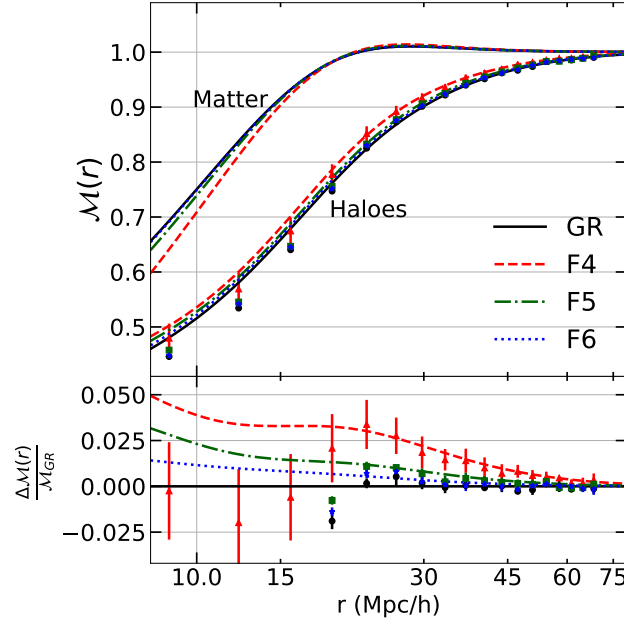


FIG. 14. The Zel'dovich-approximation prediction of the marked correlation function Eq. (30) with $\rho_* = 4$ and $p = 10$, for models F4 (shown in dashed red curves), F5 (dot-dashed green), F6 (dotted blue), and GR (solid black) at redshift $z = 0.5$. The data points are the simulation results taken from Fig 10: black data points with error bars are for GR, blue for F6, green for F5, and red for F4. The lower panel shows the relative differences in the marked correlation function for traces (and the data) with respect to the Zel'dovich approximation in GR.

model	$r_p > 0.5h^{-1}\text{Mpc}$	$r_p > 1h^{-1}\text{Mpc}$	$r_p > 4h^{-1}\text{Mpc}$	$r_p > 8h^{-1}\text{Mpc}$
F4	355	106	191	81
F5	854	52	25	30
F6	16	28	13	18
N1	28	26	67	42
N5	12	15	18	23

TABLE III. The χ^2 with 10 degrees of freedom for the mean amplitude as a function of the minimum r_p used in the calculation. The different rows corresponds to different MG models and the different columns for different minimum scale. See the text for more details and discussion.

4. Clustering in over-dense environments

The idea behind using marked correlation function to test models is that different gravity models can show clustering patterns that are density dependent in different ways. Actually, instead of estimating the marked correlation functions as done above, one can more directly estimate the clustering in different density environments and compare them to distinguish between different gravity models. Although our simulation have matched the overall projected 2PCFs, we may still find the signature of different gravity models and screening mechanisms by looking at $w_p(r_p)$ for galaxies in various environments.

Fig. 15 shows the projected correlation functions computed for the GR galaxy catalogs in different δ_8 percentile bins, where δ_8 is the environmental density contrast averaged over a $8 h^{-1}\text{Mpc}$ scale using a modified tessellation scheme. Although the results here are for periodic boxes, the method can account for survey mask and have been applied for real surveys [196]. Briefly, we count the number of randoms nearest to each galaxy and turn that into an estimate of volume occupied by the galaxy. This automatically accounts for the survey mask and incompleteness (see [196] for more details). This only requires a random catalog sampling the volume of the survey which will be anyway produced as part of the standard survey pipeline and does not require any extra information. One should think of δ_8 as over-density of galaxy field here. Once the over-density is assigned for each galaxy, we rank the galaxies based on their δ_8 values and split the catalog into 10 sub-samples each containing 10% of the whole sample. We then measure the two-point correlation functions using the Landy-Szalay estimator and project out the line-of-sight information to obtain projected correlation functions. The measurement of projected correlation function (w_p) is obtained with 20 logarithmic bins in r_p covering $0.1 - 30 h^{-1}\text{Mpc}$. We then define the ‘amplitude’ as the ratio of projected correlation function between a MG model and GR, averaged over a range of scale above a chosen r_p .

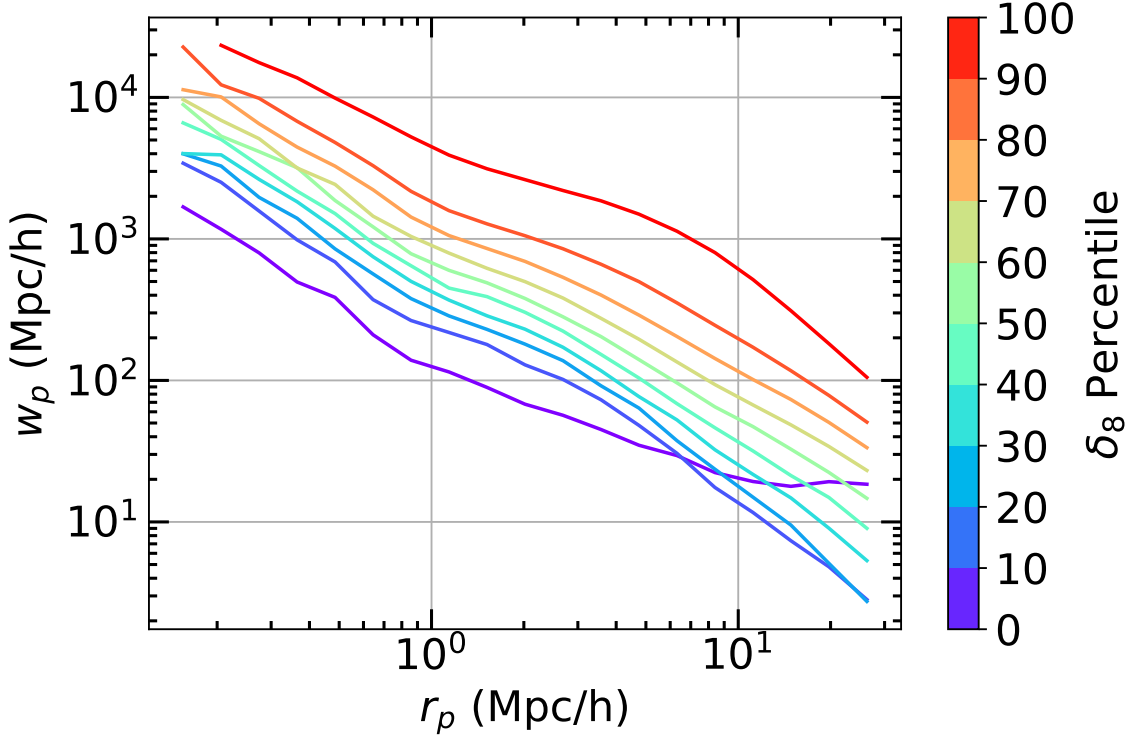


FIG. 15. The projected correlation function, $w_p(r_p)$, estimated for GR HOD catalogues at redshift $z = 0.5$. The different colored line show the w_p for over-density (δ_8) percentiles indicated by the colorbar. At small scale ($r_p < 5 h^{-1}\text{Mpc}$) the clustering amplitude increases with δ_8 monotonically. At larger scale the clustering amplitude shows non-monotonic behavior with δ_8 as expected [196].

Fig. 16 shows the mean amplitude as the function of δ_8 percentile for all the MG models at $z = 0.5$ and with minimum r_p of $4 h^{-1}\text{Mpc}$. The curves is the mean of the amplitude values from the five simulation boxes, and the errorbar denotes the error on the mean of five boxes. One can see that in high-density environments (i.e., the five largest values of δ_8) the different models all agree with each other very well, confirming the effect of screening. For smaller values of δ_8 , the amplitude generally show more deviations from 1: there are large scatters for the three $f(R)$ variants, making it hard to see any clear trend, while nDGP shows a clearer pattern. This is a consequence of the range of scale ($[4, 30] h^{-1}\text{Mpc}$) used to calculate the amplitude, because the effect of $f(R)$ gravity – especially for F6 and F5 – is restricted to smaller scales not used in the calculation, while the nDGP model deviates from ΛCDM on larger scales which are covered by the calculation. We show the results for $r_p < 4 h^{-1}\text{Mpc}$ in Fig. 16 here because, as discussed below, this is the optimal choice for N1, but we have checked other minimum values of r_p and found that decreasing it generally leads to larger amplitude values for the $f(R)$ variants, as we show more quantitatively next.

In order to combine the information from all the δ_8 subsamples, we also need to estimate the covariance between the amplitude. We generate 100 jackknife realizations for each model and each δ_8 subsample, and use these jackknife realizations to estimate the correlation matrix of the mean amplitude with δ_8 . The correlation matrix is then scaled by the error estimated from the mean of 5 boxes to get the covariance matrix $\mathbb{C}(i, j)$, where i, j represent the different δ_8 subsamples. This covariance matrix is then employed to estimate the χ^2 for each MG model with respect to GR, according to

$$\chi^2 = \sum_{i,j=1}^{10} (A_i - A_i^{\text{GR}}) \mathbb{C}(i, j) (A_j - A_j^{\text{GR}}), \quad (38)$$

where A_i denotes the amplitude value of the i th δ_8 subsample, calculated using the projected 2PCFs measured in the range of r_p between a varying minimum value and $30 h^{-1}\text{Mpc}$. Our estimated χ^2 for different models and different choices of the minimum r_p are given on Table III. As mentioned above, the $f(R)$ models show increases in χ^2 as we include smaller scales. F4, which is the model with very little screening and the largest Compton wavelength of the background scalar field, can be detected with high confidence independent of choice of scale; F5 can be detected with high confidence if at least scales up to $1 h^{-1}\text{Mpc}$ is used

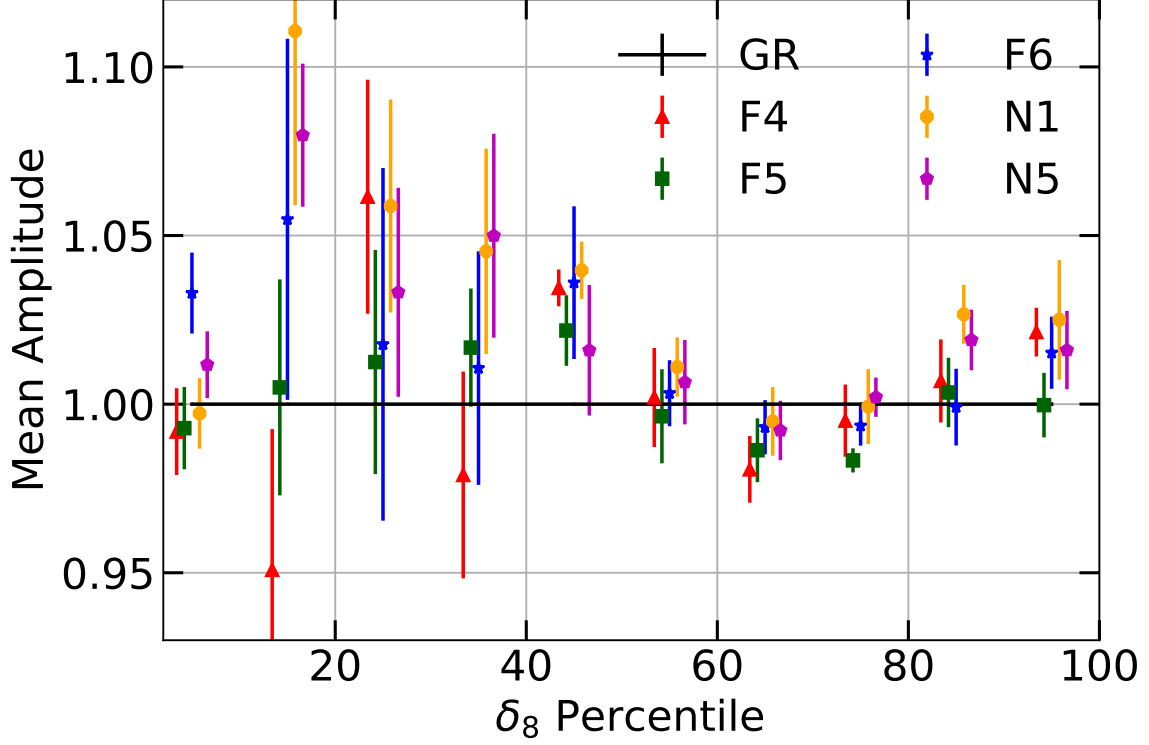


FIG. 16. The mean amplitude as a function of over-density (δ_8). The different MG models are shown with different colors, and the $1\text{-}\sigma$ errors are computed as the error on the mean of the 5 HOD catalogs. The F4, F5, and N1 models show large χ^2 values and should be detectable at better than 3σ significance but F6 and N5 are very close to GR and do not show any statistically significant difference. Note the points representing different models at same over-density are slightly shifted horizontally to avoid clutter and increase visibility.

and F6 can not be detected independent of scale used. We note that for nDGP models, including smaller scales first increases the χ^2 up to an optimal scale and then decreases it: N1 can be detected with an optimal choice of minimum scale as $4\ h^{-1}\text{Mpc}$ while N5 can not be detected independent of scale used.

Given that the number density of galaxies in future surveys such as DESI will be higher and the volume will be much larger than the simulations used, the statistical error from DESI will be much smaller. However, at the same time the real data also have systematic errors not included in the analysis here. For this particular test, the most important systematics will come from fibre collision and completeness uncertainty, both of which are major concerns for DESI main science cases and huge amounts of work have been put in to keep them under control. These conclusions about detecting the difference should hold for DESI in presence of systematics as long as they do not dominate the total error budget.

We have checked this estimator using the mock galaxy catalogs where the HOD parameters were tuned to match the real space 3D galaxy 2PCFs in different models, and found the results change slightly for the lowest five bins of δ_8 , which suggests that the uncertainty associated with HOD modelling can be a theoretical systematic which needs better control. On the other hand, we also note the MG models can be distinguished from GR using the scales above $\sim 1\ h^{-1}\text{Mpc}$, and hence these detection should be largely free from the baryonic effects.

5. Discussion

While the halo and HOD galaxy catalogs might be several steps away from the dark matter distribution, the behavior of the marked correlation functions above can be qualitatively understood. A marked correlation function quantifies the correlation between the marks. In the case of the mark in Eq. (30), for a tracer i – which can be a galaxy or a halo – with $m_i < \bar{m}$, it is likely to find another tracer j nearby with $m_j < \bar{m}$ because these tracers are in dense regions; for a tracer i with $m_i > \bar{m}$, which means that it is in a low-density region, it is less likely to find a neighbor j . This leads to $\mathcal{M} < 1$. A similar reasoning can

explain why $\mathcal{M} > 1$ in the case of gravitational potential mark.

The model differences between $f(R)$ gravity and Λ CDM can be explained as follows. An enhanced gravity means that more halos can form in regions where these tracers have low densities in the Λ CDM counterpart, while in high-density regions the increase in halo number density is less significant due to either more efficient screening (e.g., in F5 and F6) or more frequent mergers (e.g., in F4). For the mark in Eq. (30), this means that for a halo i with $m_i > \bar{m}$ (i.e., in a low-density region), it is more likely to find neighbors with $m_j > \bar{m}$ either from the same cell or from a neighboring cell, leading to a larger $\mathcal{M}(r)$ in MG than in GR. In the cases of F5 and F4, the difference is significant and can be observed to larger halo separations (cf. Fig. 10). For the gravitational potential mark, again, the enhanced gravity produces more halos with $M_{200c} > 10^{13} h^{-1} M_\odot$ which correspond to smaller halos in GR; these halos are less biased and more uniformly distributed; they also have relatively low marks, so that they increase the likelihood of finding a halo i with $m_i < \bar{m}$ near a halo j with $m_j > \bar{m}$, thereby reducing $\mathcal{M}(r)$. The strange behavior of F5 (left panel of Fig. 13) is probably a consequence of this model having many more halos of $M \sim 10^{13} h^{-1} M_\odot$, which significantly decreases $\bar{m}$, leading to an overall increase in $\mathcal{M}(r)$. This suggests that it may be possible or even preferable to define marks in ways so to pick out the region of the HMF with the most significant model differences.

We also noticed above that for the mark in Eq. (30) the marked correlation functions for galaxies show much smaller model differences, and a less clear trend, than for the gravitational potential mark. This is probably because in the former case the marks themselves are obtained from the galaxy field, which has been tuned to match in all the different models. In other words, with a similar galaxy number density and spatial distribution, the marks for the HOD galaxies in the various models are similar, and so are their correlations. In contrast, for the gravitational potential mark, the mark itself encodes external information beyond what is contained in the galaxy distribution, and what we see in the right panel of Fig. 13 is the correlation of this information. This suggests that it can prove useful to study the many possible ways to define the marks for marked correlation functions, by using complementary observational information and by varying the parameter values of the marks.

The nDGP models are very difficult to be distinguished from Λ CDM either using halos or galaxies by the marked correlation function. In these models, deviations from Λ CDM appear only at the high-mass end of the mass function (Fig. 2). In addition, the halo correlation function is very close to Λ CDM (Fig. 4) despite the fact that the underlying dark matter clustering is enhanced (Fig. 1). The difference of the HOD parameters is less prominent and the galaxy numbers as a function of the host halo mass is also very similar to Λ CDM (Fig. 3). As a consequence, irrespective of tracers and marks, the difference between nDGP and Λ CDM is always small. The environmentally-dependent clustering of galaxies shown in Figs. 15 and 16, and Table III, on the other hand, seems more promising than the marked statistics for both the $f(R)$ and nDGP models.

C. Beyond two-point statistics

In cosmological perturbation theory, see, e.g., [197], the velocity-velocity and density-velocity couplings render the evolution equations nonlinear, generating non-Gaussian late-time density fields from a Gaussian random field as the initial condition, and the non-Gaussianities may be enhanced in models with a stronger gravity. Furthermore, primordial non-Gaussianities due to the detailed properties of inflation are generically expected at some level. Both of these features indicate that the galaxy distribution cannot be fully characterized by the two-point correlation function. Higher-order correlation functions contain additional information regarding the nature of gravitational interactions, making them useful to test deviations from GR (see, e.g., [198, 199]), in particular for theories where nonlinear interactions play a significant role in screening modifications at short scales. Although there are several theoretical and observational studies showing the potential of such an estimator for LSS analysis (see, for example, the review [197] or the studies summarized in [200, 201]), it is still relatively poorly explored compared to its lower order cousin, the 2PCF.

In III C 1 and III C 2 we respectively discuss the information to investigate modifications to gravity in the three-point correlation function (3PCF) and its Fourier space counterpart, the bispectrum. In III C 3 we consider constraints using hierarchical clustering, while in III C 4 we discussed the application of Minkowski functionals, and in III C 5 we close with a consideration of the use of phase space information from stacked clusters.

1. Three-point correlation functions

Although three-point statistics in configuration and Fourier spaces have theoretically the same information, in practice their implementations are different. Measurement in configuration space is conceptually more straightforward, and does not require special effort to deal with gridding, numerical Fourier transforms or shot noise corrections [202]. Moreover, large surveys such as BOSS and DESI bias the clustering due to the targeting algorithm. There are methods to mitigate this effect, which may be more easily implemented in configuration space than Fourier space (e.g., [203–206], though this is still a subject under study. Therefore in this subsection we focus on the three-point correlation function (3PCF) first.

Large surveys such as DESI will render a brute-force 3PCF calculation computationally challenging. However, a few efficient algorithms have lately been developed that should help surmount this obstacle (for example [207–214]); of particular interest is

the one recently proposed by Slepian and Eisenstein [210], which scales as $\mathcal{O}(N^2)$ for the isotropic case, with N the number of objects. The isotropic 3PCF, $\xi^{(3)}$, which does not track the line of sight, accounts for triangular configurations parametrized by 3 variables, which can be chosen to be the two sides of the triangle, $|\vec{r}_1| = r_1$ and $|\vec{r}_2| = r_2$, and the opening angle between those sides, $\cos \theta_{12} = \hat{r}_1 \cdot \hat{r}_2$. The key idea to reduce the computational load is to use a Legendre polynomial basis [215], namely

$$\xi^{(3)}(r_1, r_2, \hat{r}_1 \cdot \hat{r}_2) = \sum_{\ell} \xi_{\ell}^{(3)}(r_1, r_2) P_{\ell}(\hat{r}_1 \cdot \hat{r}_2), \quad (39)$$

with P_{ℓ} the Legendre polynomials. While this basis does not have a straightforward geometrical interpretation, one should bear in mind that each multipole contains information about all possible shapes because the coefficients are averages over all triangles weighted by the Legendre polynomials. For example, the monopole contribution is averaged over all triangles of sides r_1 and r_2 with an equal (constant) weight on $\hat{r}_1 \cdot \hat{r}_2$. The multipole basis reduces the computational cost because the Legendre polynomial can be factored into spherical harmonics of one unit vector each, avoiding the need to explicitly construct the opening angle between the two galaxies at $\vec{r}_1$ and $\vec{r}_2$ from the vertex.

The multipole representation has additional benefits beyond the computational acceleration it offers. For instance, it displays more information on the isotropic 3PCF than using particular shapes (such as the equilateral or specific triangles with two sides fixed). In the case of a converging series, only a few multipole moments $\xi_{\ell}^{(3)}$ are needed to capture most of the 3PCF (studied in [210], seen in Fig. 8 for a triangle with $r_1, r_2 = 70, 40 h^{-1} \text{Mpc}$, and further discussed in [216]). Numerical estimates suggest that the series easily converges away from the diagonal ($r_1 \neq r_2$), while for $r_1 = r_2$ it converges on scales larger than a given value, which for our catalogs is around $10 h^{-1} \text{Mpc}$ by using the first 10 multipoles. Therefore, it is expected that the first ten multipoles capture most of the information up to some small regions in the r_1 - r_2 plane. Another benefit of this Legendre basis is the simplicity of calculating the edge corrections due to the survey boundary (for details see [210]).

For these reasons, in this subsection we shall mostly focus on the algorithm for the isotropic 3PCF proposed in Ref. [210]⁷. However, for completeness we will also show some results from considering a few triangle configurations using a full 3PCF algorithm near the end. One can indeed calculate the full 3PCF for any triangular shape using the multipole expansion Eq. (39). For fixed values in r_1 and r_2 , all the remaining 3PCF data can be compressed into a single-valued function of the opening angle between $\vec{r}_1$ and $\vec{r}_2$ (or equivalently the side $\vec{r}_3$). This approach of arbitrarily choosing two sides of the triangle has often been followed in the literature to compare three point statistics. However, each coefficient $\xi_{\ell}^{(3)}(r_1, r_2)$ is a function of two independent scales, and thus contains all the relevant information about interactions in a particular gravitational theory. In practice, the largest deviations of alternative models to GR may not be found where the strongest 3PCF signal is. Therefore, it is wise to study both the full 3PCF and particular triangular shapes when testing gravity. Moreover, we find that the previous Legendre expansion contains interesting information about gravitational clustering at each multipole, suggesting that a comparison between GR and modified gravity models for each multipole may be more useful than using the full correlation function, where often the lower multipoles comprise the dominant portion of the signal.

We run the 3PCF algorithm, up to a maximum multipole of $\ell = 9$, over the five realizations of each of the different models described in Section II. For all models we use the same random catalogs based on an equal box size as the data, and 600,000 random points. To avoid Poisson noise due to the randoms yet also elude the need to compute a large number of random pairs, we use 50 realizations of the random catalogs and average them over the data using the Szapudi-Szalay estimator for the 3PCF [218]

$$\xi^{(3)} = \frac{\overline{NNN}}{RRR}, \quad (40)$$

where $N = D - R$ and once the product is expanded, the XXX (with X either D or R) refer to the histogram of distances of triplets where each vertex belongs to the corresponding X -space. The overbar denotes an average over the different random catalog realizations.

To compare the correlation function coefficients $\xi_{\ell}^{(3)}$ between models, we use the variable

$$\Delta_{\sigma} \xi_{\ell}^{(3)} = \frac{\xi_{\ell}^{(3)}(X) - \xi_{\ell}^{(3)}(\text{GR})}{\sqrt{\sigma_{\ell}^2(X) + \sigma_{\ell}^2(\text{GR})}}, \quad (41)$$

where we have taken the mean over the 5 realizations of the modified gravity models, X , and General Relativity, GR, to obtain $\xi_{\ell}^{(3)}(X)$ and $\xi_{\ell}^{(3)}(\text{GR})$ respectively, together with their standard deviations $\sigma_{\ell}^2(X)$ and $\sigma_{\ell}^2(\text{GR})$, which we have assumed to be uncorrelated. This expression is the square root contribution of each bin to the overall χ^2 , using GR as our fiducial model. The

⁷ The code is publicly available in the n-body kit [217].



FIG. 17. 3PCF comparison of the models using the variable $\Delta_{\sigma}^{(3)}\xi_{\ell}$ defined in Eq. (41) and the multipole decomposition. Saturated regions may reach values beyond the colorbar shown. We use 10 bins for each triangle side.

results for the 5 models are shown in Fig. 17. To compare the complete 3PCF, $\xi^{(3)}$, we use a similar expression to Eq. (41), but with the complete 3PCFs and their propagated errors in place of the expressions with ℓ 's.

There is a complex structure in the resulting values for the variable $\Delta_{\sigma}^{(3)}\xi_{\ell}$, showing departures of each modified gravity model from GR on all scales and multipoles. Although each model shows a different pattern, there are some generic features shared by all. In most of the parameter space explored in Fig. 17, the differences are roughly of the same order as the scatter between model realizations, which translates into $\Delta\xi_{\ell}^{(3)} \sim 1$. However, there are departures with $\Delta\xi_{\ell}^{(3)} > 1$ in different regions of the r_1 - r_2 plane (see the high intensity or strongly saturated regions of Fig. 17). These regions come in two classes: isolated (almost point-like) areas where at least one of the scales (r_1 or r_2) is large, and the small-scale sector where both sides of the triangle are smaller than roughly $40h^{-1}\text{Mpc}$. In the later case, we expect these strong departures from GR due to the nonlinear evolution of modes, hence all models (and multipoles) have some degree of departure from GR on these scales, which are better appreciated for F4 and N1. In contrast, for the case of the isolated points, these regions only show up in particular multipoles and models, and one would need more realizations of each model to assess whether they are physical. Actually, if one constructs the HOD catalogs differently (for examples, by choosing the HOD parameters independently for each model realization, or by matching the full 2PCF between ΛCDM and the MG models) these highly-saturated points of Fig. 17 on large scales may change their saturation, suggesting that some of them are not truly physical. However, under these different HOD prescriptions the overall patterns remain invariant, supporting evidence for using the 3PCF to test gravity. It is important to stress that using an equivalent expression to Eq. (41) for the 2PCF, we find that all models depart from GR with significance less than $\Delta\xi^{(2)} < 0.4$ on scales

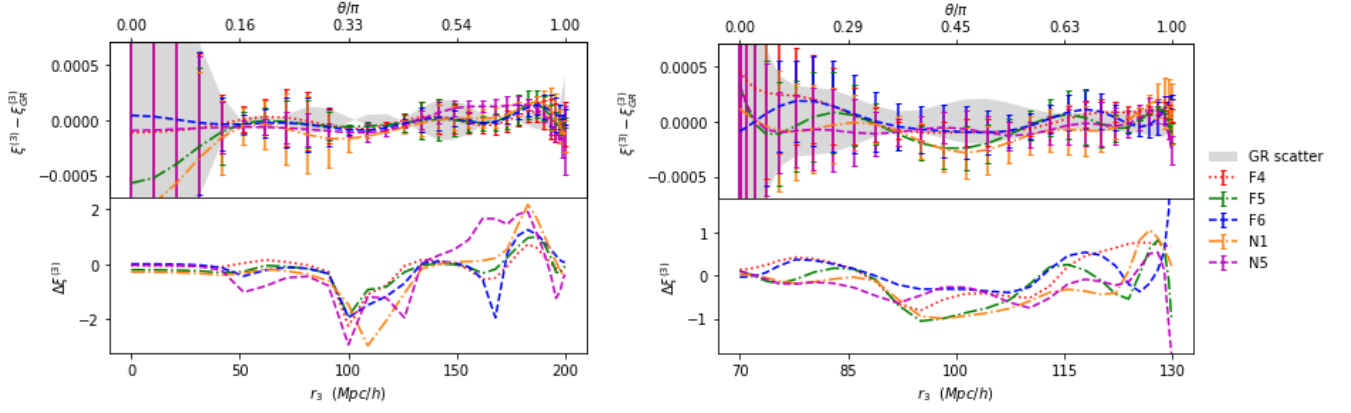


FIG. 18. Full 3PCFs for $r_1 = 100h^{-1}\text{Mpc}$ and $r_2 = r_1$ (Left) or $r_2 = 30h^{-1}\text{Mpc}$ (Right). The *top panels* show differences of the MG 3PCF with respect to GR as a function of the third triangle side r_3 , or equivalently, the opening angle θ_{12} between r_1 and r_2 . The *bottom panels* show the model differences using the analog of the variable $\Delta_\sigma \xi_\ell^{(3)}$ but for the full 3PCF. We use the first $\ell < 10$ multipole moments in Fig. 17 to construct the complete 3PCF.

between 10 and $150h^{-1}\text{Mpc}$ ⁸. This suggests strong departures of the 3PCF compared with those found in the 2PCF; however, further tests are needed to check if the 3PCF offers additional or complementary information for testing the gravity models.

The 3PCF monopole is slightly suppressed with respect to the higher multipoles, although one should have in mind that each multipole contribution to the full 3PCF is weighted by the Legendre polynomials, hence the full 3PCF is usually dominated by the first few multipoles. Moreover, largest departures from GR are found for F4, F5, and N1, as expected (see Fig. 17), and are particularly noticeable in the quadrupole. For these three models, in the small-scale region of the plots (where both sides of the triangle are smaller than roughly $40h^{-1}\text{Mpc}$), the monopole and dipole show less clustering with respect to GR, manifested by a redder color, but an enhancement of clustering (bluer color) in the same regions for the higher multipoles. As expected, models F6 and N5 are generically closer to GR in the whole parameter space. However, for particular shape configurations, specially in the small-scale region, their departures from GR may be larger than for other models, as we will discuss later. Furthermore, for certain multipoles there is no monotonic trend in the different MG models in certain regions of r_1 - r_2 , e.g., at $r_1, r_2 < 50h^{-1}\text{Mpc}$ the quadrupole is stronger in F4, F5, but weaker in F6, than in GR, which is possibly a residual of the HOD tuning; this implies that understanding the impact of the uncertainties in the galaxy-halo connection will be important for using the 3PCF multipoles to test models.

We calculate particular cases of the full 3PCF using the Legendre basis to exemplify in a different way its complex dependence on the triangle sides r_1, r_2 and the opening angle θ_{12} . When computing the full 3PCF for a specific bin pair, the signal shows convergence in the Legendre Series in most of the parameter space, except for the first couple of bins along the diagonal. The 3PCF reaches the highest values often when $\theta = 0$, and in a few other cases when $\theta = \pi$, points where the Legendre polynomials also peak (taking on values ± 1 for all ℓ). The propagated error, coming from the standard deviation among realizations, grows towards $\theta = 0$, implying that for small angles our measurement of the deviation with respect to GR (the $\Delta\xi^{(3)}$ variable) would shrink to zero while for larger angles it would be enhanced. Moreover, one should bare in mind that the strength of the signal in $\Delta\xi^{(3)}$ depends on a large difference in the 3PCFs and a small combined variance, hence tiny fluctuations in the variance using the multipole basis may lead to larger changes in $\Delta\xi^{(3)}$. As a result, this would either enlarge or reduce the significance of the departure from GR, or introduce a small shift on the angle θ where the departure occurs. Therefore, further realizations (and possibly higher multipoles) may be needed to increase the stability of our estimator $\Delta\xi^{(3)}$ for large scales. While bearing this in mind, we explore some triangle configurations.

Figure 18 shows the dependence on the opening angle (or equivalently the distance r_3 to close the triangle) by the full 3PCF, for cases where at least one of the sides is around the BAO scale ($r_1 = 100h^{-1}\text{Mpc}$). For the other side, r_2 , we consider two cases: one along the diagonal ($r_2 = r_1$), and another far from the diagonal ($r_2 = 30h^{-1}\text{Mpc}$). In both cases, the MG departures from GR show an oscillatory behavior, which is remnant of the multipole expansion, affecting both the signal and the error bars. In the diagonal case, the largest differences amongst models are found at certain values of the opening angle, with a particularly strong enhancement for the equilateral shape ($\theta = \pi/3$). Actually, for other bin pairs along the diagonal ($r_2 = r_1$) away from the BAO scale, we often find stronger departures from GR around the equilateral shape. In contrast, the off-diagonal cases show

⁸ Note that this range is different from the one within which the 2PCFs are used for the tuning of the HOD parameters.

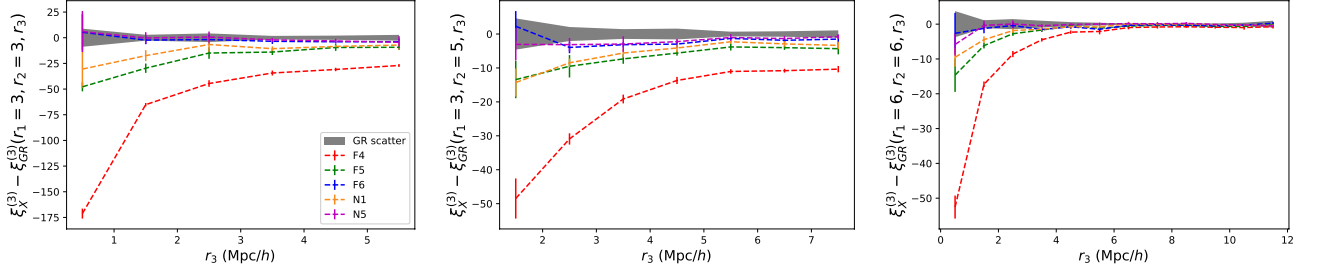


FIG. 19. The 3-point correlation function difference between GR and several modified gravity scenarios. *Left*: A triangular configuration with $r_1 = r_2 = 3 h^{-1} \text{Mpc}$. *Center*: with $r_1 = 3 h^{-1} \text{Mpc}$ and $r_2 = 5 h^{-1} \text{Mpc}$. *Right*: $r_1 = r_2 = 6 h^{-1} \text{Mpc}$. The shaded regions are the standard deviations of the five realizations of the HOD catalogs for GR. All results are for $z = 0.5$.

a non-trivial small-to-large scale mixing with no particular features over the different MG models. Further realizations would be desirable to assess strong departures from GR. A careful exploration of the non-diagonal regions may shed more light on the correlations between overdensities and underdensities, which should strongly differ from GR when screening mechanisms are at work.

For all MG models, the small-scale region (where all triangle sides are $\lesssim 40 h^{-1} \text{Mpc}$) contain important deviations from GR, as one would expect due to the nonlinear evolution of matter perturbations. This is clearly shown in Fig. 17, especially for the F4 model. Consequently, in the remaining part of this subsection we focus on the small-scale limit for both r_1 and r_2 , where the overall variation with respect to GR is larger and the structure is richer. However, here we are reaching scales where the density of objects, together with the simulation resolution and the multipole truncation, disfavors the convergence of the Legendre series Eq. (39), especially in the diagonal case ($r_1 = r_2$). Regardless of the convergence behavior, the multipole moments of the 3PCF nonetheless remain well-defined in and of themselves, hence one may still employ this decomposition truncated at a given ℓ as a discriminator to compare models. The behaviors are now rather particular to each pair of bins, with drastic changes from adjacent cells. This rapid variation implies the need for a careful choice of bin size, simulation resolution, and HOD properties. Again, further studies with higher simulation resolution, larger volume or more realizations of each model, and different HOD constructions may be required to fully assess the physical significance of the 3PCF structures due to MG on the shortest scales using the multipole expansion.

One may worry that the MG signal is degenerated with certain systematics, such as the allocation algorithm used by the DESI instrument which artificially modifies the clustering [13]. However, given the complexity in the structures of the 3PCF, it is hard for other effects to mimic a true MG signal such as the ones reviewed here. For example, in the case of previously mentioned DESI assignment effect, by using a subsample of the mocks described in Refs [203, 219], the 3PCF was computed and compared with the full sample without allocation with the yearly passes, and a very strong difference was found along the diagonal in the multipole basis for modes higher than the monopole (details will be found in a future publication). The signal is clearly stronger than the MG one, but with a very different structure from the one appreciated in Fig. 17. Therefore, it is advisable to consider the full 3PCF information, and not particular shapes, where there could be degeneracies.

Given the poor convergence of the series on small scales, one may also try to run algorithms calculating the full 3PCF. Even though they are computationally more costly with respect to the code presented previously, for triangular configurations below $10 h^{-1} \text{Mpc}$, they can indeed be realistically run even for large surveys such as DESI. Therefore, following Ref. [201], we focus on the small-scale 3PCF in redshift space, using a parallel KD-TREE algorithm⁹ for efficient neighbor matching. We computed all triplets in various triangular configurations without using any approximation. Using a random catalog 20 times larger than the galaxy data, we measured all possible triangular configurations with scales $0 < r_1 < r_2 < r_3 < 20 h^{-1} \text{Mpc}$. In Fig. 19, we show the differences between the 3PCFs of GR and the various MG models, focusing in three particular triangular configurations. One may notice that apart from the closest models to GR (F6 and N5), all others deviate significantly. The difference with respect to GR tends to decrease for larger values of r_1, r_2 and r_3 , consistent with the discussion above on the 3PCF multipoles. We also checked these results for mock galaxy catalogs where the HOD parameters were tuned by matching the 3D galaxy 2PCFs of the different models, and found them to be very insensitive to the HOD details, which confirms that the 3PCFs encode important information useful for testing gravity. The big challenges in this regime are those shared by the 2PCF on the same small scales, namely to model the signal and to clearly understand the involved systematics.

The analysis in this subsection has demonstrated that the 3PCF of galaxies contains independent information to the traditional 2-point statistics, and has the capacity to distinguish MG from GR and break degeneracies on a wide range of scales. The use

⁹ We use the public code, KSTAT, <https://bitbucket.org/csabiu/kstat>.

of the fast algorithm in [210] allows for an efficient analysis of DESI-like surveys and simulations, with an easy-to-picture comparison of models based on the Legendre series coefficients. Moreover, based on this algorithm, some authors have constructed and optimized the anisotropic 3PCF in redshift space [220, 221], which could be used in large galaxy surveys such as DESI, and capture the RSD physics to further test gravity. Other ideas, such as the recent squeezed 3PCF construction [222], may be worth exploring to test gravity in galaxy surveys as well. Finally, the potential of the 3PCF goes beyond the models and conditions explored here, and may be used to characterize other possible deviations from Λ CDM and, given the richness in the pattern of differences, it may also provide a platform to understand other systematics such as fibre collision.

2. Galaxy bispectrum

The galaxy bispectrum is the counterpart of the 3PCF of the galaxy field in Fourier space, and forms a Fourier transform pair with the configuration-space 3PCF that was discussed in Sec. III C 1. In principle, these two measures would carry the same information, but in practice this is not guaranteed as our analyses are restricted to a finite range of scales, and configuration- and Fourier-space statistics are impacted differently by systematic effects. In addition, modelling approaches of configuration- and Fourier-space quantities tend to differ and come with their own unique challenges, but models of the bispectrum have received more attention in the recent literature, putting them generally into a more mature state (see, e.g., [199, 223–226]). For these reasons, the bispectrum can provide us with a valuable complementary point of view when studying the impacts of the MG dynamics in the $f(R)$ and DGP models, and our aim of this subsection is therefore to present measurements of the bispectrum from the various galaxy mock catalogs introduced in Sec. II B, both in real and in redshift space.

The bispectrum $B(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3)$ is defined as the correlation of three density modes,

$$\langle \delta(\mathbf{k}_1)\delta(\mathbf{k}_2)\delta(\mathbf{k}_3) \rangle \equiv (2\pi)^3 B(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3), \quad (42)$$

where the three wave vectors $\mathbf{k}_1$, $\mathbf{k}_2$ and $\mathbf{k}_3$ form a closed triangle. In the absence of redshift-space distortions (or statistical anisotropies in general), the bispectrum is fully defined in terms of the length of the three triangle sides, otherwise two additional variables are required to capture the orientation of the triangle with respect to the LOS (or the direction along which isotropy is broken). We still lack a thorough understanding of whether particular triangle configurations are prominently affected in theories of modified gravity, and so in the following we are going to consider all possible triangles between the two extreme scales $k_{\min}$ and $k_{\max}$, given a specified bin width Δk for each side. A detection of a strong configuration dependence could be regarded as compelling motivation to further investigate higher-order statistics, as this would allow us to disentangle the MG signal from other potential cosmological effects, which might be degenerate in two-point statistics and other alternative measures.

For our measurements we use an implementation of the bispectrum estimator presented in Ref. [209] with fourth-order density interpolation on two interlaced cubic grids [227] of $N = 380$ cells per side. Starting from $k_{\min} = 0.025 h/\text{Mpc}$, we loop through all configurations satisfying $k_1 \geq k_2 \geq k_3$ and $k_1 \leq k_2 + k_3$ (the triangle closure condition) with bin width $\Delta k = 4k_f \approx 0.025 h/\text{Mpc}$, where k_f denotes the fundamental mode. We correct each measurement for Poissonian shot noise, finding that for the galaxy catalogs studied here the redshift-space bispectrum becomes shot noise dominated for scales $k > 0.75 h/\text{Mpc}$, which is why we choose that value for $k_{\max}$. This procedure yields a total of 2825 distinct triangle configurations.

In the upper-left and upper-right panels of Fig. 20, we show the raw real-space measurements for the $f(R)$ and nDGP models, compared to GR. The x -axis of the plots reflects the ordering of the triangle configurations corresponding to when they appear in the loop, while vertical lines and axis labels indicate the increasing values of k_1 from the left- to the right-hand side. The lower panels show the difference between the modified gravity models and GR in terms of the variable defined in Eq. (41), i.e.,

$$\Delta B(k_1, k_2, k_3) \equiv \frac{B_X(k_1, k_2, k_3) - B_{\text{GR}}(k_1, k_2, k_3)}{\sqrt{\sigma_X^2(k_1, k_2, k_3) + \sigma_{\text{GR}}^2(k_1, k_2, k_3)}}, \quad (43)$$

where σ_X and σ_{GR} are the standard deviations obtained from the five HOD catalogs for either a MG model or GR. A general trend revealed by the plots is an enhancement of the bispectrum signal for the $f(R)$ models relative to GR, whereas the modified dynamics in nDGP lead to a suppression. These effects are growing towards smaller, more non-linear scales, and the deviations from GR are strongest for the F4 and N1 models, for which ΔB takes values of ~ 3 and -1.8 when averaged over all configurations in the interval $k_1 \in [0.3, 0.75] h/\text{Mpc}$ where the effect is most significant. However, even for the F6 model we get $\Delta B \sim 1.6$ in the same interval, which is a factor ~ 25 larger than the analogue quantity we would instead obtain for the power spectrum. The behaviour of the F5 model, on the other hand, is qualitatively different from F4 and F6, and the deviations from GR are noticeably smaller. We do not find the same trend when measuring the bispectrum from the corresponding halo catalogs, which implies that the difference between F5 and F4/F6 might be driven by the HOD modelling. It is interesting to note that the 3PCF analysis in Sec. III C 1 does not seem to display a similar behaviour (see, e.g., Figs. 17 and 19), but our findings here are consistent with the measurements of the reduced cumulants in Sec. III C 3 below. A more detailed future investigation is needed to check if this non-monotonic behavior is sensitive to the HOD model employed. Since the galaxy 2PCFs in the different models have been matched to each other by the HOD tuning, the strong difference in their bispectrum hints that the latter can offer independent information of and constraints on the MG dynamics, consistent with the findings of the 3PCF subsection above.

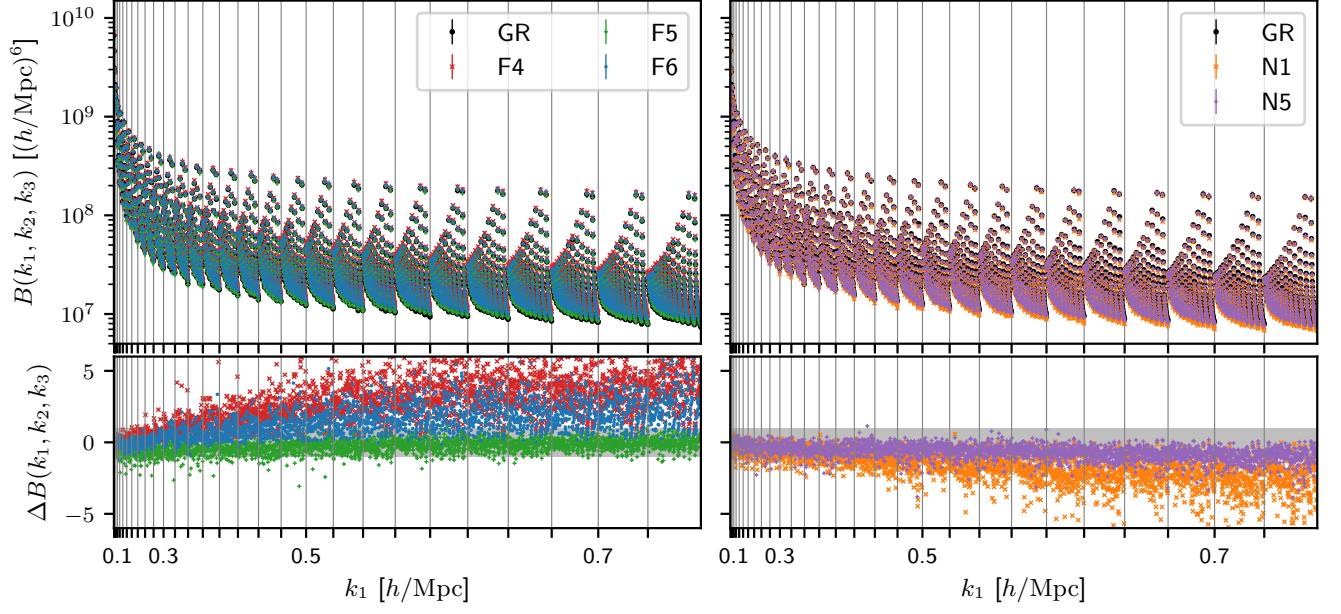


FIG. 20. Upper panels: real-space bispectrum measurements for $f(R)$ models (left) and nDGP models (right) compared to GR. Each data point corresponds to one of the 2825 triangle configurations, ordered according to when they appear in the measurement loop. Vertical lines are spaced by the bin width $\Delta k \approx 0.025 h/\text{Mpc}$ and indicate the value of k_1 , i.e., the largest triangle side. Lower panels: difference between modified gravity models and GR quantified in terms of the variable ΔB given in Eq. (43). The grey band marks the $1\text{-}\sigma$ area where deviations are compatible with the combined standard deviation of GR and a given modified gravity model.

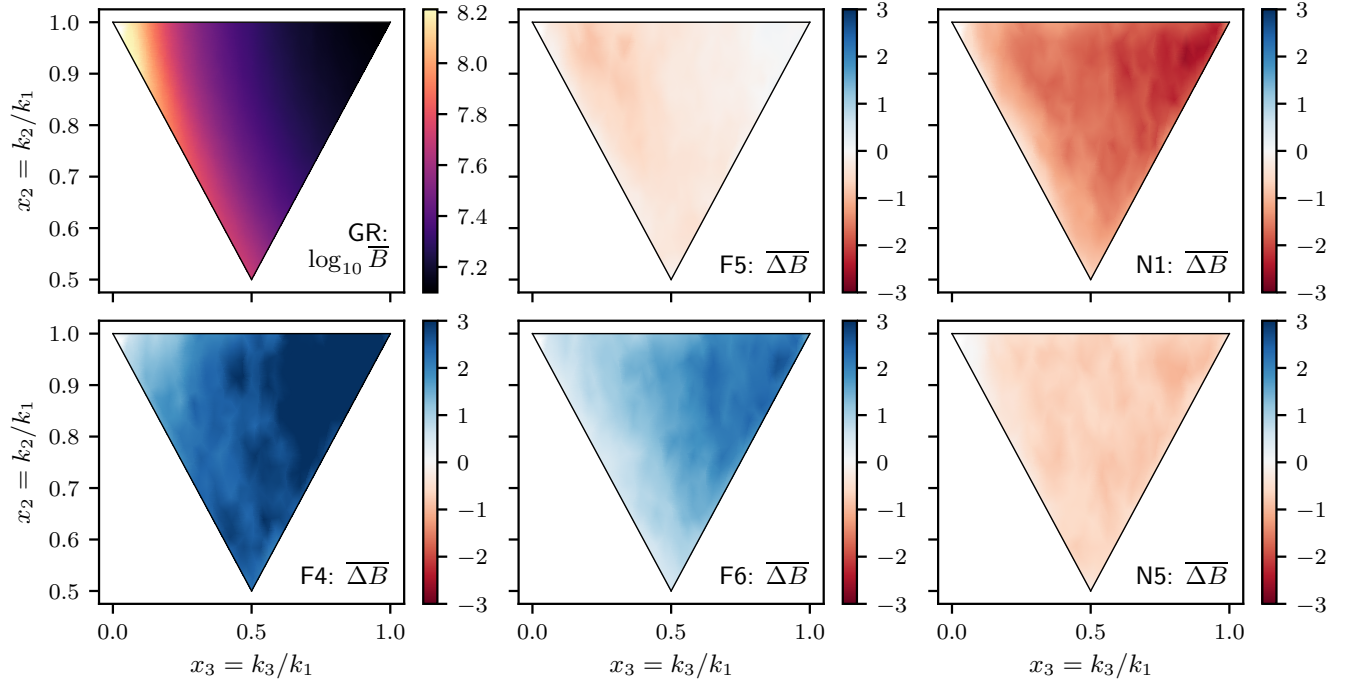


FIG. 21. Configuration dependence of the real-space bispectrum, obtained by integrating the raw measurements over k_1 (see Eq. 44) between $k_1 = 0.3 h/\text{Mpc}$ and $0.75 h/\text{Mpc}$. Configurations outside of the triangular area are forbidden according to the closure condition and the top left, top right and bottom corners correspond to squeezed, equilateral and folded configurations, respectively.

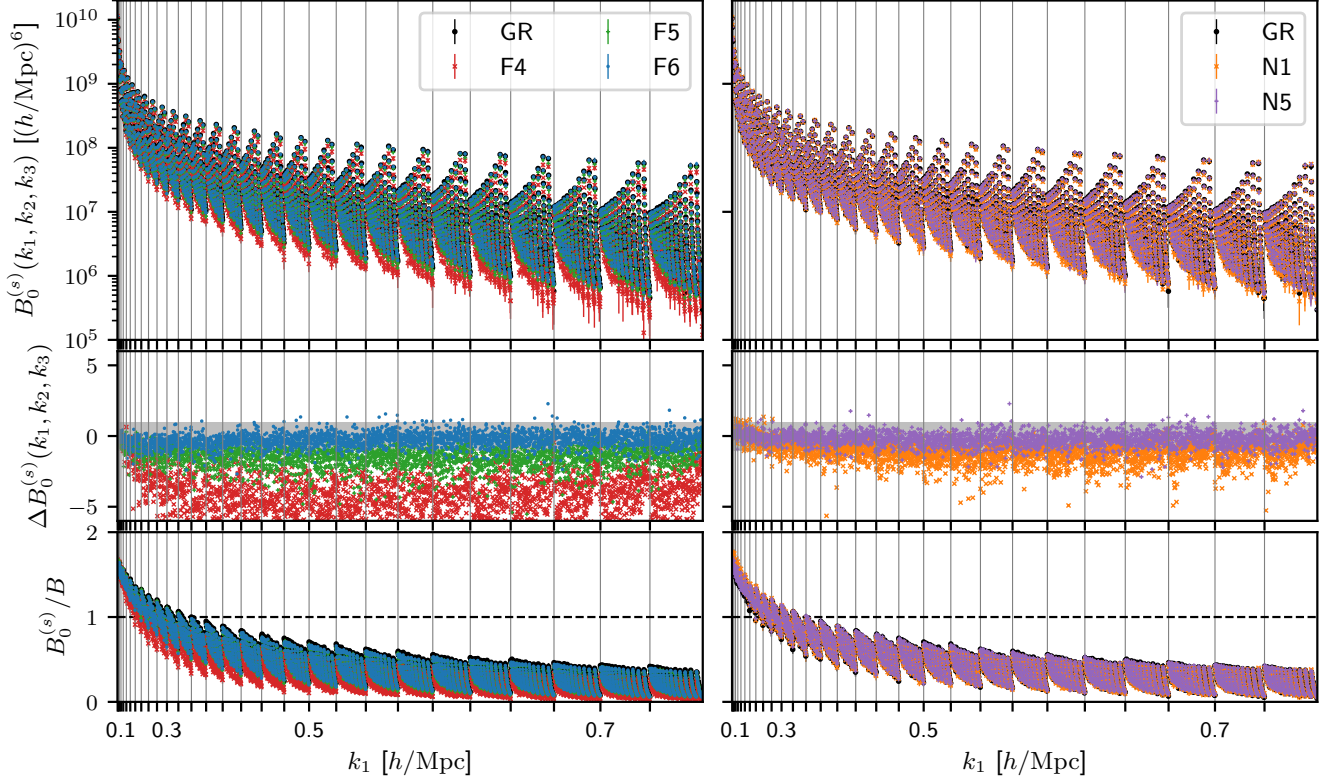


FIG. 22. Measurements of the bispectrum monopole in redshift-space. The panels in the upper two rows are the same as those in Fig. 20, while the bottom panels show the ratio between the redshift-space monopole and the real-space bispectrum.

In order to highlight the configuration dependence of the difference between modified gravity models and GR, we average the bispectrum (or ΔB) over the largest triangle side k_1 , while keeping the ratios $x_2 = k_2/k_1$ and $x_3 = k_3/k_1$ fixed:

$$\overline{B}(x_2, x_3) = \frac{1}{k_u - k_l} \int_{k_l}^{k_u} B(k, kx_2, kx_3) dk, \quad (44)$$

where we again use $k_u = 0.75 h/\text{Mpc}$ and $k_l = 0.3 h/\text{Mpc}$ for the upper and lower limits, respectively, which yields the results shown in Fig. 21 — note that in case of GR we plot $\overline{B}(x_2, x_3)$, while all other panels display the difference $\Delta \overline{B}(x_2, x_3)$. The overall amplitude and sign of $\Delta \overline{B}$ can readily identified from Fig. 20, but now we see more clearly that for the three models with the strongest deviations from GR, i.e., F4, N1 and F6, differences are maximized for configurations that are mostly equilateral (towards $x_2 = 1 = x_3$), whereas collinear configurations ($k_1 = k_2 + k_3$ or $x_2 + x_3 = 1$), which include squeezed and folded triangles, tend to be less affected. Again, F5 is qualitatively different from the other two $f(R)$ models with a preference for nearly squeezed configurations, and for N5 we do not identify any clear configuration dependence.

As discussed in Sec. III A, modified gravity not only impacts the clustering of galaxies, but also their infall and virial velocities, and as such alters the redshift-space distortions of clustering statistics. These distortions are present in any real measurement, so it is interesting to explore the difference between GR and MG models for the bispectrum in redshift space. For that we use the distant observer approximation, adopting the same LOS for all galaxies in the simulation volume, and focus on the bispectrum monopole, which averages over all orientations of the triangle with respect to the LOS, i.e.,

$$B_0^{(s)}(k_1, k_2, k_3) = \int_0^{2\pi} \frac{d\phi}{2\pi} \int_0^1 d\mu B^{(s)}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3), \quad (45)$$

where $\mu = \cos \theta$, and ϕ and θ are the angles describing the orientation.

In Fig. 22 we present the results of these measurements in a very similar fashion to Fig. 20. Compared to the real-space case, the behavior of $\Delta B_0^{(s)}$ for the $f(R)$ models has changed qualitatively: a suppression now takes the place of the previously enhanced bispectrum relative to GR with the amplitude of deviations being strongest for F4, followed by F5. They are weakest for the F6 model, which is mostly consistent with GR within the combined $1-\sigma$ scatter of the GR and F6 measurements. To

understand the origin of this change in behaviour, we plot the ratio of the redshift-space monopole and real-space bispectrum in the bottom panels of Fig. 22, which show an increased redshift-space signal on large scales that becomes heavily damped for triangles involving small-scale sides. This is simply a reflection of the two expected effects: on large scales the infall velocities of galaxies lead to a (configuration-space) squashing and thus enhanced clustering. On small scales, however, the Finger-of-God effect smears out the galaxy positions, which translates into a damping of the signal in Fourier space. In case of the $f(R)$ models the Finger-of-God effect is significantly stronger than in GR, particularly for F4, which points to an increased velocity dispersion in the $f(R)$ models. This agrees with the fact that the deviations of $B_0^{(s)}/B$ from GR show no appreciable dependence on triangle shape and that we find very similar deviations in the damping of the power spectrum. The increased Finger-of-God effect therefore reverses the real-space enhancement in clustering, which in case of F6 happens to have the right amplitude to render it consistent with GR (similar effects have been observed for the power spectrum, e.g., [228]). Further and more detailed studies with a larger number of simulations and different HOD prescriptions will need to show whether this is a physical signal or due to the HOD specifications. The ratio $B_0^{(s)}/B$ for the nDGP models, on the other hand, is mostly identical to GR on small scales, but we find a slight enhancement on large scales because of the increased growth rate, as already noted in Sec. III A 1. In total this means that the GR deviations of the bispectrum monopole tend to be similar to those observed in the real-space case, but with somewhat diminished significance because the reduced overall signal raises the relative importance of the shot noise.

In summary, our results indicate that the bispectrum can be a potentially powerful and complementary measure for discriminating theories of modified gravity. For the HOD galaxy catalogs used in this work we have found that the real-space clustering probed by the bispectrum can deviate significantly from GR on scales $k > 0.3 h/\text{Mpc}$ and these deviations tend to be at least a factor of a few larger than what an equivalent analysis implies for the power spectrum in all cases. In addition, we have seen that the difference between GR and modified gravity displays a clear dependence on triangle shape, which can further be exploited to break degeneracies with other cosmological and systematical effects. In redshift space, the small-scale bispectrum is heavily damped by the Finger-of-God effect, which leads to an increased impact of shot noise and thus less significant deviations from GR. This however is partly remedied for $f(R)$ models, which additionally differ from GR by having higher velocity dispersion.

Finally, we stress that all results shown here are based on uncertainties that correspond to the volume of a single simulation box, $\sim 1 (h/\text{Gpc})^3$. DESI will observe a much larger volume and, moreover, will have a higher number density of tracers, so that we expect all of our reported deviations from GR to grow. However, we also note that the analysis here is only based on a small number of realizations and a single HOD prescription. A future, more detailed, study is needed to assess the robustness of our results and to which degree they depend on the adopted HOD model. In addition, differentiating between GR and a given model of modified gravity using measurements such as those above will critically depend on our ability to make robust predictions in the non-linear regime. While recent progresses on the modelling of the bispectrum have been reported for chameleon and Vainshtein type models [199, 225, 229], further developments will be vital for a successful application.

3. Hierarchical clustering

The full information of all one-point correlations of cosmic density field (of matter or galaxies) is encoded in the shape and amplitude of the density probability distribution function, $pdf(\delta)$. If the density field is a Gaussian random field, then the pdf can be described simply by two numbers: the mean $\bar{\delta}$ and the variance σ^2 . However, as we have seen above, in the gravitational instability scenario, the growth of cosmic structures gives rise to significant deviations of the evolved matter distribution from the initial Gaussianity. An alternative estimator to accurately capture this information is the growth of higher-order central moments of the pdf [230–232].

If we sample N -point correlations in a big enough volume, so the *fair-sample* hypothesis is satisfied, then the central moments can be expressed as volume-averaged correlation functions [230]:

$$\bar{\xi}_n(R) \equiv \langle \delta_R^n \rangle_c = \int d^3x_1 \dots d^3x_n \xi(\mathbf{x}_1 \dots \mathbf{x}_n) W(x_1/R) \dots W(x_n/R). \quad (46)$$

Here $W()$ is the smoothing window (usually a top-hat or a Gaussian) and R is the smoothing scale. By $\langle \delta_R^n \rangle_c$ we denote here the n -th cumulants, which can be expressed in terms of central moments, $\langle \delta^n \rangle$, of the underlying density $pdf(\delta_R)$. The first few cumulants are given by

$$\begin{aligned} \langle \delta \rangle_c &= 0, \text{ (the mean)} \\ \langle \delta^2 \rangle_c &= \langle \delta^2 \rangle \equiv \sigma^2, \text{ (the variance)} \\ \langle \delta^3 \rangle_c &= \langle \delta^3 \rangle, \text{ (the skewness)} \\ \langle \delta^4 \rangle_c &= \langle \delta^4 \rangle - 3\langle \delta^2 \rangle_c^2, \text{ (the kurtosis)} \\ \langle \delta^5 \rangle_c &= \langle \delta^5 \rangle - 10\langle \delta^3 \rangle_c \langle \delta^2 \rangle_c \text{ (the hypskewness)}. \end{aligned} \quad (47)$$

In reality it is very hard to reliably estimate the full density pdf from a set of discrete field tracers (such as galaxies). However, it was established that the very first few cumulants already provide a robust characterization of the underlying density field and

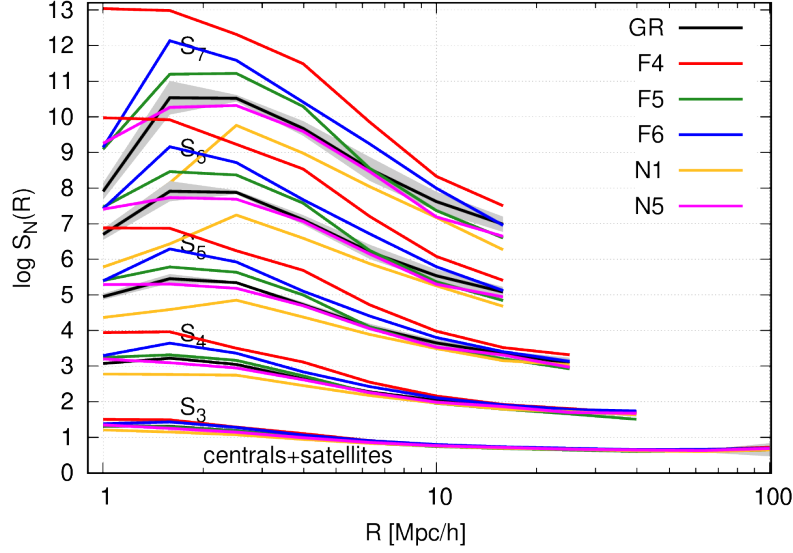


FIG. 23. The hierarchy of galaxy clustering reduced cumulants S_n 's for all models. The shaded regions represent 1σ scatters around the GR ensemble mean. The groups of lines from the bottom (S_3 or skewness) to the top (S_7) correspond to increasing cumulant order.

associated 1-point correlation statistics [232, 233]. This was extensively exploited for the benefit of cosmological analysis, since the central moments of the *pdf* can be readily extracted from observations. In this context, the particularly useful concept is the *reduced cumulants* or the so-called *hierarchical amplitudes* S_n 's, which are given by [234, 235]

$$\langle \delta^n \rangle_{(c)} = S_n \langle \delta^2 \rangle_{(c)}^{n-1} = S_n \sigma^{2n-2}. \quad (48)$$

Considerations from cosmic perturbation theory (PT) have shown that in the classical gravitational clustering scenario the higher-order cumulants are strong functions of the field variance, σ^2 . By rescaling the cumulants using the variance in a proper power to define the reduced cumulants, one can remove most of this dependence, and what is left is a statistic that is very sensitive to the higher-order and nonlinear effects of gravitational clustering. Considerations from perturbation theory [234, 236] have shown that for smoothed fields (such as count-in-cell for galaxy counts) the reduced cumulants become weak functions of the smoothing scale R and this effect is accurately captured by various combinations (set for a given cumulant order) of the logarithmic slope of the mass field variance, the so-called gamma-factors defined as

$$\gamma_n(R) \equiv \frac{d^n \log \sigma_M^2(R)}{d \log^n R}. \quad (49)$$

The above properties of the reduced cumulants in principle make them a very promising and suitable tool for testing the nature of gravitational instability [197, 237] or the nature of the initial conditions [238–240]. For the same reasons it was put forward that using S_n 's can be beneficial for testing GR and MG on cosmic scales [241–244].

In what follows we focus on the configuration-space clustering data for central+satellite HOD samples, as we have checked that for the central galaxies alone the MG effect is much weaker, typically not exceeding 10% departure from the GR case, a magnitude comparable or even smaller than the scatter around ensemble means.

We use the count-in-cell method to measure the central moments of the HOD galaxy distributions, employing the algorithm presented in [243, 244]. The count-in-cell method gives good results when the expected number counts $\langle N \rangle$ in a cell of a given size is large. Given the number density of our HOD catalogs, $n_g \sim 3.2 \times 10^{-4} (h^{-1} \text{Mpc})^{-3}$, this is guaranteed for a large enough smoothing scale, $R \gtrsim 8 h^{-1} \text{Mpc}$. At smaller scales we might expect to have a significant contamination by shot noise. To overcome this we performed the shot-noise correction for all our central moments. The method uses the moment-generating function of the Poisson model to calculate the net contribution by discrete noise (see details in [242, 245]). On the other hand, the higher-order moments are severely affected by finite volume effects [246]. The effect scales with the order n of the moment. For the simulation box size we used, the cosmic variance becomes significant at $R \sim 60 h^{-1} \text{Mpc}$ for the skewness (S_3), while for S_7 this scale drops to only $\sim 10 h^{-1} \text{Mpc}$. We expect that in a DESI-like survey both the shot noise and the cosmic variance will affect the measured cumulants to much less extent than can be noticed in our simulations.

In Fig. 23 we plot the full hierarchy of reduced cumulants from $n = 3$ (the skewness) up to seventh order (S_7). This gives us a general impression of the shapes and the amplitudes of these statistics. The solid black lines mark the fiducial GR case, while

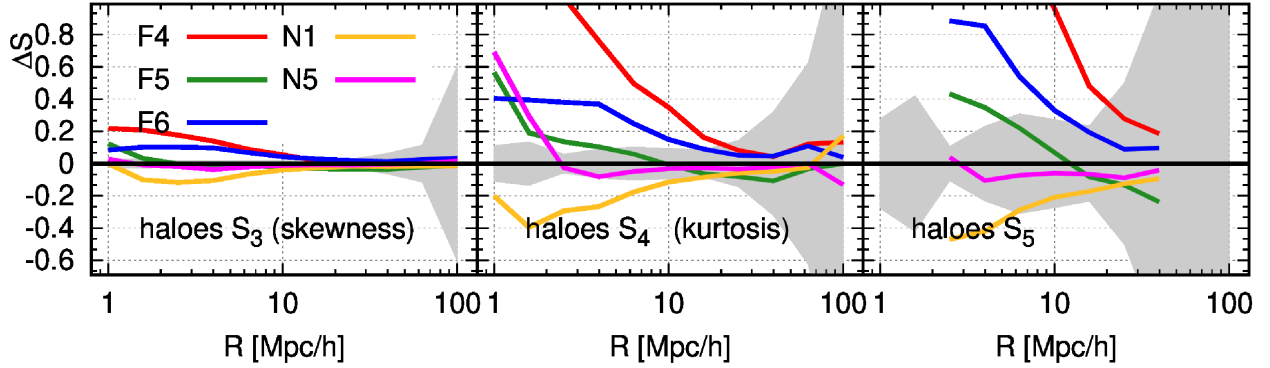


FIG. 24. The relative differences of the skewness (left panel), kurtosis (central panel) and S_5 (right panel) taken with respect to the fiducial GR case. The shaded regions mark the error round the ratio of ensemble averages.

associated shaded regions delimit 1σ dispersion around the mean from the ensemble of five GR realizations. The scatters for all the other models have an amplitude and scale dependence that is very close to that of the GR case, and for clarity we do not plot them. This figure already illustrates a couple of interesting points. First, we can observe that the higher-order moments (starting from $n = 5$) show a clear downturn of amplitudes at $R \lesssim 2.5 h^{-1}\text{Mpc}$. This is possibly caused by the limitations of the HOD scheme, which is not designed to accurately capture the galaxy clustering in the one-halo term regime [247–249]. We shall not attempt to study this in greater details, since this is beyond the scope of this work. Reassuringly, the scale and the shape of this effect seems to be very similar in all the inquired models with F4 being exception, where the effect seems to be much milder. Since this regime is strongly affected by both sampling noise and HOD accuracy, we shall stop at noting the exceptional behavior of F4 here.

Another observation is that the deviations from the GR grow with the increasing statistic's order n . Therefore, the higher-order cumulants would appear as better observables to discriminate between the models. Alas, the stronger MG signal comes at a price of greatly increased scatter, which is illustrated by much larger 1σ contours at small and large scales for higher-order cumulants. In practice, we have assessed that reduced cumulants from S_6 and higher are subjects to a scatter too large to be suitable for our analysis. Thus in our further considerations we will limit ourselves to only the first three reduced cumulants.

Figure 24 shows the relative difference with respect to GR, $\Delta S \equiv S_n/S_n^{\text{GR}} - 1$, for $n = 3, 4, 5$, as a function of the smoothing scale R . Again, the shaded regions delimit the 1σ scatters of ΔS . For nearly all scales, F4 appears to be the model deviating most strongly from GR, with the signal reaching $\sim 20\%$ for the skewness, $\sim 100\%$ for the kurtosis, and up to $\sim 200\%$ for S_5 at $R \lesssim 10 h^{-1}\text{Mpc}$. The next two standing-out models are N1 and F6: for all three S_n 's their deviations from GR show similar shape and amplitude, but differ by the signs N1 promotes lower values of amplitudes than GR, while F6, like the other $f(R)$ models, fosters higher values. The absolute effects of these two models are $\sim 10\%$ for the skewness, up to 40% for the kurtosis and even $\gtrsim 50\%$ for S_5 . The remaining two MG models (F5 and N5) exhibit considerably weaker deviations from GR, with N5 being mostly consistent with the latter.

The facts that F4 and N1 deviate most strongly from GR while N5 is close to it are as expected given that the screening effect is weaker in the former but stronger in the latter. The behavior of F6 and F5, on the other hand, is opposite to expectation, since the fifth force is more efficiently suppressed in F6. However, we note that, due to the additional complication of HOD modelling, the properties of the galaxy field may not follow exactly those of the matter field, making a physical interpretation more difficult: for example, Fig. 3 shows that the galaxies may be hosted by different halo populations in these two models. Also note that the exceptional behavior of F5 here is consistent with the findings of the real-space galaxy bispectra in the previous subsection.

To summarize, our results indicate that there is a great potential in using the reduced cumulants of galaxy distribution as strong discriminators of MG models. The relative differences we have found are quite large, from $\sim 20\%$ in the case of the skewness to $\gtrsim 100\%$ for S_5 . Indeed, while their error bars are much larger, it seems that both S_4 and S_5 offers a better prospect for testing MG and GR, since the amplitude of the signal is much higher. There are, nevertheless, uncertainties associated with our analysis which are of a statistical nature, due to the relatively low number densities and small volume of our mock galaxy catalogs. For a DESI-like survey we can expect a 20 times larger volume to be probed and also higher number density of tracers [13], which should reduce the statistical errors by a factor of at least 3 (see, e.g., [228] for more discussion). To match the specifications of such future data, larger-volume and higher-resolution simulations are needed to more accurately make the theoretical predictions for S_n .

In addition to the statistical uncertainties, like the other estimators, there are a range of theoretical and observational systematics which should be better understood. One example is the HOD method to construct mock galaxy catalogs. We have repeated the analysis using the mock galaxy catalogs where the HOD parameters were tuned to match the real-space 3D galaxy 2PCFs

amongst the different models, and the behaviors of ΔS_n are quite different, with much smaller deviations from GR in all models except N1. This suggests that galaxy modelling is an important uncertainty which should be addressed. A more accurate but also more expensive way to achieve this is to use hydrodynamical simulations which evolve the baryonic component as well. The latter can indeed be distributed in a different way than smooth dark matter at small scales [250]. Recently, however, there is a growing consensus that while nonlinear baryonic physics is crucial for understanding and modeling of galaxy internal properties, its impact on galaxy velocities and positions is very mild [251, 252]. This offers a way to test whether simplistic HOD modelling like the one used here would cause a substantial bias in the predicted S_n 's.

Another potentially important source of systematic effects lies in the fact that we have neglected the effects of redshift space distortions in our analysis. In reality, the measured line-of-sight coordinate of a galaxy is affected by its peculiar velocity, and some studies have found that in MG scenarios there exists a degeneracy between increased spatial clustering caused by the fifth force and enhanced clustering damping precipitated in redshift space by enhanced dynamics (see, e.g., [228]). The situation can be to a large extent remedied by the fact that both the nominator and denominator of the reduced cumulants ratios are affected by the RSD to a similar magnitude and the overall effect is largely canceled out [253]. This, together with the fact that the damping is limited to only small scales, suggests that at the intermediate scales of $10 \lesssim R/(h^{-1}\text{Mpc}) \lesssim 60$ the signal we have measured is not severely affected by RSD effects. Therefore, we hope that with detailed studies using future simulations we will be able to more accurately quantify these effects and extract genuine MG signals on those scales.

While the MG signal unveiled in the clustering amplitudes appears to be strong and significant, in reality, the observational data is a subject of various selection effects. To foster robust data analysis one needs to model specific survey's radial (redshift), angular and luminosity (magnitude) selection functions. Since in our analysis we did not attempt to model any of such effects, our results should be taken as an optimistic best-case scenario for an idealized perfect survey. On the other hand, we are dealing here with volume-weighted central moments estimated by count-in-cells. In contrast to pair-weighted statistics such as 2PCF, the volume-weighting makes the central moments less prone to biases induced by specific selections (especially radial and angular). As shown for example by [254], for the case of the VIPERS survey, careful modelling of the survey selection functions admits for a robust estimation of central moments and related hierarchical amplitudes. Thus, we expect that for a survey with a footprint like DESI, the observational selection effects, when modelled accurately, should not hinder the potential for obtaining strong MG constraints using the reduced cumulants of galaxy clustering.

4. The Minkowski functionals of the density field

As discussed in Sections III C 1, III C 2 and III C 3, because the observed density field is not perfectly Gaussian due to the nonlinear evolution of the cosmic structure and possibly also to primordial non-Gaussianity, one cannot extract all the information from galaxy surveys that may be relevant for cosmological analysis using the commonly used two-point correlation function or power spectrum. It is true that higher-order correlation functions, or the corresponding high-order power spectra (referred to as N-point statistics hereafter), can be complementary, given the the computational challenges to measure those quantities from observations and theoretical difficulty to model them accurately, it is worthwhile to consider other probes that encode the same information. An example of such probes is cosmic voids (see Sections III A 3 and III D 2), the statistics of which can be related to the hierarchy of N-point correlation functions (see, e.g., [255, 256]). Other alternatives include topological and morphological measures of the density field; in this subsection we will briefly describe the Minkowski functionals (MFs, [257]) as a potential test of gravity.

Compared with the standard N-point statistics, the MFs are advantageous in several aspects. First, According to Hadwiger's theorem, the morphological properties of a three-dimensional structure are completely specified by four MFs, namely the volume, the surface area, the integrated mean curvature and the genus. Second, MFs are independent of galaxy bias, which makes them an ideal tool for gravity tests. This is because galaxy bias can have a complicated scale dependence in general MG theories, which makes it challenging to build models for N-point statistics for MG theories. Last but not least, it is computationally much cheaper to measure than the N-point correlation functions from observational or mock galaxy catalogs. MFs have been applied to analyze density fields in galaxy surveys (e.g., [258–260]) and maps from CMB or weak lensing experiments (e.g., [261–266]).

A proof-of-concept study of using MFs for gravity tests was done in [267], using the dark matter fields from the simulations described in Section II B, and the results are briefly summerized here. After smoothing the dark matter density field to suppress the shot noise, the MFs V_0, V_1, V_2 and V_3 were measured as a function of the density threshold, shown as $\rho/\bar{\rho}$. Although the specific analytical form of the MFs is not needed here and can be found in, e.g., [268], their physical characterization is informative: V_0 is the volume fraction of the excursion set (structure pattern); V_1 represents the area of the surface of the excursion; V_2 is its integrated mean curvature; and V_3 is its Euler characteristic per unit volume or the genus number describing the connectedness of the isodensity contours, see, e.g., [269–271]. In the left panel of Fig. 25 the four MFs of GR (black) and four MG models – F6 (red), F5 (blue), N5 (purple) and N1 (green) – are shown, while the right panels show the differences with respect to GR.

To get a sense on why MFs are useful for testing gravity, let us consider the results in more details. Taking ΔV_0 (the difference in volume) to begin with, we can see that the volume fractions of the excursion set in MG are generally larger than that in GR,

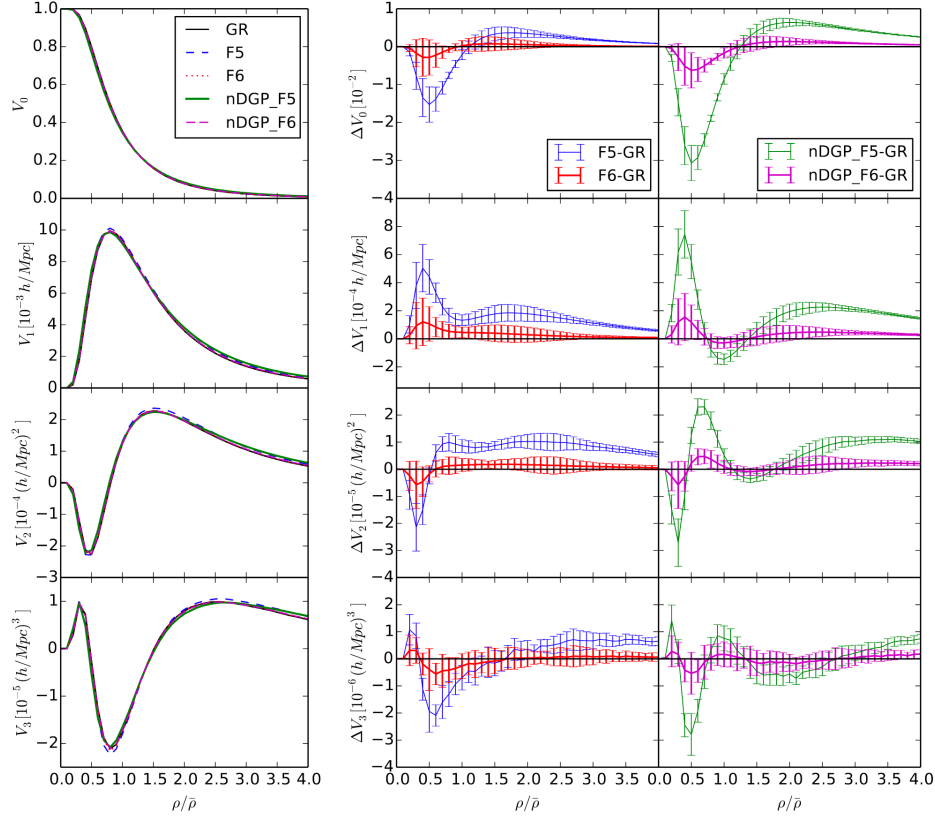


FIG. 25. Left panel: The MFs computed from N-body simulations for different models of gravity at $z = 0$, with the dark matter density field smoothed with a spherical Gaussian filter with size $R_G = 5 h^{-1} \text{Mpc}$. The two $f(R)$ models – F6 (red dotted lines), F5 (blue short dashed) – and two nDGP models – N5 (purple dashed), N1 (green solid) – are compared with the GR case (black solid) for $V_0 - V_3$ (from top to bottom). Right: The differences in the MFs between the four modified gravity models and GR, first column for $f(R)$ gravity, and second column for nDGP. The quantity $\rho/\bar{\rho}$ is the density threshold used for the MF calculations in ratio of the mean density. This plot is reproduced from [267] and so the colour scheme is different from the one adopted for most other plots in this paper. As discussed in the text, while some features of the MFs are similar for GR and MG models, their amplitude is more pronounced and can be used to distinguish between the models.

for densities above a sufficiently high threshold (i.e., $\rho > \bar{\rho}$). However, it is the opposite for under-dense regions, namely, the volume fraction above an under-dense threshold get smaller in MG, which is equivalent to having larger volume fraction below an under-dense threshold. This is as naturally expected as the halos and voids are more abundant and with larger sizes in $f(R)$ or nDGP than those in GR. Furthermore, for ΔV_1 (the surface area), the overall trend is similar to that of ΔV_0 , except for the under-dense regions ($\rho < \bar{\rho}$). This can be understood as follows: if the excursion sets are all isolated regions, as is the case for high density threshold, it is expected that the change in their surface area follows that in the volume fraction they occupy. However, regions enclosed by the low iso-density contours are no longer the excursion sets, but under-dense regions with density below the threshold, with the volume fraction $1 - V_0$. Thus, at the low density threshold region, V_1 changes in the opposite direction as V_0 , and becomes larger in both $f(R)$ gravity and nDGP models. Moreover, it can be seen from Fig. 25 that V_3 can help distinguish between GR and MG models as follows. In GR, the isodensity contours are more connected with $V_3 < 0$ for the region $0.5 \lesssim \rho/\bar{\rho} \lesssim 1.5$, but more disconnected with $V_3 > 0$ in the other regions. For the $f(R)$ and the nDGP models, this feature is overall more pronounced and can thus serve to distinguish them from GR.

To summarize, the MFs $V_0 - V_3$ capture information that is not available in the simple two-point statistics, and therefore can be useful for testing gravity and other cosmological models. However, the study in [267] was based on the dark matter, rather than galaxy, fields, which may have enhanced the model differences compared with the latter. The HOD tuning to match the (projected) galaxy 2PCFs, as used in most other probes of this paper, may further reduce the signal. Therefore, it will be useful to conduct a more detailed study using the MFs measured from realistic mock galaxy catalogs with a higher galaxy density and larger volume, allowing systematic effects from observations to be included, to fully assess the potential of this probe.

5. Stacked cluster phase spaces

In the weak-field approximation of general relativity and at sub-horizon scales, a massive particle will still feel a force from the accelerated expansion of space [272]. The effective acceleration experienced by a massive particle with zero angular momentum in the vicinity of a galaxy cluster with gravitational potential (Ψ) is given by,

$$\vec{\nabla}\Phi = \vec{\nabla}\Psi + qH^2 r \hat{r}. \quad (50)$$

The effective potential Φ therefore takes into account both the effect produced by a *matter only* density field with potential Ψ and the effect produced by the acceleration term $qH^2 r$, with q being the deceleration parameter $q \equiv -a\ddot{a}/\dot{a}^2$. From a Newtonian perspective, the latter term can be thought of as a repulsive force that opposes the inward pull of the cluster's mass distribution, caused by the accelerated expansion of space.

Given that the acceleration on a point mass is governed by the gradient of the gravitational potential, we define an equivalence radius as the point where the acceleration due to the cluster's gravity and the acceleration from the expanding space are equal to each other ($\vec{\nabla}\Phi = 0$), which yields,

$$r_{\text{eq}} = \left(\frac{GM}{-qH^2} \right)^{1/3}, \quad (51)$$

where G is the gravitational constant and M is the mass of the cluster. The escape velocity profile inferred from the observed phase space data can be modeled with a function of the mass distribution of a specific cluster, as specified by its gravitational potential, in our case we use the Einasto profile with its standard three free parameters: $\alpha, r_{-2}, \rho_{-2}$ [273]. In the concordance Λ CDM cosmology, the escape profile is also a function of redshift z and cosmological parameters, Ω_m, h , etc.. Therefore in standard Λ CDM, the escape velocity radial profile is given by a function of the cosmology and cluster parameters combined,

$$v_{\text{esc}} = v_{\text{esc}}(r, z, \Omega_m, h, \dots, \alpha, r_{-2}, \rho_{-2}) \quad (52)$$

where we include cosmological parameters like Ω_m as well as cluster observables like α, r_{-2} and ρ_{-2} .

In terms of escape speeds, we now recognize that the gravitational potential at r_{eq} plays an important role and that the distance to escape a cluster is well-defined and finite in an accelerating Universe. In other words, we set the boundary condition such that the radial component of the escape velocity with respect to the cluster is zero at the equivalence radius, $-2\Phi(r_{\text{eq}}) = v_{\text{esc}}^2(r_{\text{eq}}) = 0$. Using the Newtonian analogy that $v_{\text{esc}}^2 = -2\Phi$, in Ref. [274] it was noted that

$$v_{\text{esc}}(r, z) = \sqrt{-2(\Psi(r) - \Psi(r_{\text{eq}})) - qH^2(r^2 - r_{\text{eq}}^2)}. \quad (53)$$

Eq. (53) yields a radial escape speed of 0, relative to the cluster center, at the equivalence radius given the gravitational potential profile Ψ . Note also that this equation applies to an accelerating universe for any choice of gravitational potential Ψ [275]. For instance, modified gravity theories where the Poisson equation holds still satisfy Eq. (53) [177].

The observed escape velocity is readily identified by an edge in the radius-velocity phase-space data. This edge is typically suppressed statistically from under-sampling of the phase-space. Ref. [276] showed that this suppression (Z_v) is dependent only on the number of galaxies which sample the phase space, such that when $\mathcal{O}(10^5)$ tracers are available, the statistical suppression term $Z_v \approx 1$ and a practically exact tracing of the 3D escape velocity profile can be made. In practice and even in stacked cluster phase-spaces, we typically have a few hundred (in well-sampled single clusters) to a few thousand (stacked) available tracers, and so $1.1 \lesssim Z_v \lesssim 1.5$. Thus, we measure the line-of-sight (LOS) escape velocity and correct for this statistical suppression to infer the 3D escape velocity: $v_{\text{esc}} = Z_v v_{\text{esc,LOS}}$.

Let us consider the $f(R)$ model as introduced in Section II. The gravitational potential which massive particles experience is given by [277],

$$\Phi_{f(R)}(r) = \Phi_{\text{GR}}(r) - \frac{1}{2}\delta f_R(r), \quad (54)$$

where the δ signifies that the background has been subtracted from the scalar field: $\delta f_R = f_R - \bar{f}_R$. In an expanding chameleon $f(R)$ gravity Universe, instead of Eq. (50), we have to modify the effective potential and its relation to the escape velocity as,

$$-2\Phi_{f(R)}(r) = v_{\text{esc},f(R)}^2(r) = v_{\text{esc,GR}}^2(r) + [\delta f_R(r) - \delta f_R(r_{\text{eq}})]. \quad (55)$$

In practice, the gravitational potential in Eqs. (53) and (55) is constrained by weak lensing shear profiles around galaxy clusters. The LOS escape velocity surface can be observed using highly multiplexed spectroscopic instruments like DESI.

We expect the BGS data to provide spectroscopic redshifts for ~ 25 member galaxies per cluster in a typical $\sim 10^{15} M_\odot$ cluster, a value much lower than the actual cluster richnesses at this mass. This is because DESI suffers from significant fiber

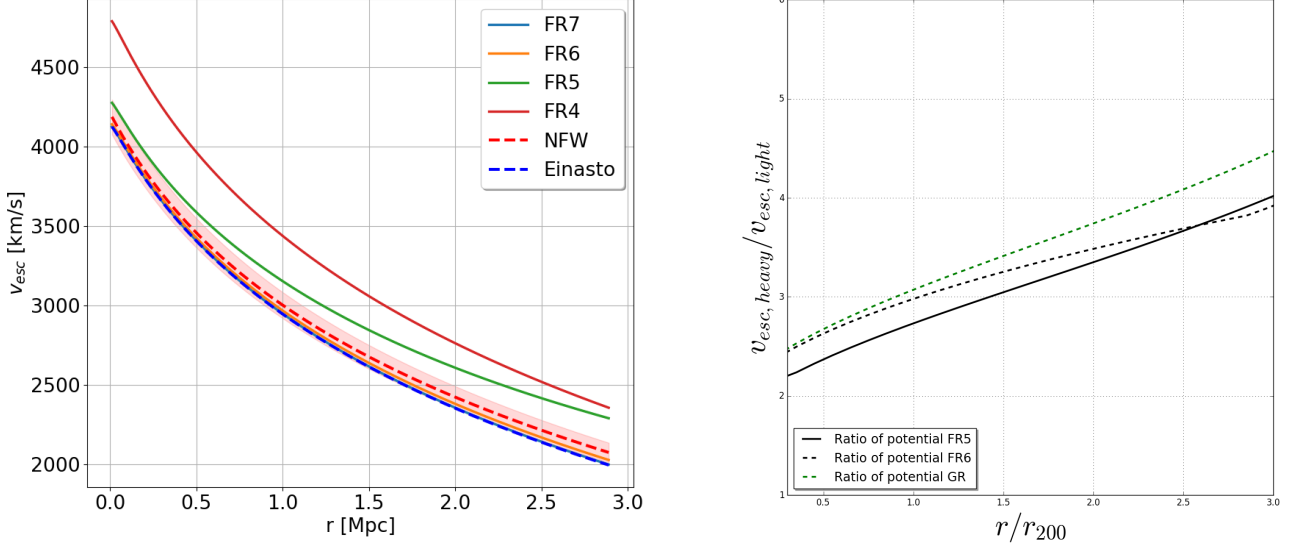


FIG. 26. (Color Online) *Left*: The radial escape velocity profile for a stacked high mass ($10^{15} M_{\odot}$) cluster with ~ 5000 member galaxies. The shaded region includes nominal 5% error from the weak-lensing inference of the gravitational potential. *Right*: The same as in the left panel, but shows the ratio between the high mass and low mass clusters escape velocities. This probe utilizes the ratio for two reasons: (1) we expect the inner ~ 0.5 Mpc of high mass clusters to look like GR, which sets a lever-arm for the detection of $f(R)$ signals in lower mass systems; (2) this removes a primary systematic, the under-sampling which suppresses the true escape edge. Note that the color schemes are different in the two panels, as indicated in the legends.

collisions in dense regions like the cores of galaxy clusters. However, due to the area and depth of the survey, we can expect ~ 100 s of massive galaxy clusters to lie within the survey footprint. Therefore, DESI will provide useful line-of-sight velocities for a few thousand galaxies in a stacked cluster phase-space that has an average cluster mass of $\sim 10^{15} M_{\odot}$. The cluster masses themselves need to be measured via an independent weak lensing analysis. Note that unlike most of the other probes discussed in this paper which require some knowledge of the physical distances of the tracers, the escape velocity analysis is conducted purely in proper units, such as km/s and angular separations [278].

In the left panel of Fig. 26, we show the escape velocity for a stack of high mass ($10^{15} M_{\odot}$) clusters for different variants of HS $f(R)$ gravity including F4, F5, F6 and as well as F7 ($|f_{R0}| = 10^{-7}$); note that the line colors are different from other plots of this paper. The dashed lines representing Einasto and NFW profiles are for GR and highlight the difference one would expect from an incorrect choice of mass profile when inferring the gravitational potential from the weak lensing data. The shaded region includes error from weak lensing (5%). For F4 and F5, we find significant differences from GR in both the amplitude and the shape at all radii. For F6, it becomes difficult to differentiate against GR using a single stacked phase-space, and for F7 the result is practically indistinguishable from GR.

The right panel of Fig. 26 shows the ratio between the escape edges of high- and low-mass clusters. Taking the ratio enables us to divide out the effect of the statistical suppression, on the assumption that the phase-space sampling is similar between them. In this case, we see that the difference between F6 and GR becomes appreciable at large radii. This is a result of the chameleon screening mechanism, which means that MG deepens the potential in the outskirts of low-mass galaxy clusters with respect to GR, but leaves the potential of high-mass clusters relatively unaffected.

Therefore, this offers two ways to constrain gravity: using the radial escape velocity profile based on $v_{\text{esc}} = Z_v v_{\text{esc,LOS}}$, cf. left panel of Fig. 26, or using the ratio between stacked high- and low-mass clusters, cf. right panel of Fig. 26. We have found that the first approach, assuming a stack of massive ($\langle 10^{15} M_{\odot} \rangle$) clusters having ~ 5000 tracers leads to a constraint of $|f_{R0}| \lesssim 5 \times 10^{-6}$ when we take into account the uncertainty in Z_v and the weak lensing mass errors. However, the latter approach which utilizes the ratio of the escape edge for a high and low mass stack produces superior constraints.

Since the galaxy number in the mock HOD galaxy catalogs described in Section II B 3 is too low, we use dark matter particles from the same simulations as described in Section II B to test the ratio of the high- and low-mass cluster potential ratios. Ref. [275] showed that galaxies are unbiased tracers of the dynamical escape velocity edge when compared to particles in a Λ CDM dark matter only simulation. Since our MG simulations are dark matter only, we assume galaxies in a MG universe will also be unbiased and that our constraints based on the particles will match expectations for real galaxies. We choose halos in the simulation such that the average masses in the high- and low-mass bins are nearly the same. For GR, the low mass bin is

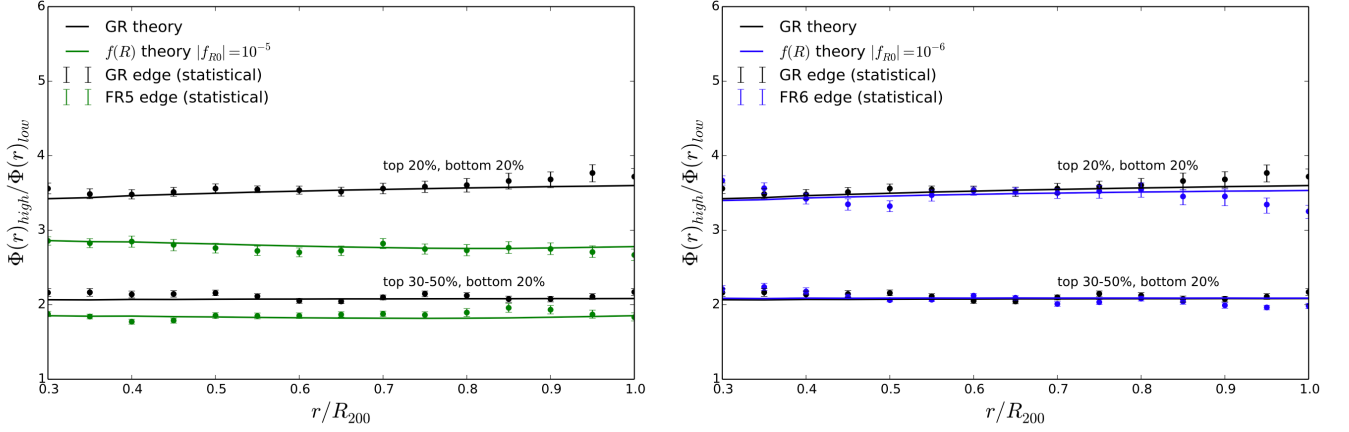


FIG. 27. The $z = 0$ gravitational potential ratio between high and low mass bins (of 20 halos) for GR (black), and the F5 (left) (green) and F6 (right) parametrization of modified gravity (blue). The points are the averages of the square of the measured radial escape velocities using the simulation particles for each bin in radius and mass. The errors are 1σ on the mean from boot-strap re-sampling. The solid lines represent the theoretical predictions using the NFW density parameter. The separation between GR and $f(R)$ potential ratios increases with increasing separation in the mass bins and we show two different mass binning schemes in each panel where the percentages denote the percentiles of cluster masses we keep in the sample. Note the large difference between the GR and F5 ratios at either mass binning scheme. Note the precise and accurate agreement between the theory and the measured escape profile (or Φ) profile ratios. These figures justify our use of only theory in making predictions without the need for additional simulations at differing levels of f_{R0} .

$9.10 \times 10^{13} \sim 1.96 \times 10^{14} h^{-1} M_{\odot}$ and the high mass bin is $7.48 \times 10^{14} - 1.58 \times 10^{15} h^{-1} M_{\odot}$. For F6, the low-mass bin is $9.13 \times 10^{13} - 1.97 \times 10^{14} h^{-1} M_{\odot}$ and the high-mass bin is $7.34 \times 10^{14} - 1.58 \times 10^{15} h^{-1} M_{\odot}$. For F5, the low-mass bin is $1.16 \times 10^{14} - 1.94 \times 10^{14} h^{-1} M_{\odot}$ and the high-mass bin is $7.49 \times 10^{14} - 1.58 \times 10^{15} h^{-1} M_{\odot}$. We then measure the radial escape velocity for each cluster in these bins and take the average, and calculate the error on the theory through bootstrap re-sampling of the clusters in each bin. As shown in Fig. 27, we find excellent agreement with theory, and this allows us to use the analytical prescriptions for making predictions for $f(R)$ gravity variants other than F5 and F6. We also note that there are detectable differences in the shape between GR and F6 in the outskirts, in agreement with our analytical result in Fig. 26.

We then utilize the probe

$$\frac{\Phi_{\text{high,LOS}}(r)}{\Phi_{\text{low,LOS}}(r)} \equiv \frac{\langle Z_v(\text{low}) \rangle}{\langle Z_v(\text{high}) \rangle} \frac{\langle v_{\text{esc,high}}^2(r) \rangle}{\langle v_{\text{esc,low}}^2(r) \rangle} \approx \frac{\Phi_{\text{high}}(r)}{\Phi_{\text{low}}(r)}, \quad (56)$$

which was tested extensively on lightcone data from Ref. [279]. A forecast analysis was carried out in Ref. [177], where it was found that using this probe DESI BGS data can place a constraint $|f_{R0}| < 4 \times 10^{-6}$ at 5σ .

In reality, we expect to be able to make multiple independent measures of the escape velocity ratio, each with different average masses in their high- and low-mass bins. We will have enough clusters to choose from to ensure that there are no clusters residing in more than one binned measurement of the escape ratio. This additional power to differentiate between $f(R)$ and GR models lies in both more clusters and more mass bins. We have checked that the escape ratio linearly depends on the mass ratio of the high- and low-mass bins. Using the large sample from the GR simulations described in Section II B, we are able to create up to four more high- and low-mass bins, each with a different mass ratio. In practice, we treat each of these as independent measures as a function of radius, thus increasing our statistical sample by a factor of 5. In doing so, we are able to differentiate between GR and $f_{R0} = 6 \times 10^{-7}$ at $> 5\sigma$.

The observables for this test involve the galaxy redshifts and radial positions for the phase-spaces and the weak-lensing shear profiles. We assume that the systematics from these measurements are well-controlled and minimized in the data. For the phase-spaces, this should not be an issue, since the galaxy sky positions and redshifts should be determined at percent level accuracy compared to the escape surface. Recent advances in the methods to constrain weak lensing mass profiles claim control on the systematics at $< 10\%$ [280]. Nominal cluster-based systematic issues like the selection function and mis-centering are not issues for this probe, simply because we require high signal-to-noise measurements of the phase-space density and the weak-lensing signal for individual clusters.

It should be noted that while the ratio test exploits the unique properties of the $f(R)$ model (e.g. the chameleon screening mechanism), the test of the radial escape velocity profile does not require any specific characteristics from the gravity models as it places constraints based on the difference in gravitational potential alone. That allows to utilize galaxy cluster's phase spaces to test other MG and alternative gravity models such as emergent gravity [281] in addition to constraints based on mass profiles

of galaxy clusters [282–284].

In the end, the primary observational systematic when using high signal-to-noise clusters and parameterized mass profiles is the weak lensing shear. It is likely that the DESI Bright Galaxy Survey will provide the required amount of phase-space sampling for over 1000 clusters. Many of these will have (albeit noisy) weak lensing profiles from the deep ground-based imaging from DES and DECaLS (as well as from the Hyper-SuprimeCam Survey).

The story is different when we stack phase-space data and weak-lensing shears. Stacking is the presumed scenario for this probe. DESI does a poor job of spatial sampling within high density regions. A concern is that weak-lensing profiles are hard to measure for clusters $\lesssim 10^{14}M_\odot$, where the $f(R)$ signal becomes interesting. However, in order to make further in-roads into how well this probe will work for realistic DESI stacks, we require a new generation of mock galaxy catalogs in light-cones with estimated shear profiles and high enough galaxy sampling to measure phase-spaces with many 10s of galaxies per cluster.

D. Gravitational lensing statistics

While the paper principally focuses on utilizing density and velocity statistics coming out of spectroscopic and photometric galaxy surveys, in this section we discuss the potential to use complementary lensing information to test the properties of gravity directly through the evolution of the gravitational potential. For a spectroscopic survey that has substantial sky coverage overlap with imaging surveys, lensing can be a useful probe for testing gravity. In this section we discuss two distinct environments that can be considered. In III D 1 we consider galaxy-galaxy lensing, while in III D 2 we discuss the potential of lensing measurement in low density, void environments.

1. Galaxy-galaxy lensing

Galaxy-galaxy lensing (GGL) describes the distortions of images of background (source) galaxies around foreground (lensing) galaxies, and detects the matter distribution around the latter up to radii which typically go well beyond the dark matter halos of the lensing galaxies. It is an ideal probe to study properties of dark matter halos such as the mass profiles [285, 286], galaxy-matter cross correlation and galaxy bias [287, 288], and to constrain cosmological parameters [289]. It has been applied in multiple lensing surveys, such as CFHTLENS [290–292], KIDS [293–295] and DES [296].

In regions well outside foreground galaxies, the screening mechanisms are expected to work less efficiently. Such regions can experience substantially stronger gravitational force, and consequently enhanced clustering of matter, in the MG models studied here, making GGL a potentially useful probe to test them. However, as in void lensing (see §III D 2), because individual galaxies generally do not produce strong enough lensing effect, one has to stack the tangential shear around many foreground galaxies to detect signals at high signal-to-noise. GGL in the context of MG models has been studied previously in, e.g., [297–299].

The tangential shear around a foreground galaxy is related to the average excess surface density profile, $\Delta\Sigma(r_p)$ [300, 301], given by the following integration of the galaxy-matter cross correlation function, $\xi_{gm}(r)$:

$$\Delta\Sigma(r_p) = \rho_{\text{crit}}\Omega_m \frac{2}{r_p^2} \int_{-\infty}^{\infty} d\chi \int_0^{r_p} dr \cdot r \xi_{gm}(\sqrt{r^2 + \chi^2}) - \rho_{\text{crit}}\Omega_m \int_{-\infty}^{\infty} d\chi \xi_{gm}(r_p, \chi), \quad (57)$$

where r and r_p are respectively the 3D distance and projected distance from the lensing galaxy, and χ is the comoving distance from it. We have measured ξ_{gm} using a modified version of the publicly-available code CUTE [87], which counts pairs of HOD galaxies and simulation particles. To carry out the integrations in Eq. (57), we first interpolate $\xi_{gm}(r = \sqrt{r_p^2 + \chi^2})$ onto a grid in $(\log(r_p), \chi)$ using cubic spline, and then do discrete summations of the integrands evaluated at the grid points. We use $\pm\chi_{\text{max}}$, with $\chi_{\text{max}} = 90 h^{-1}\text{Mpc}$, as the integral limit and have checked that this choice leads to converged results.

The results of $\Delta\Sigma$ and the relative differences between modified gravity models and GR are shown in Fig. 28, considering only lenses at $z = 0.5$ for simplicity. The MG curves display the expected trends with respect to GR – within the $f(R)$ family, F4 shows the largest deviation and F6 the smallest, and within the nDGP family, N1 shows stronger difference than N5; these are in the same order as decreasing screening efficiency and increasing strength of the fifth force. The differences between $f(R)$ gravity and GR decrease on larger scales, while for nDGP the deviation from GR remains scale-independent at r_p greater than a few $h^{-1}\text{Mpc}$. This behavior was expected since $f(R)$ gravity has scale-dependent linear growth, with the fifth force suppressed outside the Compton wavelength of the scalar field, while in DGP the linear growth rate is scale-independent. On the other hand, on small scales ($r_p \lesssim 1 h^{-1}\text{Mpc}$) the nDGP predictions become close to the GR ones, which is a result of the efficient Vainshtein screening near and inside dark matter halos.

We have checked the results at $z = 0$ and found similar qualitative and quantitative conclusions as $z = 0.5$, with the maximum deviation from GR at the level of $\sim 28\%$, $\sim 12\%$ and $\sim 9\%$ for F4, F5 and N1 respectively (the F6 and N5 models are, on the other hand, very close to GR). The model differences also have a weak dependence on the galaxy number density [299]. The GGL results are not sensitive to the way in which the HOD catalogs are produced for the MG models; for example, we did a test

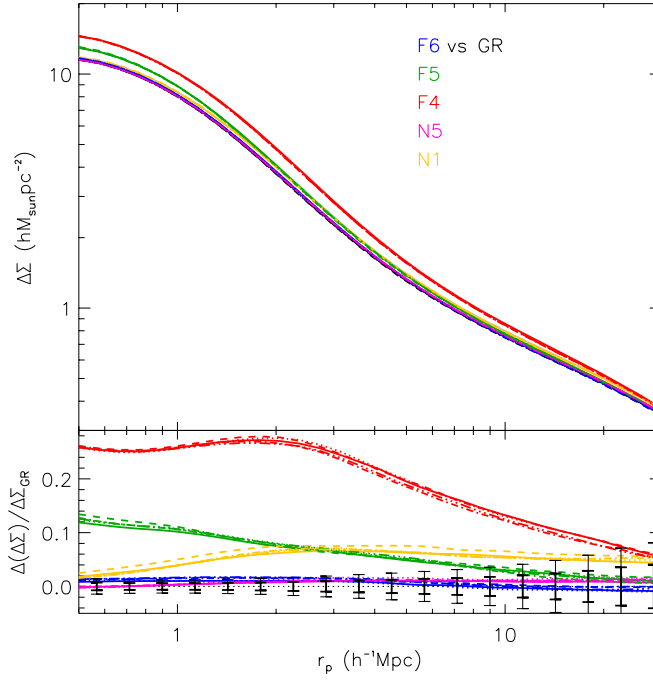


FIG. 28. (Color Online) The excess surface mass density profiles, $\Delta\Sigma$, measured at $z = 0.5$, for the six models studied here: GR (black), F6 (blue), F5 (green), F4 (red), N5 (magenta) and N1 (orange). The lower sub-panel shows the relative differences with respect to GR, where the black horizontal dotted line is zero, and the thick and thin error bars indicate respectively the statistic uncertainties in an optimistic and a pessimistic case, in which an LSST-like imaging survey has 6,000 and 1,500 sq. deg. overlap with a DESI-like survey. We have used different lines styles to represent the five realisations, instead of showing their mean, to highlight the small scatter across the different realizations.

by tuning the respective HOD parameters so that these models match the real-space 3D galaxy correlation function of GR, and found almost identical conclusions as in Fig. 28. Finally, note that in Fig. 28, instead of the average results from all 5 simulation realizations, we have shown the curve for each of them using different line styles, to see the scatter between them is very small even for the stronger MG models (F4 and N1); we also did a test by tuning the HOD parameters individually for each simulation realization instead of doing this for all 5 boxes together, and while the HOD parameters are now slightly different, the realization scatter was again negligible. This makes sample variance a lesser concern for testing models using GGL, especially for surveys like DESI which will give higher galaxy number densities in a larger volume than used in the analyses here.

To forecast the constraining power of GGL, we have calculated the signal-to-noise (S/N) of the distinguishability of the MG models from GR, which is defined as

$$(S/N)^2 \equiv \delta\Delta\Sigma^T(r_{p,i})\mathbb{C}^{-1}(r_{p,i}, r_{p,j})\delta\Delta\Sigma(r_{p,j}), \quad (58)$$

in which $\delta\Delta\Sigma(r_{p,i})$ is the model difference of the excess surface mass density in the i th r_p bin, and $\mathbb{C}(r_{p,i}, r_{p,j})$ the covariance matrix between the i -th and j -th r_p bins. The covariance matrix is calculated following the analytical prescription of [302], based on halo model predictions of the shear-shear, galaxy-galaxy and shear-galaxy correlation functions (for which we have used the same cosmological and HOD parameters as in the simulation and galaxy catalogs). The calculation takes into account contributions from cosmic variance, the Poisson noise of lens galaxies and source shape noise σ_γ , and assumes a single source redshift $z_S = 1.0$ and lens redshift $z_L = 0.5$. The original covariance matrix, $\mathbb{C}(\theta, \theta')$ is calculated for the tangential shear $\gamma_t(\theta)$, where θ is the angular separation from the lens, and then converted to the covariance matrix for $\Delta\Sigma$ using

$$\mathbb{C}(r_{p,i}, r_{p,j}) = \Sigma_{\text{crit}}^2(z_L)\mathbb{C}(\theta, \theta'), \quad (59)$$

where Σ_{crit} is the critical surface mass density, and $\gamma_t(\theta) = \Delta\Sigma[r_p = D_A(z_L)\theta] / \Sigma_{\text{crit}}$. We consider GGL measured using the synergy of a DESI-like spectroscopic survey and an LSST-like imaging survey, with two cases of overlapping sky areas – an optimistic case of 6,000 and a pessimistic case of 1,500 squared degrees. In both cases we adopt a value $\sigma_\gamma = 0.22$ and assume that the source galaxy number density is $n_S = 40 \text{ arcmin}^{-2}$. The error bars in the lower panel of Fig. 28 show the square roots of the diagonal elements of the covariance matrix, $\mathbb{C}(r_{p,i}, r_{p,i})$; even in the optimistic case the uncertainty is significantly larger than the model differences for F6 and N5, but for the other models the reverse is true.

In addition to the statistical uncertainties discussed above, in real observations the total error budget must also include contributions from a range of systematic effects. Following [296], we calculate the total covariance matrix as $\mathbb{C}_{ij}^{\text{tot}} = \mathbb{C}_{ij}^{\text{stat}} + \mathbb{C}_{ij}^{\text{syst}}$,

	f_{syst}	F6	F5	F4	N5	N1
	0.0	1.0	5.1	18.9	0.9	5.3
Pessimistic case	0.1	0.4	1.4	4.8	0.1	0.8
	0.0	2.0	10.6	39.2	1.8	11.1
Optimistic case	0.1	0.9	2.7	9.0	0.2	1.1

TABLE IV. The S/N values at which GGL from the synergy of a DESI-like spectroscopic and an LSST-like imaging survey can distinguish the different modified models studied in this paper from GR. See the main text for more details.

where the statistical contribution is as above, while the systematic contribution is given by

$$\mathbb{C}_{ij}^{\text{syst}} = \mathbb{C}^{\text{syst}}(r_{p,i}, r_{p,j}) = f_{\text{syst}}^2 \Delta\Sigma(r_{p,i}) \Delta\Sigma(r_{p,j}), \quad (60)$$

where $\Delta\Sigma(r_{p,i})$ is the excess surface mass density for GR in the i -th r_p bin, and f_{syst} is a multiplicative factor accounting for systematic uncertainties caused by photo- z bias, shear calibration, stellar contamination, intrinsic alignments of galaxies, etc., see, e.g., [296, 303].

We have done the calculation of S/N in Eq. (58) using the full covariance matrix, with two values of f_{syst} : 0.0 (negligible systematic error) and 0.1, and the results are summarized in Table IV. Because of the relatively poor resolution of our simulations, on small scales the matter clustering and galaxy distribution they predict could be inaccurate; to be conservative, we only used the $\Delta\Sigma$ data within $r_p \in [2, 30]h^{-1}\text{Mpc}$ in the forecast (this will also make the result less affected by astrophysical uncertainties such as the impact of baryons on $\Delta\Sigma$). As one can see from Table IV, including systematic uncertainties substantially reduces the power of GGL in distinguishing the various modified gravity models. However, GGL with a LSST-like imaging survey can still tell apart F5 and F4 from GR.

2. Void lensing

We have discussed that while MG models are usually screened in halos and high-density regions, departures from GR can be substantial in underdense (void) regions. The fifth force present in many such models (including $f(R)$ gravity and nDGP) can lead to a more efficient evacuation of underdense regions, and therefore emptier voids than in GR [50, 60, 304–307]. Once the galaxy populations were matched to have the same two-point galaxy correlation function across models, voids in MG models such as $f(R)$ gravity, despite being emptier of matter, have essentially the same void abundances and void galaxy number density profiles as their GR counterparts [50, 305]. However, the weak lensing signal of voids show significant differences with respect to GR, with voids in $f(R)$ and nDGP having a larger tangential shear signal than GR ones [50, 305, 308, 309]. The weak lensing imprint of voids has already been measured by current surveys [310–312], and future observational campaigns would greatly improve the quantity and quality of void lensing data. Voids have an additional advantage, namely their properties are largely insensitive to the baryonic and galaxy formation physics – which is still a major uncertainty – and are well reproduced by dark matter only simulations [313]. As a result, void lensing can be an appealing technique for testing MG models. Voids are also highly versatile [50] in that there can be various different void definitions which trace different aspects of the cosmic web and can be tailored to maximize the potential of extracting certain features of a particular model.

Here, we test the potential of two galaxy void finders to constrain the MG models studied in this paper. The first, the Watershed Void Finder [314, hereafter WVF] identifies underdensities in the 3D distribution of galaxies. The voids are determined by the watershed basins of a given large-scale galaxy density field without imposing any constraints on the shape, size or underdensity of these objects. The method constructs a volume-filling galaxy density field using the Delaunay Tessellation Field Estimator [315, 316], which is based on Delaunay triangulations. The resulting density is defined on a 1024^3 regular grid with a grid cell size of $1 h^{-1}\text{Mpc}$. To reduce small-scale structures that could give rise to artificial voids, the method smooths the density field with a Gaussian filter of $2 h^{-1}\text{Mpc}$ radius – this filter size corresponds to the typical width of the filaments and sheets that form the void edges [e.g. 317, 318]. The smoothed density field is then segmented into watershed basins. This process is equivalent to following the path of a rain drop along a landscape: each volume element, in our case the voxel of a regular grid, is connected to the neighbor with the lowest density, with the same process repeated for each neighbor until a minimum of the density field is reached. Finally, a watershed basin is composed of all the voxels whose paths end at the same density minimum. The void centers are chosen as the volume-weighted barycentre of all the voxels associated to each void, and the void radius is the radius of a sphere with the same volume as the void volume.

The second method identifies tunnels, which are 2D underdensities in the distribution of galaxies projected onto the plane of the sky. The tunnels are defined as circular regions that are devoid of galaxies and consist of elongated line-of-sight regions that intersect one or more voids without passing through overdense regions [50]. To identify tunnels, we build a Delaunay tessellation of the projected galaxy distribution since, by definition, the circumcircle of every Delaunay triangle is empty of galaxies, with the closest galaxies being the ones that give the triangle vertices and that are found exactly on the circumcircle. The tunnels

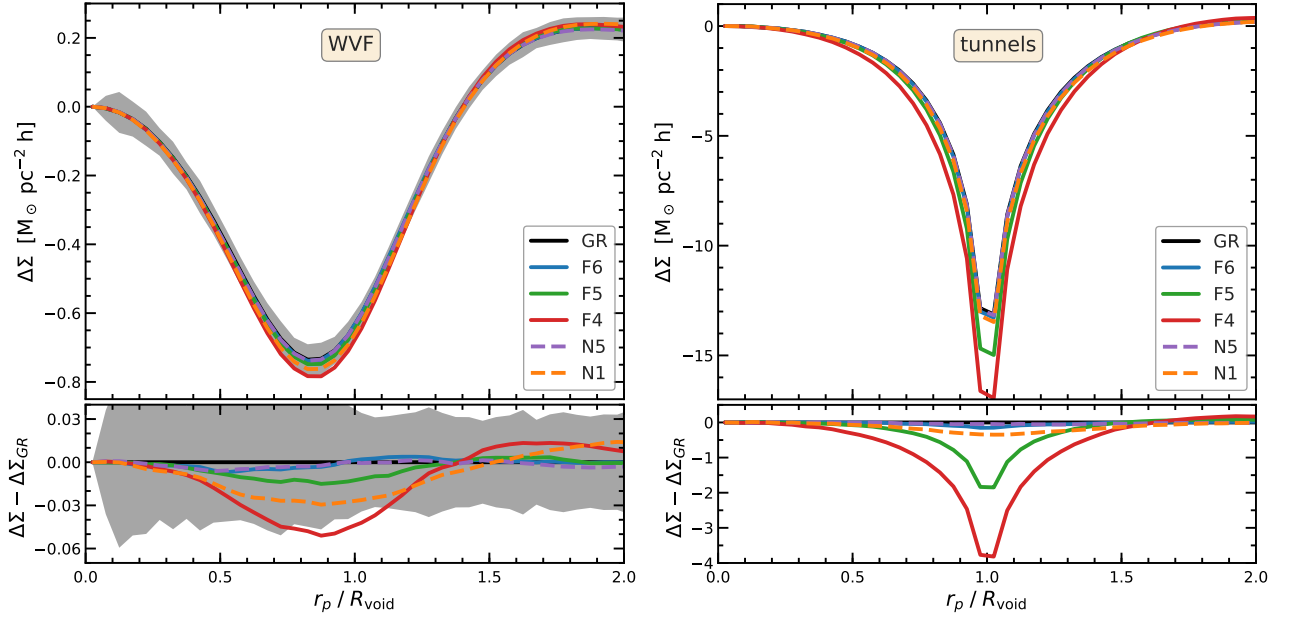


FIG. 29. (Color Online) The excess surface mass density profiles, $\Delta\Sigma$, of 3D voids (left) and 2D tunnels (right) at $z = 0.5$, as a function of scaled radial distance, r/R_{void} . The lines show the predictions for the six models studied here: GR (black), F6 (blue), F5 (green), F4 (red), N5 (cyan) and N1 (orange). The bottom panels show the difference between the MG models and GR, with the grey shaded region showing the 1σ sample variance uncertainties for a 6,000 sq.deg. overlap with the LSST survey (for tunnels, the uncertainty range is very small and hence not easily visible).

consist of the circumcircles whose centres are not inside a larger circumcircle. We are interested in the modified gravity signature of underdense regions, and so we select only the tunnels with radii above $1 h^{-1}\text{Mpc}$, which correspond to underdense regions in projection [50]. We project the entire simulation box, since its length roughly corresponds to the comoving distance between redshift 0.3 and 0.7.

To obtain the void tangential shear, we compute the mean excess surface density profile around each void. We have followed the procedure described in Ref. [50], where the $\Delta\Sigma(r_p)$ for the 3D underdensities, that is WVF objects, was computed similarly to GGL, that is using Eq. (57), but with the galaxy-matter cross correlation function replaced by the void-matter cross correlation one, ξ_{vm} . To average over the density profiles of voids of different sizes, we computed ξ_{vm} as a function of the scaled radial distance, r/R_{void} , with R_{void} the radius of each void. The mean excess surface density of WVF voids was computed as

$$\Delta\Sigma(\eta_p) = \rho_{\text{crit}}\Omega_m\bar{R}_{\text{void}} \left[\frac{2}{\eta_p^2} \int_{-3}^{+3} d\chi \int_0^{\eta_p} d\eta \cdot \eta \xi_{vm} \left(\sqrt{\eta^2 + \chi^2} \right) - \int_{-3}^{+3} d\chi \xi_{vm} \left(\sqrt{\eta_p^2 + \chi^2} \right) \right], \quad (61)$$

where $\bar{R}_{\text{void}}$ is the mean void radius. The symbols η ¹⁰ and χ denote the spatial coordinates perpendicular to and along the line-of-sight, respectively, with both coordinates representing scaled distances, that is in units of the void radius, R_{void} . The value of 3 in the $d\chi$ line-of-sight integral comes from limiting the integral to three times the void radius, which is sufficient for calculating the void lensing signal [50, 305]. In the case of the 2D underdensities, we compute the 2D void-matter cross correlation function, $\xi_{vm\ 2D}(\eta)$, again as a function of the scaled distance, $\eta = r/R_{\text{void}}$, by projecting the dark matter particle distribution along the same axis along which we projected the HOD galaxy distribution. Then, the excess surface density is given by

$$\Delta\Sigma(\eta_p) = \rho_{\text{crit}}\Omega_m \left[\frac{2}{\eta_p^2} \int_0^{\eta_p} d\eta \cdot \eta \xi_{vm\ 2D}(\eta) - \xi_{vm\ 2D}(\eta_p) \right]. \quad (62)$$

We have measured the void-matter cross correlation functions using a brute force algorithm, similarly to §III D 1.

The excess surface density profiles of WVF voids and tunnels at $z = 0.5$ are shown in Fig. 29. Since voids are underdense, they have negative $\Delta\Sigma$ values, which means that they give rise to a similar effect as a divergent lens. Of the two void finders, the

¹⁰ Note that here we use η instead of r_p for the distance transverse to the line of sight, to highlight the fact that the former is rescaled by R_{void} .

3D WVF voids							2D tunnels						
	f_{syst}	F6	F5	F4	N5	N1		f_{syst}	F6	F5	F4	N5	N1
Pessimistic case	0.00	0.1	0.3	1.0	0.1	0.7	Pessimistic case	0.00	1.4	15	32	0.8	4.3
	0.10	0.1	0.2	0.6	0.1	0.4		0.10	0.5	3.1	4.5	0.7	2.7
Optimistic case	0.00	0.3	0.6	2.0	0.3	1.3	Optimistic case	0.00	2.8	30	63	1.7	8.6
	0.10	0.3	0.3	0.7	0.3	0.6		0.10	0.9	5.8	7.5	1.5	5.4

TABLE V. The S/N values for using void tangential shear measurements to distinguish the various modified gravity models from GR. These are based on forecasts using the synergy between DESI and two overlapping imaging surveys as described in the text. We present results for two void identification methods: 3D WVF voids and 2D tunnels.

tunnels have lensing signals that are nearly 20 times larger, demonstrating that selecting underdensities in projection results in a larger lensing signal than 3D underdensities. Void lensing would be even larger if we select the voids using the weak lensing convergence field, as shown in [319, 320]. Compared with GR, voids in the two MG models studied here have systematically larger lensing signals, i.e., show more divergent lensing effects. The differences are largest for F4 within the $f(R)$ family and for N1 in the nDGP family. The differences vary with redshift (not shown here), being larger at lower z [50], because in both MG models studied here the screening becomes weaker at late times, leading to larger deviations from GR.

Interestingly, for 3D voids the F4 and N1 models show a similarly sized difference with respect to GR. However, for tunnels the N1 model – which is the nDGP variant that deviates from GR more strongly – is nearly as close to GR as the F6 model, while the F4 model shows a much larger difference. The tunnels are selected to maximize the transition contrast between underdense and overdense regions, with $\Delta\Sigma$ being proportional to the steepness of this transition. In contrast to $f(R)$ gravity, the overdense regions in nDGP models have a similar density as in GR thanks to the more efficient screening (see §III D 1), and the differences between nDGP and GR lensing come only from underdense regions, which explains why tunnel lensing is better for testing $f(R)$ models than nDGP ones. This is another example to show how the different screening mechanisms can leave distinct imprints in the observed large-scale matter distribution, which can in turn be picked up or maximized by a suitably-tailored void finding algorithm [308].

To estimate the power of WVF and tunnels to constrain modified gravity models, we have calculated the signal-to-noise (S/N) with which various models can be distinguished from GR. For this, we follow the same approach as in §III D 1 and compute the S/N using Eq. (58). The covariance matrix contains the contribution from sample variance, shape noise and systematic effects. We estimate this by considering a single source redshift of $z_s = 1.0$ and a distribution of void lenses between redshifts $z = 0.3$ and $z = 0.7$. Considering lower redshift lenses makes little difference since the $z < 0.3$ volume is small; considering lenses with $z > 0.7$ adds little constraining power due to a combination of decreasing lensing kernel and smaller differences between modified gravity models and GR. For simplicity, we consider that all void lenses between $0.3 \leq z \leq 0.7$ have the same excess surface density given by the mean value at $z = 0.5$. Similar to §III D 1, we consider two cases in which the DESI spectroscopic survey have different overlaps – 6,000 deg² for the *optimistic case* and 1,500 deg² for the *pessimistic case* – with the galaxy imaging survey LSST. In both cases we adopt an intrinsic source shape noise $\sigma_\gamma = 0.22$ and a source galaxy number density $n_S = 40 \text{ arcmin}^{-2}$.

We estimate the sample variance using the 5 realizations of GR simulations. For each realization, we split the volume into 4³ non-overlapping regions and compute $\Delta\Sigma$ for each of these regions; we do so using Eq. (62) for both 3D voids and tunnels (see [50] for a discussion of why we cannot use Eq. 61). Then, we generate 100 bootstrap samples over these regions and calculate the mean $\Delta\Sigma$ of each of the bootstrap samples. The procedure leads to 5×100 samples which we use to calculate the sample variance of $\Delta\Sigma$. The resulting uncertainties are shown as a grey shaded region in Fig. 29. To estimate the sample variance for the survey, we scale it by the ratio of the box volume to the survey volume between $z = 0.3$ and 0.7 . We estimate the shape noise covariance matrix, $\mathbb{C}_{\text{SN}}$, by first calculating it for the tangential shear, $\gamma_t(\theta)$, using angular coordinates, θ , and then converting it to a covariance matrix for $\Delta\Sigma$ similarly to Eq. (59):

$$\mathbb{C}_{\text{SN}}(R_{\text{void}}\eta_{p,i}, R_{\text{void}}\eta_{p,j}) = \Sigma_{c;\text{eff}}^2 \mathbb{C}_{\text{SN}}(\theta_i, \theta_j), \quad (63)$$

where $\Sigma_{c;\text{eff}}$ is the effective critical surface density for lensing, computed using Eq. (19) of [50] taking into account the variation of the lensing kernel across the considered redshift range. The total statistical uncertainty is given by the sum of the sample variance and shape noise covariance matrices. The lensing measurements can be affected by systematic effects, which we model as a fraction, f_{syst} , of the total lensing signal, cf. Eq. (60), and which we add to the covariance matrix describing the statistical uncertainties similar to the case of GGL above.

The S/N values with which 3D WVF voids and 2D tunnels can discriminate the various MG models are given in Table V. A combination of 3D WVF voids and the pessimistic LSST overlap scenario is unable to constrain any of the studied MG models even before considering potential systematic errors in the lensing measurement. Having a survey with a larger sky coverage and deeper imaging, like the optimistic LSST overlap scenario, results in 3D voids being able to constrain only the F4 model to a modest 2σ level. However, including a 10% systematic lensing error degrades again very much the constraining power of WVF

voids and results in $S/N \lesssim 1$. The situation is much better for 2D tunnels, which can probe F5, F4 and N1 models even for the pessimistic overlap case. The constraining power of tunnels is even better for the optimistic overlap scenario, which would result in $S/N \gtrsim 3$ for all models studies here except N5 which has a $S/N = 1.7$. Including a 10% systematic error leads to S/N values of ~ 6 for F5, F4 and N1 (all models end up having the same S/N although they have very different S/N values in the absence of systematic errors), and S/N value of 1.5 for N5 and 0.9 for F6.

Comparing the S/N values for GGL and tunnels, we notice that, amongst the $f(R)$ variants, tunnels give better constraints for F6 and F5; which is possibly because GGL probes the lensing of high-density regions while void lensing measures the effect of low-density regions where the effect of the fifth force is stronger. The case of F4 is special, in that chameleon screening is quite inefficient in this model so that the fifth force effect can be strong both in voids and near halos, and our result suggests that GGL is boosted more than void lensing in this case. Adding systematic errors downgrades constraints for all $f(R)$ models but does not qualitatively change this observation. In the case of nDGP, it is known that screening is strong in the vicinity of dark matter halos but weaker far away. Since GGL probes regions far beyond the halo virial radius, in the absence of systematic errors GGL and void lensing give similar S/N values, while the systematics seem to affect tunnel lensing more. We remark, however, that the GGL S/N values are calculated for distances larger than $2 h^{-1} \text{Mpc}$ and including the inner regions may boost the S/N values, especially for $f(R)$ models (though we would need to worry about uncertainties caused by the impact of baryons in that case), and that the ways to calculate the sample variance contributions to the statistic error budgets for GGL and void lensing are not exactly the same. A more detailed and rigorous comparison of the constraining powers of these two probes will be left for future work.

IV. CONCLUSIONS

In this paper, we have undertaken an initial study of estimators that could potentially be employed to extract information about the properties of gravity from DESI large-scale structure data, leveraging the variety of environments DESI will observe, from voids to densely populated galaxy clusters. We have considered one, two and multi-point statistics of positional clustering and relative motions of galaxies in both configuration and Fourier spaces, utilizing mock galaxy samples from simulations that currently cover smaller volumes than will be covered in the full DESI survey.

This work is a preparatory step to both assess the relative potential of different statistics to put constraints on the properties of gravity and to determine the requirements for future larger-scale cosmological simulations, in terms of both the necessary fidelity in reproducing modified gravity phenomenology, and in reproducing the realities of the DESI survey extent, completeness and instrumental and astrophysical systematic effects.

To make accurate predictions for how gravity might be constrained by DESI, when fully leveraging all observed spatial scales, and linear to nonlinear clustering, numerical simulations and mock galaxy catalogs that include specific MG phenomenology are essential.

In this work, we have considered two examples of scalar-tensor theories, which have played an important role in the development of Solar System tests of GR, and serve as excellent case studies for this. The chameleon $f(R)$ gravity and DGP braneworld models considered exhibit different ‘screening’ mechanisms, that allow their predictions to mimic GR in the Solar System and pass existing local gravity tests, and both predict the same speeds of photons and gravitational waves, a key feature that makes them compatible with detections of gravitational waves with electromagnetic counterparts. These models exhibit subtle differences from GR and from each other, however, on cosmic scales, that one can hope to use to distinguish or constrain them with observations. Their features are expected to arise in a broader range of theories, making the simulations more broadly applicable.

This paper primarily uses existing N-body simulations that, while fully incorporating the complex physics of the scalar-tensor theory, are over smaller volumes and with lower mass resolution than will ultimately be sought for a full simulation of DESI. We also briefly commented on faster, approximate, simulation methods such as MG-COLA, which will be valuable complements to the full simulations, giving the capability of running a large number of realizations for covariance matrix estimates etc.. The dark matter particle simulations have been used to construct mock galaxy catalogs using simulated halo catalogs and applying simple HOD prescriptions. The HOD parameters have been tuned so that they have approximately the same galaxy number density and galaxy two-point clustering properties (equivalent to fixing the number density and projected galaxy 2-point correlation function of the galaxy sample to those from observations), and then studying other statistics to see if they show any appreciable difference between the different gravity models.

Of the statistics considered, we find that the redshift space distortions and higher order correlation functions offer the greatest immediate potential for distinguishing modifications to GR with the first two years of DESI data.

- Prospective large scale RSD constraints on the growth rate from DESI, parameterized by β , are expected to be markedly tighter than current constraints. This can be most useful in constraining models such as nDGP, which modifies the growth rate on large, linear, scales. The SHAM analysis presented here demonstrates a promising potential to differentiate between GR and some MG models, such as $f(R)$ gravity, whose effect is most prominent on small, nonlinear, scales; the inclusion

of small-scale RSD offers complementary information that could improve the ability of differentiating the various models. Finally, void RSD is a relatively new probe complementary to RSD around galaxies or galaxy clusters, and is of particular interest since near voids the effect of modified gravity is usually stronger. We discussed the challenges facing these probes.

- The 3PCF offers complementary information about the nonlinear evolution of matter to the traditional two-point statistics. We find deviations of the MG models from GR on all scales, with different degrees of strength depending on the triangular shape searched. On large scales, a multipole decomposition in one of the triangle angles allows for an efficient comparison of models, showing particular patterns in the two triangle sides that define the angle: the signal is relatively small on scales larger than $\sim 40 h^{-1}\text{Mpc}$, but the complexity in the patterns at each multipole is unique for each tested model, and cannot be easily reproduced by other mechanisms. On scales $\sim 10 h^{-1}\text{Mpc}$, the full 3PCF can be calculated and the signal is strong enough to provide a robust estimator to test gravity – this is further confirmed by looking at the galaxy bispectrum at $k_1 \gtrsim 0.1 h\text{Mpc}^{-1}$ and the higher order central moments with a smoothing scale of $\lesssim 20 h^{-1}\text{Mpc}$. However, detailed studies on the systematics involved in the different high-order statistics should be carried on, together with an appropriate modeling of the signal including effects such as RSD, either through perturbation theory or by using simulations, to assess the findings of these estimators when applying to broader classes of MG models.

Other statistics appear to have a strong variation in outcomes depending on whether one uses the halo catalog or a HOD tuned galaxy catalog, or on the details of the HOD model. Our theoretical understanding of these estimators is dependent on how well we can model their predictions and assumptions of the galaxy formation model. The utility of these will therefore require further analysis to assess if they offer strong constraining potential for DESI:

- Marked correlation functions, with appropriate choices of marks, can in principle up-weight galaxies in environments where MG signals are stronger. Although the results depend on the details of the implementation, we find that if the mark is defined using the galaxy number density, which along with the galaxy two-point clustering has been tuned to match in the different models, the distinguishing power is degraded in general. Using additional information, e.g., the Newtonian potentials of the host halos of galaxies, is found to lead to increased model differences, which suggests that in future works other possibilities of marks should be explored. It is also important to understand, and properly account for, the systematic errors of the marks themselves, which are (directly or indirectly) related to observables.
- Other measures of the information beyond two-point statistics, such as the Minkowski functionals, have the potential of offering complementary constraints on models or breaking parameter degeneracies. However, their study is not yet in a mature state and further efforts are needed to assess their model-constraining power.

In addition to statistics based solely on DESI data, we have also considered the potential to constrain gravity from combined results using lensing statistics in areas with DESI overlap. While the results shown are promising, further analyses using galaxy clustering simulations with concurrent lensing predictions are required to fully understand the constraining potential:

- The MG models studied here predict stronger matter clustering around galaxies, with the enhancement of galaxy galaxy lensing with respect to GR predictions strongest near (far from) the galaxies for the $f(R)$ (nDGP) models as a result of their different screening mechanisms. We find that for a LSST-like imaging survey overlapping with DESI, with reasonably well-controlled systematics ($\lesssim 10\%$), the enhanced galaxy-galaxy lensing signals can be used to distinguish F5, F4 and N1 from GR.
- Voids in modified gravity models are emptier and have a larger lensing signal. Of the two methods tested here, i.e., the 3D Watershed Void Finder and the 2D tunnels, the latter applied to a LSST-like deep lensing survey overlapping the DESI region can distinguish all the models investigated in this paper from GR (at more than 3σ level even when considering 10% systematic errors).
- Phase space statistics probe both position and velocity data and photometric data for weak-lensing mass (or mass-richness relation). This probe compares the observations directly to the potential (GR or modified), as traced by the dynamics. If the modified gravity model affects the dynamics and not light-travel, this becomes a powerful probe. It can be used both to rule out specific non-GR models (e.g., by using $f(R)$ as the null hypothesis) or it can be a goodness-of-fit to GR (where GR is the null hypothesis). After removing the dominant systematic in our tests by using potential ratios and by using a large DESI-like sample of clusters, this probe can constrain $|\overline{f_{R0}}| = 6 \times 10^{-7}$ at $> 5\sigma$.

To further assess the most promising statistics and obtain accurate forecasts of their constraining power with DESI data, it would be useful to have a single simulation with sufficient volume, resolution and redshift coverage to construct realistic mocks for all types of objects targeted by DESI. However, despite the latest technical developments, a full simulation of this kind with modified gravity effects is likely still too expensive. Therefore, based on the findings of this paper, we recommend the following alternatives:

(i) multiple simulations with higher mass and force resolutions than the ones used in this paper, e.g., $L_{\text{box}} \sim 500 h^{-1}\text{Mpc}$ and $N_p = 1024^3$ or $L_{\text{box}} \sim 800\text{--}1000 h^{-1}\text{Mpc}$ and $N_p = 2048^3$. These should be run for a large number of models enough for building emulators of various statistics in a higher-dimensional parameters space spanned by not just the modified gravity but also cosmological parameters.

(ii) even higher-resolution simulations with $L_{\text{box}} \sim 500 h^{-1}\text{Mpc}$ and $N_p = 2048^3$. These are roughly of the same size and resolution as the Millennium Simulation [174], and some of these simulations (for nDGP and ΛCDM) have already been running [321]. The high resolution will make such simulations useful for resolving even small halos which host galaxies such as ELGs, and the box size is still large enough to study clusterings on nonlinear and some linear scales. Their large cost means that they can not be run for a large number of models, unless a new generation of much more efficient full simulation codes come to existence in the near future.

At this point, we consider it important to prioritize resolution and coverage of the model/parameter space over volume, since most of the estimators studied in this paper still lack reliable theoretical predictions, and for some of them simulations offer a more promising and accurate way.

These will be complemented by a large suite of fast approximate simulations using (appropriately tested and tuned versions of) MG-COLA that can be used to estimate the covariance matrices for some of the estimators studied here. In ΛCDM models, various approximate methods were compared in Ref. [322] and their validity of creating galaxy mocks to estimate the covariance matrices established on BAO scales [323]. In this paper, we are interested in estimators that go into the non-linear scales and the validity of the COLA approach needs to be tested carefully in this regime. In ΛCDM , it was shown that by re-calibrating halo masses and/or velocities, it is possible to bring accuracy in clustering to one percent level on large scales [324], and this analysis needs to be extended to low-mass dark matter halos and small-scale clustering. In addition, MG-COLA uses an approximate method for screening based on spherically symmetric solutions [41], which was shown to have a sufficient accuracy up to $k = 1h\text{Mpc}^{-1}$ in describing deviations from ΛCDM in the matter power spectrum [43] but again a calibration might be required. Once it is calibrated and validated against full N-body simulations, MG-COLA could provide a promising way to compute the covariance matrices for some of the estimators considered in this paper.

The (purely dark matter) simulations mentioned above need to be populated with galaxies so that they can be compared with data. In this paper, we have primarily used mock galaxy catalogs based on HOD in our study, but other methods to make mock galaxies such as SHAM or using sub-sampled dark matter particles have also been used. As we pointed out when describing a few estimators, uncertainties in such galaxy-halo connections may have a non-negligible impact on their theoretical predictions. A possible way to quantify this is by assessing the impact of different models of galaxy-halo connection on every estimator; where hydrodynamics simulations of galaxy formation exist, it is also useful to check the effect of the subgrid baryonic models on galaxy clustering observables, or use them to calibrate the galaxy populating recipe. It is perhaps more advisable to directly use observational data, e.g., of the two-point galaxy clustering, to calibrate the HOD model (e.g., [325, 326]), a simplified variant of which is we have followed to create galaxy mocks in this paper.

Moreover, during the past few years there has been a rise of interests in the use of emulators to constrain cosmological parameters directly from N-body simulations. For instances, the Aemulus [151, 327] and Dark Quest [328] projects presented accurate estimators for the halo mass function and galaxy clustering, among other summary statistics, as functions of the cosmological parameters in ΛCDM . In general, emulators learn to mimic the behavior of a physical model that is slow and expensive to evaluate. They provide predictions of the model outputs at input parameters where the expensive model has not been evaluated. By far, the most common emulation technique applied to cosmology has been Gaussian Process Emulation (see [329] for a comprehensive review), due to this model's ability to provide uncertainty estimations together with emulator outputs, although neural network approaches have also been used [152]. Directly leveraging N-body simulations to constrain gravity on intermediate to small scales can potentially play a big role in testing MG scenarios, where screening mechanisms introduce further nonlinearities that strongly affect the small scales. Although there has been some initial work done in this direction [160], a more extensive investigation including a variety of models will be material for future studies.

Beyond the simulation creation, another important thing is to include critical systematic effects when constraining MG from small scales. Missing galaxies due to fiber collisions, i.e., the finite size of the fibers preventing their placement close to each other within the focal plane, generally affects all scales but – relevant to our work — leads to a factor of two discrepancy in the 2-point clustering on small scales if not accounted for properly. We would want to incorporate fiber collision modeling into the mock data we analyze. One way to account for the effect of collisions is to introduce a probability based weighting scheme using the target algorithm itself [330]. While this method was developed within the framework of 2-point statistics, in principle it can be applied to other estimators as well. Nonetheless, it is necessary to test for potential bias and efficiency when the method is applied to other statistics and in particular the interplay with statistics to detect the environment dependent effects of modified gravity.

Beyond this, one needs a pipeline that produces multiple mocks on a light-cone that incorporates the DESI survey geometry and imposes survey masks which incorporate information about the observing conditions, such as targeting (e.g., galaxy magnitude, color, surface brightness), placement, galactic extinction, atmospheric extinction, seeing, cloud-cover, zodiacal light, and so on. Mitigation strategies for these observational systematics, including a forward modeling approach and weighting methods, will be developed by the DESI clustering working group and are in detail discussed in a companion paper elsewhere. We do

not expect the measurement and observational effects, such as extinction or seeing, to particularly affect small scales, but rather mainly be important for large to intermediate scales.

To close, DESI, a Stage IV state-of-the-art galaxy survey, will open up a wide range of opportunities to test the theory of gravity on cosmological scales with unprecedented precision. This work provides a useful first step for the DESI collaboration, and the wider theoretical and observational communities, to plan joint efforts to ensure we can fully exploit the wealth of future observations and contribute to the tests of fundamental theories in physics.

ACKNOWLEDGMENTS

We thank Peder Norberg and Martin White for useful comments on the early results of this work.

RB and GV's work is supported by DOE Grant DE-SC0011838, NASA ATP Grant NASA ATP Grants NNX14AH53G and 80NSSC18K0695G, NASA ROSES Grant 12-EUCLID12-0004 and NASA support for the Nancy Grace Roman Space Telescope High Latitude Survey Science Investigation Team. MC was supported by Science and Technology Facilities Council (STFC) [ST/L00075X/1, ST/P000451/1], ERC Advanced Investigator Grant, DMIDAS [GA 786910], and the EU Horizon 2020 research and innovation programme under a Marie Skłodowska-Curie Grant agreement 794474 (DancingGalaxies). JH is supported by a Durham co-fund Junior Research Fellowship, and the European Research Council (ERC-StG-716532-PUNCA). WAH is supported by the Polish National Science Center Grant No. UMO-2018/30/E/ST9/00698 and UMO-2018/31/G/ST9/03388. This project has also benefited from numerical computations performed at the Interdisciplinary Centre for Mathematical and Computational Modelling (ICM) University of Warsaw under Grants Nos. GA67-17 and GA65-30. CC-L is supported by a PhD Studentship from the Durham Centre for Doctoral Training in Data Intensive Science, funded by the UK Science and Technology Facilities Council (STFC, ST/P006744/1) and Durham University. CH-A is supported by the Excellence Cluster ORIGINS which is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC-2094 - 390783311. MI acknowledges support by the Department of Energy, Office of Science, under Award Number DE-SC0019206. KK is supported by the STFC Grant No. ST/N000668/1. The work of KK and HAW has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement 646702 "CosTesGrav"). BL is supported by the European Research Council (ERC-StG-716532-PUNCA) under the H2020 Framework of the European Commission, and STFC Consolidated Grants ST/L00075X/1, ST/P000451/1. CGS acknowledges support from the National Research Foundation of South Korea (NRF, 2020R111A1A01073494). GN and AGM acknowledge support from CONACyT's Grants 179208, 286897 and Fronteras de la Ciencia 281, Instituto Avanzado de Cosmología (IAC), DAIP-UG and the computer infrastructure of the DCI-UG DataLab. MV is partially supported by Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica (PAPIIT) No. IA102516, No. IA101518, and from Proyecto LANCAD-UNAM-DGTIC-319 and LANCAD-UNAM-DGTIC-136. PN is funded by 'Centre National d'Études Spatiales' (CNES). SF is supported by Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica (PAPIIT) No. IA101619. AA and JLCC are supported by CONACyT project 283151. GBZ is supported by the National Key Basic Research and Development Program of China (No. 2018YFA0404503), and by NSFC Grants 11720101004 and 11673025. CM and VH is partially supported by the National Science Foundation under Grant No. 181273.

O.V. and N. C. Devi are partially supported by a PAPIIT UNAM Grants IN112518 and AG101620. NCDevi acknowledges support from the European Commission's Framework Programme 7, Marie Curie International Research Staff Exchange Scheme LACEGAL (PIRSES-GA-2010-269264). This study also used computer and human resources from the LAMOD supercomputing initiative.

This research used resources of the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility operated under Contract No. DE-AC02-05CH11231. This research was partially supported through computational and human resources provided by the LAMOD UNAM project through the clusters Atocatl and Tochtli. LAMOD is a collaborative effort between the IA, ICN and IQ institutes at UNAM. The N-body simulations used in this work were carried out on two supercomputer systems: (1) the DiRAC Data Centric system at Durham University, operated by the Institute for Computational Cosmology (ICC) on behalf of the STFC DiRAC HPC Facility (www.dirac.ac.uk). This equipment was funded by BIS National E-infrastructure capital Grant ST/K00042X/1, STFC capital Grants ST/H008519/1 and ST/K00087X/1, STFC DiRAC Operations Grant ST/K003267/1 and Durham University. DiRAC is part of the National E-Infrastructure. (2) the Interdisciplinary Center for Mathematical and Computational Modelling (ICM) at University of Warsaw, under Grants No. GA67-17 and No. GA65-30.

[1] J. M. Taylor and W. J. M., *Astrophys. J.* **253**, 908 (1982).

[2] Virgo, LIGO Scientific, B. P. Abbott *et al.*, *Phys. Rev. Lett.* **116**, 061102 (2016), arXiv:1602.03837.

[3] Virgo, Fermi-GBM, INTEGRAL, LIGO Scientific, B. P. Abbott *et al.*, *Astrophys. J.* **848**, L13 (2017), arXiv:1710.05834.

- [4] GROND, SALT Group, OzGrav, DFN, DES, INTEGRAL, Virgo, Insight-Hxmt, Fermi-LAT, J-GEM, RATIR, ATLAS, IceCube, CAAS-TRO, LWA, TZAC, GRAWITA, RIMAS, SKA South Africa/MeerKAT, H.E.S.S., 1M2H Team, IKI-GW Follow-up, Fermi GBM, Pi of Sky, MASTER, ePESSTO, AstroSat Cadmium Zinc Telluride Imager Team, Swift, HAWC, Pierre Auger, ASKAP, VINROUGE, JAGWAR, Chandra Team at McGill University, TTU-NRAO, GROWTH, DWF, AGILE Team, MWA, ATCA, AST3, TOROS, Pan-STARRS, NuSTAR, BOOTES, CaltechNRAO, LIGO Scientific, High Time Resolution Universe Survey, Nordic Optical Telescope, Las Cumbres Observatory Group, LOFAR, IPN, DLT40, Texas Tech University, MAXI, ANTARES, KU, Dark Energy Camera GW-EM, CALET, Euro VLBI Team, ALMA, LIGO, *Astrophys. J.* **848**, L12 (2017), arXiv:1710.05833.
- [5] Supernova Cosmology Project, S. Perlmutter *et al.*, *Astrophys. J.* **517**, 565 (1999), arXiv:astro-ph/9812133.
- [6] Supernova Search Team, A. G. Riess *et al.*, *Astron. J.* **116**, 1009 (1998), arXiv:astro-ph/9805201.
- [7] E. J. Copeland, M. Sami, and S. Tsujikawa, *Int. J. Mod. Phys. D* **15**, 1753 (2006), arXiv:hep-th/0603057.
- [8] T. Clifton, P. G. Ferreira, A. Padilla, and C. Skordis, *Phys. Rept.* **513**, 1 (2012), arXiv:1106.2476.
- [9] A. Joyce, B. Jain, J. Khoury, and M. Trodden, *Reports on Progress in Physics* **568**, 1 (2015), arXiv:1407.0059.
- [10] K. Koyama, *Reports on Progress in Physics* **79**, 046902 (2016), arXiv:1504.04623.
- [11] M. Ishak, *Living Rev. Rel.* **22**, 1 (2019), arXiv:1806.10122.
- [12] A. Albrecht *et al.*, (2006), arXiv:astro-ph/0609591.
- [13] DESI, A. Aghamousa *et al.*, (2016), arXiv:1611.00036.
- [14] S. M. Carroll, V. Duvvuri, M. Trodden, and M. S. Turner, *Phys. Rev. D* **70**, 043528 (2004), arXiv:astro-ph/0306438.
- [15] S. M. Carroll *et al.*, *Phys. Rev. D* **71**, 063513 (2005), arXiv:astro-ph/0410031.
- [16] G. R. Dvali, G. Gabadadze, and M. Porrati, *Phys. Lett. B* **485**, 208 (2000), arXiv:hep-th/0005016.
- [17] P. Ntelis, arXiv e-prints, arXiv:2010.06707 (2020), arXiv:2010.06707.
- [18] J. Khoury and A. Weltman, *Phys. Rev. Lett.* **93**, 171104 (2004), astro-ph/0309300.
- [19] J. Khoury and A. Weltman, *Physical Review D* **69**, 044026 (2004), astro-ph/0309411.
- [20] A. I. Vainshtein, *Physics Letters B* **39**, 393 (1972).
- [21] B. Li *et al.*, *Mon. Not. R. Astron. Soc.* **428**, 743 (2013), arXiv:1206.4317.
- [22] S. Bose, W. A. Hellwing, and B. Li, *J. Cosmo. Astropart. Phys.* **2**, 034 (2015), arXiv:1411.6128.
- [23] H. Oyaizu, *Phys. Rev. D* **78**, 123523 (2008), arXiv:0807.2449.
- [24] B. Li and H. Zhao, *Phys. Rev. D* **80**, 044027 (2009), arXiv:0906.3880.
- [25] K. C. Chan and R. Scoccimarro, *Phys. Rev. D* **80**, 104005 (2009), arXiv:0906.4548.
- [26] B. Li and J. D. Barrow, *Phys. Rev. D* **83**, 024007 (2011), arXiv:1005.4231.
- [27] B. Li, D. F. Mota, and J. D. Barrow, *Astrophys. J.* **728**, 109 (2011), arXiv:1009.1400.
- [28] G.-B. Zhao, B. Li, and K. Koyama, *Phys. Rev. D* **83**, 044007 (2011), arXiv:1011.1257.
- [29] A.-C. Davis, B. Li, D. F. Mota, and H. A. Winther, *Astrophys. J.* **748**, 61 (2012), arXiv:1108.3081.
- [30] B. Li, G.-B. Zhao, R. Teyssier, and K. Koyama, *J. Cosmo. Astropart. Phys.* **1201**, 051 (2012), arXiv:1110.1379.
- [31] E. Puchwein, M. Baldi, and V. Springel, *Mon. Not. Roy. Astron. Soc.* **436**, 348 (2013), arXiv:1305.2418.
- [32] B. Li, G.-B. Zhao, and K. Koyama, *J. Cosmo. Astropart. Phys.* **1305**, 023 (2013), arXiv:1303.0008.
- [33] B. Li *et al.*, *J. Cosmo. Astropart. Phys.* **1311**, 012 (2013), arXiv:1308.3491.
- [34] C. Llinares, D. F. Mota, and H. A. Winther, *Astron. Astrophys.* **562**, A78 (2014), arXiv:1307.6748.
- [35] A. Barreira, B. Li, W. A. Hellwing, C. M. Baugh, and S. Pascoli, *J. Cosmo. Astropart. Phys.* **1310**, 027 (2013), arXiv:1306.3219.
- [36] G.-B. Zhao *et al.*, in prep. (2019).
- [37] C. Arnold, M. Leo, and B. Li, *Nat. Astron.* **3**, 945 (2019), arXiv:1907.02977.
- [38] C. Hernández-Aguayo, C. Arnold, B. Li, and C. M. Baugh, (2020), arXiv:2006.15467.
- [39] H. A. Winther *et al.*, *Mon. Not. R. Astron. Soc.* **454**, 4208 (2015), arXiv:1506.06384.
- [40] J. Khoury and M. Wyman, *Phys. Rev. D* **80**, 064023 (2009), arXiv:0903.1292.
- [41] H. A. Winther and P. G. Ferreira, *Phys. Rev. D* **91**, 123507 (2015), arXiv:1403.6492.
- [42] G. Valogiannis and R. Bean, *Phys. Rev. D* **95**, 103515 (2017), arXiv:1612.06469.
- [43] H. A. Winther, K. Koyama, M. Manera, B. S. Wright, and G.-B. Zhao, *J. Cosmo. Astropart. Phys.* **1708**, 006 (2017), arXiv:1703.00879.
- [44] A. Barreira, S. Bose, and B. Li, *J. Cosmo. Astropart. Phys.* **1512**, 059 (2015), arXiv:1511.08200.
- [45] S. Bose *et al.*, *J. Cosmo. Astropart. Phys.* **1702**, 050 (2017), arXiv:1611.09375.
- [46] Y.-S. Song *et al.*, *Phys. Rev. D* **92**, 043522 (2015), arXiv:1507.01592.
- [47] A. Barreira, A. G. Sánchez, and F. Schmidt, *Phys. Rev. D* **94**, 084022 (2016), arXiv:1605.03965.
- [48] E.-M. Mueller *et al.*, (2016), arXiv:1612.00812.
- [49] R. K. Sheth, A. J. Connolly, and R. Skibba, ArXiv e-prints (2005), arXiv:astro-ph/0511773.
- [50] M. Cautun *et al.*, *Mon. Not. R. Astron. Soc.* **476**, 3195 (2018), arXiv:1710.01730.
- [51] D. Shi, B. Li, J. Han, L. Gao, and W. A. Hellwing, *Mon. Not. Roy. Astron. Soc.* **452**, 3179 (2015), arXiv:1503.01109.
- [52] H. A. Winther *et al.*, *Mon. Not. Roy. Astron. Soc.* **454**, 4208 (2015), arXiv:1506.06384.
- [53] C. Burrage and J. Sakstein, (2017), arXiv:1709.09071.
- [54] L. Lombriser and A. Taylor, *J. Cosmo. Astropart. Phys.* **1603**, 031 (2016), arXiv:1509.08458.
- [55] P. Creminelli and F. Vernizzi, (2017), arXiv:1710.05877.
- [56] J. Sakstein and B. Jain, (2017), arXiv:1710.05893.
- [57] J. M. Ezquiaga and M. Zumalacárregui, (2017), arXiv:1710.05901.
- [58] A. Barreira, B. Li, C. Baugh, and S. Pascoli, *Phys. Rev. D* **90**, 023528 (2014), arXiv:1404.1365.
- [59] A. Barreira, B. Li, C. Baugh, and S. Pascoli, *J. Cosmo. Astropart. Phys.* **1408**, 059 (2014), arXiv:1406.0485.
- [60] A. Barreira, M. Cautun, B. Li, C. Baugh, and S. Pascoli, *J. Cosmo. Astropart. Phys.* **1508**, 028 (2015), arXiv:1505.05809.
- [61] J. Renk, M. Zumalacárregui, F. Montanari, and A. Barreira, *J. Cosmo. Astropart. Phys.* **1710**, 020 (2017), arXiv:1707.02263.

- [62] J. Wang, L. Hui, and J. Khoury, *Phys. Rev. Lett.* **109**, 241301 (2012), arXiv:1208.4612.
- [63] P. Brax, C. van de Bruck, A.-C. Davis, and D. J. Shaw, *Phys. Rev. D* **78**, 104021 (2008), arXiv:0806.3415.
- [64] J. J. Ceron-Hurtado, J.-h. He, and B. Li, *Phys. Rev. D* **94**, 064052 (2016), arXiv:1609.00532.
- [65] L. Lombriser, *Int. J. Mod. Phys. D* **27**, 1848002 (2018), arXiv:1908.07892.
- [66] P. Ntelis *et al.*, **2017**, 019 (2017), arXiv:1702.02159.
- [67] P. Ntelis *et al.*, arXiv e-prints, arXiv:1810.09362 (2018), arXiv:1810.09362.
- [68] A. De Felice and S. Tsujikawa, *Living Rev. Rel.* **13**, 3 (2010), arXiv:1002.4928.
- [69] W. Hu and I. Sawicki, *Physical Review D* **76**, 064004 (2007), arXiv:0705.1158.
- [70] C. Charmousis, R. Gregory, N. Kaloper, and A. Padilla, *J. High Energy Phys.* **10**, 066 (2006), arXiv:hep-th/0604086.
- [71] R. Gregory, N. Kaloper, R. C. Myers, and A. Padilla, *J. High Energy Phys.* **10**, 069 (2007), arXiv:0707.2666.
- [72] D. Gorbunov, K. Koyama, and S. Sibiryakov, *Phys. Rev. D* **73**, 044016 (2006), arXiv:hep-th/0512097.
- [73] F. Schmidt, *Phys. Rev. D* **80**, 123003 (2009), arXiv:0910.0235.
- [74] F. Schmidt, *Phys. Rev. D* **81**, 103002 (2010), arXiv:1003.0409.
- [75] B. Falck, K. Koyama, and G.-B. Zhao, *J. Cosmo. Astropart. Phys.* **1507**, 049 (2015), arXiv:1503.06673.
- [76] R. Teyssier, *Astron. Astrophys.* **385**, 337 (2002), arXiv:astro-ph/0111367.
- [77] G. Hinshaw *et al.*, *Astrophys. J. Suppl.* **208**, 19 (2013).
- [78] S. Prunet *et al.*, *Astrophys. J. Suppl.* **178**, 179 (2008), arXiv:0804.3536.
- [79] S. Colombi, A. H. Jaffe, D. Novikov, and C. Pichon, *Mon. Not. Roy. Astron. Soc.* **393**, 511 (2009), arXiv:0811.0313.
- [80] P. S. Behroozi, R. H. Wechsler, and H.-Y. Wu, *Astrophys. J.* **762**, 109 (2013).
- [81] B. Li and H. Zhao, *Phys. Rev. D* **81**, 104047 (2010).
- [82] A. A. Berlind *et al.*, *Astrophys. J.* **593**, 1 (2003), arXiv:astro-ph/0212357.
- [83] Z. Zheng *et al.*, *Astrophys. J.* **633**, 791 (2005), arXiv:astro-ph/0408564.
- [84] Z. Zheng, A. L. Coil, and I. Zehavi, *Astrophys. J.* **667**, 760 (2007), arXiv:astro-ph/0703457.
- [85] J. F. Navarro, C. S. Frenk, and S. D. M. White, *Astrophys. J.* **462**, 563 (1996), arXiv:astro-ph/9508025.
- [86] J. F. Navarro, C. S. Frenk, and S. D. M. White, *Astrophys. J.* **490**, 493 (1997), arXiv:astro-ph/9611107.
- [87] D. Alongso, arXiv:1210.1833.
- [88] M. Manera *et al.*, *Mon. Not. R. Astron. Soc.* **428**, 1036 (2013), arXiv:1203.6609.
- [89] N. Kaiser, *Mon. Not. R. Astron. Soc.* **227**, 1 (1987).
- [90] J. C. Jackson, *Mon. Not. R. Astron. Soc.* **156**, 1P (1972), arXiv:0810.3908.
- [91] A. Cabre and E. Gaztanaga, *Mon. Not. Roy. Astron. Soc.* **393**, 1183 (2009), arXiv:0807.2460.
- [92] A. G. Sánchez, M. Crocce, A. Cabré, C. M. Baugh, and E. Gaztañaga, *Mon. Not. R. Astron. Soc.* **400**, 1643 (2009), arXiv:0901.2570.
- [93] N. Hamaus, P. M. Sutter, G. Lavaux, and B. D. Wandelt, *J. Cosmo. Astropart. Phys.* **1511**, 036 (2015), arXiv:1507.04363.
- [94] N. Hamaus *et al.*, *J. Cosmo. Astropart. Phys.* **1707**, 014 (2017), arXiv:1705.05328.
- [95] Y.-C. Cai, A. Taylor, J. A. Peacock, and N. Padilla, *Mon. Not. Roy. Astron. Soc.* **462**, 2465 (2016), arXiv:1603.05184.
- [96] J. Hou *et al.*, *Mon. Not. R. Astron. Soc.* **480**, 2521 (2018).
- [97] H. Gil-Marín *et al.*, *Mon. Not. R. Astron. Soc.* **477**, 1604 (2018).
- [98] P. Zarrouk *et al.*, *Mon. Not. R. Astron. Soc.* **477**, 1639 (2018).
- [99] C. Hernández-Aguayo, J. Hou, B. Li, C. M. Baugh, and A. G. Sánchez, *Mon. Not. Roy. Astron. Soc.* **485**, 2194 (2019), arXiv:1811.09197.
- [100] B. S. Wright, K. Koyama, H. A. Winther, and G.-B. Zhao, *J. Cosmo. Astropart. Phys.* **1906**, 040 (2019), arXiv:1902.10692.
- [101] F. Beutler *et al.*, *Mon. Not. Roy. Astron. Soc.* **423**, 3430 (2012), arXiv:1204.4725.
- [102] C. Howlett, A. Ross, L. Samushia, W. Percival, and M. Manera, *Mon. Not. Roy. Astron. Soc.* **449**, 848 (2015), arXiv:1409.3238.
- [103] 2dFGRS, W. J. Percival *et al.*, *Mon. Not. Roy. Astron. Soc.* **353**, 1201 (2004), arXiv:astro-ph/0406513.
- [104] C. Blake *et al.*, *Mon. Not. Roy. Astron. Soc.* **436**, 3089 (2013), arXiv:1309.5556.
- [105] C. Blake *et al.*, *Mon. Not. Roy. Astron. Soc.* **415**, 2876 (2011), arXiv:1104.2948.
- [106] L. Samushia, W. J. Percival, and A. Raccanelli, *Mon. Not. Roy. Astron. Soc.* **420**, 2102 (2012), arXiv:1102.1014.
- [107] H. Gil-Marín *et al.*, *Mon. Not. Roy. Astron. Soc.* **460**, 4188 (2016), arXiv:1509.06386.
- [108] E. Hawkins *et al.*, *Mon. Not. R. Astron. Soc.* **346**, 78 (2003), astro-ph/0212375.
- [109] M. Crocce and R. Scoccimarro, *Phys. Rev. D* **73**, 063519 (2006), astro-ph/0509418.
- [110] J. N. Grieb, A. G. Sánchez, S. Salazar-Albornoz, and C. Dalla Vecchia, *Mon. Not. R. Astron. Soc.* **457**, 1577 (2016), arXiv:1509.04293.
- [111] K. C. Chan, R. Scoccimarro, and R. K. Sheth, *Phys. Rev. D* **85**, 083509 (2012), arXiv:1201.3614.
- [112] C. Alcock and B. Paczynski, *Nature* **281**, 358 (1979).
- [113] A. G. Sánchez *et al.*, *Mon. Not. R. Astron. Soc.* **464**, 1640 (2017), arXiv:1607.03147.
- [114] C. Hernández-Aguayo, F. Prada, C. M. Baugh, and A. Klypin, (2020), arXiv:2006.00612.
- [115] C. Conroy, R. H. Wechsler, and A. V. Kravtsov, *Astrophys. J.* **647**, 201 (2006), arXiv:astro-ph/0512234.
- [116] R. M. Reddick, R. H. Wechsler, J. L. Tinker, and P. S. Behroozi, *The Astrophys. J.* **771**, 30 (2013).
- [117] B. P. Moster *et al.*, *The Astrophys. J.* **710**, 903 (2010).
- [118] J.-h. He, L. Guzzo, B. Li, and C. M. Baugh, *Nat. Astron.* **2**, 967 (2018), arXiv:1809.09019.
- [119] J. Schaye *et al.*, *Mon. Not. Roy. Astron. Soc.* **446**, 521 (2015), arXiv:1407.7040.
- [120] J. Chaves-Montero *et al.*, *Mon. Not. Roy. Astron. Soc.* **460**, 3100 (2016), arXiv:1507.01948.
- [121] J.-h. He, A. J. Hawken, B. Li, and L. Guzzo, *Phys. Rev. Lett.* **115**, 071306 (2015), arXiv:1501.00846.
- [122] J.-h. He, B. Li, and A. J. Hawken, *Phys. Rev. D* **92**, 103508 (2015), arXiv:1505.03656.
- [123] J.-H. He, B. Li, and C. M. Baugh, *Phys. Rev. Lett.* **117**, 221101 (2016), arXiv:1605.04709.
- [124] A. Klypin, G. Yepes, S. Gottlober, F. Prada, and S. Hess, *Mon. Not. Roy. Astron. Soc.* **457**, 4340 (2016), arXiv:1411.4001.

- [125] SDSS, M. A. Strauss *et al.*, *Astron. J.* **124**, 1810 (2002), arXiv:astro-ph/0206225.
- [126] J.-h. He, *Mon. Not. R. Astron. Soc.* **493**, 4453 (2020), arXiv:1905.01612.
- [127] Y.-C. Cai, A. Taylor, J. A. Peacock, and N. Padilla, *Mon. Not. R. Astron. Soc.* **462**, 2465 (2016), arXiv:1603.05184.
- [128] N. Hamaus *et al.*, *Phys. Rev. Lett.* **117**, 091302 (2016), arXiv:1602.01784.
- [129] L. Hui, A. Nicolis, and C. W. Stubbs, *Phys. Rev. D* **80**, 104002 (2009), arXiv:0905.2966.
- [130] J. Clampitt, Y.-C. Cai, and B. Li, *Mon. Not. R. Astron. Soc.* **431**, 749 (2013), arXiv:1212.2216.
- [131] T. Y. Lam, J. Clampitt, Y.-C. Cai, and B. Li, *Mon. Not. R. Astron. Soc.* **450**, 3319 (2015), arXiv:1408.5338.
- [132] Y.-C. Cai, N. Padilla, and B. Li, *Mon. Not. R. Astron. Soc.* **451**, 1036 (2015), arXiv:1410.1510.
- [133] M. C. Neyrinck, *Mon. Not. R. Astron. Soc.* **386**, 2101 (2008), arXiv:0712.3049.
- [134] S. Nadathur and W. J. Percival, *Mon. Not. R. Astron. Soc.* **483**, 3472 (2019), arXiv:1712.07575.
- [135] R. Scoccimarro, H. M. P. Couchman, and J. A. Frieman, *Astrophys. J.* **517**, 531 (1999), arXiv:astro-ph/9808305.
- [136] R. Scoccimarro, *Phys. Rev. D* **70**, 083007 (2004), astro-ph/0407214.
- [137] A. Taruya, T. Nishimichi, and S. Saito, *Phys. Rev. D* **82**, 063522 (2010), arXiv:1006.0699.
- [138] M. Lewandowski, L. Senatore, F. Prada, C. Zhao, and C.-H. Chuang, *Phys. Rev. D* **97**, 063526 (2018), arXiv:1512.06831.
- [139] L. F. de la Bella, D. Regan, D. Seery, and D. Parkinson, (2018), arXiv:1805.12394.
- [140] B. Bose, A. Pourtsidou, K. Marković, and F. Beutler, (2019), arXiv:1905.05122.
- [141] B. Bose, K. Koyama, and H. A. Winther, (2019), arXiv:1905.05135.
- [142] V. Desjacques, D. Jeong, and F. Schmidt, *Phys. Rept.* **733**, 1 (2018), arXiv:1611.09787.
- [143] B. A. Reid and M. White, *Mon. Not. Roy. Astron. Soc.* **417**, 1913 (2011), arXiv:1105.4165.
- [144] J. Carlson, B. Reid, and M. White, *Mon. Not. Roy. Astron. Soc.* **429**, 1674 (2013), arXiv:1209.0780.
- [145] D. Bianchi, M. Chiesa, and L. Guzzo, *Mon. Not. Roy. Astron. Soc.* **446**, 75 (2015), arXiv:1407.4753.
- [146] D. Bianchi, W. Percival, and J. Bel, *Mon. Not. Roy. Astron. Soc.* **463**, 3783 (2016), arXiv:1602.02780.
- [147] C. Uhlemann, M. Kopp, and T. Haugg, *Phys. Rev. D* **92**, 063004 (2015), arXiv:1503.08837.
- [148] J. Kuruvilla and C. Porciani, *Mon. Not. Roy. Astron. Soc.* **479**, 2256 (2018), arXiv:1710.09379.
- [149] C. Cuesta-Lazaro *et al.*, *Mon. Not. Roy. Astron. Soc.* **498**, 1175 (2020), arXiv:2002.02683.
- [150] Y.-S. Song, Y. Zheng, A. Taruya, and M. Oh, *J. Cosmo. Astropart. Phys.* **1807**, 018 (2018), arXiv:1801.04950.
- [151] Z. Zhai *et al.*, *Astrophys. J.* **874**, 95 (2019), arXiv:1804.05867.
- [152] Y. Kobayashi, T. Nishimichi, M. Takada, R. Takahashi, and K. Osato, *Phys. Rev. D* **102**, 063504 (2020), arXiv:2005.06122.
- [153] A. Taruya, K. Koyama, T. Hiramatsu, and A. Oka, *Phys. Rev. D* **89**, 043509 (2014), arXiv:1309.6783.
- [154] A. Taruya, T. Nishimichi, F. Bernardeau, T. Hiramatsu, and K. Koyama, *Phys. Rev. D* **90**, 123515 (2014), arXiv:1408.4232.
- [155] B. Bose and K. Koyama, *J. Cosmo. Astropart. Phys.* **1608**, 032 (2016), arXiv:1606.02520.
- [156] B. Bose and K. Koyama, *J. Cosmo. Astropart. Phys.* **1708**, 029 (2017), arXiv:1705.09181.
- [157] B. Bose, K. Koyama, M. Lewandowski, F. Vernizzi, and H. A. Winther, *J. Cosmo. Astropart. Phys.* **1804**, 063 (2018), arXiv:1802.01566.
- [158] G. Valogiannis, R. Bean, and A. Aviles, (2019), arXiv:1909.05261.
- [159] H. Winther *et al.*, *Phys. Rev. D* **100**, 123540 (2019), arXiv:1903.08798.
- [160] LSST Dark Energy Science, N. Ramachandra, G. Valogiannis, M. Ishak, and K. Heitmann, (2020), arXiv:2010.00596.
- [161] M. C. Neyrinck, I. Szapudi, and A. S. Szalay, *Astrophys. J., Letters* **698**, L90 (2009), arXiv:0903.4693.
- [162] X. Wang *et al.*, *Astrophys. J.* **735**, 32 (2011), arXiv:1103.2166.
- [163] J. Carron, *Phys. Rev. Lett.* **108**, 071301 (2012), arXiv:1201.1000.
- [164] F. Simpson, J. B. James, A. F. Heavens, and C. Heymans, *Phys. Rev. Lett.* **107**, 271301 (2011), arXiv:1107.5169.
- [165] F. Simpson, A. F. Heavens, and C. Heymans, *Phys. Rev. D* **88**, 083510 (2013), arXiv:1306.6349.
- [166] L. Lombriser, F. Simpson, and A. Mead, *Phys. Rev. Lett.* **114**, 251101 (2015), arXiv:1501.04961.
- [167] M. White, *J. Cosmo. Astropart. Phys.* **1611**, 057 (2016), arXiv:1609.08632.
- [168] G. Valogiannis and R. Bean, *Phys. Rev. D* **97**, 023535 (2018), arXiv:1708.05652.
- [169] J. Armijo, Y.-C. Cai, N. Padilla, B. Li, and J. A. Peacock, *Mon. Not. Roy. Astron. Soc.* **478**, 3627 (2018), arXiv:1801.08975.
- [170] C. Hernández-Aguayo, C. M. Baugh, and B. Li, *Mon. Not. Roy. Astron. Soc.* **479**, 4824 (2018), arXiv:1801.08880.
- [171] K. A. Olive and M. Pospelov, *Phys. Rev. D* **77**, 043524 (2008), arXiv:0709.3825.
- [172] K. Hinterbichler and J. Khoury, *Phys. Rev. Lett.* **104**, 231301 (2010).
- [173] J. Coupon *et al.*, *Astronomy and Astrophysics* **542**, A5 (2012), arXiv:1107.0616.
- [174] V. Springel, *Mon. Not. Roy. Astron. Soc.* **364**, 1105 (2005), arXiv:astro-ph/0505010.
- [175] C. Park *et al.*, *The Astrophys. J.* **658**, 898 (2007).
- [176] A. Cabre, V. Vikram, G.-B. Zhao, B. Jain, and K. Koyama, *J. Cosmo. Astropart. Phys.* **1207**, 034 (2012), arXiv:1204.6046.
- [177] A. Stark *et al.*, *Physical Review D* **93**, 084036 (2016), arXiv:1603.00056.
- [178] D. Shi, B. Li, and J. Han, *Mon. Not. Roy. Astron. Soc.* **469**, 705 (2017), arXiv:1702.03940.
- [179] L. Lombriser, K. Koyama, G.-B. Zhao, and B. Li, *Phys. Rev. D* **85**, 124054 (2012), arXiv:1203.5125.
- [180] A. Barreira *et al.*, *J. Cosmo. Astropart. Phys.* **1404**, 029 (2014), arXiv:1401.1497.
- [181] A. Barreira, B. Li, W. A. Hellwing, C. M. Baugh, and S. Pascoli, *J. Cosmo. Astropart. Phys.* **1409**, 031 (2014), arXiv:1408.1084.
- [182] M. A. Mitchell, C. Arnold, J.-h. He, and B. Li, *Mon. Not. Roy. Astron. Soc.* **487**, 1410 (2019), arXiv:1901.06392.
- [183] S. Cole and C. G. Lacey, *Mon. Not. Roy. Astron. Soc.* **281**, 716 (1996), arXiv:astro-ph/9510147.
- [184] E. L. Łokas and G. A. Mamon, *Mon. Not. Roy. Astron. Soc.* **321**, 155 (2001), arXiv:astro-ph/0002395.
- [185] N. A. Henden, E. Puchwein, and D. Sijacki, *Mon. Not. Roy. Astron. Soc.* **489**, 2439 (2019), arXiv:1905.00013.
- [186] C. Bradshaw, A. Leauthaud, A. Hearin, S. Huang, and P. Behroozi, *Mon. Not. Roy. Astron. Soc.* **493**, 337 (2020), arXiv:1905.09353.
- [187] X. Yang *et al.*, *Astrophys. J.* **671**, 153 (2007), arXiv:0707.4640.
- [188] S. H. Lim, H. J. Mo, Y. Lu, H. Wang, and X. Yang, *Mon. Not. R. Astro. Soc.* **470**, 2982 (2017), arXiv:1706.02307.

- [189] K. Wang, H. Mo, C. Li, J. Meng, and Y. Chen, (2020), arXiv:2006.05426.
- [190] A. Aviles and J. L. Cervantes-Cota, Phys. Rev. D **96**, 123526 (2017), arXiv:1705.10719.
- [191] A. Aviles, M. A. Rodriguez-Meza, J. De-Santiago, and J. L. Cervantes-Cota, J. Cosmo. Astropart. Phys. **1811**, 013 (2018), arXiv:1809.07713.
- [192] G. Valogiannis and R. Bean, Phys. Rev. D **99**, 063526 (2019), arXiv:1901.03763.
- [193] T. Matsubara, Phys. Rev. D **78**, 083519 (2008), arXiv:0807.1733, [Erratum: Phys. Rev. D **78**, 109901 (2008)].
- [194] A. Aviles, Phys. Rev. D **98**, 083541 (2018), arXiv:1805.05304.
- [195] A. Aviles, K. Koyama, J. L. Cervantes-Cota, H. A. Winther, and B. Li, J. Cosmo. Astropart. Phys. **2001**, 006 (2020), arXiv:1911.06362.
- [196] S. Alam, Y. Zu, J. A. Peacock, and R. Mandelbaum, Mon. Not. R. Astron. Soc. **483**, 4501 (2019), arXiv:1801.04878.
- [197] F. Bernardeau, S. Colombi, E. Gaztanaga, and R. Scoccimarro, Phys. Rept. **367**, 1 (2002), arXiv:astro-ph/0112551.
- [198] A. Borisov and B. Jain, Phys. Rev. D **79**, 103506 (2009), arXiv:0812.0013.
- [199] B. Bose, J. Byun, F. Lacasa, A. Moradinezhad Dizgah, and L. Lombriser, J. Cosmo. Astropart. Phys. **02**, 025 (2020), arXiv:1909.02504.
- [200] H. Guo *et al.*, Mon. Not. Roy. Astron. Soc. **449**, L95 (2015), arXiv:1409.7389.
- [201] C. G. Sabiu, D. F. Mota, C. Llinares, and C. Park, Astron. Astrophys. **592**, A38 (2016), arXiv:1603.05750.
- [202] K. Hoffmann, E. Gaztañaga, R. Scoccimarro, and M. Crocce, Mon. Not. R. Astron. Soc. **476**, 814 (2018), arXiv:1708.08941.
- [203] A. Burden, N. Padmanabhan, R. N. Cahn, M. J. White, and L. Samushia, J. Cosmo. Astropart. Phys. **1703**, 001 (2017), arXiv:1611.04635.
- [204] H. Guo, I. Zehavi, and Z. Zheng, Astrophys. J. **756**, 127 (2012), arXiv:1111.6598.
- [205] D. Bianchi *et al.*, (2018), arXiv:1805.00951.
- [206] A. Smith *et al.*, Mon. Not. Roy. Astron. Soc. **484**, 1285 (2019), arXiv:1809.07355.
- [207] A. W. Moore *et al.*, Fast Algorithms and Efficient Statistics: N-Point Correlation Functions, in *Mining the Sky*, edited by A. J. Banday, S. Zaroubi, and M. Bartelmann, p. 71, 2001, astro-ph/0012333.
- [208] G. Chen and I. Szapudi, Astrophys. J. **635**, 743 (2005), astro-ph/0508316.
- [209] R. Scoccimarro, Phys. Rev. D **92**, 083532 (2015), arXiv:1506.02729.
- [210] Z. Slepian and D. J. Eisenstein, Mon. Not. R. Astron. Soc. **454**, 4142 (2015), arXiv:1506.02040.
- [211] Z. Slepian and D. J. Eisenstein, Mon. Not. R. Astron. Soc. **455**, L31 (2016), arXiv:1506.04746.
- [212] C. G. Sabiu, B. Hoyle, J. Kim, and X.-D. Li, Astrophys. J. **242**, 29 (2019), arXiv:1901.00296.
- [213] C. Sabiu, B. Hoyle, J. Kim, and X.-D. Li, Gramsci: GRAPh Made Statistics for Cosmological Information, 2019, ascl:1908.004.
- [214] F. Sosa Nuñez and G. Niz, (2020), arXiv:2006.05434.
- [215] I. Szapudi, Astrophys. J. **605**, L89 (2004), arXiv:astro-ph/0404476.
- [216] Z. Slepian and D. J. Eisenstein, Mon. Not. R. Astron. Soc. **469**, 2059 (2017), arXiv:1607.03109.
- [217] N. Hand *et al.*, Astron. J. **156**, 160 (2018), arXiv:1712.05834.
- [218] I. Szapudi and A. S. Szalay, Astrophys. J. Lett. **494**, L41 (1998).
- [219] M. White, J. L. Tinker, and C. K. McBride, Mon. Not. Roy. Astron. Soc. **437**, 2594 (2014), arXiv:1309.5532.
- [220] B. Friesen *et al.*, (2017), arXiv:1709.00086.
- [221] Z. Slepian and D. J. Eisenstein, (2017), arXiv:1709.10150.
- [222] S. Yuan, D. J. Eisenstein, and L. H. Garrison, Mon. Not. Roy. Astron. Soc. **478**, 2019 (2018), arXiv:1802.10115.
- [223] I. Hashimoto, Y. Rasera, and A. Taruya, Phys. Rev. D **96**, 043526 (2017), arXiv:1705.02574.
- [224] V. Desjacques, D. Jeong, and F. Schmidt, J. Cosmo. Astropart. Phys. **2018**, 035 (2018), arXiv:1806.04015.
- [225] B. Bose and A. Taruya, J. Cosmo. Astropart. Phys. **2018**, 019 (2018), arXiv:1808.01120.
- [226] A. Eggemeier, R. Scoccimarro, and R. E. Smith, Phys. Rev. D **99**, 123514 (2019), arXiv:1812.03208.
- [227] E. Sefusatti, M. Crocce, R. Scoccimarro, and H. M. P. Couchman, Mon. Not. R. Astron. Soc. **460**, 3624 (2016), arXiv:1512.07295.
- [228] B. Bose, K. Koyama, W. A. Hellwing, G.-B. Zhao, and H. A. Winther, Phys. Rev. D **96**, 023519 (2017), arXiv:1702.02348.
- [229] B. C., E. A., C. T. Davies, and B. Li, J. Cosmo. Astropart. Phys. **submitted** (2020), arXiv:2010.xxxxx.
- [230] P. J. E. Peebles, *The large-scale structure of the universe* (Research supported by the National Science Foundation. Princeton, N.J., Princeton University Press, 1980. 435 p., 1980).
- [231] J. N. Fry, Astrophys. J. **279**, 499 (1984).
- [232] F. Bernardeau, Astrophys. J. **392**, 1 (1992).
- [233] F. Bernardeau, Astrophys. J. **433**, 1 (1994), astro-ph/9312026.
- [234] R. Juszkiewicz, F. R. Bouchet, and S. Colombi, Astrophys. J. **412**, L9 (1993), arXiv:astro-ph/9306003.
- [235] E. L. Łokas, R. Juszkiewicz, D. H. Weinberg, and F. R. Bouchet, Mon. Not. R. Astro. Soc. **274**, 730 (1995), arXiv:astro-ph/9407095.
- [236] F. Bernardeau, Astron. Astrophys. **291**, 697 (1994), arXiv:astro-ph/9403020.
- [237] R. Juszkiewicz, W. A. Hellwing, and R. van de Weygaert, Mon. Not. R. Astro. Soc. **429**, 1206 (2013), arXiv:1205.6163.
- [238] X. Luo and D. N. Schramm, Astrophys. J. **408**, 33 (1993).
- [239] M. J. Chodorowski and F. R. Bouchet, Mon. Not. R. Astron. Soc. **279**, 557 (1996), astro-ph/9507038.
- [240] R. Durrer, R. Juszkiewicz, M. Kunz, and J.-P. Uzan, Phys. Rev. D **62**, 021301 (2000), astro-ph/0005087.
- [241] E. Gaztañaga and J. A. Lobo, Astrophys. J. **548**, 47 (2001), astro-ph/0003129.
- [242] W. A. Hellwing, R. Juszkiewicz, and R. van de Weygaert, Phys. Rev. D **82**, 103536 (2010), arXiv:1008.3930.
- [243] W. A. Hellwing, B. Li, C. S. Frenk, and S. Cole, Mon. Not. R. Astro. Soc. **435**, 2806 (2013), arXiv:1305.7486.
- [244] W. A. Hellwing, K. Koyama, B. Bose, and G.-B. Zhao, Phys. Rev. D **96**, 023515 (2017), arXiv:1703.03395.
- [245] E. Gaztanaga, Mon. Not. R. Astron. Soc. **268**, 913 (1994), astro-ph/9309019.
- [246] S. Colombi, F. R. Bouchet, and R. Schaeffer, Astron. Astrophys. **281**, 301 (1994).
- [247] A. V. Kravtsov *et al.*, Astrophys. J. **609**, 35 (2004), astro-ph/0308519.
- [248] X. Yang, H. J. Mo, Y. P. Jing, and F. C. van den Bosch, Mon. Not. R. Astron. Soc. **358**, 217 (2005), astro-ph/0410114.

- [249] X. Yang, H. J. Mo, and F. C. van den Bosch, *Astrophys. J.* **676**, 248 (2008), arXiv:0710.5096.
- [250] T. Guillet, R. Teyssier, and S. Colombi, *Mon. Not. R. Astro. Soc.* **405**, 525 (2010), arXiv:0905.2615.
- [251] M. P. van Daalen, J. Schaye, I. G. McCarthy, C. M. Booth, and C. Dalla Vecchia, *Mon. Not. R. Astro. Soc.* **440**, 2997 (2014), arXiv:1310.7571.
- [252] W. A. Hellwing *et al.*, *Mon. Not. R. Astro. Soc.* **461**, L11 (2016), arXiv:1603.03328.
- [253] E. Hivon, F. R. Bouchet, S. Colombi, and R. Juszkiewicz, *Astro. Astrophys.* **298**, 643 (1995), astro-ph/9407049.
- [254] A. Cappi *et al.*, *Astron. Astrophys.* **579**, A70 (2015), arXiv:1505.05347.
- [255] S. D. M. White, *Mon. Not. Roy. Astron. Soc.* **186**, 145 (1979).
- [256] J. N. Fry, *Astrophys. J.* **306**, 358 (1986).
- [257] K. R. Mecke, T. Buchert, and H. Wagner, *Astron. Astrophys.* **288**, 697 (1994), arXiv:astro-ph/9312028.
- [258] SDSS, C. Hikage *et al.*, *Publ. Astron. Soc. Jap.* **55**, 911 (2003), arXiv:astro-ph/0304455.
- [259] C. Blake, J. B. James, and G. B. Poole, *Mon. Not. Roy. Astron. Soc.* **437**, 2488 (2014), arXiv:1310.6810.
- [260] A. Wiegand, T. Buchert, and M. Ostermann, *Mon. Not. Roy. Astron. Soc.* **443**, 241 (2014), arXiv:1311.3661.
- [261] J. Schmalzing and K. M. Gorski, *Mon. Not. Roy. Astron. Soc.* **297**, 355 (1998), arXiv:astro-ph/9710185.
- [262] C. Hikage *et al.*, *Mon. Not. Roy. Astron. Soc.* **389**, 1439 (2008), arXiv:0802.3677.
- [263] W. Fang, A. Becker, D. Huterer, and E. Lim, *Phys. Rev. D* **88**, 041302 (2013), arXiv:1303.4381.
- [264] Planck, P. A. R. Ade *et al.*, *Astron. Astrophys.* **594**, A17 (2016), arXiv:1502.01592.
- [265] C. Ling *et al.*, *Phys. Rev. D* **92**, 064024 (2015), arXiv:1410.2734.
- [266] D. Munshi, B. Hu, T. Matsubara, P. Coles, and A. Heavens, *J. Cosmo. Astropart. Phys.* **1604**, 056 (2016), arXiv:1602.00965.
- [267] W. Fang, B. Li, and G.-B. Zhao, *Phys. Rev. Lett.* **118**, 181301 (2017), arXiv:1704.02325.
- [268] J. Schmalzing and T. Buchert, *Astrophys. J. Lett.* **482**, L1 (1997), arXiv:astro-ph/9702130.
- [269] I. Gott, J. R., M. Dickinson, and A. Melott, *Astrophys. J.* **306**, 341 (1986).
- [270] P. Coles, A. G. Davies, and R. C. Pearson, *Mon. Not. R. Astron. Soc.* **281**, 1375 (1996), arXiv:astro-ph/9603139.
- [271] T. Matsubara, (2000), arXiv:astro-ph/0006269.
- [272] R. Nandra, A. N. Lasenby, and M. P. Hobson, *Mon. Not. R. Astro. Soc.* **422**, 2945 (2012), arXiv:1104.4458.
- [273] J. Einasto, *Trudy Astrofizicheskogo Instituta Alma-Ata* **5**, 87 (1965).
- [274] P. S. Behroozi, A. Loeb, and R. H. Wechsler, *J. Cosmo. Astropart. Phys.* **2013**, 019 (2013), arXiv:1208.0334.
- [275] C. J. Miller, A. Stark, D. Gifford, and N. Kern, *Astrophys. J.* **822**, 41 (2016), arXiv:1612.05565.
- [276] V. Halenka, C. J. Miller, and P. Vansickle, arXiv e-prints, arXiv:2003.02733 (2020), arXiv:2003.02733.
- [277] L. Lombriser, K. Koyama, G.-B. Zhao, and B. Li, *Physical Review D* **85**, 124054 (2012), arXiv:1203.5125.
- [278] A. Stark, C. J. Miller, and D. Huterer, *Phys. Rev. D* **96**, 023543 (2017).
- [279] B. M. B. Henriques *et al.*, *Monthly Notices of the Royal Astronomical Society* **421**, 2904 (2012), arXiv:1109.3457.
- [280] D. E. Applegate *et al.*, *Mon. Not. R. Astron. Soc.* **439**, 48 (2014), arXiv:1208.0605.
- [281] E. P. Verlinde, *SciPost Phys.* **2**, 016 (2017), arXiv:1611.02269.
- [282] V. Halenka and C. J. Miller, *Phys. Rev. D* **102**, 084007 (2020), arXiv:1807.01689.
- [283] S. Ettori *et al.*, *Astron. Astrophys.* **621**, A39 (2019), arXiv:1805.00035.
- [284] A. Tamosiunas, D. Bacon, K. Koyama, and R. C. Nichol, *J. Cosmo. Astropart. Phys.* **2019**, 053 (2019), arXiv:1901.05505.
- [285] R. Mandelbaum *et al.*, *Mon. Not. Roy. Astron. Soc.* **372**, 758 (2006), arXiv:astro-ph/0605476.
- [286] E. van Uitert *et al.*, *Mon. Not. R. Astron. Soc.* **459**, 3251 (2016), arXiv:1601.067911710.01730.
- [287] SDSS, E. S. Sheldon *et al.*, *Astron. J.* **127**, 2544 (2004), arXiv:astro-ph/0312036.
- [288] A. Dvornik *et al.*, *Mon. Not. R. Astron. Soc.* **479**, 1240 (2018), arXiv:1802.00734.
- [289] M. Cacciato *et al.*, *Mon. Not. Roy. Astron. Soc.* **394**, 929 (2009), arXiv:0807.4932.
- [290] B. R. Gillis *et al.*, *Mon. Not. Roy. Astron. Soc.* **431**, 1439 (2013), arXiv:1301.7421.
- [291] M. Velander *et al.*, *Mon. Not. Roy. Astron. Soc.* **437**, 2111 (2014), arXiv:1304.4265.
- [292] M. J. Hudson *et al.*, *Mon. Not. Roy. Astron. Soc.* **447**, 298 (2015), arXiv:1310.6784.
- [293] C. Sifón *et al.*, *Mon. Not. Roy. Astron. Soc.* **454**, 3938 (2015), arXiv:1507.00737.
- [294] KiDS, M. Viola *et al.*, *Mon. Not. Roy. Astron. Soc.* **452**, 3529 (2015), arXiv:1507.00735.
- [295] A. Dvornik *et al.*, *Mon. Not. R. Astron. Soc.* **468**, 3251 (2017), arXiv:1703.06657.
- [296] DES, J. Clampitt *et al.*, *Mon. Not. Roy. Astron. Soc.* **465**, 4204 (2017), arXiv:1603.05790.
- [297] Y. Park and M. Wyman, *Phys. Rev. D* **91**, 064012 (2015), arXiv:1408.4773.
- [298] A. Leauthaud *et al.*, *Mon. Not. Roy. Astron. Soc.* **467**, 3024 (2017), arXiv:1611.08606.
- [299] B. Li and M. Shirasaki, in prep. (2017).
- [300] J. Miralda-Escude, *Astrophys. J.* **370**, 1 (1991).
- [301] G. Wilson, N. Kaiser, G. A. Luppino, and L. L. Cowie, *Astrophys. J.* **555**, 572 (2001), arXiv:astro-ph/0008504.
- [302] D. Jeong, E. Komatsu, and B. Jain, *Phys. Rev. D* **80**, 123527 (2009).
- [303] R. Mandelbaum, *Ann. Rev. Astron. Astrophys.* **56**, 393 (2018), arXiv:1710.03235.
- [304] B. Li, *Mon. Not. Roy. Astron. Soc.* **411**, 2615 (2011), arXiv:1009.1406.
- [305] Y.-C. Cai, N. Padilla, and B. Li, *Mon. Not. R. Astro. Soc.* **451**, 1036 (2015), arXiv:1410.1510.
- [306] M. Cautun, Y.-C. Cai, and C. S. Frenk, *Mon. Not. R. Astro. Soc.* **457**, 2540 (2016), arXiv:1509.00010.
- [307] B. Falck, K. Koyama, G.-B. Zhao, and M. Cautun, *Mon. Not. R. Astron. Soc.* **475**, 3262 (2018), arXiv:1704.08942.
- [308] E. Paillas *et al.*, *Mon. Not. R. Astron. Soc.* **484**, 1149 (2019), arXiv:1810.02864.
- [309] C. T. Davies, M. Cautun, and B. Li, *Mon. Not. R. Astron. Soc.* **490**, 4907 (2019), arXiv:1907.06657.
- [310] J. Clampitt and B. Jain, *Mon. Not. R. Astro. Soc.* **454**, 3357 (2015), arXiv:1404.1834.
- [311] D. Gruen *et al.*, *Mon. Not. R. Astro. Soc.* **455**, 3367 (2016), arXiv:1507.05090.

- [312] DES, C. Sanchez *et al.*, Mon. Not. R. Astro. Soc. **465**, 746 (2017), arXiv:1605.03982.
- [313] E. Paillas *et al.*, Mon. Not. R. Astro. Soc. **470**, 4434 (2017), arXiv:1609.00101.
- [314] E. Platen, R. van de Weygaert, and B. J. T. Jones, Mon. Not. R. Astro. Soc. **380**, 551 (2007), arXiv:0706.2788.
- [315] W. E. Schaap and R. van de Weygaert, Astronomy and Astrophysics **363**, L29 (2000), arXiv:astro-ph/0011007.
- [316] M. C. Cautun and R. van de Weygaert, ArXiv e-prints (2011), arXiv:1105.0370.
- [317] M. Cautun, R. van de Weygaert, and B. J. T. Jones, Mon. Not. R. Astro. Soc. **429**, 1286 (2013), arXiv:1209.2043.
- [318] M. Cautun, R. van de Weygaert, B. J. T. Jones, and C. S. Frenk, Mon. Not. R. Astro. Soc. **441**, 2923 (2014), arXiv:1401.7866.
- [319] C. T. Davies, M. Cautun, and B. Li, Mon. Not. R. Astron. Soc. **480**, L101 (2018), arXiv:1803.08717.
- [320] C. T. Davies, E. Paillas, M. Cautun, and B. Li, arXiv e-prints, arXiv:2004.11387 (2020), arXiv:2004.11387.
- [321] W. A. e. a. Hellwing, *in prep.* (2020).
- [322] E. Munari *et al.*, J. Cosmo. Astropart. Phys. **1707**, 050 (2017), arXiv:1704.00920.
- [323] F.-S. Kitaura *et al.*, Mon. Not. Roy. Astron. Soc. **456**, 4156 (2016), arXiv:1509.06400.
- [324] A. Izard, M. Crocce, and P. Fosalba, Mon. Not. Roy. Astron. Soc. **459**, 2327 (2016), arXiv:1509.04685.
- [325] I. Zehavi *et al.*, Astrophys. J. **736**, 59 (2011), arXiv:1005.2413.
- [326] A. Smith *et al.*, Mon. Not. Roy. Astron. Soc. **470**, 4646 (2017), arXiv:1701.06581.
- [327] T. McClintock *et al.*, Astrophys. J. **872**, 53 (2019), arXiv:1804.05866.
- [328] T. Nishimichi *et al.*, Astrophys. J. **884**, 29 (2019), arXiv:1811.09504.
- [329] C. E. Rasmussen and C. K. I. Williams, *Gaussian processes for machine learning*. Adaptive computation and machine learning (MIT Press, 2006).
- [330] D. Bianchi and W. J. Percival, Mon. Not. Roy. Astron. Soc. **472**, 1106 (2017), arXiv:1703.02070.