

## The constraining power of the marked power spectrum: an analytical study

Marco Marinucci <sup>a,b</sup> Gabriel Jung <sup>c</sup> Michele Liguori <sup>a,b</sup> Andrea Ravenni <sup>a,b</sup>  
Francesco Spezzati <sup>a,b</sup> Adam Andrews <sup>d,e</sup> Marco Baldi <sup>d,e,f</sup>  
William R. Coulton <sup>g,h</sup> Dionysios Karagiannis <sup>i,l,m</sup> Francisco Villaescusa-Navarro <sup>n</sup>  
and Benjamin D. Wandelt <sup>o,p,q,n</sup>

<sup>a</sup>Dipartimento di Fisica e Astronomia “G. Galilei”, Università degli Studi di Padova,  
via Marzolo 8, I-35131, Padova, Italy

<sup>b</sup>INFN, Sezione di Padova, via Marzolo 8, I-35131, Padova, Italy

<sup>c</sup>Université Paris-Saclay, CNRS, Institut d’Astrophysique Spatiale, 91405 Orsay, France

<sup>d</sup>INAF/OAS Bologna, via Piero Gobetti 101, I-40129 Bologna, Italy

<sup>e</sup>INFN, Sezione di Bologna, via Irnerio 46, I-40126 Bologna, Italy

<sup>f</sup>Dipartimento di Fisica e Astronomia, Alma Mater Studiorum — University of Bologna,  
Via Piero Gobetti 93/2, 40129 Bologna BO, Italy

<sup>g</sup>Kavli Institute for Cosmology Cambridge, Madingley Road, Cambridge CB3 0HA, U.K.

<sup>h</sup>DAMTP, Centre for Mathematical Sciences, University of Cambridge,  
Wilberforce Road, Cambridge CB3 0WA, U.K.

<sup>i</sup>Dipartimento di Fisica e Scienze della Terra, Università degli Studi di Ferrara,  
via Giuseppe Saragat 1, 44122 Ferrara, Italy

<sup>l</sup>INFN, Sezione di Ferrara, via Giuseppe Saragat 1, 44122 Ferrara, Italy

<sup>m</sup>Department of Physics & Astronomy, University of the Western Cape,  
Cape Town 7535, South Africa

<sup>n</sup>Center for Computational Astrophysics, 160 5th Avenue, New York, NY, 10010, U.S.A.

<sup>o</sup>Department of Physics and Astronomy, Johns Hopkins University,  
3400 North Charles Street, Baltimore, MD, 21218, U.S.A.

<sup>p</sup>Department of Applied Mathematics and Statistics, Johns Hopkins University,  
3400 North Charles Street, Baltimore, MD, 21218, U.S.A.

<sup>q</sup>CNRS & Sorbonne Université, Institut d’Astrophysique de Paris (IAP), UMR 7095, 98 bis bd  
Arago, F-75014 Paris, France

E-mail: [marco.marinucci@unipd.it](mailto:marco.marinucci@unipd.it), [gabriel.jung@protonmail.com](mailto:gabriel.jung@protonmail.com),  
[michele.liguori@unipd.it](mailto:michele.liguori@unipd.it), [andrea.ravenni@unipd.it](mailto:andrea.ravenni@unipd.it),  
[francesco.spezzati@studenti.unipd.it](mailto:francesco.spezzati@studenti.unipd.it), [adam.andrews@inaf.it](mailto:adam.andrews@inaf.it),  
[marco.baldi5@unibo.it](mailto:marco.baldi5@unibo.it), [william.coulton@gmail.com](mailto:william.coulton@gmail.com), [dakaragian@gmail.com](mailto:dakaragian@gmail.com),  
[villaescusa.francisco@gmail.com](mailto:villaescusa.francisco@gmail.com), [benwandelt@gmail.com](mailto:benwandelt@gmail.com)

**ABSTRACT:** The marked power spectrum — a two-point correlation function of a weighted density field — has emerged as a promising tool for extracting cosmological information from the large-scale structure of the Universe. In this work, we present the first comprehensive analytical study of the marked power spectrum’s sensitivity to primordial non-Gaussianity (PNG) of the non-local type. We extend previous effective field theory frameworks to incorporate PNG, developing a complete theoretical model that we validate against the Quijote simulation suite. Through a systematic Fisher analysis, we compare the constraining power of the marked power spectrum against traditional approaches combining the power spectrum and bispectrum (P+B). We explore different choices of mark parameters to evaluate their impact on parameter constraints, particularly focusing on equilateral and orthogonal PNG as well as neutrino masses. Our analysis shows that while marking up underdense regions yields optimal constraints in the low shot-noise regime, the marked power spectrum’s performance for discrete tracers with BOSS-like number densities does not surpass that of P+B analysis at mildly non-linear scales ( $k \lesssim 0.25 h/\text{Mpc}$ ). However, the marked approach offers several practical advantages, including simpler estimation procedures and potentially more manageable systematic effects. Our theoretical framework reveals how the marked power spectrum incorporates higher-order correlation information through terms resembling tree-level bispectra and power spectrum convolutions. This work establishes a robust foundation for applying marked statistics to future large-volume surveys.

**KEYWORDS:** cosmic web, cosmological perturbation theory, inflation, cosmological parameters from LSS

**ARXIV EPRINT:** [2411.14377](https://arxiv.org/abs/2411.14377)

---

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Theory modeling for the marked spectrum</b>	<b>3</b>
2.1	Eulerian perturbation theory	3
2.2	Counterterms, shot noise and IR-resummation	7
2.3	Primordial non-gaussianities of the non-local type	8
2.4	Low- $k$ behavior	11
<b>3</b>	<b>Results</b>	<b>12</b>
3.1	Comparison to simulations	12
3.2	Analysis specifics	14
3.3	Dark matter	16
3.4	Galaxies	18
<b>4</b>	<b>Discussion and conclusion</b>	<b>22</b>
<b>A</b>	<b>Effect of mark parameters</b>	<b>24</b>

---

## 1 Introduction

Galaxy surveys have made remarkable progress in recent years, enabling the study of the large-scale structure (LSS) of the Universe with unprecedented precision. Notably, current and forthcoming LSS observations will give us access to a significant amount of new cosmological information from non-linear scales in the galaxy density field. Extracting this information is however challenging, as it involves either the study of beyond power spectrum statistics or full field-level analysis. To address this, the most direct approach consists in measuring higher order correlators of the field, starting with the bispectrum (three-point function in Fourier space). This approach has been adopted in many studies, e.g., [1–13]. Its primary problem is however that the bispectrum remains a nearly lossless statistic only at mildly non-linear scales; extending it to higher orders moments, on the other hand, presents significant numerical and computational challenges, especially when accounting for realistic data analysis and observational issues, such as the necessity to evaluate covariances and include window functions.

For this reason, there has been increasing interest in the study of alternative summary statistics that can efficiently compress information even in the strongly non-linear regime while remaining computationally manageable. A partial list of these summaries includes methods such as the Wavelet Scattering Transform [14–19],  $k$ -nearest neighbors [20, 21], skew-spectra [22, 23], clipping [24] and Minkowski functionals [25, 26]. So far, these summaries have been either explored through Fisher analysis or used as inputs for likelihood-free inference.

In this work, we focus on another promising statistic that has received significant attention in the literature: the marked power spectrum. This is defined as the two-point correlator

of a weighted version of the density field, where the weight (“mark”) can represent galaxy properties, halo merger history or — as considered in this work — the average density of the surrounding environment.

The marked power spectrum was first introduced in [27]. In the context of Cosmology, it was recently pointed out in [28] that a density-weighted mark could be a powerful probe of Modified Gravity. Its effectiveness at constraining many cosmological parameters — and notably neutrino masses — was shown by the authors of [29, 30] in a simulation-based study. Its sensitivity to primordial non-Gaussianity was instead recently pointed out in [31].

Previous works [28, 30, 32–34] emphasized two main factors that explain why the marked power spectrum is so efficient at extracting cosmological information, compared to the standard power spectrum. The former is that, by correlating the field with the environment, the marked power spectrum is effectively incorporating information from higher-order correlation functions. The latter is that the mark can be chosen to up-weight low-density regions. These carry significant information but, at the same time, give very little contribution to the conventional power spectrum.

The primary aim of this work is to shed further insight into these issues by significantly extending previous analytical studies of the marked power spectrum in several directions.

First, we extend the EFTofLSS-based formalism introduced by [32], developing for the first time a theoretical framework that incorporates primordial non-Gaussianity (PNG) into the perturbative description of the marked power spectrum. Additionally, we conduct a Fisher matrix analysis to give an analytical quantitative assessment of the marked power spectrum’s sensitivity to various parameters, including PNG and neutrino masses; in this assessment, we consider different choices of the density-weight parameters of the mark in order to verify the impact of under/over-weighting low-density regions. Finally, we make a detailed comparison between the constraining power of the marked power spectrum and that of a combination of the conventional power spectrum and bispectrum ( $P + B$ ). In previous works, the marked 2-point statistic was typically compared to the standard power spectrum, but  $P + B$  looks like a more natural benchmark, considering that the marked field inherently includes higher-order information, as noted earlier. This will be made more transparent in our analytical model, which displays new terms in the one-loop power spectrum of the marked fields that are very similar to the tree-level bispectrum and to the convolution of linear power spectra.

In the current study, we focus specifically on the form of the mark introduced in [28], which is widely used in cosmological studies on this topic. Recently, [35] highlighted that this ansatz can be optimized in a parameter-dependent manner to enhance estimation efficiency. We leave this optimization — especially in the context of PNG analysis, which is a primary focus of our study — for future work. Additionally, we restrict our analysis to biased tracers in real space, with an extension to redshift space also planned for a forthcoming publication.

The paper is structured as follows. In section 2, we present our theoretical modeling of the marked power spectrum, developing the perturbative framework and including the effects of primordial non-Gaussianities. We discuss the necessary counterterms, shot noise contributions, and IR-resummation techniques required for a complete theoretical description. Section 3 contains our main results, beginning with a detailed comparison of our theoretical predictions against the Quijote simulations. We then present Fisher forecasts comparing the

constraining power of the marked power spectrum to standard P+B analyses, examining both the dark matter field and biased tracers. We explore how different choices of mark parameters affect constraints on neutrino masses and non-local PNG, and investigate the role of bias parameters in degrading these constraints. In section 4, we discuss the implications of our findings, comparing with previous work and highlighting the practical advantages and limitations of the marked power spectrum approach. We also outline several promising directions for future work, including the extension to redshift space and the possibility of optimizing the mark for specific cosmological parameters.

## 2 Theory modeling for the marked spectrum

The standard statistic employed in Large Scale Structure analysis is the power spectrum of the galaxy number density perturbation<sup>1</sup>

$$n_g(\mathbf{x}) \equiv \bar{n}_g(1 + \delta_g(\mathbf{x})), \quad (2.1)$$

where  $\bar{n}_g \equiv \langle n_g(\mathbf{x}) \rangle$  is the mean galaxy density and  $\delta_g(\mathbf{x})$  is the usual overdensity field. In this work, we are interested in studying an alternative summary statistic to extract more information from the galaxy clustering, the *Marked Power Spectrum* (see refs. [31, 36] for other examples of summary statistics), in which the galaxy number density of eq. (2.1) is weighted by a (space-dependent) mark

$$\rho_M(\mathbf{x}) = m(\mathbf{x})n_g(\mathbf{x}) = m(\mathbf{x})\bar{n}[1 + \delta_g(\mathbf{x})]. \quad (2.2)$$

The mark function usually adopted in literature is [28, 29, 32, 33, 37]

$$m(\mathbf{x}) = \left( \frac{1 + \delta_s}{1 + \delta_s + \delta_{g,R}(\mathbf{x})} \right)^p, \quad (2.3)$$

which is a function of the smoothed galaxy overdensity field,  $\delta_{g,R}(\mathbf{x})$ ,<sup>2</sup> and  $R$ ,  $\delta_s$ , and  $p$  are hyper-parameters of the model that can be used to enhance underdense ( $p > 0$ ) and overdense ( $p < 0$ ) regions. Marking the overdensity field with this non-linear function of the smoothed field will make the marked density field more non-Gaussian, allowing us to access information on the large-scale structure distribution that is usually included in higher-order correlation functions.

### 2.1 Eulerian perturbation theory

By defining the mean mark  $\bar{m}$  as

$$\langle \rho_M(\mathbf{x}) \rangle = \langle m(\mathbf{x})n_g(\mathbf{x}) \rangle \equiv \bar{m}\bar{n} \quad (2.4)$$

<sup>1</sup>Throughout this work, we will always refer to galaxies, but the present approach refers to any biased tracer of the dark matter density field.

<sup>2</sup>We will use a Gaussian smoothing function (in Fourier space) which reads

$$W_R(k) = e^{-\frac{k^2 R^2}{2}}, \quad \delta_{g,R}(\mathbf{k}) = W_R(k)\delta_g(\mathbf{k}).$$

we can consider the marked overdensity field

$$\delta_M(\mathbf{x}) \equiv \frac{\rho_M(\mathbf{x}) - \langle \rho_M \rangle}{\langle \rho_M \rangle} = \frac{m(\mathbf{x})}{\bar{m}} [1 + \delta_g(\mathbf{x})] - 1, \quad (2.5)$$

and expand it perturbatively in the non-linear fields  $\delta_g$  and  $\delta_{g,R}$ . Expanding the mark, we first obtain

$$\begin{aligned} m(\mathbf{x}) &= \left( \frac{1 + \delta_s}{1 + \delta_s + \delta_{g,R}} \right)^p = \sum_{n=0}^{\infty} (-1)^n \frac{p(p+1) \dots (p+n-1)}{n!} \frac{\delta_{g,R}^n}{(1 + \delta_s)^n} \\ &\equiv \sum_{n=0}^{\infty} (-1)^n C_n \delta_{g,R}^n(\mathbf{x}), \end{aligned} \quad (2.6)$$

with

$$C_n \equiv \frac{p(p+1) \dots (p+n-1)}{n!} \frac{1}{(1 + \delta_s)^n}. \quad (2.7)$$

For biased tracers, such as galaxies, we need to be careful about the convergence of the series in (2.6). Taking a large enough  $R$ , the time evolution of the smoothed field is  $\delta_{g,R}(z) \simeq b_1(z) D(z) \delta_0$ , with  $b_1$  and  $D$  being the linear bias and the growth factor; the convergence of the series is guaranteed simply by  $|\delta_{g,R}(\mathbf{x})/(1 + \delta_s)| < 1$ , and, noting that the fluctuation of the smoothed field is  $\sigma_{RR}^2(z) \equiv \langle \delta_{m,R}^2(\mathbf{x}) \rangle$ , we expect convergence if

$$\sigma_{RR}(0) < \frac{1 + \delta_s}{b_1(z) D(z)}. \quad (2.8)$$

We can proceed by expanding perturbatively as

$$\delta_g(\mathbf{x}) = \sum_{n=0}^{\infty} \delta_g^{(n)}(\mathbf{x}), \quad \delta_{g,R}(\mathbf{x}) = \sum_{n=0}^{\infty} \delta_{g,R}^{(n)}(\mathbf{x}), \quad (2.9)$$

where, in Fourier space, we have  $\delta_{g,R}^{(n)}(\mathbf{k}) = W_R(k) \delta_g^{(n)}(\mathbf{k})$ . Now, inserting (2.9) inside the marked density field definition, eq. (2.5), we obtain

$$\begin{aligned} \bar{m}(1 + \delta_M(\mathbf{x})) &= m(\mathbf{x})(1 + \delta_g(\mathbf{x})) \\ &= \sum_{n=0}^{\infty} (-1)^n C_n \left( \sum_{l=0}^{\infty} \delta_{g,R}^{(l)}(\mathbf{x}) \right)^n \left[ 1 + \sum_{j=1}^{\infty} \delta_g^{(j)}(\mathbf{x}) \right] \\ &\equiv 1 + \sum_{n=0}^{\infty} \delta_M^{(n)}(\mathbf{x}), \end{aligned} \quad (2.10)$$

from which we obtain

$$\delta_M^{(1)}(\mathbf{x}) = C_0 \delta_g^{(1)}(\mathbf{x}) - C_1 \delta_{g,R}^{(1)}(\mathbf{x}), \quad (2.11)$$

$$\delta_M^{(2)}(\mathbf{x}) = C_0 \delta_g^{(2)}(\mathbf{x}) - C_1 \delta_{g,R}^{(1)} \delta_g^{(1)}(\mathbf{x}) - C_1 \delta_{g,R}^{(2)}(\mathbf{x}) + C_2 \delta_{g,R}^{(1)} \delta_{g,R}^{(1)}(\mathbf{x}), \quad (2.12)$$

$$\begin{aligned} \delta_M^{(3)}(\mathbf{x}) &= C_0 \delta_g^{(3)}(\mathbf{x}) - C_1 \delta_{g,R}^{(1)} \delta_g^{(2)}(\mathbf{x}) - C_1 \delta_{g,R}^{(2)} \delta_g^{(1)}(\mathbf{x}) - C_1 \delta_{g,R}^{(3)}(\mathbf{x}) \\ &\quad C_2 \delta_{g,R}^{(1)} \delta_{g,R}^{(1)} \delta_g^{(1)}(\mathbf{x}) + 2C_2 \delta_{g,R}^{(1)} \delta_{g,R}^{(2)}(\mathbf{x}) - C_3 \delta_{g,R}^{(1)} \delta_{g,R}^{(1)} \delta_{g,R}^{(1)}(\mathbf{x}). \end{aligned} \quad (2.13)$$

To proceed, we expand the various fields involved into power of the linear matter perturbation field  $\delta^{(1)}(\mathbf{k})$ . For biased tracers in real space, we have

$$\delta_g^{(n)}(\mathbf{k}, z) = \mathcal{I}_{\mathbf{k}, \mathbf{q}_1, \dots, \mathbf{q}_n} K_n(\mathbf{q}_1, \dots, \mathbf{q}_n; z) \delta^{(1)}(\mathbf{q}_1) \dots \delta^{(1)}(\mathbf{q}_n), \quad (2.14)$$

where we have defined

$$\mathcal{I}_{\mathbf{k}, \mathbf{q}_1, \dots, \mathbf{q}_n} \equiv \int \frac{d^3 \mathbf{q}_1}{(2\pi)^3} \dots \int \frac{d^3 \mathbf{q}_n}{(2\pi)^3} (2\pi)^3 \delta_D(\mathbf{k} - \mathbf{q}_{1\dots n}), \quad (2.15)$$

and  $K_n$  are the real space kernels for a generic biased tracer [38–40]. Analogously, we can define new kernels for the marked field as

$$\delta_M^{(n)}(\mathbf{k}) = \mathcal{I}_{\mathbf{k}, \mathbf{q}_1, \dots, \mathbf{q}_n} H_n(\mathbf{q}_1, \dots, \mathbf{q}_n; z) \delta^{(1)}(\mathbf{q}_1) \dots \delta^{(1)}(\mathbf{q}_n), \quad (2.16)$$

with

$$\begin{aligned} H_1(\mathbf{k}) &= C_{\delta_M}(k) K_1(\mathbf{k}), \\ H_2(\mathbf{k}_1, \mathbf{k}_2) &= C_{\delta_M}(k) K_2(\mathbf{k}_1, \mathbf{k}_2) + C_{\delta_M^2}(k_1, k_2) K_1(\mathbf{k}_1) K_1(\mathbf{k}_2), \\ H_3(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) &= C_{\delta_M}(k) K_3(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) + 2C_{\delta_M^2}(k_1, k_{23}) K_1(\mathbf{k}_1) K_2(\mathbf{k}_2, \mathbf{k}_3) \\ &\quad + C_{\delta_M^3}(k_1, k_2, k_3) K_1(\mathbf{k}_1) K_1(\mathbf{k}_2) K_1(\mathbf{k}_3), \end{aligned} \quad (2.17)$$

where the  $H_3$  still needs to be symmetrized of the three momenta. We have defined the functions

$$C_{\delta_M}(k) = C_0 - C_1 W_R(k), \quad (2.18)$$

$$C_{\delta_M^2}(k) = C_2 W_R(k_1) W_R(k_2) - \frac{C_1}{2} [W_R(k_1) + W_R(k_2)], \quad (2.19)$$

$$\begin{aligned} C_{\delta_M^3}(k_1, k_2, k_3) &= -C_3 W_R(k_1) W_R(k_2) W_R(k_3) \\ &\quad + \frac{C_2}{3} [W_R(k_2) W_R(k_3) + W_R(k_1) W_R(k_3) + W_R(k_1) W_R(k_2)]. \end{aligned} \quad (2.20)$$

Notice that these functions implicitly contain information about the choice of the adopted mark through the coefficients  $C_i$ . Finally, from this expansion, we can compute the power spectrum of the marked density field up to third perturbative order, i.e. the one loop marked power spectrum. We have

$$\bar{m}^2 M(\mathbf{k}) = \bar{m}^2 |\delta_M(\mathbf{k})|^2 \equiv M_{11}(\mathbf{k}) + 2M_{13}(\mathbf{k}) + M_{22}(\mathbf{k}), \quad (2.21)$$

with

$$M_{11}(\mathbf{k}) = H_1^2(\mathbf{k}) P_L(k), \quad (2.22)$$

$$M_{13}(\mathbf{k}) = 3H_1(\mathbf{k}) P_L(k) \int \frac{d^3 \mathbf{q}}{(2\pi)^3} H_3(\mathbf{k}, \mathbf{q}, -\mathbf{q}) P_L(q), \quad (2.23)$$

$$M_{22}(\mathbf{k}) = 2 \int \frac{d^3 \mathbf{q}}{(2\pi)^3} |H_2(\mathbf{k} - \mathbf{q}, \mathbf{q})|^2 P_L(q) P_L(|\mathbf{k} - \mathbf{q}|), \quad (2.24)$$

where  $P_L$  is the linear power spectrum. By writing explicitly the expression for the 22 term in eq. (2.24) we obtain

$$\begin{aligned}
 M_{22}(k) &= 2C_{\delta_M}^2(k) \int_{\mathbf{q}} K_2^2(\mathbf{k} - \mathbf{q}, \mathbf{q}) P_{\text{lin}}(|\mathbf{k} - \mathbf{q}|) P_{\text{lin}}(q) \\
 &+ 2 \int_{\mathbf{q}} C_{\delta_M}^2(|\mathbf{k} - \mathbf{q}|, q) K_1^2(\mathbf{k} - \mathbf{q}) K_1^2(\mathbf{q}) P_{\text{lin}}(|\mathbf{k} - \mathbf{q}|) P_{\text{lin}}(q) \\
 &+ 4C_{\delta_M}(k) \int_{\mathbf{q}} C_{\delta_M}^2(|\mathbf{k} - \mathbf{q}|, q) K_1(\mathbf{k} - \mathbf{q}) K_1(\mathbf{q}) K_2(\mathbf{k} - \mathbf{q}, \mathbf{q}) P_{\text{lin}}(|\mathbf{k} - \mathbf{q}|) P_{\text{lin}}(q).
 \end{aligned} \tag{2.25}$$

The first line is a simple rescaling of the usual  $P_{22}$  term of the galaxy one-loop power spectrum. This piece should contain, at most, the same amount of information of the  $P_{22}$  term, modulo the dependence on the scale of the  $C_{\delta_M}$  function. The additional terms in the second and third line of eq. (2.25), which are generated by the presence of the smoothed field in the mark function, are convolutions of bispectrum- and (disconnected) trispectrum-like terms with some scale-dependent kernels given by the marking procedure. These terms arise because the marked field is by definition more non-gaussian, and already at the two-point function level, it contains information from higher order correlation functions of the unmarked field.

The kernels  $K_n$  that appear in eq. (2.17) need to be evaluated using a given prescription for the bias expansion. We will follow the usual bias expansion described in [41] up to third order

$$\delta_g(\mathbf{x}) = b_1 \delta(\mathbf{x}) + \frac{b_2}{2} \delta^2(\mathbf{x}) + b_{\mathcal{G}_2} \mathcal{G}_2(\mathbf{x}) + \frac{b_3}{6} \delta^3(\mathbf{x}) + b_{\mathcal{G}_3} \mathcal{G}_3(\mathbf{x}) + b_{\mathcal{G}_2 \delta} \mathcal{G}_2(\mathbf{x}) \delta(\mathbf{x}) + b_{\Gamma_3} \Gamma_3(\mathbf{x}), \tag{2.26}$$

where we have defined the tidal operators as

$$\mathcal{G}_2(\mathbf{x}) \equiv (\partial_i \partial_j \Phi(\mathbf{x}))^2 - (\partial^2 \Phi(\mathbf{x}))^2 \tag{2.27}$$

$$\mathcal{G}_3(\mathbf{x}) \equiv -\partial_i \partial_j \Phi \partial_j \partial_k \Phi \partial_k \partial_i \Phi - \frac{1}{2} (\partial^2 \Phi)^2 + \frac{3}{2} (\partial_i \partial_j \Phi)^2 \partial^2 \Phi \tag{2.28}$$

$$\Gamma_3(\Phi_g, \Phi_v)(\mathbf{x}) \equiv \mathcal{G}_2(\Phi_g)(\mathbf{x}) - \mathcal{G}_2(\Phi_v)(\mathbf{x}), \tag{2.29}$$

with  $\Phi, \Phi_v$  being the gravitational and velocity potential perturbations. The only operators that will enter the one-loop power spectrum are the four renormalized operators  $\delta$ ,  $[\delta^2]$ ,  $[\mathcal{G}_3]$  and  $[\Gamma_3]$ , related to the usual fields via

$$[\delta^2](\mathbf{x}) = \delta^2 - \sigma^2, \quad [\mathcal{G}_2](\mathbf{x}) = \mathcal{G}_2(\mathbf{x}), \tag{2.30}$$

with  $\sigma^2$  being the variance of the unsmoothed field

$$\sigma^2 = \int_{\mathbf{q}} P_L(q). \tag{2.31}$$

As pointed out in [32, 33], the marked power spectrum requires more caution compared to the standard case as the inclusion of renormalized operators leads to the addition of a zero-lag contribution, which, in Fourier space, manifests as

$$\delta_g(\mathbf{k}) \rightarrow \delta_g(\mathbf{k}) - \frac{b_2}{2} \sigma^2 (2\pi)^3 \delta_D(\mathbf{k}), \tag{2.32}$$

or, for the marked field at second and third perturbative orders

$$\delta_M^{(2)}(\mathbf{k}) \rightarrow \delta_M^{(2)} - \frac{b_2}{2} \sigma^2 C_{\delta_M}(k) (2\pi)^3 \delta_D(\mathbf{k}), \quad (2.33)$$

$$\delta_M^{(3)}(\mathbf{k}) - b_2 \sigma^2 C_{\delta_M^2}(k, 0) \delta^{(1)}(\mathbf{k}). \quad (2.34)$$

The inclusion of the third-order contribution is crucial for the small-scale safety of the one-loop marked power spectrum.

In this work, we will compare the information content of a standard analysis that involves galaxy power spectrum and bispectrum against an analysis performed by adopting only two-point correlation functions, namely the one-loop power spectrum and marked power spectrum. The model for the galaxy bispectrum in real space is given by

$$B(k_1, k_2, k_3) = 2K_1(\mathbf{k}_1)K_1(\mathbf{k}_2)K_2(\mathbf{k}_1, \mathbf{k}_2)P_L(k_1)P_L(k_2) + 2 \text{ perms.}, \quad (2.35)$$

where we stopped at the tree level. As for the one-loop contribution, one would need to go to the fourth perturbative order. The tree-level approximation has been shown to work properly up to  $k_{\text{max}}^B \simeq 0.08/0.10 h/\text{Mpc}$  [9, 42, 43], while the one-loop is needed to extend the maximum  $k$  and get unbiased results [6, 7, 44].

## 2.2 Counterterms, shot noise and IR-resummation

The perturbative results presented in the previous section need to be corrected in order to achieve better agreement with smaller, more non-linear scales. The Effective Field Theory of Large Scale Structure (EFTofLSS), originally presented in [45–47], proposes itself as a completion treatment of standard perturbation theory, addressing its main issues: UV divergence and the treatment of imperfections in the cosmological fluid. For the standard one-loop power spectrum in real space, this is achieved with the introduction of a third-order counterterm, which, in Fourier space, reads  $\delta^{(ct)}(\mathbf{k}) = -c_s^2 k^2 \delta^{(1)}(\mathbf{k})$ . This term accounts for the backreaction of small-scale physics on long-wavelength modes. The speed of sound  $c_s^2$  is a free parameter of the theory and will be fitted with observational data. For the marked power spectrum, it was shown that this term is enough to capture the UV divergences in real space; all other terms are manifestly convergent for hard loop momenta  $p \gg k$  due to the presence of smoothing windows that depend on the physical scale  $R$ , used in the definition of the mark. Refs. [32, 33] showed that the marked theory does not require additional counterterms, and the relevant counterterm in  $M(\mathbf{k})$  is

$$M_{ct}(\mathbf{k}) = -2c_s^2 k^2 C_{\delta_M^2}^2(k) P_L(k), \quad (2.36)$$

with  $c_s^2$  being redshift dependent. In the case of biased tracers, additional UV divergences arise

$$M_{13}(\mathbf{k}) \supset 2C_{\delta_M}(k)C_{\delta_M^2}(k, 0)K_1^2(\mathbf{k})P_L(k) \int_{\mathbf{q}} \frac{b_2}{2} P_L(q), \quad (2.37)$$

$$\begin{aligned} M_{22}(\mathbf{k}) &\supset \frac{b_2^2}{2} C_{\delta_M^2}^2(k) \int_{\mathbf{q}} P_L(q) P_L(|\mathbf{k} - \mathbf{q}|) \\ &= \frac{b_2^2}{2} C_{\delta_M^2}^2(k) \int_{\mathbf{q}} [P_L(q) P_L(|\mathbf{k} - \mathbf{q}|) - P_L^2(q)] + \frac{b_2^2}{2} C_{\delta_M^2}^2(k) \int_{\mathbf{q}} P_L^2(q). \end{aligned} \quad (2.38)$$

The  $M_{13}$  contribution vanishes when including the proper renormalized bias operator as in eq. (2.34), while the second term contains a divergent part in the last line that can be completely absorbed when considering the shot noise term. Indeed, the stochastic part of the marked power spectrum

$$M_{\text{stoch}}(k) = C_{\delta_M}^2(k) P_{\text{shot}}, \quad (2.39)$$

for constant  $P_{\text{shot}}$  reabsorbs the UV divergence present in the  $M_{22}$  term. For this reason, following [33], we will assume this shape for the noise contribution henceforth.

For the infrared resummation, that accounts for the correct treatment of long-wavelength (IR) displacements on the baryon acoustic peak [48, 49], we will follow the approximation of [32, 33]: for the terms in the marked spectrum involving  $C_{\delta_M}^2(k) P_{\text{NL}}(k)$  we replace  $P_{\text{NL}}$  with its IR-resummed form, while for the other terms that involve the linear power spectrum  $P_L(k)$ , we replace it with the IR-resummed one. The IR-resummed power spectrum for matter has the form [50]

$$P_{\text{LO}}^{\text{IR-res}}(k) = P_L^{nw}(k) + e^{-k^2 \Sigma^2} P_L^w \quad (2.40)$$

$$P_{\text{NLO}}^{\text{IR-res}}(k) = P_L^{nw}(k) + P_{1\text{-loop}}^{nw}(k) + e^{-k^2 \Sigma^2} \left[ (1 + k^2 \Sigma^2) P_L^w + P_{1\text{-loop}}^w \right], \quad (2.41)$$

where  $nw$  and  $w$  parts are the broadband (no-wiggle) and the oscillatory parts of the power spectrum, and  $\Sigma$  is the velocity dispersion. Similar relations hold for biased tracers. The effectiveness of this approximation has been shown to fit the simulated data without any residual wiggles appropriately.

Summarizing, the one-loop EFT model for the marked power spectrum of biased tracers in real space has the following form

$$\bar{m}^2 M(k) = M_{11}(k) + M_{22}(k) + 2M_{13}(k) + M_{ct}(k) + M_{\text{stoch}}(k), \quad (2.42)$$

which contains the following five parameters<sup>3</sup>

$$\{b_1, b_2, b_{\mathcal{G}_2}, c_s^2, P_{\text{shot}}\}. \quad (2.43)$$

At the perturbative order that we are considering, the galaxy bispectrum presents only the following stochastic terms

$$B_{\text{stoch}}(k_1, k_2, k_3) = P_{\text{shot}} [P_L(k_1) + P_L(k_2) + P_L(k_3)] + B_{\text{shot}}, \quad (2.44)$$

adding the parameter  $B_{\text{shot}}$  to the list of varied parameters in eq. (2.43).

### 2.3 Primordial non-gaussianities of the non-local type

The main goal of this work is to investigate the constraining power of marked statistics on primordial non-gaussianities (PNG). In particular, we are interested in primordial non-gaussianities of the non-local type, meaning the non-gaussianities that are suppressed in the squeezed limit and that can arise from quadratic or higher order interaction in the inflationary field [52].

<sup>3</sup>Notice that we are fixing  $b_{\Gamma_3} = 0$ , as the power spectrum alone is poorly sensitive to this parameter [39, 51].

In this work, we focus on non-local PNG because the potential gain from analyzing the bispectrum is maximized in this scenario, particularly due to the absence of the scale-dependent bias peaked at large scales that appear in local PNGs [53–55]. Some works have recently explored non-local PNG in the context of machine-learning base approaches to enhance signal extraction [56, 57]. Additionally, within the framework of the Cosmological Collider [58], recently adopted on the BOSS survey [59], several theoretical models involving massive particles during inflation can be effectively approximated by equilateral and orthogonal shapes in the non-local non-Gaussianity regime. This approximation allows for a more accurate representation of interactions in the early universe and provides a pathway for distinguishing between different inflationary models.

It is useful to represent non-local PNG using a linear combination of two shapes, equilateral and orthogonal [60], depending on the triangular configuration over which the bispectrum is peaked. Concretely, the two parameters for the amplitudes of different shapes,  $f_{\text{NL}}^{\text{eq}}$  and  $f_{\text{NL}}^{\text{ortho}}$ , can be directly linked to physical observables of the inflaton field. The Lagrangian of general single-field models in the context of the effective field theory of inflation is given by [52, 60]

$$S = \int d^4x \sqrt{-g} \left[ -\frac{M_{\text{Pl}}^2 \dot{H}}{c_{\pi,s}^2} \left( \dot{\pi}^2 - c_{\pi,s}^2 \frac{(\partial_i \pi)^2}{a^2} \right) - M_{\text{Pl}}^2 \dot{H} (1 - c_{\pi,s}^{-2}) \dot{\pi} \frac{(\partial_i \pi)^2}{a^2} \right. \\ \left. + \left( M_{\text{Pl}}^2 \dot{H} (1 - c_{\pi,s}^{-2}) - \frac{4}{3} M_3^4 \right) \dot{\pi}^3 \right], \quad (2.45)$$

with the scalar perturbation  $\pi$  being related to the curvature perturbation  $\zeta = -H\pi$ . The two leading interactions in eq. (2.45) are given by the third order terms  $\dot{\pi}(\partial_i \pi)^2$  and  $\dot{\pi}^3$ , which produce specific bispectra with amplitudes [61]

$$f_{\text{NL}}^{(\dot{\pi}(\partial_i \pi)^2)} = -(85/324)(c_{\pi,s}^{-2} - 1), \quad \text{and} \quad f_{\text{NL}}^{\dot{\pi}^3} = -(10/243)(c_{\pi,s}^{-2} - 1)[\tilde{c}_3 + (3/2)c_{\pi,s}^2], \quad (2.46)$$

where we have defined the dimensionless parameter  $\tilde{c}_3$  through  $\tilde{c}_3(c_{\pi,s}^{-2} - 1) = 2M_3^4 c_{\pi,s}^2 / (\dot{H} M_{\text{Pl}}^2)$ . The two EFT shapes can be projected onto the equilateral and orthogonal shapes as

$$f_{\text{NL}}^{\text{equil}} = \frac{1 - c_{\pi,s}^2}{c_{\pi,s}^2} [-0.275 - 0.0780c_s^2 - 0.53\tilde{c}_3], \quad (2.47)$$

$$f_{\text{NL}}^{\text{ortho}} = \frac{1 - c_{\pi,s}^2}{c_{\pi,s}^2} [0.0159 - 0.0167c_{\pi,s}^2 - 0.01113\tilde{c}_3]. \quad (2.48)$$

$$B_\zeta(k_1, k_2, k_3) = \frac{18}{5} f_{\text{NL}} \Delta_\zeta^4 \frac{\mathcal{S}(k_1, k_2, k_3)}{k_1^2 k_2^2 k_3^2}, \quad (2.49)$$

with the equilateral and orthogonal templates defined as [60, 62]

$$\mathcal{S}_{\text{eq}}(k_1, k_2, k_3) = \left( \frac{k_1}{k_2} + 5 \text{ perms.} \right) - \left( \frac{k_1^2}{k_2 k_3} + 2 \text{ perms.} \right) - 2, \quad (2.50)$$

$$\mathcal{S}_{\text{ort}}(k_1, k_2, k_3) = (1 + p) \frac{\Delta}{e_3} - p \frac{\Gamma^3}{e_3^2}, \quad (2.51)$$

where  $p = 8.52587$ ,  $\Delta = (k_T - 2k_1)(k_T - 2k_2)(k_T - 2k_3)$

$$k_T = k_1 + k_2 + k_3, \quad e_2 = k_1 k_2 + k_1 k_3 + k_2 k_3, \quad e_3 = k_1 k_2 k_3, \quad \Gamma = \frac{2}{3} e_2 - \frac{1}{3} (k_1^2 + k_2^2 + k_3^2). \quad (2.52)$$

Here we are making use of the so-called *orthogonal-LSS* shape, that, contrary to the *orthogonal-CMB* shape, does not present the usual scale dependent bias effect on large scales [60]. This choice is motivated by many inflationary models that can physically produce such signal [60, 63].

The PNG affects the late-time dark matter distribution, inducing an additional bispectrum term given by

$$\langle \delta^{(1)} \delta^{(1)} \delta^{(1)} \rangle = f_{\text{NL}} B_{111}(k_1, k_2, k_3) (2\pi)^3 \delta_D(\mathbf{k}_{123}) \quad (2.53)$$

with

$$f_{\text{NL}} B_{111}(k_1, k_2, k_3) = \mathcal{T}(k_1) \mathcal{T}(k_2) \mathcal{T}(k_3) B_\zeta(k_1, k_2, k_3), \quad (2.54)$$

where we have introduced the transfer function  $\mathcal{T}(k) \equiv \delta^{(1)}(\mathbf{k})/\zeta(\mathbf{k}) = (P_{11}(k)/P_\zeta(k))^{1/2}$ . Notice that this can be easily calculated via the definition of the power spectrum of initial fluctuations

$$P_\zeta(k) = \Delta_\zeta^2 k^{-3} \left( \frac{k}{k_*} \right)^{n_s - 1}. \quad (2.55)$$

Analysis of *Planck* data finds  $\Delta_\zeta^2 \simeq 4.1 \times 10^{-8}$ ,  $n_s \simeq 0.96$  for the pivot scale  $k_* = 0.05 \text{ Mpc}^{-1}$  [64]. The initial bispectrum also enters the 1-loop galaxy power spectrum through the 1-2 term

$$f_{\text{NL}} P_{12}(\mathbf{k}) = 2f_{\text{NL}} K_1(\mathbf{k}) \int_{\mathbf{q}} K_2(\mathbf{k} - \mathbf{q}, \mathbf{q}) B_{111}(k, q, |\mathbf{k} - \mathbf{q}|). \quad (2.56)$$

In addition, non-local PNG modulates galaxy formation, which is captured by the scale-dependent galaxy bias [3, 63, 65]

$$\delta_g = b_1 \delta + f_{\text{NL}} b_\zeta \left( \frac{k}{k_{\text{NL}}} \right)^2 \zeta + \text{nonlinear}, \quad (2.57)$$

so that the final model for the galaxy power spectra and bispectra in real space is

$$P_g(\mathbf{k}) = P_{g,G}(k) + f_{\text{NL}} \left( P_{12}(\mathbf{k}) + 2b_\zeta K_1(k) \frac{k^2}{k_{\text{NL}}^2} \frac{P_L(k)}{\mathcal{T}(k)} \right), \quad (2.58)$$

$$B_g(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) = B_{g,G}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) + f_{\text{NL}} K_1(k_1) K_1(k_2) K_1(k_3) B_{111}(k_1, k_2, k_3), \quad (2.59)$$

where  $P_{g,G}$  and  $B_{g,G}$  are the standard Gaussian power spectrum and bispectrum models. Following [3], we neglect all higher order correction terms, such as  $\mathcal{O}(f_{\text{NL}}^2)$  PNG terms, non-Gaussian two-loop terms and contributions generated by higher order terms such as  $\delta \nabla^2 \zeta$ , as they are subdominant for the scales considered in this work. Notice that for PNG of the local type some of these terms should be included as in that case PNG bias terms would dominate on large scales.

For the marked field, we have the additional term coming from the initial bispectrum contribution

$$f_{\text{NL}} M_{12}(\mathbf{k}) = 2f_{\text{NL}} H_1(\mathbf{k}) \int_{\mathbf{q}} H_2(\mathbf{k} - \mathbf{q}, \mathbf{q}) B_{111}(k, q, |\mathbf{k} - \mathbf{q}|), \quad (2.60)$$

so that the marked power spectrum and bispectrum for galaxies in real space are

$$M_g(\mathbf{k}) = M_{g,G}(\mathbf{k}) + f_{\text{NL}} \left( M_{12}(\mathbf{k}) + 2b_\zeta K_1(k) \frac{k^2}{k_{\text{NL}}^2} C_{\delta_M}^2(k) \frac{P_L(k)}{\mathcal{T}(k)} \right) \quad (2.61)$$

$$BM_g(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) = BM_{g,G}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) + H_1(\mathbf{k}_1)H_1(\mathbf{k}_2)H_1(\mathbf{k}_3)f_{\text{NL}}B_{111}(k_1, k_2, k_3), \quad (2.62)$$

with

$$BM_{g,G}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) = 2H_1(\mathbf{k}_1)H_1(\mathbf{k}_2)H_2(\mathbf{k}_1, \mathbf{k}_2)P_L(k_1)P_L(k_2) + 2 \text{ perms.} . \quad (2.63)$$

Writing down  $M_{12}(\mathbf{k})$  explicitly:

$$\begin{aligned} M_{12}(\mathbf{k}) = & 2C_{\delta_M}(k)^2 b_1 \int_{\mathbf{q}} \left[ b_1 F_2(\mathbf{k} - \mathbf{q}, \mathbf{q}) + \frac{b_2}{2} + b_{\mathcal{G}_2} \sigma^2(\mathbf{k} - b\mathbf{q}, \mathbf{q}) \right] B_{111}(k, |\mathbf{k} - \mathbf{q}|, q) \\ & + 2C_{\delta_M}(k) b_1^3 \int_{\mathbf{q}} C_{\delta_M}^2(|\mathbf{k} - \mathbf{q}|, q) B_{111}(k, |\mathbf{k} - \mathbf{q}|, q). \end{aligned} \quad (2.64)$$

Through this work, we will assume that the following relation holds for the primordial non-Gaussian bias and the linear bias

$$b_\zeta = \frac{18}{5} \delta_c (b_1 - 1), \quad (2.65)$$

motivated by the peak-background split argument [65], with  $\delta_c$  being the critical overdensity for the spherical collapse.

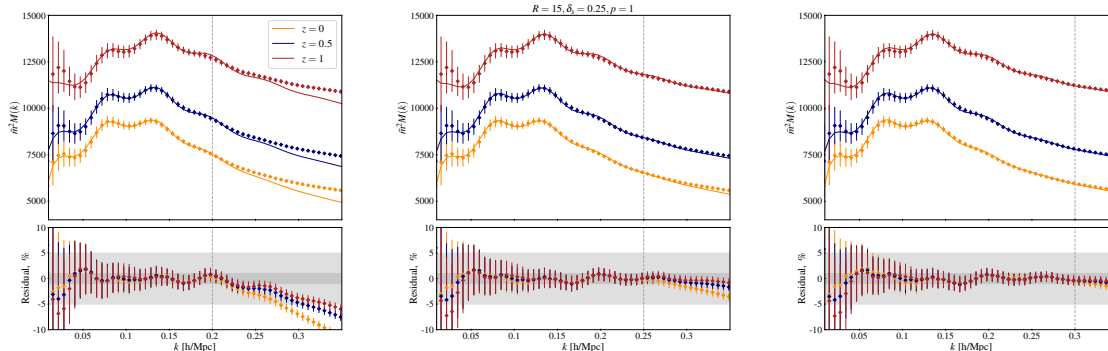
It is worth noticing that, at the power spectrum level, the PNG contributions are entered only at the one-loop level. Moreover, as shown in [3], the effect of equilateral and orthogonal non-Gaussianities is very degenerate with the second order biases  $b_2$  and  $b_{\mathcal{G}_2}$ . These two effects conspire to make power spectrum only analyses very inefficient in measuring non-local PNG, making the inclusion of the bispectrum essential in order to constrain them. The marked power spectrum seems very promising in this direction by bringing information from higher-order functions back to a two-point function, as also shown in [34]. This could naively explain the results of ref. [31], where the authors find that the marked power statistic helps in constraining non-local PNG much more than that of the local type.

## 2.4 Low- $k$ behavior

Ref. [33] showed that, differently from the standard perturbative treatment of the power spectrum, when considering marked statistics, the perturbative series converges more slowly, and loop terms can give non-zero contributions also to linear scales. The authors came up with a low- $k$  correction for the marked power spectrum that, in real space, is

$$M|_{\text{low-}k}(k) = C_{\delta_M}^2(k) [a_0 P_L(k) + c_0], \quad (2.66)$$

with the addition of two more parameters. We can already exclude the second one since it is completely degenerate with the shot noise term, leaving only the free parameter  $a_0$ . One should also vary this parameter, but we will fix it for the following reasons. In [33], they showed that for biased tracers in real space this should not matter much, which also confirmed by our fit against simulations shown in section 3.1; in this work we want to explore what is the actual gain of the marked power spectrum compared to standard perturbative analyses and fixing  $a_0 = 0$  represents the optimistic scenario where we can completely trust our theory model and exploit all of its information content.



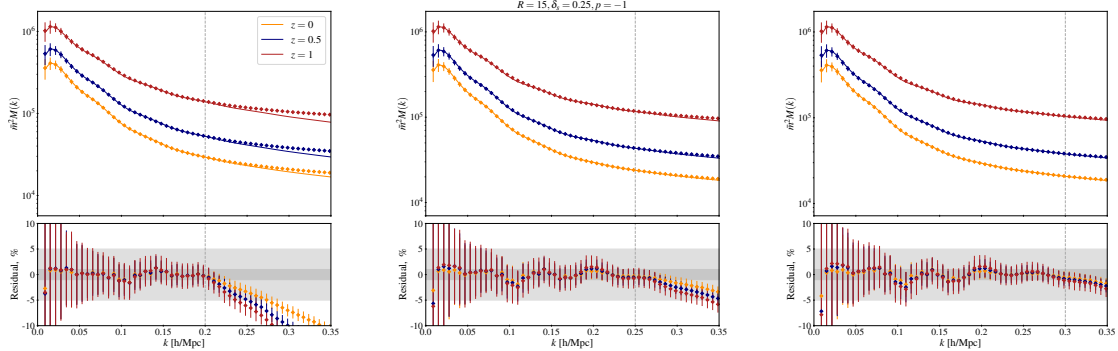
**Figure 1.** Results of the fitting of the marked power spectrum from the Quijote-PNG simulations with equilateral PNG,  $f_{\text{NL}}^{\text{eq}} = 100$  for different redshifts. The left plot is obtained fitting the simulations up to  $k_{\text{max}}^M = 0.2 h/\text{Mpc}$ , the central one with  $k_{\text{max}}^M = 0.25 h/\text{Mpc}$  and the right one with  $k_{\text{max}}^M = 0.3 h/\text{Mpc}$ . The diamonds represent the mean marked power spectrum from the simulations, while the solid lines are the results obtained from the EFT fit.

### 3 Results

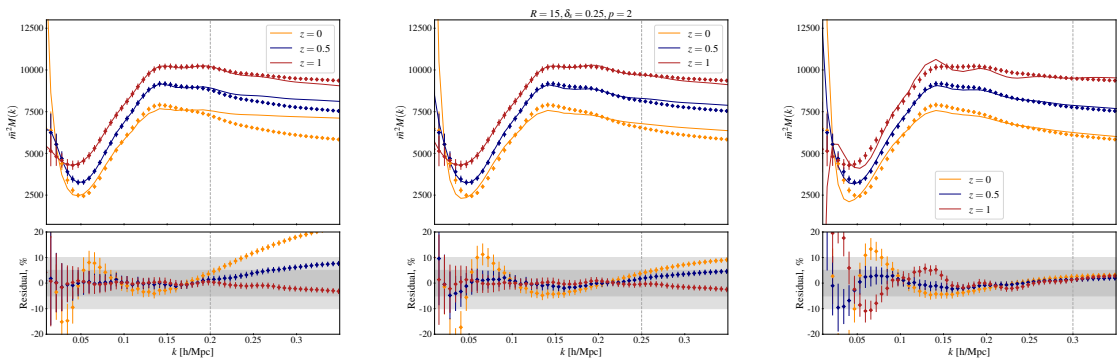
#### 3.1 Comparison to simulations

Before proceeding with the Fisher forecast, we compare our predictions with the Quijote [66] and Quijote-PNG [63] simulations. These are N-body simulations of volume  $1 (\text{Gpc}/h)^3$ , containing  $512^3$  particles each, and run using the TreePM code GADGET-III from initial conditions generated at  $z = 127$  by the codes 2LPTIC [67] and 2LPTPNG [63, 68], for the simulations with and without PNG respectively. We will compare our PT predictions with real space dark matter halos identified in each simulation by the standard Friends-of-friends algorithm [69] by setting the linking length parameter to  $b = 0.2$  and considering halos with more than 20 dark matter particles. We compute the marked power spectrum for halos with minimum mass  $M_{\text{min}} = 3.2 \times 10^{13} M_{\odot}/h$ , following the procedure described in [13, 70]. The fiducial cosmology is  $\{\sigma_8, n_s, \Omega_b, \Omega_m, h\} = \{0.834, 0.9624, 0.049, 0.3175, 0.6711\}$ . We compare with the simulations at redshifts  $z = \{0, 0.5, 1\}$  and mark parameters  $\{R = 15 h/\text{Mpc}, \delta_s = 0.25\}$  and  $p = \{-1, 1, 2\}$  to test our PT model and numerical implementation by fixing the cosmological parameters to their fiducial values and fitting the bias, small-scales and stochastic parameters  $\{b_1, b_2, b_{\mathcal{G}_2}, c_s^2, P_{\text{shot}}\}$  to the simulation power spectrum. The results of these fits are shown in figures 1, 2, 3, where we display the mean marked power spectrum obtained from the PNG simulations with equilateral shape and  $f_{\text{NL}} = +100$  fitted up to  $k_{\text{max}} = \{0.20, 0.25, 0.30\} h/\text{Mpc}$ . To perform this comparison, we estimate the data covariance numerically using the scattering among the different realizations of the simulations with the same cosmology.

We observe that the theoretical predictions with the EFT of LSS are in excellent agreement with the simulations for the cases with  $p = 1$  and  $p = -1$ , within 1%, up to the maximum  $k_{\text{max}}$  adopted for the fitting and for different redshifts. These results confirm again the additional terms introduced in [33] to control the low- $k$  behavior of the marked power spectrum are not needed for biased tracers in real space, also in the presence of non-local PNG. Moreover, we do not see any residual wiggles, corroborating the validity of the approximation adopted for



**Figure 2.** Same as figure 1, with  $p = -1$ .



**Figure 3.** Same as figure 1 and 2, with  $p = 2$ .

the IR-resummation. For the case  $p = 2$  more caution is needed. Looking at the expression for the PT kernels for the marked power spectrum, eq. (2.17) and (2.20) we see that the linear contribution can go to zero for some values of  $k$  in some  $\{R, \delta_s, p\}$  configuration. The condition for these terms to go to zero is given by

$$W_R(k) = \frac{1 + \delta_s}{p}, \quad (3.1)$$

where we used  $C_0 = 1$  and  $C_1 = p/(1 + \delta_s)$ . For  $\{R, \delta_s, p\} = \{15, 0.25, 2\}$ , this condition is satisfied for  $k \simeq 0.06 h/\text{Mpc}$ , implying that around that scales some contribution of the marked power spectrum go to zero. For the halo marked spectrum, the agreement between data and our model decreases to 10 – 20%, as shown in figure 3. We can see that, especially for low redshifts, the scales around  $k \simeq 0.06 h/\text{Mpc}$  are poorly fitted, and the disagreement is more pronounced when the  $k_{\text{max}}$  is higher. We still observe an agreement of  $\sim 5\%$  for the redshift and the scales considered for this analysis, and hence we will use this configuration to perform our forecast. In appendix A we also show the agreement of the model with different configurations of the mark parameters. In particular we explore variations of  $R$ ,  $\delta_s$  and  $p$  at redshift  $z = 0.5$  for the fiducial  $\Lambda\text{CDM}$  cosmology and find excellent agreement with the data.

### 3.2 Analysis specifics

We are now ready to study the amount of information accounted for by the marked power spectrum and compare it with a standard power spectrum plus bispectrum analysis. This can be readily addressed by assuming, at leading order, that the statistics considered follow a multivariate Gaussian distribution and compute the Fisher matrix defined as

$$F_{\mu\nu} = \sum_{i,j} \frac{\partial D_i}{\partial \theta_\mu} C_{ij}^{-1} \frac{\partial D_j}{\partial \theta_\nu}, \quad (3.2)$$

where  $\mathbf{D} = \{P(\mathbf{k}), M(\mathbf{k}), B(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3), \dots\}$  is the data vector, a collection of the different observables employed in the analysis measured at different wavevectors  $\mathbf{k}$ .  $\theta_\mu$  are the cosmological and EFT parameters that are varied in the analysis, and  $C$  is the covariance matrix among the different observables. The Fisher matrix in eq. (3.2) gives the variance of an optimal unbiased estimator for the parameter  $\theta_\mu$ , with the marginalized error satisfying  $\sigma^2(\theta_\mu) \geq (F^{-1})_{\mu\mu}$ : the lower the marginalized error, the better a certain parameter can be constrained with the statistics employed. The statistics that we will consider are the power spectrum, the marked power spectrum and the bispectrum for galaxies in real space. We will compute these statistics by varying the value of  $k_{\max}$  for P and for M, while we will keep fixed  $k_{\max}^B = 0.10h/\text{Mpc}$  for the bispectrum. This has been shown to be a mildly conservative choice as the tree-level model for the bispectrum is valid up to these scales for the volume that we will consider in this work [39, 42].

Previous analyses showed that the marked power spectrum always outperforms the standard power spectrum because it is able to keep track of the cosmic web through the non-linear mark in eq. (2.3): different values of the exponent  $p$  can enhance voids ( $p > 0$ ) or nodes ( $p < 0$ ). [29] finds that positive values of  $p$  are optimal to constrain the total neutrino mass and modified gravity theories, given that voids are more sensitive to these two ingredients. In this work, we will perform a Fisher forecast up to mildly non-linear scales using  $p = 2, 1$  to enhance underdense regions and  $p = -1$  to mark the overdense ones to constrain massive neutrinos and non-local-primordial non-gaussianities and compare with the standard analysis of power spectrum combined with bispectrum. The fiducial cosmology adopted here is one of the Quijote simulations for the standard cosmological parameters, with the additional parameters  $\{M_\nu, f_{\text{NL}}^{\text{eq}}, f_{\text{NL}}^{\text{or}}\} = \{0.06, 0., 0.\}$ . We will run the analysis separately for massive neutrinos and primordial non-gaussianities.

For the fiducial of the galaxy biases, we adopt a semi-analytic model of galaxy formation to fix them; specifically, for the linear bias we adopt the following model<sup>4</sup>

$$b_1(z) = 0.9 + 0.4z. \quad (3.3)$$

For the higher order biases we adopt the following fitting formulae [71, 72], obtained from a combination of N-body simulations and halo occupation modeling

$$b_2(z) = -0.704 - 0.208z + 0.183z^2 - 0.00771z^3. \quad (3.4)$$

<sup>4</sup>Notice that this model should describe Euclid-like H $\alpha$  targets [51], but we will employ this model for BOSS-like volumes as well since we are only interested in the relative change between different statistics.

For  $b_{\mathcal{G}_2}$  and  $b_{\Gamma_3}$  we use instead the co-evolution model [71, 73], which gives

$$b_{\mathcal{G}_2}(z) = -\frac{2}{7}(b_1(z) - 1), \quad b_{\Gamma_3}(z) = \frac{23}{42}(b_1(z) - 1). \quad (3.5)$$

For the higher derivative term, we adjusted the prescription of [51] in order to match with the BOSS data

$$c_s = -D^2(z)[\text{Mpc}/h]^2, \quad (3.6)$$

with  $D(z)$  being the growth factor for linear perturbations. Finally, the fiducial values for the stochastic terms are set to the Poisson sampling prediction

$$P_{\text{shot}} = \bar{n}_g^{-1} \quad B_{\text{shot}} = \bar{n}_g^{-2}. \quad (3.7)$$

To compute the one-loop prediction of the standard and the marked power spectrum we implement the public FFTLog-based code CLASS-PT [74], an extension of the Boltzmann solver CLASS [75] that includes the EFTofLSS as non-linear model. For massive neutrino cosmologies we just modify the linear power spectrum and input it inside the one-loop integrals, neglecting the non-linear effects in the perturbative kernels and the scale dependent growth function as this approximation has been shown to perform well for the current constraints on neutrino masses [51, 76].

We perform a forecast for BOSS CMASS2 redshift  $z = 0.61$  and total volume  $V_{\text{BOSS}} = 3.83 (\text{Gpc}/h)^3$ , with number density  $\bar{n}_g^{\text{BOSS}} = 3 \times 10^{-4} (h/\text{Mpc})^3$ . We consider diagonal Gaussian covariances, after having tested that this approximation correctly reproduces the ones obtained with the Quijote simulations. They are defined as

$$C_P(k) = \frac{2(2\pi)^3}{VV_s} P^2(k), \quad C_M(k) = \frac{2(2\pi)^3}{VV_s} M^2(k), \quad (3.8)$$

$$C_B(k_1, k_2, k_3) = \frac{(2\pi)^6 s_{123}}{VV_{123}} P(k_1)P(k_2)P(k_3) \quad (3.9)$$

where  $V_s = 4k^2\delta k$  and  $V_{123} = 8\pi^2 k_1 k_2 k_3 \delta k^3$  are the shell volumes for the power spectrum and the bispectrum,  $\delta k = 0.01 h/\text{Mpc}$ ,  $s_{123}$  is the triangular symmetry factor and  $V$  is the volume of the survey. To avoid double counting of the information content in the different spectra, we also include the cross covariance between the marked and the standard power spectrum

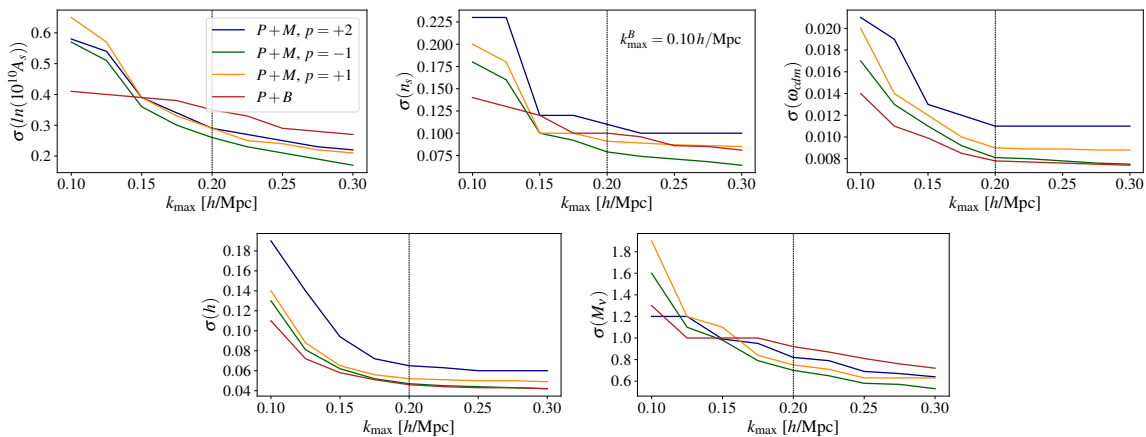
$$C_{PM}(k) = \frac{2(2\pi)^3}{VV_s} P(k)M(k). \quad (3.10)$$

In ref. [29] it has been shown that the Gaussian approximation for the covariance matrix is particularly good for the marked power spectrum, while for the power spectrum and the bispectrum one should, in principle, also include non-Gaussian terms. These are expected to give a negligible contribution up to the scales considered, although for the bispectrum one should include them to properly estimate the forecasted errorbars, especially for primordial non-Gaussianities [77].

In standard LSS analyses, the abundance of baryons is usually constrained using BBN priors [39], and hence we decide to fix it to its fiducial value. We vary the cosmological parameters  $\{\ln(10^{10} A_s), n_s, \omega_{\text{cdm}}, h, M_\nu, f_{\text{NL}}^{\text{Equil}}, f_{\text{NL}}^{\text{Ortho}}\}$ , with the total neutrino mass and the two

$z$	$b_1$	$b_2$	$b_{\mathcal{G}_2}$	$b_{\Gamma_3}$	$c_s^2[\text{Mpc}/h]^2$	$P_{\text{shot}}[\text{Mpc}/h]^3$	$B_{\text{shot}}[\text{Mpc}/h]^6 \times 10^{-3}$
0.61	1.14	-0.38	-0.041	0.079	0.53	3333	11111

**Table 1.** Table of fiducial values for the redshift, bias parameters, counterterm and shot-noise parameters adopted for the forecast.



**Figure 4.**  $2\sigma$  forecasted errorbars for dark matter only in presence of massive neutrinos with BOSS CMASS volume and redshift (see main text). The dashed vertical line represents the maximum wave vector adopted in standard analyses [39].

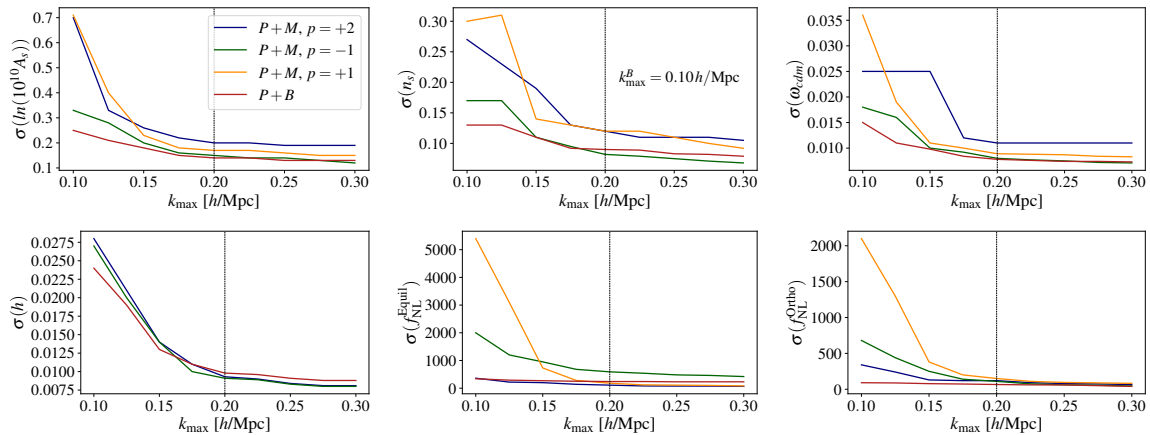
PNG parameters varied separately. For galaxies, we fix the third order bias parameter  $b_{\Gamma_3}$  to its theoretical prediction, as it is poorly constrained [51], and we vary  $\{b_1, b_2, b_{\mathcal{G}_2}, c_s, P_{\text{shot}}, B_{\text{shot}}\}$ .

The list of fiducial parameters is listed in table 1

In the following we will present our results and discuss various effects that contribute to the cosmological parameter measurement from the standard and marked power spectrum and their combination with the bispectrum. We first discuss the results obtained analyzing the dark matter only case and then we will go through the results for galaxies.

### 3.3 Dark matter

LSS observation always involve luminous sources (galaxies, quasars, ...), hence studying the dark matter field is a merely academic exercise; nevertheless, it can still be worthy as it can give useful insights on where the physical information is. It also represents an optimal case, since the unknown bias parameters will degrade the constraints and the discrete nature of galaxies will diminish the signal-to-noise ratio due to the shot-noise. The results of our analysis are shown in figure 4 for the neutrino mass  $M_\nu$  and figure 5 for non-local PNG. We decide not to show the comparison between the standard power spectrum and the marked one as it has been shown that, with the same volumes and sky-cuts, the latter gives better constraints on almost all the cosmological parameters. We compare here two different mark configuration to understand if the information is coming from different structures of LSS web (voids as underdense regions or filaments and nodes for overdense ones) or it is just information taken from higher order correlation functions. The effects of massive neutrinos on LSS are mainly a rescaling of the growth of linear perturbation and a erasing



**Figure 5.** Same as figure 4 but for cosmologies with primordial non-gaussianities.

of small scale clustering below the free streaming scale [78]. The former is the most visible and detectable, while the latter is slightly degenerate with the small scales parameter, the counterterm  $c_s^2$  that already appears at level of matter, and the higher order biases that are present for biased tracers. For all the parameters, in the analysis where the total neutrino mass is varied, we can see that different marks,  $p = 2, 1$  for voids and  $p = -1$  for nodes, give constraints that are almost comparable, with slightly better results for positive  $p$ . Most importantly, we can see that these constraints are comparable with the  $P + B$  analysis, which, we remind, has a fixed  $k_{\max}^B = 0.10 h/\text{Mpc}$ : this confirms that the additional information of the marked power spectrum is mostly a bispectrum-like contribution. In particular, for the neutrino mass  $M_\nu$  we can see that for  $k_{\max} \gtrsim 0.17 h/\text{Mpc}$ , the constraints obtained using the marked spectra are slightly better compared to the  $P + B$  analysis, most probably due to the smaller scales probed with  $M$ .

For the primordial non-gaussianities the outcomes are slightly different. Non-local non-gaussianities affect the standard power spectrum only through one-loop contribution (differently from the well-know large scales effect for non-gaussianities of the local type [53, 54]) and hence their effect is either tiny and degenerate with higher order biases or counterterms. This means that the bispectrum plays a crucial role in constraining this kind of non-gaussianities as it is the main source of signal. Figure 5 shows that the errors  $\sigma(f_{\text{NL}}^{\text{Equil}}), \sigma(f_{\text{NL}}^{\text{Ortho}})$  from the  $P + M$  analysis are always bigger than the  $P + B$  one, at least on the scales at which EFT of LSS is still reliable for BOSS volumes. We see that the case with  $p = 2$  slightly outperforms the  $P + B$  analysis for PNG of equilateral shape. As we already discussed in the previous section, the theoretical results for this mark are less accurate compared to the other two and hence these results might be slightly affected by this. We see that the differences between a positive and a negative  $p$  are negligible and that the results we obtain are very similar to those obtained with the bispectrum: this again confirms that the information content of the marked power spectrum is in fact a subset of the information contained in higher order correlation functions. While for the total neutrino mass this was not obvious as the contribution of the bispectrum to its constraints are not essential, in the case of non-local primordial non-gaussianities this results to be evident.

At this stage, it is worth comparing our findings with the results of [29], where the authors investigated the constraining power of the marked power spectrum on massive neutrinos using a simulation based approach. Using the  $cb$  (cold dark matter + baryons) component they find that using the marked power spectrum produces an improvement of  $\sim 3$  over the standard one at  $k_{\max} = 0.5 h/\text{Mpc}$ . If consider a direct comparison between  $M$  and  $P$ , we find a similar improvement in our Fisher forecast — at scales ( $k_{\max} = 0.20 - 0.30 h/\text{Mpc}$  — when we consider the same volume, redshift and mark parameters. This again confirms the agreement with previous results, even though here we focus on the comparison with the standard  $P + B$  analysis, which seems a more natural one, considering that  $M$  does incorporate higher-order information.

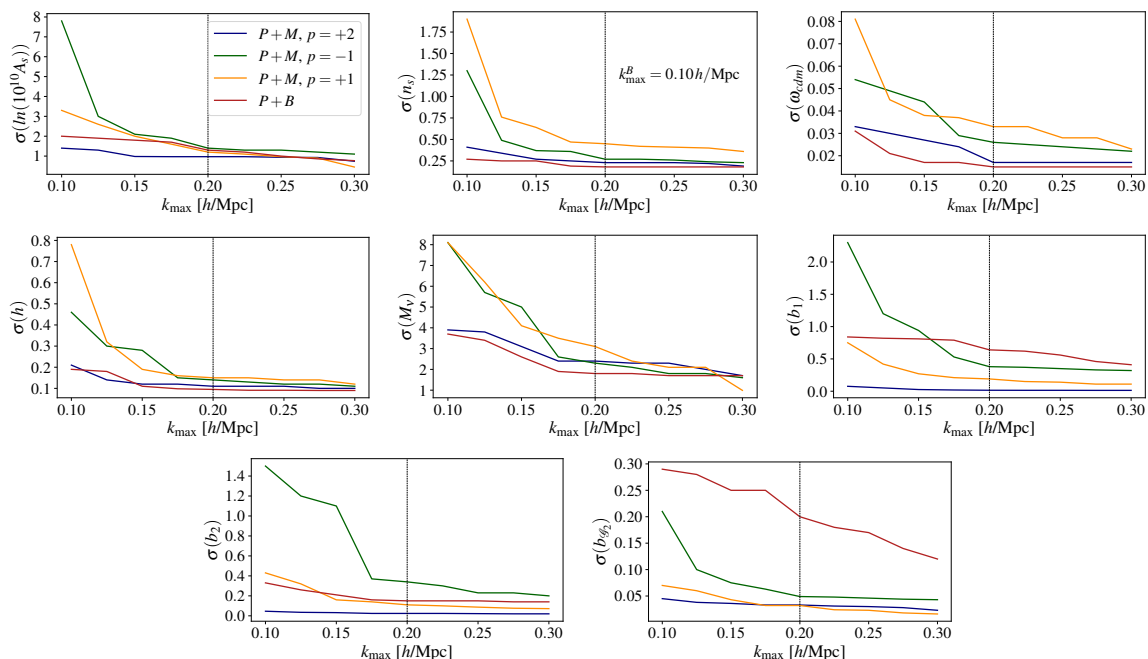
The previous analysis in [29] also emphasizes a void-enhancing mark is the optimal choice for constraining neutrino masses. This result is again qualitatively consistent with our findings for the matter field analysis, see figure 4. Moreover, we find that a similar conclusion holds for equilateral PNG’s: as pointed out in previous works, see e.g. [79], the signal from equilateral PNG is maximized for non-linear filamentary structure and this could explain why a “void-enhanced” marked power spectrum could be effective in constraining this shape.

### 3.4 Galaxies

Biased tracers, such as galaxies, introduce a certain degree of freedom through the biases, that parametrize our incomplete knowledge about the connection between dark matter and galaxies. They are the main source of degradation of cosmological information and indeed much effort has been recently made to constrain them using hydro-dynamical simulations [80–82]. For what concerns our work, we know that, in real space, specific biases are almost perfectly degenerate with cosmological parameters. When the power spectrum alone is considered, the linear bias is completely degenerate with the primordial amplitude  $A_s$ , and this degeneracy can be slightly broken by using the information from the IR-resummed BAO. This degeneracy can be broken by combining the power spectrum and the bispectrum, at the price of introducing a new slight degeneracy among the scalar amplitude and the second order bias  $b_2$ . Including modes from smaller scales could indeed alleviate these degeneracies, allowing to constrain higher order biases through the one-loop terms. This eventually results in worse signal-to-noise for the cosmological parameters. For massive neutrinos we observe in figure 6 that the marked power spectra with  $p = \pm 1$  perform poorly for small  $k_{\max}$ ’s, but they reach the level of constraints of  $p = 2$  when the loop terms are more important around  $k_{\max} \simeq 0.2 h/\text{Mpc}$ .

Interestingly, we see that  $\sigma(M_\nu)$  for the  $P+M$  case, for maximum  $k$  around the non-linear scale, can reach the level of the  $P+B$  one, confirming what was already found in redshift space in [30] and [83]. It is also interesting to notice that the main effect of adding the bispectrum is usually a better measurement of the second order biases  $b_2$  and  $b_{\mathcal{G}_2}$ , which are usually less constrained from  $P$  alone analyses. We see that the inclusion of the marked power spectrum allows to measure these biases at the same level of  $P+B$  analyses, or even better in the case of  $b_{\mathcal{G}_2}$ .

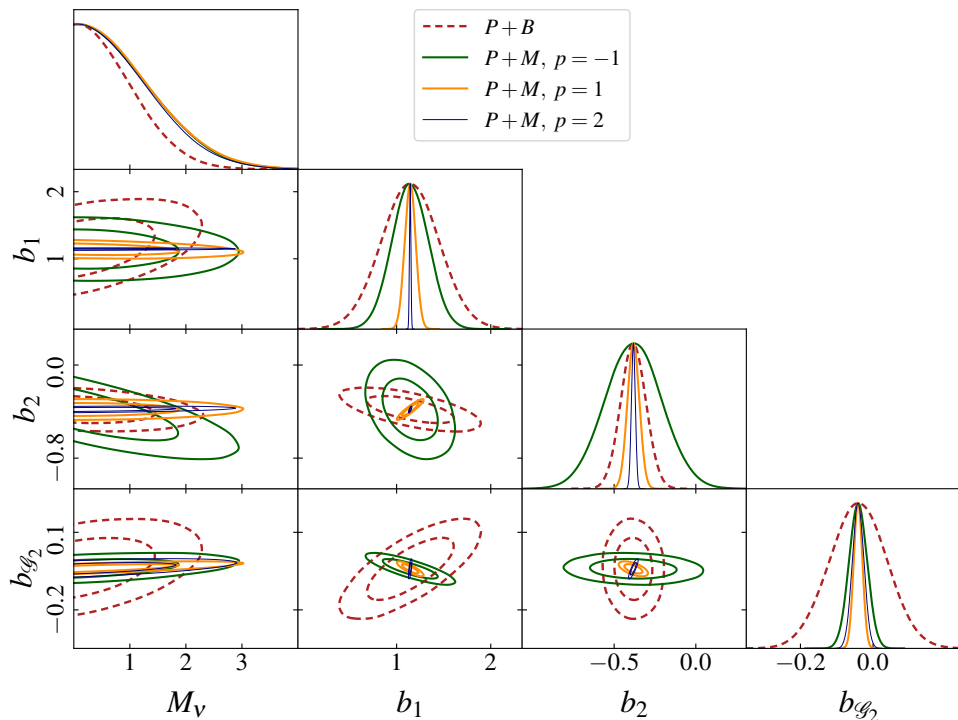
Neutrino mass affects the linear power spectrum through a modification of the linear growth and a damping around and below the free streaming scale, which is a nonlinear



**Figure 6.**  $2\sigma$  constraints from galaxies for massive neutrinos cosmologies and BOSS CMASS volume, number density and redshift. The dashed vertical line represent the maximum wave vector adopted in standard analysis [39].

effect. The galaxy power spectrum alone is mostly sensitive to the former effects, while the latter is degenerate with higher order biases and counterterms. As shown in figure 7, the marked power spectrum, especially for marks with positive  $p$ , breaks some of the degeneracies among the total neutrino mass  $M_\nu$  and the biased tracers, improving the constraints on the bias parameters. This should not change dramatically when we will consider redshift space distortions since, at first order, these will allow to have a better measurement of the linear bias  $b_1$ , which does not appear to be degenerate with the neutrino total mass.

For non-local primordial non-gaussianities, we see a similar behavior to what we found with the dark matter field. We already mentioned that for the shapes studied here, the bispectrum is essential in order to constrain primordial non-gaussianities, differently from the case of massive neutrinos cosmologies, where most of the information is taken from linear scales. The results of our analysis are shown in figure 8. We notice that in both cases where  $p = \pm 1$ , the constraints on  $f_{\text{NL}}^{\text{Equil}}$  and  $f_{\text{NL}}^{\text{Ortho}}$  obtained combining  $P$  and  $M$  never reach those obtained with  $P + B$ , even for very non-linear  $k_{\text{max}}$  and the same is true for almost all the other parameters varied in this analysis, with the exception of the higher order biases. The most promising results are obtained using  $p = 2$ , where still some caution is required since the theoretical modeling is less accurate due to the particular features of the marked power spectrum for this case, see figure 3. This is again a confirmation of what we mentioned earlier for massive neutrinos: if the information about a cosmological parameter within the galaxy power spectrum comes both from linear and non-linear effects, then the marked power spectrum can improve the constraints obtained when combining with the



**Figure 7.** Triangle plot for massive neutrinos cosmology and biased tracers. We compare  $P + M$  with  $k_{\max}^{P,M} = 0.20 h/\text{Mpc}$  and  $P + B$  with  $k_{\max}^B = 0.10 h/\text{Mpc}$ .

bispectrum, as it is for  $M_\nu$ ,<sup>5</sup> on the other hand, non-local PNG affect the galaxy power spectrum only through a one-loop effect, and basically all the information is contained in the bispectrum, hence the marked power spectrum can only do as well as the bispectrum does, for the scales considered. It is important however to stress that our results are obtained in real space and without accounting for redshift space distortions, survey mask effects, binning, which are expected to degrade even more our results.

For biased tracers we also see, as shown in figure 9, a small degeneracy among the PNG parameters and the linear bias  $b_1$ : this is a consequence of the universality relation assumed for the non-Gaussian bias in eq. (2.65).

The corner plots presented in this work suggest that combining the power spectrum and the bispectrum with different marks would give even better results either for massive neutrinos and primordial non-gaussianities. This kind of combination is beyond the scope of this work and we will investigate it in future work accounting also for redshift space distortions and other observational effects.

<sup>5</sup>We stress again that usual analyses as in ref. [31] always compare  $P$ ,  $M$ ,  $B$  and various combinations of them at the same  $k_{\max} = 0.5 h/\text{Mpc}$ . Here we are using the maximum scale at which the EFTofLSS is valid, since going to higher  $k_{\max}$  would give biased results in realistic analyses due to exclusion of higher order loops entering in the analytical model. We are, however, in agreement with their findings.

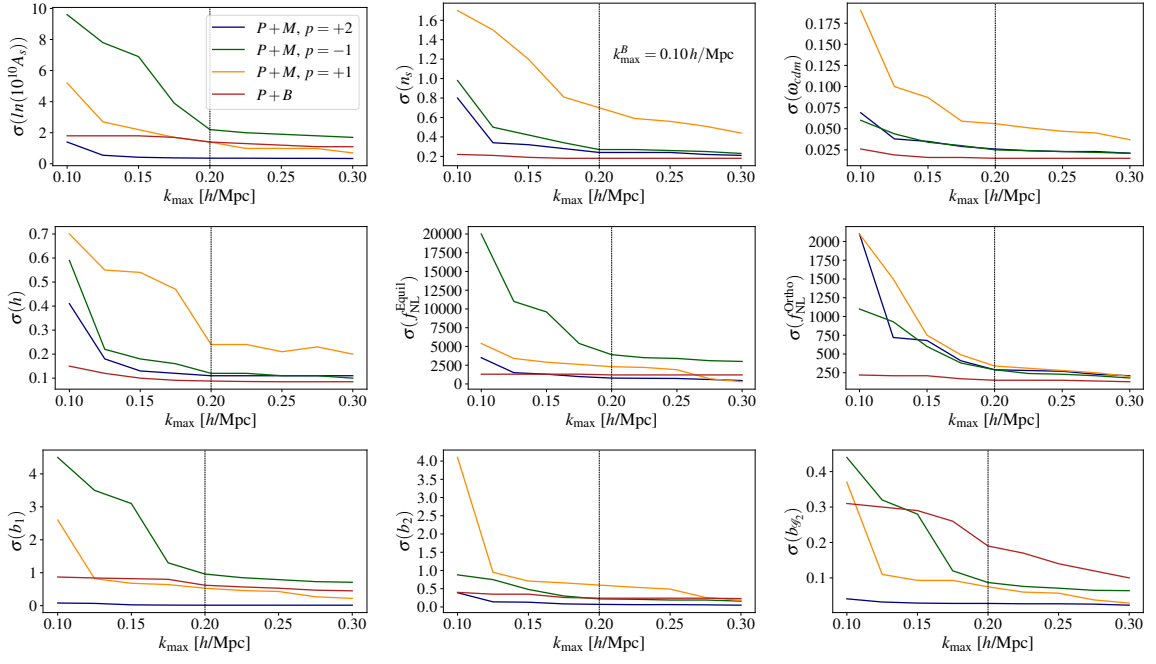


Figure 8. Same as figure 6 for cosmologies with primordial non-gaussianities.

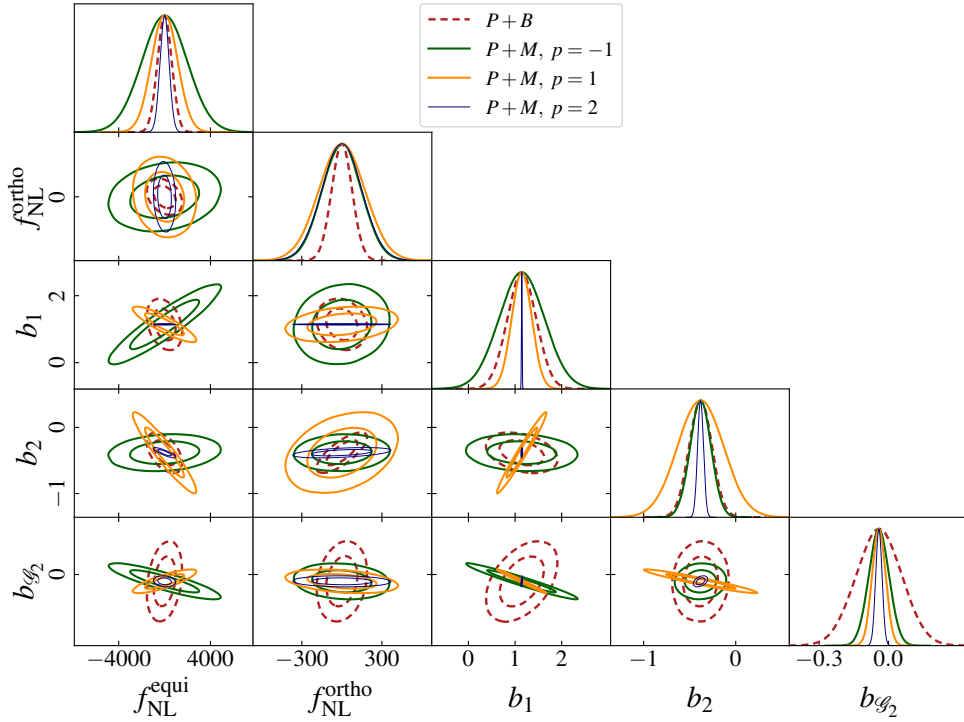


Figure 9. Same as figure 7 but for primordial non-gaussianities.

## 4 Discussion and conclusion

In this work we present the first analytical study of the marked power spectrum in the context of primordial non-Gaussianities (PNG) of the non-local type, contributing new insights into the behavior and potential applications of marked statistics in cosmology. Following the work of [32, 33] we develop a theoretically consistent model for the marked power spectrum at one-loop order adopting the Effective Field Theory of Large Scale Structure to reach mildly non-linear scales and analytically establish the information content of such statistic. When compared to the Quijote-PNG simulations [63], our quantitative results reveal a strong agreement between analytical predictions and simulations, supporting the robustness of the theoretical method adopted to describe the marked power spectrum. The model displays an excellent agreement with simulated data from low redshifts ( $z = 0$ ) up to higher ones ( $z = 1$ ), suggesting that it can be reliably adopted also for on-going and future large volume surveys, such as the DESI [84] and Euclid [85, 86] missions. The power of marked power spectrum resides in the possibility to enhance some features of the non-linear galaxy distribution, in particular by up-weighting underdense regions such as voids, where some cosmological information is expected to be found.

In this work, we compare the constraining power on non-local PNG and massive neutrinos — obtained through a combined analysis of power spectrum and marked power spectrum,  $P+M$  — with the traditional and well-established power spectrum and bispectrum,  $P+B$ , approach. We find that marking underdense regions produces the highest signal, in terms of error bars on  $M_\nu$  and both  $f_{\text{NL}}^{\text{eq}}$  and  $f_{\text{NL}}^{\text{of}}$ , in comparison to up-weighting higher density regions. Neither choice however improves over the  $P+B$  method, within the mildly non-linear scale range investigated in this work. More precisely, we find that up-weighting voids is efficient in a low shot-noise context, as it saturates the  $P+B$  signal in our pure dark matter field analysis. On the other hand, when considering realistic, discrete, biased tracers with BOSS-like galaxy number density and volume, the gain is limited and  $P+B$  turns out to be the most optimal approach. This is in agreement with previous forecasts [29, 30] and with the recent data analysis performed on the BOSS data using a simulation-based approach [87]. These results are, however, expected to improve with future low-shot-noise surveys, where underdense regions would play an important role.

Even though the  $P+M$  combination does not produce improvements over  $P+B$ , the marked power spectrum is a promising tool to constrain beyond- $\Lambda$ CDM physics, such as PNG, massive neutrinos, and modified gravity [88]. A  $P+B$  analysis based on the EFT approach still requires the implementation of an estimator for the bispectrum and to take into account a number of complex observational effects, such as binning and convolution with the survey window. Since the marked power spectrum is a two-point statistic, these effects have a less severe and well understood impact, which is much easier to keep under control. Moreover, it was recently shown that not including off-diagonal terms in the bispectrum covariance induces an underestimation of the error bars for cosmological parameters, especially for primordial non-gaussianities [77], which are the focus of this work. The diagonal approximation for the marked power spectra covariance is, on the other hand, more accurate, as tested in [29, 30]. This makes the marked power spectrum potentially more powerful at extracting NG information at non-linear scales; at the same time, this

is the reason why we adopt a Gaussian approximation for the marked power spectrum covariance in our forecasts.

In the present work we adopt a perturbative model whose validity is truncated to mildly non-linear scales  $k_{\text{max}} \simeq 0.25 h/\text{Mpc}$ , while simulation-based approaches [31, 87] could potentially reach very non-linear scales  $k_{\text{max}} = 0.50 h/\text{Mpc}$  where more information from the cosmic web could be captured. This is relevant for neutrino mass detection, modified gravity and PNG, where the small-scale, non-linear behavior offers significant leverage for parameter constraints. The main goal of this paper was to explore the information content of the marked power spectrum adopting a theoretical non-linear prescription build on perturbation theory and effective theory of large scale structure compared to standard 2- and 3-point function analyses. For this reason, we did not combine power spectra with different marks, even though this is certainly worth investigating. The combination of different marks to constrain the mass of neutrinos was explored in [30], which that combining the power spectrum with different marked power spectra can improve the constrains on the neutrino mass by a factor of 3 compared to power spectrum alone analysis. These results have to be compared with the  $P+B$  forecast performed with the same methodology and dataset in [83] where the authors find that  $P+B$  outperforms  $P + \sum_i M_i$ , in redshift space, for almost all the parameters.

Our findings also align with the recent work [34], where the authors explore new expressions for the mark function and investigate the degeneracy-breaking power of the marked power spectrum, compared to a power spectrum only analysis. In this work, we further extend this by finding, indeed, that this constraining power on higher order biases, such as  $b_2$ , of the marked power spectrum, is indeed descending from the inclusion of bispectrum-like convolution terms, that help in breaking some of the crucial degeneracies. This comparison reinforces the value of continued development in this analytical framework and further highlights the marked approach as a promising complement to established methodologies.

In future research, it will be natural to extend the present approach to redshift space, even if this will most likely not modify the core findings of this work about the amount of information gained with marked statistics. We will also need to include all the observational effects, such as Alcock-Paczynski, binning, and survey geometry, in order to build a full pipeline to use in actual data analysis, to further examine the capabilities and limitations of marked statistics in cosmological contexts. Recently, an analysis with the marked power spectrum of a subvolume of the BOSS galaxy survey was performed [87] using a simulation based approach yielding promising results. The authors accounted for survey selection effects, showing that this kind of analysis does not present major obstacles when a marked density field is used. We leave the inclusion of observational effects to future works.

Another important direction is the optimization of the mark function, as it has been performed in [35] for  $\sigma_8$  and  $\Omega_m$ . Regarding the mark itself, we limited our exploration to three values for the parameter  $p$ , in order to up-weight either voids or nodes in the cosmic structure, using a fixed expression for the mark, eq. (2.3). Our analytical approach — possibly combined with numerical techniques (such as the Gaussian process-based optimization in [35]) — should lead to a generalization of the choice of the mark function, able to maximize PNG information. Such optimization could significantly boost the utility of marked statistics for addressing key questions in cosmology and will be investigated more in future works.

$R$	15	15	15	15	15	15	15	15	15	5	10	30
$\delta_s$	0.25	1	2	3	4	0.25	0.25	0.25	0.25	0.25	0.25	0.25
$p$	1	2	2	2	2	2	-1	1.25	1.5	1	1	1

**Table 2.** Mark parameters used in this appendix.

In summary, our study contributes to provide a foundational framework for the application of marked statistics in cosmological analyses — with emphasis on PNG studies — demonstrating their compatibility with simulation data, potential for parameter estimation, and promise for further refinement. By offering a distinct perspective from traditional  $P + B$  analysis, marked statistics expand the cosmologist’s toolkit, opening new paths toward probing the properties of the primordial perturbations and the detailed composition of the universe.

## Acknowledgments

We are grateful to Thomas Flöss, Elena Massara, Azadeh Moradinezhad, Massimo Pietroni and Marco Peloso for useful discussions during the development of this work. MM is particularly grateful to Marko Simonović for his thorough support with the FFTLog. MM and ML acknowledge support by the MUR Progetti di Ricerca di Rilevante Interesse Nazionale (PRIN) Bando 2022 — grant 20228RMX4A, funded by the European Union — Next generation EU, Mission 4, Component 1, CUP C53D23000940006. GJ acknowledges support from the ANR LOCALIZATION project, grant ANR-21-CE31-0019 / 490702358 of the French Agence Nationale de la Recherche. The Flatiron Institute is supported by the Simons Foundation.

## A Effect of mark parameters

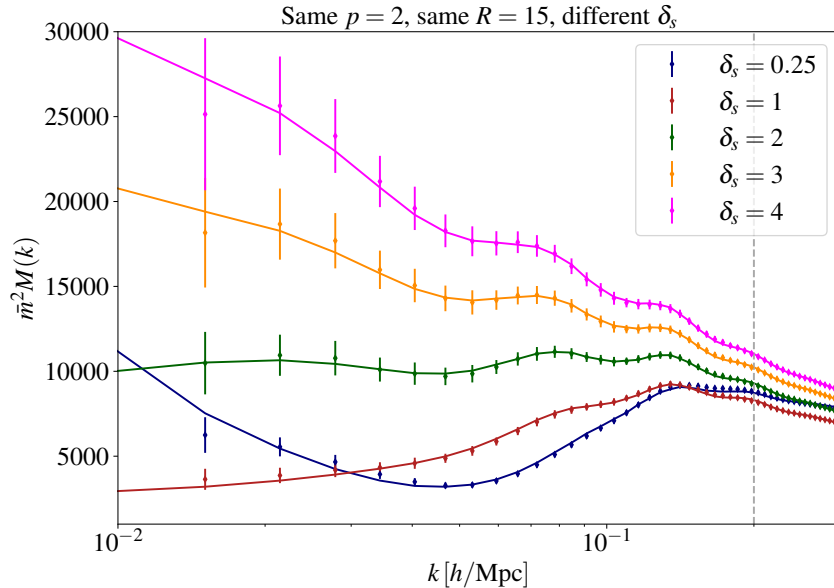
In this appendix we investigate how the marked power spectrum  $M(k)$  responds to variations in the mark parameters. We also assess the effectiveness of **perturbation theory (PT)** in reproducing the simulation data for various combinations of these parameters. We recall that we are adopting the usual prescription for the mark, defined as:

$$m(\mathbf{x}) = \left( 1 + \left( \frac{\delta_{g,R}(\mathbf{x})}{1 + \delta_s} \right) \right)^{-p} \quad (\text{A.1})$$

where  $\delta_{g,R}$  is the local smoothed density contrast at scale  $R$ ,  $\delta_s$  is the density scaling parameter that sets the effective normalization, also referred as bias, and  $p$  is the hyper-parameter controlling the strength of the down-weighting in high-density regions.

As fully explained in the main text, by tuning  $R$ ,  $\delta_s$ , and  $p$ , the mark selectively enhances the contribution from underdense or overdense regions, where nonlinear effects could be weaker and perturbation theory is expected to be more reliable. We explore separately the effect of each parameters, as they play different roles in the non-linear modeling adopted in this work. We measure the marked power spectrum from the Quijote simulations using the mark parameters combinations showed in table 2.

For each configuration, we compute the marked power spectrum  $M(k)$ , and its covariance  $C_M(k)$  from the scattering among different realizations of the same simulations. We



**Figure 10.** Effect of varying the bias  $\delta_s$ , fixing the other mark parameters  $p$  and  $R$ .

show here only the redshift  $z = 0.5$  as it is of main interest for the results presented in this paper, but we stress that the same arguments are general for each redshift considered.

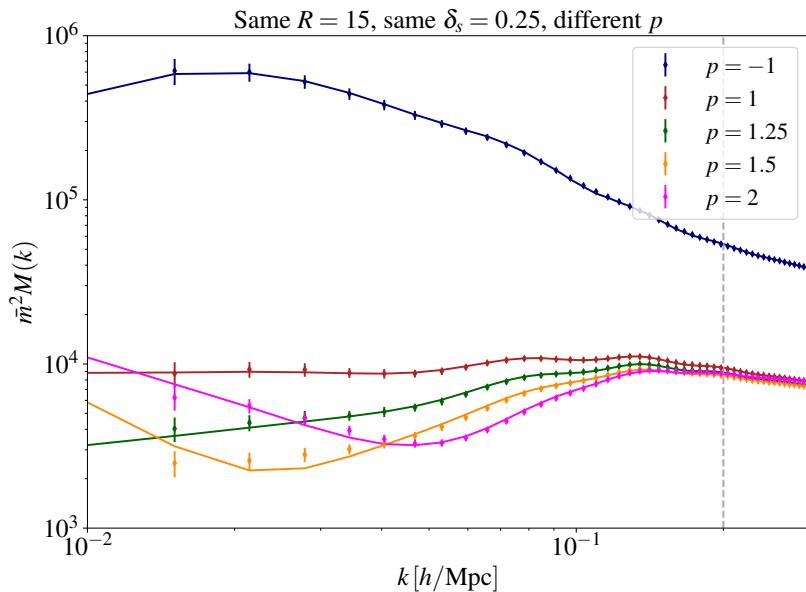
We first investigate the problematic case of  $p = 2$  that can present peculiar features that could be hard to capture using an analytical prescription, as already discussed in the main text. Eqs. (2.20) show that for some combination of the mark parameters, there are scales for which some contributions of the marked power spectrum get canceled exactly: in particular this is the case when the function  $C_{\delta_M}(k)$  goes to zero. This condition translates to

$$1 - C_1 W_R(k) = 0, \quad (\text{A.2})$$

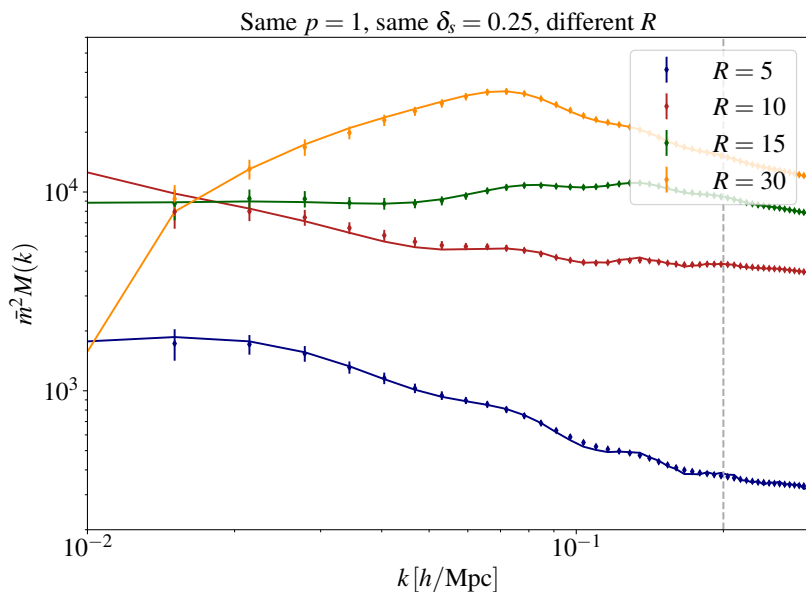
with  $C_1 = p/(1 + \delta_s)$ . This equation is always satisfied for some value of  $k$  when the combination of parameters  $\{p, \delta_s\}$  satisfies  $p \geq 1 + \delta_s$  and  $\delta_s \geq -1$ . For negative values of  $p$  and positive for  $\delta_s$  the marked power spectrum is always positive and never goes to zero. On the other hand, when  $p = 2$ , if  $-1 \leq \delta_s \leq 1$ , then some contributions of the marked power spectrum go to zero and the theoretical modeling becomes more challenging and less accurate. This is shown explicitly in figure 10, where  $p = 2$  and  $R = 15$  are kept fixed and  $\delta_s$  is varied. For  $\delta_s = 0.25$ ,  $C_{\delta_M}$  goes zero for  $k \simeq 0.06 h/\text{Mpc}$ , which are the scales that are poorly fitted by our theoretical model. For values  $\delta_s \geq 1$ , we can see that the perturbative modeling fits perfectly with the simulation data, up to mildly non-linear scales.

We also consider the effect of varying the exponent  $p$ , keeping  $\delta_s = 0.25$  fixed, as shown in figure 11. The condition on  $p$  to not have unwanted zero-crossing of the marked power spectrum is  $p \leq 1.25$ . In fact, we see that the cases with  $p = 2$  and  $p = 1.5$  are problematic and the theoretical model is not accurate enough to fit the larger scales. For smaller values of  $p$ , the theoretical predictions sits almost perfectly on top of the simulations data.

It is also worth investigating the role of the smoothing scale  $R$ . The condition to have a well-behaved perturbative series up to the scales we are considering, is given by eq. (2.8), which



**Figure 11.** Effect of varying the exponent scale  $p$ , fixing the other mark parameters  $R$  and  $\delta_s$ . We note that for  $p \geq 1.5$  the theoretical model is less accurate due to the marked power spectrum features that appear in this mark combinations.



**Figure 12.** Effect of varying the smoothing scale  $R$ , fixing the other mark parameters  $p$  and  $\delta_s$ .

is strongly affected by  $R$ . Smaller  $R$  values make the fluctuations larger and, consequently, give a larger value for  $\delta_{RR}$ . This implies that smaller  $R$  cause the perturbative series to converge more slowly. Despite this, we observe that the agreement within theoretical model and data is in exquisite agreement for the scales analyzed, as shown in figure 12.

## References

- [1] H. Gil-Marín et al., *The power spectrum and bispectrum of SDSS DR11 BOSS galaxies — I. Bias and gravity*, *Mon. Not. Roy. Astron. Soc.* **451** (2015) 539 [[arXiv:1407.5668](#)] [[INSPIRE](#)].
- [2] D. Gualdi, H. Gil-Marín and L. Verde, *Joint analysis of anisotropic power spectrum, bispectrum and trispectrum: application to N-body simulations*, *JCAP* **07** (2021) 008 [[arXiv:2104.03976](#)] [[INSPIRE](#)].
- [3] G. Cabass et al., *Constraints on Single-Field Inflation from the BOSS Galaxy Survey*, *Phys. Rev. Lett.* **129** (2022) 021301 [[arXiv:2201.07238](#)] [[INSPIRE](#)].
- [4] G. Cabass et al., *Constraints on multifield inflation from the BOSS galaxy survey*, *Phys. Rev. D* **106** (2022) 043506 [[arXiv:2204.01781](#)] [[INSPIRE](#)].
- [5] G. D’Amico, M. Lewandowski, L. Senatore and P. Zhang, *Limits on primordial non-Gaussianities from BOSS galaxy-clustering data*, *Phys. Rev. D* **111** (2025) 063514 [[arXiv:2201.11518](#)] [[INSPIRE](#)].
- [6] G. D’Amico et al., *The BOSS bispectrum analysis at one loop from the Effective Field Theory of Large-Scale Structure*, *JCAP* **05** (2024) 059 [[arXiv:2206.08327](#)] [[INSPIRE](#)].
- [7] G. D’Amico et al., *The one-loop bispectrum of galaxies in redshift space from the Effective Field Theory of Large-Scale Structure*, *JCAP* **07** (2024) 041 [[arXiv:2211.17130](#)] [[INSPIRE](#)].
- [8] C.H. Hahn et al., *Cosmological constraints from non-Gaussian and nonlinear galaxy clustering using the SimBIG inference framework*, *Nature Astron.* **8** (2024) 1457 [[arXiv:2310.15246](#)] [[INSPIRE](#)].
- [9] M.M. Ivanov et al., *Cosmology with the galaxy bispectrum multipoles: Optimal estimation and application to BOSS data*, *Phys. Rev. D* **107** (2023) 083515 [[arXiv:2302.04414](#)] [[INSPIRE](#)].
- [10] SIMBIG collaboration, *Cosmological constraints from the nonlinear galaxy bispectrum*, *Phys. Rev. D* **109** (2024) 083534 [[arXiv:2310.15243](#)] [[INSPIRE](#)].
- [11] J. Byun and E. Krause, *Modal compression of the redshift-space galaxy bispectrum*, *Mon. Not. Roy. Astron. Soc.* **525** (2023) 4854 [[arXiv:2205.04579](#)] [[INSPIRE](#)].
- [12] W.R. Coulton et al., *Quijote-PNG: The Information Content of the Halo Power Spectrum and Bispectrum*, *Astrophys. J.* **943** (2023) 178 [[arXiv:2206.15450](#)] [[INSPIRE](#)].
- [13] G. Jung et al., *Quijote-PNG: Quasi-maximum Likelihood Estimation of Primordial Non-Gaussianity in the Nonlinear Halo Density Field*, *Astrophys. J.* **948** (2023) 135 [[arXiv:2211.07565](#)] [[INSPIRE](#)].
- [14] SIMBIG collaboration, *Galaxy clustering analysis with SimBIG and the wavelet scattering transform*, *Phys. Rev. D* **109** (2024) 083535 [[arXiv:2310.15250](#)] [[INSPIRE](#)].
- [15] G. Valogiannis and C. Dvorkin, *Towards an optimal estimation of cosmological parameters with the wavelet scattering transform*, *Phys. Rev. D* **105** (2022) 103534 [[arXiv:2108.07821](#)] [[INSPIRE](#)].
- [16] G. Valogiannis, F. Villaescusa-Navarro and M. Baldi, *Towards unveiling the large-scale nature of gravity with the wavelet scattering transform*, *JCAP* **11** (2024) 061 [[arXiv:2407.18647](#)] [[INSPIRE](#)].
- [17] G. Valogiannis, S. Yuan and C. Dvorkin, *Precise cosmological constraints from BOSS galaxy clustering with a simulation-based emulator of the wavelet scattering transform*, *Phys. Rev. D* **109** (2024) 103503 [[arXiv:2310.16116](#)] [[INSPIRE](#)].

- [18] M. Eickenberg et al., *Wavelet Moments for Cosmological Parameter Estimation*, [arXiv:2204.07646](#) [INSPIRE].
- [19] M. Peron, G. Jung, M. Liguori and M. Pietroni, *Constraining primordial non-Gaussianity from large scale structure with the wavelet scattering transform*, *JCAP* **07** (2024) 021 [[arXiv:2403.17657](#)] [INSPIRE].
- [20] A. Banerjee and T. Abel, *Tracer-field cross-correlations with  $k$ -nearest neighbour distributions*, *Mon. Not. Roy. Astron. Soc.* **519** (2023) 4856 [[arXiv:2210.05140](#)] [INSPIRE].
- [21] W.R. Coulton, T. Abel and A. Banerjee, *Small-scale signatures of primordial non-Gaussianity in  $k$ -nearest neighbour cumulative distribution functions*, *Mon. Not. Roy. Astron. Soc.* **534** (2024) 1621 [[arXiv:2309.15151](#)] [INSPIRE].
- [22] M. Schmittfull and A. Moradinezhad Dizgah, *Galaxy skew-spectra in redshift-space*, *JCAP* **03** (2021) 020 [[arXiv:2010.14267](#)] [INSPIRE].
- [23] J. Hou et al., *Cosmological constraints from the redshift-space galaxy skew spectra*, *Phys. Rev. D* **109** (2024) 103528 [[arXiv:2401.15074](#)] [INSPIRE].
- [24] F. Simpson, A.F. Heavens and C. Heymans, *Clipping the cosmos. II. Cosmological information from nonlinear scales*, *Phys. Rev. D* **88** (2013) 083510 [[arXiv:1306.6349](#)] [INSPIRE].
- [25] M. Lippich and A.G. Sánchez, *MEDUSA: Minkowski functionals estimated from Delaunay tessellations of the three-dimensional large-scale structure*, *Mon. Not. Roy. Astron. Soc.* **508** (2021) 3771 [[arXiv:2012.08529](#)] [INSPIRE].
- [26] W. Liu, A. Jiang and W. Fang, *Probing massive neutrinos with the Minkowski functionals of the galaxy distribution*, *JCAP* **09** (2023) 037 [[arXiv:2302.08162](#)] [INSPIRE].
- [27] D. Stoyan, *On Correlations of Marked Point Processes*, *Math. Nachr.* **116** (1984) 197.
- [28] M. White, *A marked correlation function for constraining modified gravity models*, *JCAP* **11** (2016) 057 [[arXiv:1609.08632](#)] [INSPIRE].
- [29] E. Massara et al., *Using the Marked Power Spectrum to Detect the Signature of Neutrinos in Large-Scale Structure*, *Phys. Rev. Lett.* **126** (2021) 011301 [[arXiv:2001.11024](#)] [INSPIRE].
- [30] E. Massara et al., *Cosmological Information in the Marked Power Spectrum of the Galaxy Field*, *Astrophys. J.* **951** (2023) 70 [[arXiv:2206.01709](#)] [INSPIRE].
- [31] G. Jung et al., *Quijote-PNG: Optimizing the Summary Statistics to Measure Primordial Non-Gaussianity*, *Astrophys. J.* **976** (2024) 109 [[arXiv:2403.00490](#)] [INSPIRE].
- [32] O.H.E. Philcox, E. Massara and D.N. Spergel, *What does the marked power spectrum measure? Insights from perturbation theory*, *Phys. Rev. D* **102** (2020) 043516 [[arXiv:2006.10055](#)] [INSPIRE].
- [33] O.H.E. Philcox, A. Aviles and E. Massara, *Modeling the Marked Spectrum of Matter and Biased Tracers in Real- and Redshift-Space*, *JCAP* **03** (2021) 038 [[arXiv:2010.05914](#)] [INSPIRE].
- [34] H. Ebina and M. White, *An analytically tractable marked power spectrum*, *JCAP* **01** (2025) 150 [[arXiv:2409.17133](#)] [INSPIRE].
- [35] J.A. Cowell, D. Alonso and J. Liu, *Optimizing marked power spectra for cosmology*, *Mon. Not. Roy. Astron. Soc.* **535** (2024) 3129 [[arXiv:2409.05695](#)] [INSPIRE].
- [36] BEYOND-2PT collaboration, *A Parameter-Masked Mock Data Challenge for Beyond-Two-Point Galaxy Clustering Statistics*, [arXiv:2405.02252](#) [INSPIRE].

- [37] A. Aviles et al., *Marked correlation functions in perturbation theory*, *JCAP* **01** (2020) 006 [[arXiv:1911.06362](#)] [[INSPIRE](#)].
- [38] F. Bernardeau, S. Colombi, E. Gaztanaga and R. Scoccimarro, *Large scale structure of the universe and cosmological perturbation theory*, *Phys. Rept.* **367** (2002) 1 [[astro-ph/0112551](#)] [[INSPIRE](#)].
- [39] M.M. Ivanov, M. Simonović and M. Zaldarriaga, *Cosmological Parameters from the BOSS Galaxy Power Spectrum*, *JCAP* **05** (2020) 042 [[arXiv:1909.05277](#)] [[INSPIRE](#)].
- [40] G. D’Amico, M. Marinucci, M. Pietroni and F. Vernizzi, *The large scale structure bootstrap: perturbation theory and bias expansion from symmetries*, *JCAP* **10** (2021) 069 [[arXiv:2109.09573](#)] [[INSPIRE](#)].
- [41] V. Assassi, D. Baumann, D. Green and M. Zaldarriaga, *Renormalized Halo Bias*, *JCAP* **08** (2014) 056 [[arXiv:1402.5916](#)] [[INSPIRE](#)].
- [42] M.M. Ivanov et al., *Precision analysis of the redshift-space galaxy bispectrum*, *Phys. Rev. D* **105** (2022) 063512 [[arXiv:2110.10161](#)] [[INSPIRE](#)].
- [43] O.H.E. Philcox and M.M. Ivanov, *BOSS DR12 full-shape cosmology:  $\Lambda$ CDM constraints from the large-scale galaxy power spectrum and bispectrum monopole*, *Phys. Rev. D* **105** (2022) 043517 [[arXiv:2112.04515](#)] [[INSPIRE](#)].
- [44] O.H.E. Philcox et al., *Cosmology with the redshift-space galaxy bispectrum monopole at one-loop order*, *Phys. Rev. D* **106** (2022) 043530 [[arXiv:2206.02800](#)] [[INSPIRE](#)].
- [45] D. Baumann, A. Nicolis, L. Senatore and M. Zaldarriaga, *Cosmological Non-Linearities as an Effective Fluid*, *JCAP* **07** (2012) 051 [[arXiv:1004.2488](#)] [[INSPIRE](#)].
- [46] J.J.M. Carrasco, M.P. Hertzberg and L. Senatore, *The Effective Field Theory of Cosmological Large Scale Structures*, *JHEP* **09** (2012) 082 [[arXiv:1206.2926](#)] [[INSPIRE](#)].
- [47] M. Pietroni, G. Mangano, N. Saviano and M. Viel, *Coarse-Grained Cosmological Perturbation Theory*, *JCAP* **01** (2012) 019 [[arXiv:1108.5203](#)] [[INSPIRE](#)].
- [48] L. Senatore and M. Zaldarriaga, *The IR-resummed Effective Field Theory of Large Scale Structures*, *JCAP* **02** (2015) 013 [[arXiv:1404.5954](#)] [[INSPIRE](#)].
- [49] T. Baldauf, M. Mirbabayi, M. Simonović and M. Zaldarriaga, *Equivalence Principle and the Baryon Acoustic Peak*, *Phys. Rev. D* **92** (2015) 043514 [[arXiv:1504.04366](#)] [[INSPIRE](#)].
- [50] M.M. Ivanov and S. Sibiryakov, *Infrared Resummation for Biased Tracers in Redshift Space*, *JCAP* **07** (2018) 053 [[arXiv:1804.05080](#)] [[INSPIRE](#)].
- [51] A. Chudaykin and M.M. Ivanov, *Measuring neutrino masses with large-scale structure: Euclid forecast with controlled theoretical error*, *JCAP* **11** (2019) 034 [[arXiv:1907.06666](#)] [[INSPIRE](#)].
- [52] C. Cheung et al., *The Effective Field Theory of Inflation*, *JHEP* **03** (2008) 014 [[arXiv:0709.0293](#)] [[INSPIRE](#)].
- [53] S. Matarrese and L. Verde, *The effect of primordial non-Gaussianity on halo bias*, *Astrophys. J. Lett.* **677** (2008) L77 [[arXiv:0801.4826](#)] [[INSPIRE](#)].
- [54] V. Desjacques, U. Seljak and I. Iliev, *Scale-dependent bias induced by local non-Gaussianity: A comparison to N-body simulations*, *Mon. Not. Roy. Astron. Soc.* **396** (2009) 85 [[arXiv:0811.2748](#)] [[INSPIRE](#)].
- [55] N. Dalal, O. Dore, D. Huterer and A. Shirokov, *The imprints of primordial non-gaussianities on large-scale structure: scale dependent bias and abundance of virialized objects*, *Phys. Rev. D* **77** (2008) 123514 [[arXiv:0710.4560](#)] [[INSPIRE](#)].

- [56] U. Giri, M. Münchmeyer and K.M. Smith, *Robust neural network-enhanced estimation of local primordial non-Gaussianity*, *Phys. Rev. D* **107** (2023) L061301 [[arXiv:2205.12964](#)] [[INSPIRE](#)].
- [57] Y. Kvasiuk, M. Münchmeyer and K. Smith, *Two-field formalism for a neural network-enhanced non-Gaussianity search with halos*, *Phys. Rev. D* **112** (2025) 023540 [[arXiv:2410.01007](#)] [[INSPIRE](#)].
- [58] N. Arkani-Hamed and J. Maldacena, *Cosmological Collider Physics*, [arXiv:1503.08043](#) [[INSPIRE](#)].
- [59] G. Cabass et al., *BOSS constraints on massive particles during inflation: The cosmological collider in action*, *Phys. Rev. D* **111** (2025) 063510 [[arXiv:2404.01894](#)] [[INSPIRE](#)].
- [60] L. Senatore, K.M. Smith and M. Zaldarriaga, *Non-Gaussianities in Single Field Inflation and their Optimal Limits from the WMAP 5-year Data*, *JCAP* **01** (2010) 028 [[arXiv:0905.3746](#)] [[INSPIRE](#)].
- [61] PLANCK collaboration, *Planck 2018 results. IX. Constraints on primordial non-Gaussianity*, *Astron. Astrophys.* **641** (2020) A9 [[arXiv:1905.05697](#)] [[INSPIRE](#)].
- [62] D. Babich, P. Creminelli and M. Zaldarriaga, *The shape of non-Gaussianities*, *JCAP* **08** (2004) 009 [[astro-ph/0405356](#)] [[INSPIRE](#)].
- [63] W.R. Coulton et al., *Quijote-PNG: Simulations of Primordial Non-Gaussianity and the Information Content of the Matter Field Power Spectrum and Bispectrum*, *Astrophys. J.* **943** (2023) 64 [[arXiv:2206.01619](#)] [[INSPIRE](#)].
- [64] PLANCK collaboration, *Planck 2018 results. X. Constraints on inflation*, *Astron. Astrophys.* **641** (2020) A10 [[arXiv:1807.06211](#)] [[INSPIRE](#)].
- [65] F. Schmidt and M. Kamionkowski, *Halo Clustering with Non-Local Non-Gaussianity*, *Phys. Rev. D* **82** (2010) 103002 [[arXiv:1008.0638](#)] [[INSPIRE](#)].
- [66] F. Villaescusa-Navarro et al., *The Quijote simulations*, *Astrophys. J. Suppl.* **250** (2020) 2 [[arXiv:1909.05273](#)] [[INSPIRE](#)].
- [67] M. Crocce, S. Pueblas and R. Scoccimarro, *Transients from Initial Conditions in Cosmological Simulations*, *Mon. Not. Roy. Astron. Soc.* **373** (2006) 369 [[astro-ph/0606505](#)] [[INSPIRE](#)].
- [68] R. Scoccimarro, L. Hui, M. Manera and K.C. Chan, *Large-scale Bias and Efficient Generation of Initial Conditions for Non-Local Primordial Non-Gaussianity*, *Phys. Rev. D* **85** (2012) 083002 [[arXiv:1108.5512](#)] [[INSPIRE](#)].
- [69] M. Davis, G. Efstathiou, C.S. Frenk and S.D.M. White, *The Evolution of Large Scale Structure in a Universe Dominated by Cold Dark Matter*, *Astrophys. J.* **292** (1985) 371 [[INSPIRE](#)].
- [70] G. Jung et al., *Quijote-PNG: The Information Content of the Halo Mass Function*, *Astrophys. J.* **957** (2023) 50 [[arXiv:2305.10597](#)] [[INSPIRE](#)].
- [71] V. Desjacques, D. Jeong and F. Schmidt, *Large-Scale Galaxy Bias*, *Phys. Rept.* **733** (2018) 1 [[arXiv:1611.09787](#)] [[INSPIRE](#)].
- [72] V. Yankelevich and C. Porciani, *Cosmological information in the redshift-space bispectrum*, *Mon. Not. Roy. Astron. Soc.* **483** (2019) 2078 [[arXiv:1807.07076](#)] [[INSPIRE](#)].
- [73] M.M. Abidi and T. Baldauf, *Cubic Halo Bias in Eulerian and Lagrangian Space*, *JCAP* **07** (2018) 029 [[arXiv:1802.07622](#)] [[INSPIRE](#)].
- [74] A. Chudaykin, M.M. Ivanov, O.H.E. Philcox and M. Simonović, *Nonlinear perturbation theory extension of the Boltzmann code CLASS*, *Phys. Rev. D* **102** (2020) 063533 [[arXiv:2004.10607](#)] [[INSPIRE](#)].

- [75] D. Blas, J. Lesgourgues and T. Tram, *The Cosmic Linear Anisotropy Solving System (CLASS) II: Approximation schemes*, *JCAP* **07** (2011) 034 [[arXiv:1104.2933](#)] [[INSPIRE](#)].
- [76] M.M. Ivanov, M. Simonović and M. Zaldarriaga, *Cosmological Parameters and Neutrino Masses from the Final Planck and Full-Shape BOSS Data*, *Phys. Rev. D* **101** (2020) 083504 [[arXiv:1912.08208](#)] [[INSPIRE](#)].
- [77] T. Flöss, M. Biagetti and P.D. Meerburg, *Primordial non-Gaussianity and non-Gaussian covariance*, *Phys. Rev. D* **107** (2023) 023528 [[arXiv:2206.10458](#)] [[INSPIRE](#)].
- [78] J. Lesgourgues and S. Pastor, *Massive neutrinos and cosmology*, *Phys. Rept.* **429** (2006) 307 [[astro-ph/0603494](#)] [[INSPIRE](#)].
- [79] A. Lewis, *The real shape of non-Gaussianities*, *JCAP* **10** (2011) 026 [[arXiv:1107.5431](#)] [[INSPIRE](#)].
- [80] M.M. Ivanov et al., *Full-shape analysis with simulation-based priors: Constraints on single field inflation from BOSS*, *Phys. Rev. D* **110** (2024) 063538 [[arXiv:2402.13310](#)] [[INSPIRE](#)].
- [81] M.M. Ivanov, A. Obuljen, C. Cuesta-Lazaro and M.W. Toomey, *Full-shape analysis with simulation-based priors: Cosmological parameters and the structure growth anomaly*, *Phys. Rev. D* **111** (2025) 063548 [[arXiv:2409.10609](#)] [[INSPIRE](#)].
- [82] H. Zhang et al., *HOD-informed prior for EFT-based full-shape analyses of LSS*, *JCAP* **04** (2025) 041 [[arXiv:2409.12937](#)] [[INSPIRE](#)].
- [83] C. Hahn and F. Villaescusa-Navarro, *Constraining  $M_\nu$  with the bispectrum. Part II. The information content of the galaxy bispectrum monopole*, *JCAP* **04** (2021) 029 [[arXiv:2012.02200](#)] [[INSPIRE](#)].
- [84] DESI collaboration, *The DESI Experiment Part I: Science, Targeting, and Survey Design*, [arXiv:1611.00036](#) [[INSPIRE](#)].
- [85] A. Cimatti et al., *Euclid Assessment Study Report for the ESA Cosmic Visions*, [arXiv:0912.0914](#) [[INSPIRE](#)].
- [86] EUCLID collaboration, *Euclid. I. Overview of the Euclid mission*, *Astron. Astrophys.* **697** (2025) A1 [[arXiv:2405.13491](#)] [[INSPIRE](#)].
- [87] E. Massara et al., *SIMBIG: Cosmological Constraints using Simulation-Based Inference of Galaxy Clustering with Marked Power Spectra*, [arXiv:2404.04228](#) [[INSPIRE](#)].
- [88] M. Kärcher, J. Bel and S. de la Torre, *Towards an optimal marked correlation function analysis for the detection of modified gravity*, *Astron. Astrophys.* **694** (2025) A253 [[arXiv:2406.02504](#)] [[INSPIRE](#)].