

Void probability function inside cosmic voids: evidence for hierarchical scaling of high-order correlations in real space

Federico Dávila-Kurbán ¹★, Andrés N. Ruiz ^{1,2}, Dante J. Paz ^{1,2} and Diego Garcia Lambas^{1,2}

¹*Instituto de Astronomía Teórica y Experimental, CONICET-UNC, Laprida 854, X5000BGR, Córdoba, Argentina*

²*Observatorio Astronómico, Universidad Nacional de Córdoba, Laprida 854, X5000BGR, Córdoba, Argentina*

Accepted 2023 November 22. Received 2023 November 22; in original form 2023 September 29

ABSTRACT

We compare the reduced void probability function (VPF) inside and outside of cosmic voids in the TNG300-1 simulation, both in real and simulated redshift space. The VPF is a special case of the Counts-in-Cells approach for extracting information of high-order clustering that is crucial for a full understanding of the distribution of galaxies. Previous studies have validated the hierarchical scaling paradigm of galaxy clustering moments, in good agreement with the ‘negative binomial’ model, in redshift surveys but have also reported that this paradigm is not valid in real space. However, in this work, we find that hierarchical scaling can indeed be found in real space inside cosmic voids. This is well fitted by the negative binomial model. We find this result to be robust against changes in void identification, galaxy mass, random dilutions, and redshift. We also obtain that the VPF in real space at high redshift approaches the negative binomial model, and, therefore, it is similar to the VPF inside voids at the present time. This study points, for the first time, towards evidence of hierarchical scaling of high-order clustering of galaxies in real space inside voids, preserving the pristine structure formation processes of the Universe.

Key words: methods: numerical – methods: statistical – cosmology: theory – large-scale structure of Universe.

1 INTRODUCTION

The void probability function (VPF) estimates the probability of finding zero galaxies in a given volume, and it is a special case of the Counts-in-Cells (CiC) analysis. This approach to clustering statistics has been used to estimate higher order correlation functions (Coles & Jones 1991; Baugh et al. 2004; Croton et al. 2004b), which are increasingly difficult to obtain, in order to extract more information than is available in the more traditionally used two-point correlation function. A further understanding of these higher-orders of correlation is needed for a full description of the clustering of galaxies and the small- and large-scale structures it originates. This work aims to characterize the VPF inside and outside cosmic voids, in real and redshift space, and to study what information this underused statistic can provide.

An alternative method for characterizing the clustering of a sample has been the CiC analysis, which yields the count probability distribution function (CPDF). This approach focusses on the computation of the probability of finding N objects in a volume V , $P_N(V)$. If one can calculate this statistic for all N , then, in principle, one can fully characterize the distribution of the sample. The VPF is a particular case of this approach where $N = 0$, and its main appeal has been its theoretical relation to all high-order correlation functions (White 1979; Sharp 1981; Fry 1986; Balian et al. 1989; Gaztanaga & Yokoyama 1993; Vogeley et al. 1994; Benson et al. 2003). Whenever this relation is valid, the VPF can provide a pathway

to the information encoded in every correlation order that would otherwise be very costly to obtain.

The calculation of the VPF of a given set of data has been often contrasted in the literature with known hierarchical clustering models (e.g. Croton et al. 2004a; Mekjian 2007; Fry & Colombi 2013). These models can be understood as recipes for writing the scaling coefficients that parametrize the hierarchy of the correlation functions. In particular the Negative Binomial model (NB, hereafter) has been shown to be a very good fit for the VPF of galaxies in redshift surveys (Elizalde & Gaztanaga 1992; Conroy et al. 2005; Croton et al. 2006, and references above), while the Thermodynamic model provides a less accurate fit for some dark matter samples and galaxies in certain cosmological models (Vogeley et al. 1994; Croton et al. 2004a). However, despite evidence of the success of the NB model as a fit for the VPF of galaxies in redshift surveys, and therefore confirming the hierarchical scaling of the clustering correlations, one main problem remains: this behaviour does not hold in real space. For example, Vogeley et al. (1994) simulate real and redshift space with different cosmologies and confirms the idea (previously reported as well by Lahav et al. 1993) that the hierarchical scaling is only valid in redshift space, i.e. when the clustering pattern is distorted by peculiar velocities. However, Croton et al. (2006) find that fainter and redder galaxies move away from the NB model and so disputes the idea that the distortion of peculiar velocities is the main reason for the success of the NB model.

Similar reports for the success of the NB model have been presented for the entire CiC treatment. Using data from the Sloan Digital Sky Survey (SDSS), Hurtado-Gil et al. (2017) conclude that the NB model is the best fit for the CiC distribution: Not only for the probability P_0 of finding zero galaxies but also for the probability P_N

* E-mail: fdavilakurban@unc.edu.ar

of finding N galaxies with $N \geq 1$, while the log-normal distribution (a model with similar behaviour to that of the Thermodynamic) is, with some modifications, a justifiable alternative for cells spanning large scales.

However, there are some deviations from the NB model. Croton et al. (2006) find that red galaxies of the SDSS have stronger clustering strength than blue galaxies at the orders they measured and, therefore, stray upwards from the NB fit. In agreement with this result Tinker et al. (2008) discovered that occupation functions resulting in correlation functions that significantly deviate from a power law tend to stray from the NB model. They argue that the reason for this departure is the prevalence of satellite galaxies within the red occupation function, which leads to a distinct shift from the one- to two-halo regime. On the other hand, the correlation function for the blue sample closely follows a power law, and is therefore well fitted by the NB model.

It is also important to note that the Ansatz of hierarchical scaling is obtained with two key assumptions: self-similarity and stable clustering (Bernardeau et al. 2002). Stable clustering implies that the coefficients are constant with time, while self-similarity leads to coefficients that are constant with scale. Both of these approximations are simplistic, but they have been useful in understanding the behaviour of small-scale correlations. Therefore, it is to be expected that the Ansatz of hierarchical scaling is true only in regimes in which these approximations have reasonable validity.

Authors such as Conroy et al. (2005) and Tinker, Weinberg & Warren (2006) argue that the VPF is completely determined by the number density and the volume-averaged two-point correlation function of the sample, so it does not necessarily provide more information than these statistics and should be used as a complementary method. Nevertheless, the CiC is a simple but powerful statistical approach to characterize the distribution of galaxies in space, since it contains statistical information about voids and subdense regions, clusters of various sizes and shapes, filaments, about the probability of finding an arbitrary number of neighbours around random positions, about counts of galaxies in randomly positioned cells with arbitrary shapes and random sizes, and about the correlation function of galaxies of all orders (Yang & Saslaw 2011, and references therein). And despite the wealth of information about galaxy clustering this statistic contains, it has not received as much attention as more traditional clustering statistics such as the two-point correlation function, not in small part because of the difficulty in extracting from it a straightforward interpretation. The VPF alone may be insufficient to fully describe the galaxy distribution, but it is a complementary statistic that is part of the CiC approach that holds great potential. Moreover, the void distribution is relatively insensitive to information on large cell sizes because large cells are unlikely to be completely empty (Yang & Saslaw 2011). This makes cosmic voids a particularly good environment for using the VPF.

With this work, we aim to shed some light on the information encoded in the VPF on different environments, in particular inside void regions in comparison with the general Universe. We focus on how this statistic is sensitive to different tracers, different redshifts and the different expansion history provided by void environments.

Following this introduction, Section 2 provides a brief overview of the statistics needed for the calculations performed, including necessary definitions and uncertainties estimations. Section 3 details the simulation used and the void identification algorithm. Results are shown and analysed in Section 4 and finally, in Section 5, we summarize this work while discussing our findings.

2 METHODS

2.1 Statistics overview

Given a distribution of galaxies, the count probability distribution function (CPDF), $P_N(R)$, describes the probability of finding N galaxies in a radius R . It can be shown that the VPF, or the probability of finding zero galaxies $P_0(R)$, can be related to the hierarchy of p -order correlation functions (White 1979):

$$P_0(R) = \exp \left\{ \sum_{p=1}^{\infty} \frac{[-\bar{N}(R)]^p}{p!} \bar{\xi}_p(R) \right\}, \quad (1)$$

where $\bar{N}(R)$ is the mean number of galaxies in a sphere with radius R , and $\bar{\xi}_p$ is the volume-averaged p th-order correlation function:

$$\bar{\xi}_p = \frac{\int \xi_p dV}{\int dV}. \quad (2)$$

The expression in equation (1) can be simplified when inserting the ‘hierarchical scaling’ Ansatz:

$$\bar{\xi}_p(R) = S_p \bar{\xi}_2^{p-1}(R), \quad (3)$$

which then gives the expression:

$$P_0(R) = \exp \left[\sum_{p=1}^{\infty} \frac{[-\bar{N}(R)]^p}{p!} S_p \bar{\xi}_2^{p-1}(R) \right]. \quad (4)$$

The aforementioned Ansatz postulates that the higher order correlations ξ_p are related to the lowest order of correlation ξ_2 (omitting the subscript hereafter for simplicity) by way of constant coefficients S_p . The concept of hierarchical emergence of the higher-order clustering from the two-point correlation function is natural in both perturbation theory and the highly non-linear regime of gravitational clustering (Peebles 1980). This idea has been strongly reinforced by a wealth of observational evidence (see, e.g. the review of Bernardeau et al. 2002 and references therein).

Finally, it can be seen from equation (4) that P_0 can be written in terms of \bar{N} and $\bar{\xi}$ alone. This is the so-called ‘scaling variable’, $\bar{N}\bar{\xi}$. This quantity can be roughly thought of as the number of objects in excess with respect to a randomly distributed sample for a given volume. Naturally, this variable $\bar{N}\bar{\xi}$ has a finite range, depending on the sample, e.g. this number can never grow too large on a sample with no clustering. Fry (1986) formalizes the concept of writing the VPF in terms of the scaling variable by introducing χ , the *reduced* VPF (RVPF), defined as

$$\chi = -\ln(P_0)/\bar{N}, \quad (5)$$

so that when combined with equation (4), we obtain

$$\chi(\bar{N}\bar{\xi}) = \sum_{p=1}^{\infty} \frac{S_p}{p!} (-\bar{N}\bar{\xi})^{p-1}. \quad (6)$$

This expression shows that, under the assumption of the scaling relation, we can anticipate a convergence of different galaxy samples with varying densities and clustering strengths onto a universal curve. This convergence occurs because of the shared dependence on a common scaling variable. If this holds, it also follows that differences in the shape of VPFs of galaxy samples must arise only from differences between their higher order correlation functions, i.e. differences in their S_p . Conversely, if changing the density, correlation strength or some other parameter of a sample, yields vastly different VPFs, that could indicate that the coefficients S_p are not constant and indeed depend strongly on the parameter that is

being tested. This scenario would invalidate the hierarchical Ansatz of equation (3).

Different ways of fixing the scaling coefficients S_p constitute different models of hierarchical scaling. For this work, we will use the two models that more closely fit our data: the negative binomial model, and the thermodynamic model. Previous works have shown that the NB model is a very good fit for the VPF of galaxies in redshift surveys, while the thermodynamic model has been proposed as a loose fit for some dark matter samples and galaxies in a few obsolete cosmological models (Vogeley et al. 1994; Croton et al. 2004a, 2006; Conroy et al. 2005; Fry & Colombi 2013). The RVPFs and scaling coefficients of the NB and thermodynamical models are, respectively,

$$\chi_{\text{NB}} = \ln(1 + \bar{N}\bar{\xi})/\bar{N}\bar{\xi}; \text{ with } S_p = (p-1)!,$$

and

$$\chi_{\text{T}} = [(1 + 2\bar{N}\bar{\xi})^{1/2} - 1]/\bar{N}\bar{\xi}; \text{ with } S_p = (2p-3)!!.$$

2.2 Main quantities

Computing the VPF requires the measurement of three quantities: $\bar{N}(R)$, $\bar{\xi}(R)$, and $P_0(R)$. These quantities are determined directly from the CiC approach (their dependency on R is hereafter made implicit). The procedure can be summarized as follows: a large number N_{sph} of probing spheres of radius R are randomly placed throughout the data, the TNG300-1 in our case, and the number of galaxies in each sphere is counted. Finally, we calculate the probability of finding an empty sphere. This is then repeated for a number of radii in a given range.

Here, we briefly define the main quantities mentioned above. Let \bar{N} be the mean number of galaxies in a sphere (equation (7)), and P_0 be the probability that this volume is empty, as shown in equation (8), where N_0 is the number of spheres with zero galaxies and N_{sph} , as mentioned above, is the total number of probing spheres. Finally, let $\bar{\xi}$ be the variance in the number of galaxies per sphere (equation 9):

$$\bar{N} = \frac{1}{N_{\text{sph}}} \sum_{i=1}^{N_{\text{sph}}} N_i \quad (7)$$

$$P_0 = \frac{N_0}{N_{\text{sph}}}, \quad (8)$$

$$\bar{\xi} = \frac{(N - \bar{N})^2 - \bar{N}}{\bar{N}^2}. \quad (9)$$

In the limit of $N_{\text{sph}} \rightarrow \infty$, the CiC approach to determine the volume-averaged correlation function, $\bar{\xi}$, is mathematically equivalent to using the Landy–Szalay estimator (Szapudi 1998; Conroy et al. 2005).

We studied the sensitivity of these quantities to the total number of spheres used, and find that the quantities converge well for $N_{\text{sph}} > 10^4$ (see Appendix A). To ensure a reliable estimate of them we use $N_{\text{sph}} = 10^5$ for the calculation throughout the simulation. The maximum probing radius throughout the box was chosen so that the resulting $\chi(\bar{N}\bar{\xi})$ value is close to 0; we found this value to be $r_{\text{max}} = 4 h^{-1} \text{Mpc}$. This way we probe the largest $\bar{N}\bar{\xi}$ range possible.

For the calculation of the VPF inside voids, the choice of N_{sph} is non-trivial and is dependent on the maximum probing radius if we want to minimize oversampling. The maximum amount of non-overlapping spherical volumes $V(r)$ that can be fit inside a bigger spherical volume $V(R)$ is $n_{\text{max}} = V(R)/V(r) = (R/r)^3$ in the limit where $R/r \rightarrow \infty$. However, for $1 \ll R/r \ll \infty$, the most efficient sphere packing with randomly placed spheres is approximately 64 per cent (Jaeger & Nagel 1992), so we can approximate $n_{\text{max}} \simeq 0.64(R/r)^3$. Considering this relation we can derive our maximum probing radius within voids: r cannot grow so large so as to make n drop below the

needed 10^4 (see Section 2.3 and Appendix A) across all voids. We find that $N_{\text{sph}} = 2 \times 10^4$ is a good compromise between a reliable estimation of the CiC statistics, and a large enough maximum probing radius that minimizes chances of oversampling.

When choosing the smallest void radius of the sample as R , this relation yields $r_{\text{max}} \simeq 1.6 h^{-1} \text{Mpc}$ as the largest probing radius within voids. Independent runs (with different random seeds) of the VPF calculation algorithm within voids with this choice of N_{sph} (and therefore, of r_{max}) yields statistically identical results.

Finally, since the simulation is periodic in volume, its centre and edges are equally sampled, regardless of the radius of the sampling spheres. In practice we ensure this by extending the simulation periodically to the bounds $[L_{\text{box}} - R_{\text{max}}, L_{\text{box}} + R_{\text{max}}]$, where R_{max} is the radius of the largest sampling sphere. Finally, to simulate redshift space in TNG300-1, we use the distant observer approximation using the $H_0 = 67.74 \text{ km s}^{-1} \text{ Mpc}^{-1}$ parameter of the simulation and making the conversion $x^s = x + v_x/H_0$, where x and v_x are the coordinate and velocity in a given axis, respectively, and apply periodic conditions when appropriate. We have calculated the statistics presented in this section modifying the three axes of the box and have found that the results are indistinguishable from each other.

2.3 Uncertainties

The uncertainties of \bar{N} and $\bar{\xi}$ were estimated by Jackknife resampling of the total volume into thirds, resulting in 27 equal subvolumes. Subsequently, we used analytically derived expressions to calculate the uncertainty of P_0 and χ (Hamilton 1985; Maurogordato & Lachieze-Rey 1987; Colombi, Bouchet & Schaeffer 1994):

$$\Delta P_0 = \sqrt{\frac{P_0(1-P_0)}{N_{\text{sph}}}}, \quad (10)$$

and

$$\left(\frac{\Delta\chi}{\chi}\right) \simeq \left| \frac{\Delta P_0}{P_0 |\ln P_0|} - \frac{\Delta\bar{N}}{\bar{N}} \right|, \quad (11)$$

since when calculating the uncertainty of $\chi = -\ln(P_0)/\bar{N}$, the denominator and numerator are not independent but almost exactly anticorrelated (see Colombi, Bouchet & Schaeffer 1994; Fry & Colombi 2013).

To calculate the statistics within the cosmic voids identified in TNG300-1, we uniformly sampled their spherical volume. Jackknife estimates of the statistics were done by ignoring a subset of voids from the complete sample so that we have the same number of jackknife resamplings as in the computation of the VPF throughout the box.

3 DATA

In this work, we employ galaxy data from the ILLUSTRISTNG¹ project (Marinacci et al. 2018; Naiman et al. 2018; Nelson et al. 2018, 2019; Pillepich et al. 2018; Springel et al. 2018; Pillepich et al. 2019). ILLUSTRISTNG is a set of cosmological magneto-hydrodynamical simulations obtained with the AREPO moving-mesh code (Springel 2010), and adopting a Planck cosmology (Planck Collaboration XIII 2016): $\Omega_{\text{m}} = 0.3089$, $\Omega_{\text{b}} = 0.0486$, $\Omega_{\Lambda} = 0.6911$, $\sigma_8 = 0.8159$, $n_s = 0.9667$, and $h = 0.6774$. These simulations present comprehensive models for the physics of galaxy formation, and improve upon its predecessor, ILLUSTRIS, by including magnetic fields and improving

¹<https://www.tng-project.org/>.

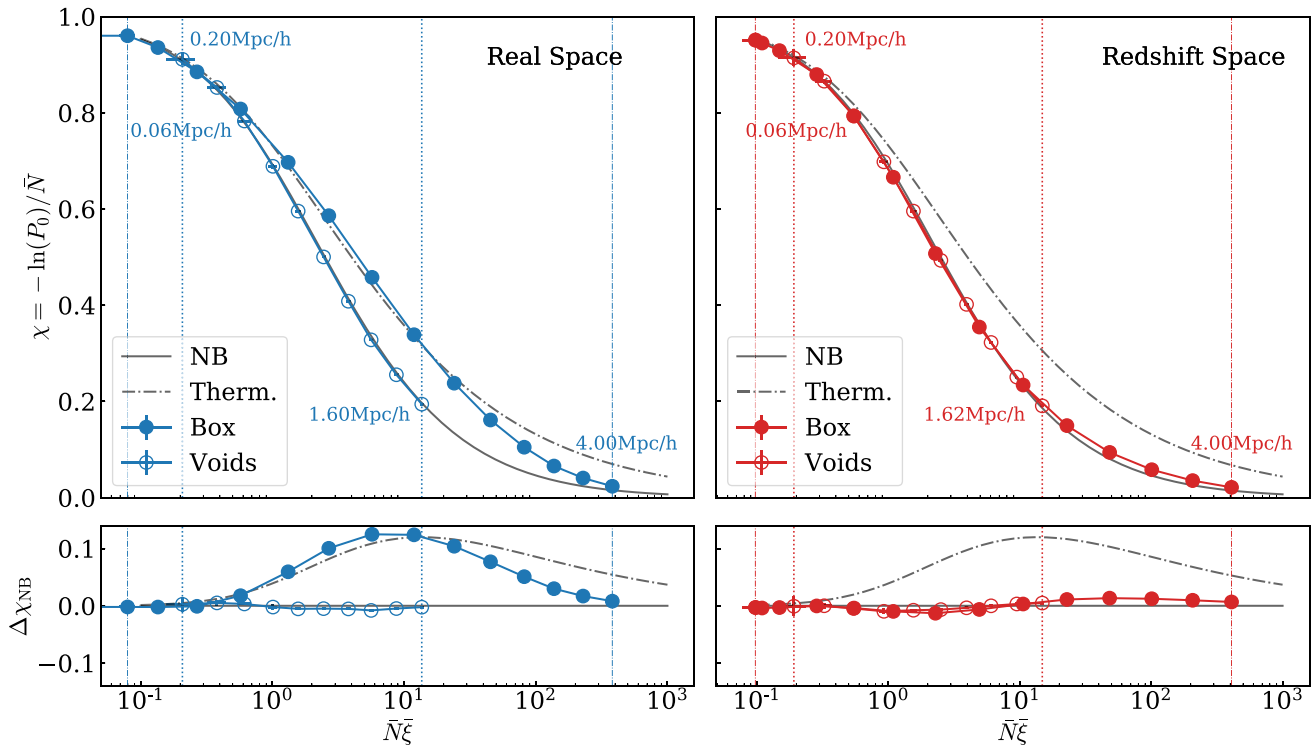


Figure 1. Reduced VPF, $\chi(\bar{N}\bar{\xi})$, inside (empty circles) and outside (filled circles) of cosmic voids, in real and redshift space (blue and red symbols on left- and right-hand panels, respectively). Vertical lines with scales corresponding to the first and last data values of $\bar{N}\bar{\xi}$ for each sample are provided for reference. The negative binomial model of clustering (NB, grey solid line) is known to be a good fit for the VPF of galaxies in redshift space. However, we find a striking agreement between this model and the VPF of galaxies within voids in real space (empty blue circles), which is an indication of a previously unknown clustering scaling relation in real space. The bottom panels show the difference between each of the reduced VPFs with that of the NB model, where the agreement between the data and the fit can be seen more clearly.

models of the galactic wind and AGN feedback. The ILLUSTRITNG project encompasses three different volumes with identical initial conditions and physical models: TNG50, TNG100, and TNG300. In particular, we employ TNG300-1, which has periodic box of $205 h^{-1}\text{Mpc}$, the largest box with the highest resolution in the suite. The haloes (clusters) and subhaloes (galaxies) are found with a standard ‘friends of friends’ (FoF) algorithm with link length $b = 0.2$ (in units of the mean interparticle spacing) run on the dark matter particles, and with the SUBFIND algorithm (Springel et al. 2001), respectively. The latter detects substructure within clusters and defines locally overdense, self-linked particle clusters, where the baryonic component in the substructure is defined as a galaxy. We analyse this simulation at redshift $z = 0, 0.5, 1$ and 2 , and use galaxies in the mass range $10^9 \leq M/M_{\odot} \leq 10^{13}$.

The identification of voids in the simulation follows the algorithm described in Ruiz et al. (2015), a modified version of previous algorithms presented in Padilla, Ceccarelli & Lambas (2005) and Ceccarelli et al. (2006). The algorithm estimates the density profile with a Voronoi tessellation over density tracers, which in this work are galaxies from the TNG300-1 simulation. Subdense regions are obtained by selecting Voronoi cells below a density threshold and selected as cosmic void candidates. Centred on these cells, the integrated density contrast $\Delta(r)$ is calculated at increasing values of r . Void candidates are selected as the largest spheres satisfying the condition $\Delta(R_v) = -0.9$, where R_v is the radius of the void. Next, the centres of the voids are randomly shifted so that the spheres can grow. This is done because the algorithm is likely to

produce spherical voids where their shells do not accurately match the surrounding structures, and the recentring procedure provides structures with edges that better match the surrounding local density field. Finally, the catalogue of voids comprises the largest subdense, non-overlapping spheres of radius R_v .

The tracers we use to identify voids are galaxies of TNG300-1 with a lower total mass cutoff of $M \geq 10^{11}M_{\odot}$ to emulate observable voids. For the proper VPF calculations, as stated above, we include galaxies down to an $M \geq 10^9M_{\odot}$ threshold. After applying this algorithm to TNG300-1 and eliminating shot-noise voids, we are left with a sample of 174 voids in real space and 182 in redshift space with radii in the range of $9\text{--}18 h^{-1}\text{Mpc}$.

4 RESULTS

As a general reference, we show results in real space in blue colours and redshift space in red colours. Statistics corresponding to cosmic void environments are represented with empty symbols while results for the entire simulation box are shown with filled symbols.

We plot the RVPF, χ , as a function of the scaling variable, $\bar{N}\bar{\xi}$, in Fig. 1. The right-side panels show the RVPFs in redshift space calculated for the entire box (solid red circles) and inside voids (empty red circles). We find that there are no differences between the two environments, despite the low densities in voids by their definition. Both curves closely follow the NB model of hierarchical clustering (solid grey line). The panels on the left, however, show that the RVPF in real space is indeed different inside voids (empty blue

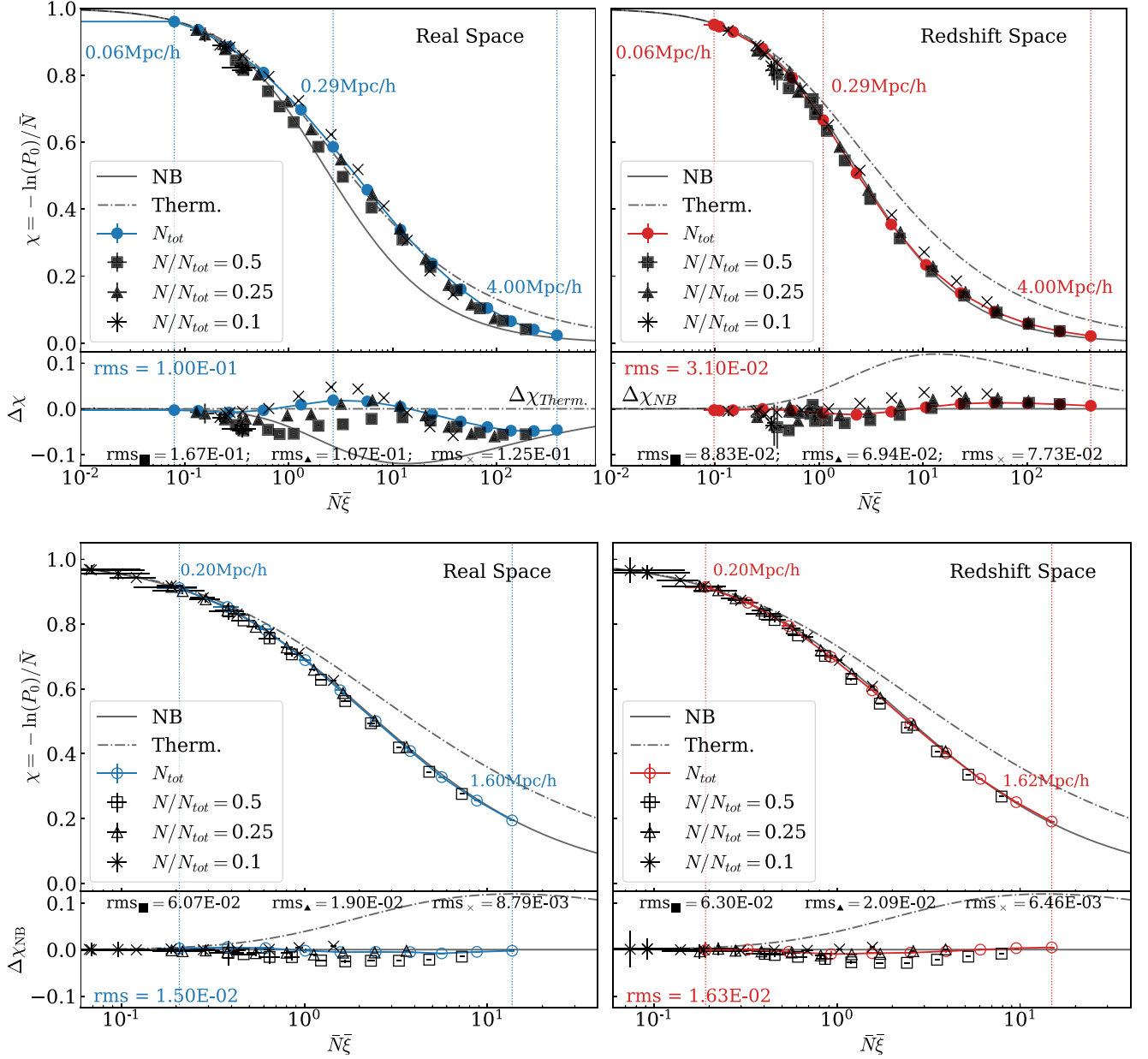


Figure 2. Reduced VPF of galaxies within cosmic voids in TNG300-1 in redshift (right-hand panels) and real space (left-hand panels) with random dilutions of 50, 25, and 10 per cent. The bottom panels show the subtractions between the RVVPF of the samples with the RVVPF of the thermodynamic model ($\Delta\chi_{Therm.}$, bottom left-hand panel) and with that of the negative binomial model ($\Delta\chi_{NB}$, bottom right-hand panel). In contrast to previous results, the low scatter in the diluted RVVPFs in the lower left-hand panel indicates a possible detection of hierarchical scaling of galaxies within cosmic voids in real space.

circles) as opposed to elsewhere in the box. Throughout the box in real space (solid blue circles) the RVVPF appears to somewhat follow the thermodynamical model (dash-dotted grey line) but the fit is not thorough as it clearly overestimates the data for values of $\bar{N}\xi \gtrsim 30$. On the other hand, the RVVPF inside voids in real space follow the NB model surprisingly well.

We also plot the scales corresponding to the first and last values of the RVVPF of each curve. These scales are indicated with vertical lines as a reference for the reader, since the corresponding scale for each value of $\bar{N}\xi$, i.e. the relation $\bar{N}\xi(R)$, will vary from sample to sample. The bottom panels show the quantity $\Delta\chi_{NB} \equiv \chi - \chi_{NB}$, corresponding to the difference between the RVVPF of the data and NB model for clarity.

4.1 Random dilutions test

Following the analysis of Croton et al. (2004a), we test the robustness of this result by randomly diluting the sample to verify whether the scaling relations found in Fig. 1 hold. Results with random dilutions are shown in Fig. 2. In both spaces (left and right-hand panels), the top panels display the full sample throughout the simulation in coloured filled circles, and dilutions of 50, 25, and 10 per cent with squares, triangles, and crosses with incremental transparency respectively. The quantity $\Delta\chi_T$ shown in the top left-hand panel is defined in an analogous way as $\Delta\chi_{NB}$ described above. It can be seen that the dispersion between dilutions is larger in real space than in redshift space. This means that the scaling coefficients S_p are not completely

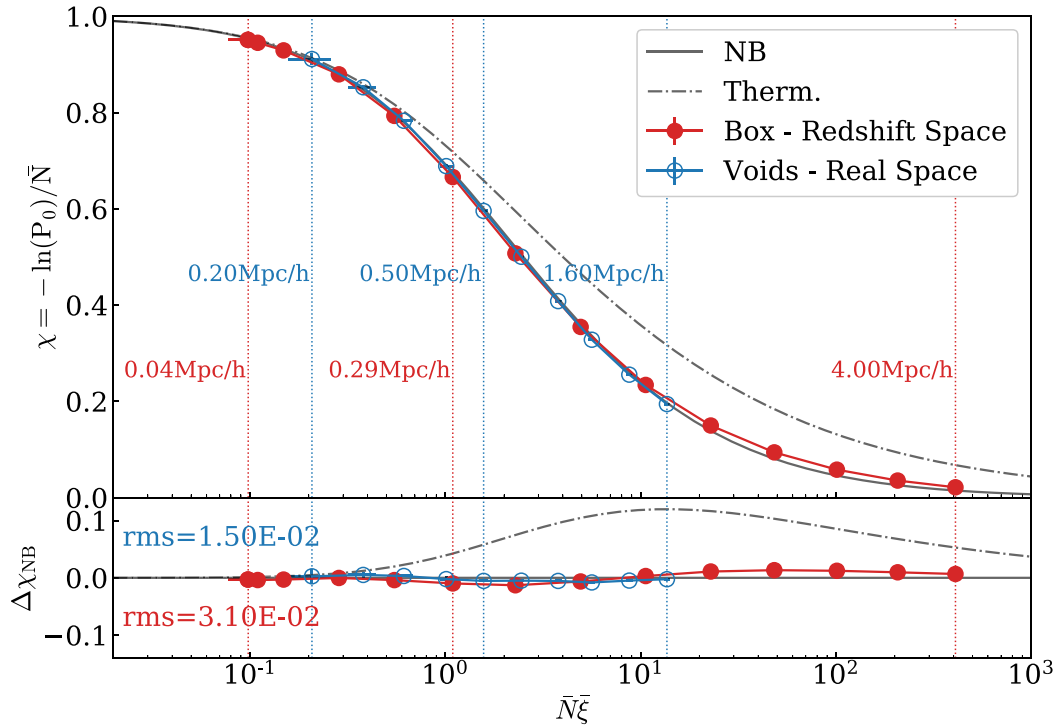


Figure 3. Reduced VPF, $\chi(\bar{N}\bar{\xi})$, inside voids (empty circles) and elsewhere in the simulation (filled circles), in real and redshift space (blue and red lines, respectively). The negative binomial model of clustering (NB, black solid line) is known to be a good fit for the VPF of galaxies in redshift space. However, we find a striking agreement between this model, and the VPF of galaxies within voids in real space (empty blue circles), which is an indication of a previously unknown hierarchical scaling relation in real space.

independent of the sample density, and therefore, hierarchical scaling in real space does not hold as rigorously as it does in redshift space.

We calculated the root-mean-square deviations of the data from the closest fit (the NB or thermodynamical model when computing $\Delta\chi_{\text{NB}}$ or $\Delta\chi_{\text{T}}$, respectively). For this calculation, we only use data corresponding to $\bar{N}\bar{\xi} \geq 0.1$, given that, for smaller values, every model yields approximately the same $\chi(\bar{N}\bar{\xi})$, which is close to 1. Notably, the rms deviations of the RVPF of galaxies in the general redshift space is one order of magnitude lower than that of real space galaxies with respect to their closest models.

It should be noted that larger dilutions shifts the range of $\bar{N}\bar{\xi}$ to lower values, where the different models are more similar to each other and, consequently, rms values will tend to be lower. This means that larger dilutions probe a less statistically significant range of $\bar{N}\bar{\xi}$ in the sense that it becomes more difficult to differentiate between models. Therefore, rms values should be regarded as a reference and compared amongst samples with the same dilution between both spaces. This is true also for the different analyses that follow in the next subsections.

We show the RVPF inside voids represented in the lower panels of Fig. 2 with empty symbols. We can see that both spaces exhibit lower scatter within voids than elsewhere in the box, with similar rms values for each dilution in real and redshift space. This robustness with respect to dilutions reinforces the evidence in favour of hierarchical scaling within voids in both spaces. This is unexpected given that previous works indicate that any hierarchical scaling found in redshift space breaks down in real space.

Thus far we have found that the RVPFs for galaxies inside and outside voids are markedly different in real space, with the NB model being a very good fit for the RVPF inside voids. On the

other hand, we find no such difference in redshift space: both inside and outside voids, the RVPF is well fitted by the NB model. There is a striking agreement between the RVPFs of galaxies in the general redshift space and galaxies within real space voids (shown in Fig. 3). This result indicates that within cosmic voids there is evidence of hierarchical scaling not found elsewhere in the Universe.

4.2 VPF dependence on void identification parameters

From this section onwards, for clarity and visibility, we only show the zoomed-in differences between the calculated RVPFs and the NB or thermodynamic model accordingly ($\Delta\chi_{\text{NB}}$ and $\Delta\chi_{\text{T}}$), as shown in the lower panels of Figs 1–3.

In this subsection, we show how the RVPF changes with variations in the physical parameters that define a void: the minimum threshold of mass of the density tracers, M_t , and the integrated density contrast at one void radius, $\Delta(R_v)$. This allows us to gain insight into the effects on the RVPF of gradual change in the environment that is defined by a cosmic void.

4.2.1 Changing minimum mass of density tracers in void identification

We test the sensitivity of the RVPF to M_t , the mass of the density tracers used in void identification. This is done by calculating the RVPF inside voids identified with different tracers (Fig. 4). The main void sample (open circles) was identified using galaxies with masses greater than $10^{11}M_{\odot}$. The other lower mass thresholds we chose are 10^{10} (open squares) and 10^9M_{\odot} (open triangles). The cutoff radius is 9, 7, and $6h^{-1}\text{Mpc}$ for each void sample, respectively. These choices

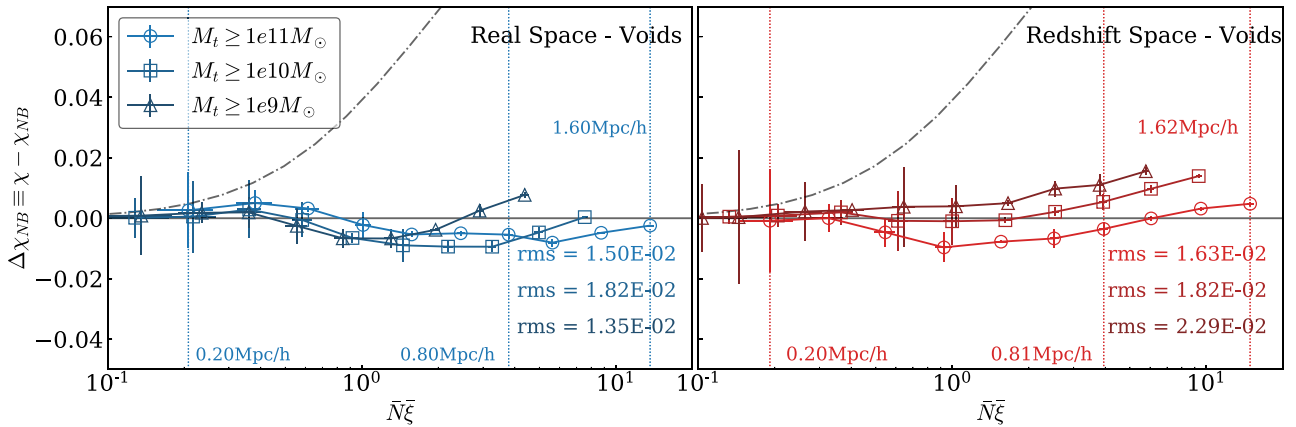


Figure 4. Reduced VPFs, plotted as differences to the NB model (solid grey line), of voids identified with different void tracer mass, M_t . rms values are calculated for each sample with respect to the NB model and shown with the shade of colour of the corresponding curve (light to dark for larger to smaller M_t , respectively). We find no significant changes in the reduced VPFs, in either space, around the NB model.

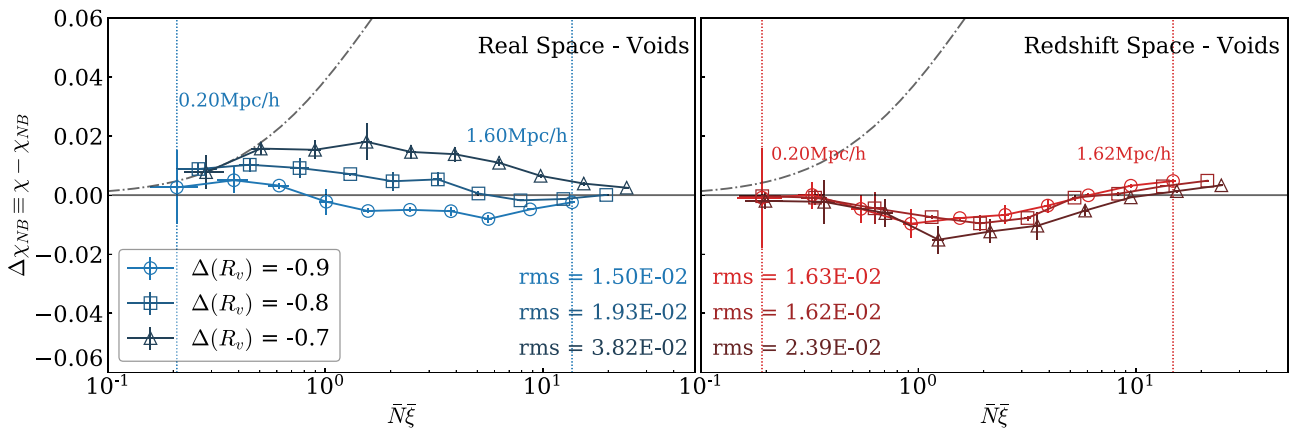


Figure 5. Reduced VPF plotted as differences to the NB model (solid grey line) in real and redshift space within voids identified with different integrated density contrast $\Delta(R_v)$. rms values are calculated for each sample with respect to the NB model and shown with the shade of colour of the corresponding curve (light to dark for larger to smaller $\Delta(R_v)$, respectively). In real space, we find a trend for the reduced VPF of voids with larger $\Delta(R_v)$ to be more like the reduced VPF throughout the box. This effect is not observed in redshift space, where the curves are statistically identical.

in minimum void radii prevents including spurious or shot-noise voids in our void sample (Correa et al. 2021). We find only small differences in the RVPFs when calculating them in voids traced by galaxies with different masses, with small variations of rms values with respect to the NB model in both spaces ($\text{rms} \simeq 1$). The RVPF and its agreement with the NB model seems to be insensitive to void tracer mass.

4.2.2 Changing integrated density contrast

Another important parameter in void identification is the integrated density contrast $\Delta(R_v)$. It is standard practice to define a void as a sphere with $\Delta(R_v) = -0.9$ (see Section 3) so, in addition, we tested values of $\Delta(R_v) = -0.8$ and -0.7 . Results for the RVPF in voids with each of these integrated deltas are shown in Fig. 5.

In real space, the structure as traced by voids with a lower density contrast, naturally yields higher values of RVPF, more similar to those found in the general Universe. There is no such effect found in redshift space: The RVPF is robust against these integrated density changes used in the definition of voids. These results are consistent with Fig. 2.

4.3 VPF dependence on mass

Next, in Fig. 6, we study the sensitivity of the VPF to galaxy mass. We create galaxy subsamples according to the mass ranges 10^9 – 10^{10} (circles), 10^{10} – 10^{11} (squares), and $M \geq 10^{11} M_\odot$ (triangles). Then we calculate the RVPF throughout the box and inside the main void sample.

There is a clear difference in the effects of selecting galaxy mass between real and redshift space. While different subsamples fall on the same curve in redshift space, they have a wide spread in real space, further affirming the notion that hierarchical scaling is not found in the general Universe in real space. The scaling coefficients S_p appear to have a strong dependence on mass in real space.

Within voids, however, the scenario is similar both in real and redshift space. Cosmic voids seem to provide a scenario where the S_p appear to be independent of mass.

4.4 VPF dependence on redshift

Finally, we want to study how the RVPF evolves with redshift. We show the RVPF in real and redshift space at different redshift values,

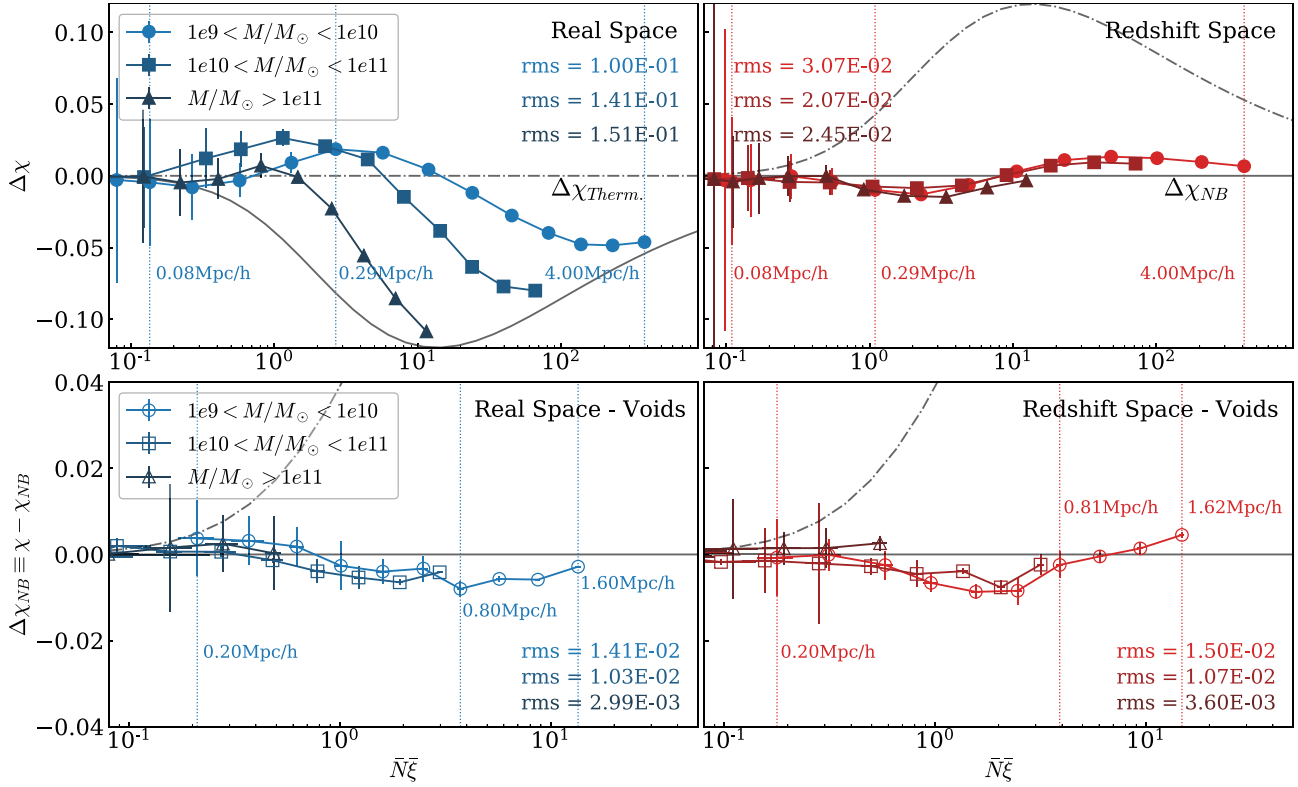


Figure 6. Reduced VPF plotted as the difference with respect to the NB model (solid grey line) calculated with different galaxy subsamples selected by mass ranges. rms values are calculated for each sample with respect to the NB model (solid grey line) and shown with the shade of colour of the corresponding curve (light to dark for smaller to larger mass values, respectively). Throughout the box (top panels), the differences between real and redshift space in selecting mass is stark: The hierarchical scaling detected in redshift space completely breaks down in real space. However, within voids (bottom panels), we detect hierarchical scaling in both spaces roughly consistent with the NB model.

namely $z = 0, 0.5, 1,$ and 2 , throughout the box and inside voids (Fig. 7). Following the spherical collapse model (Gunn & Gott 1972; Lilje & Lahav 1991; Sheth & van de Weygaert 2004), for the different redshifts we identified voids with a time-dependent integrated density contrast value that corresponds to a perturbation with $\Delta(R_v) = -0.9$ at $z = 0$. For the mentioned redshift values, these are $\Delta(R_v) \simeq -0.87, -0.85,$ and -0.80 , respectively.

Throughout the box (top panels, Fig. 7), we find a similar effect as with mass (Section 4.3, Fig. 6): The changes on the RVPF seen in redshift space are minor and all the curves are roughly fitted by the NB model, while in real space the differences are quite large. Interestingly, we detect a trend in real space for the RVPFs to be more similar to the NB model with increasing redshift. Between $z = 0$ and 2 (light blue circles and dark blue crosses, respectively) we have a difference of one order of magnitude in rms with respect to the NB model. The RVPF of galaxies at high redshift in real space is similar to the RVPF of galaxies within voids at present time.

The computation of the RVPF inside voids at different redshifts (bottom panels) is more self-similar in both spaces, which would indicate that the S_p have remained approximately constant with time. Naturally, the best fit within voids is still the NB model.

5 SUMMARY AND DISCUSSION

The VPF can be a pathway to higher-order correlations, but a number of conditions need to be met. When the solutions of ‘self-similarity’ and ‘stable clustering’ are valid, then the scaling coefficients S_p

of equation (3) are constant, and therefore the hierarchical scaling Ansatz holds and the reduced VPF, χ , can be expressed in terms of the scaling variable, $\tilde{N}\xi$, alone (equation 6). As a consequence, different samples fall on the same curve as long as they follow the same hierarchical clustering model. The constancy of S_p can be tested with random dilutions or by selecting subsamples with a given parameter of the sample and testing whether the RVPFs collapse onto the same curve. If they do not, it could be evidence that the S_p have a strong dependence on said parameter, and therefore hierarchical clustering of the high-order correlations cannot be confirmed.

We have calculated the VPF in the TNG300-1 simulation, inside and outside of cosmic voids, both in real and redshift space. We compared the RVPF of galaxies in these samples with two popular clustering models: the negative binomial (NB) and the thermodynamic model. We have focused on the effects of the void environment on the VPF, and the sensitivity of this statistic with void definition, random sample dilution, galaxy mass, and structure evolution via redshift. Previous works have shown that galaxies in redshift surveys are well fitted by the NB model of hierarchical clustering (e.g. Croton et al. 2004a; Conroy et al. 2005), but that this fit breaks down in real space (Lahav et al. 1993; Vogelely et al. 1994).

We find that in redshift space, galaxies indeed appear to follow the NB model regardless of the environment. However, we report a strong difference in the RVPFs inside and outside cosmic voids in real space.

Throughout the simulation, galaxies seem to follow the thermodynamic model up to $\tilde{N}\xi \simeq 30$, with an overall rms deviation of

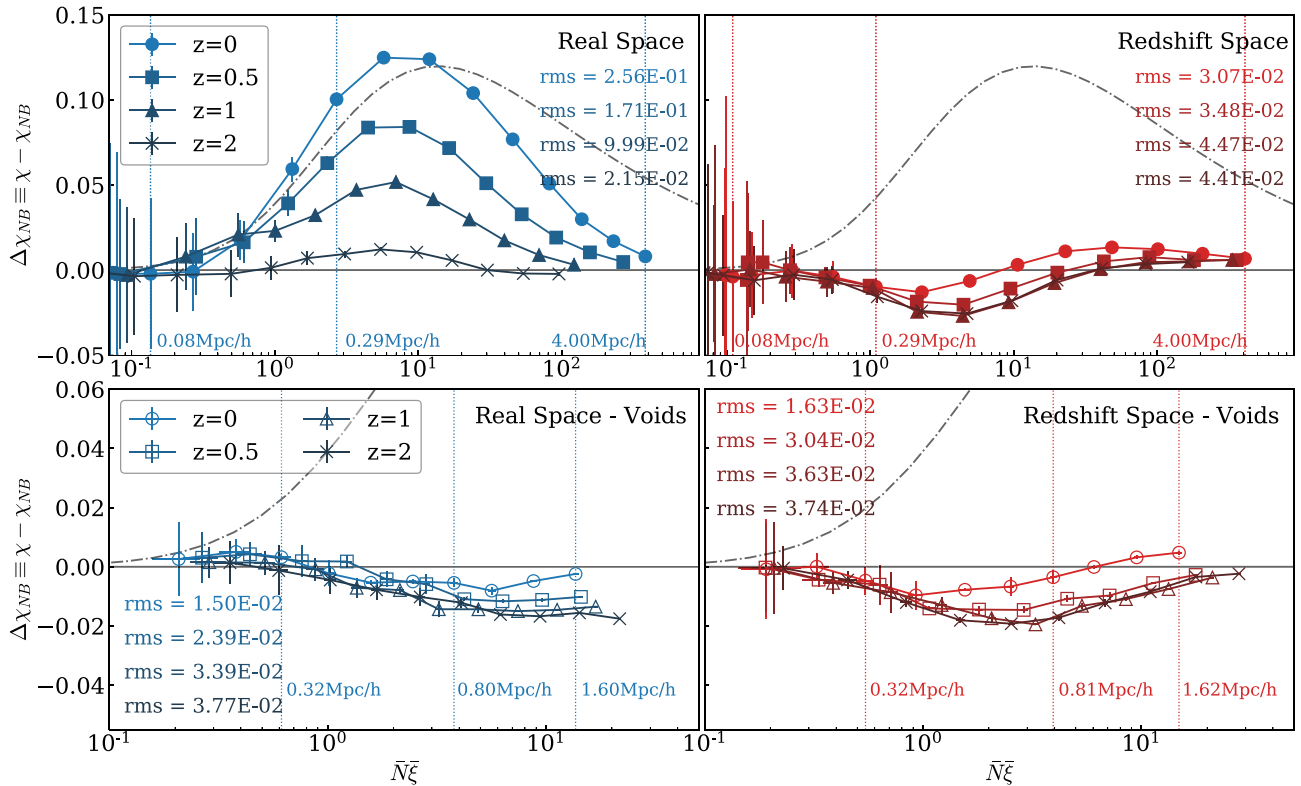


Figure 7. Reduced VPF plotted as differences to the NB model (solid grey line) calculated for the simulation different redshifts inside voids identified in different redshifts with their corresponding integrated density contrast. rms values are calculated for each sample with respect to the NB model and shown with the shade of colour of the corresponding curve (light to dark for smaller to larger redshift values, respectively). Analogous to Fig. 6, we detect a great difference throughout the simulation (top panels) between real and redshift space for different redshifts where the RVPF in real space is much more affected by the evolution of structure than in redshift space. Within voids (bottom panels), the variation in the RVPFs is much smaller in both spaces, staying close the NB model in each case.

0.10, although this fit does not hold when the sample is diluted. On the other hand, inside cosmic voids, we find an excellent agreement between the RVPF of galaxies and the NB model, with an rms of 0.01. This good fit is maintained even with dilutions. For comparison, we find similar rms values (0.03 and 0.02) for the RVPFs of galaxies in redshift space throughout the box and inside voids respectively.

From this results it follows that the RVPF of galaxies inside cosmic voids in real space is very similar to the RVPF of galaxies throughout the box in redshift space, with both of them being well fitted by the NB model. That is, the Ansatz of hierarchical scaling of the high orders of the correlation function that has been shown to be valid for galaxies in redshift space can also be found in real space for galaxies within voids.

Following this analysis we study how the RVPF changes when changing what defines a void environment, which means calculating the RVPF inside voids for different integrated density contrasts, $\Delta(R_v)$, and different tracer mass, M_t , used to trace the underlying density when identifying cosmic voids. We find that the RVPF is not sensitive to changes in M_t in neither redshift nor real space, with the NB model still being the best fit. On the other hand, while the RVPF remains insensitive to $\Delta(R_v)$ in redshift space, we notice a trend in real space where the data deviates from the NB model with larger density contrasts (i.e. more dense voids), and moves towards the RVPF of the general Universe in real space. Since we find that the RVPF in redshift space is the same regardless of the environment, it

makes sense that we would find no significant changes in the RVPF with different $\Delta(R_v)$.

Next, we analysed the sensitivity of the RVPF to galaxy mass. In accordance to Croton et al. (2004a), we find that subsamples with different mass follow the NB model in redshift space, hinting towards hierarchical scaling. In real space, this agreement between the subsamples breaks down. However, we notice that the subsamples still follow the NB model inside voids both in real and redshift.

Finally, we studied how the RVPF changes with redshift, where, again, we detected the largest effects on the RVPF throughout the box in real space, and only small changes inside voids and everywhere in redshift space. Interestingly, we find that the RVPF in real space moves towards the NB model with increasing redshift. This leads to the NB model being a close fit for galaxies in cosmic voids at present time and galaxies at high redshift. This points towards the idea that the way in which high-order clustering emerges from the two-point correlation function in high redshifts is preserved within voids today.

The most widely used argument for the validity of hierarchical scaling in redshift space is that small-scale distortions dampen the amplitude of the coefficients S_p so that they appear to be constant with scale (a necessary condition for hierarchical scaling), even though they are not in real space (see fig. 49 of Bernardeau et al. 2002). This argument still warrants further exploration as to why the coefficients are affected by distortions in such a way that allows for apparent evidence of hierarchical scaling to arise. It could be the case that

redshift space distortions blur strong non-linear effects at small scales and any other clustering behaviour that would break the hierarchical scaling relation. Cosmic voids and higher redshifts (where the VPF is closely fitted by the NB model, see Figs 3 and 7) could provide an environment where velocity dispersions are low and, therefore, high non-linearities are still not a main agent in the clustering process, allowing for hierarchical scaling to be preserved.

It is also interesting to contrast our results with the discussion of Croton et al. (2006) in which they argue that the peculiar motion of galaxies should not be the main agent in generating hierarchical scaling consistent with the NB model, as the results from Vogeley et al. (1994) might suggest. Instead, they find that faint red galaxies, which would be more affected by peculiar motions since they are in denser environments, are driven further away from (instead of towards) the NB model.

Regarding the random dilutions analyses, it should be noted that the cosmic voids, by definition, are the least affected structures by this test. While this could partly be the reason why the statistics are less affected by dilutions within the cosmic voids, it should not be the main cause of the good agreement with hierarchical scaling. The RVPF by definition is disaffected from the mean density in each sampling sphere, which in principle makes it a good statistic to use in different environments. Particularly, the maximum dilution is 10 per cent of the total sample, and is therefore comparable to the internal density of a cosmic void (see identification criteria and algorithms in Ruiz et al. 2015); and yet, even with this dilution, the RVPF outside cosmic voids is far from being well fitted by the NB model that seems to characterize the distribution inside voids in real space, as well as everywhere in redshift space.

Additionally, the success of the NB model in fitting the data is not well understood. One one hand, it has been shown that it is not a physically possible model since it breaks the second law of thermodynamics despite being the best fit for the data (Saslaw & Fang 1996; Yang & Saslaw 2011), although some authors find its use justified (Carruthers & Duong-van 1983; Elizalde & Gaztanaga 1992; Betancort-Rijo et al. 2009). Furthermore, the derivation of this phenomenological model (or any of the others) does not explicitly take redshift space distortions into account, so it has been surprising that it is such a good fit for the VPF of galaxies in redshift surveys and not in real space. However, in this work we find that this good fit is not a particularity of the redshift space. We show that it can also be found in real space within cosmic voids.

The theoretical background of the CiC statistics yields that if two VPFs are different, it is due not necessarily to the difference in sample density or two-point clustering amplitude (they could actually be identical), but a consequence of how the intrinsic clustering of the samples gives rise to a different hierarchical emergence of the higher-order correlations. For example, we find the VPFs in redshift space to be the same inside or outside the voids, although the clustering inside the voids is weaker and the density by definition is lower. However, in real space, the voids and the general Universe give markedly different VPFs. Under this framework, this difference is an effect of the high-order correlations.

Moreover, our results can also be understood as a consequence of the smaller changes in density contrast within voids. The integrated density profiles of our voids are not completely flat within one void radius because we are identifying them with galaxies of higher mass ($M_t \geq 10^{11} M_\odot$) than we are ultimately considering in the calculations ($M \geq 10^9 M_\odot$). Nevertheless, the changes in the density contrasts are smaller inside than outside voids. Therefore, the solutions of self-similarity and stable clustering may be appropriate approximations in these underdense regimes.

Ultimately, in cosmic voids, we find evidence of hierarchical scaling that unravels further away in denser environments. The similarities between the RVPF in real space at higher redshifts and inside voids in present day indicate that clustering can be found inside voids that more pristinely preserve the initial conditions of the Universe. Upcoming surveys, such as *Euclid* (Laureijs et al. 2011) and the *Legacy Survey of Space and Time* (Ivezić et al. 2019) at Vera Rubin Observatory, will have the required resolution and volume for a suitable statistical analysis with voids at higher redshifts, to contrast these results with observations.

ACKNOWLEDGEMENTS

This work was partially supported by the Consejo Nacional de Investigaciones Científicas y Técnicas, Argentina (PIP 11220200102832CO), the Secretaría de Ciencia y Tecnología, Universidad Nacional de Córdoba, Argentina, and the Agencia Nacional de Promoción de la Investigación, el Desarrollo Tecnológico y la Innovación, Ministerio de Ciencia, Tecnología e Innovación, Argentina. The ILLUSTRATING project used in this work (TNG300-1) have been run on the HazelHen Cray XC40-system at the High Performance Computing Center Stuttgart as part of project GCS-ILLU of the Gauss centres for Supercomputing (GCS). This research has made use of NASA's Astrophysics Data System. Visualizations made use of PYTHON packages.

DATA AVAILABILITY

The data underlying this paper will be shared on reasonable request to the corresponding author.

REFERENCES

- Balian R., Schaeffer R., Balian R., Schaeffer R., 1989, *A&A*, 220, 1
 Baugh C. M. et al., 2004, *MNRAS*, 351, L44
 Benson A. J., Hoyle F., Torres F., Vogeley M. S., 2003, *MNRAS*, 340, 160
 Bernardeau F., Colombi S., Gaztañaga E., Scoccimarro R., 2002, *Phys. Rep.*, 367, 1
 Betancort-Rijo J., Patiri S. G., Prada F., Romano A. E., 2009, *MNRAS*, 400, 1835
 Carruthers P., Duong-van M., 1983, *Phys. Lett. B*, 131, 116
 Ceccarelli L., Padilla N. D., Valotto C., Lambas D. G., 2006, *MNRAS*, 373, 1440
 Coles P., Jones B., 1991, *MNRAS*, 248, 1
 Colombi S., Bouchet F. R., Schaeffer R., 1994, *ApJS*, 96, 401
 Conroy C. et al., 2005, *ApJ*, 635, 990
 Correa C. M., Paz D. J., Sánchez A. G., Ruiz A. N., Padilla N. D., Angulo R. E., 2021, *MNRAS*, 500, 911
 Croton D. J. et al., 2004a, *MNRAS*, 352, 828
 Croton D. J. et al., 2004b, *MNRAS*, 352, 1232
 Croton D. J., Norberg P., Gaztanaga E., Baugh C. M., 2006, *MNRAS*, 379, 1562
 Elizalde E., Gaztanaga E., 1992, *MNRAS*, 254, 247
 Fry J. N., 1986, *ApJ*, 308, L71
 Fry J. N., Colombi S., 2013, *MNRAS*, 433, 581
 Gaztanaga E., Yokoyama J., 1993, *ApJ*, 403, 450
 Gunn J. E., Gott J. Richard I., 1972, *ApJ*, 176, 1
 Hamilton D., 1985, *ApJ*, 297, 371
 Hurtado-Gil L., Martínez V. J., Arnalte-Mur P., Pons-Bordería M. J., Pareja-Flores C., Paredes S., 2017, *A&A*, 601, A40
 Ivezić Ž. et al., 2019, *ApJ*, 873, 111
 Jaeger H. M., Nagel S. R., 1992, *Science*, 255, 1523
 Lahav O., Itoh M., Inagaki S., Suto Y., 1993, *ApJ*, 402, 387
 Laureijs R. et al., 2011, preprint (arXiv:1110.3193)

- Lilje P. B., Lahav O., 1991, *ApJ*, 374, 29
 Marinacci F. et al., 2018, *MNRAS*, 480, 5113
 Maugorodato S., Lachieze-Rey M., 1987, *ApJ*, 320, 13
 Mekjian A. Z., 2007, *ApJ*, 655, 1
 Naiman J. P. et al., 2018, *MNRAS*, 477, 1206
 Nelson D. et al., 2018, *MNRAS*, 475, 624
 Nelson D. et al., 2019, *ComAC*, 6, 2
 Padilla N. D., Ceccarelli L., Lambas D. G., 2005, *MNRAS*, 363, 977
 Peebles P. J. E., 1980, *The Large-Scale Structure of the Universe*. Princeton University Press, Princeton, New Jersey, p. 120
 Pillepich A. et al., 2018, *MNRAS*, 473, 4077
 Pillepich A. et al., 2019, *MNRAS*, 490, 3196
 Planck Collaboration et al., 2016, *A&A*, 594, A13
 Ruiz A. N., Paz D. J., Lares M., Luparello H. E., Ceccarelli L., Lambas D. G., 2015, *MNRAS*, 448, 1471
 Saslaw W. C., Fang F., 1996, *ApJ*, 460, 16
 Sharp N. A., 1981, *MNRAS*, 195, 857
 Sheth R. K., van de Weygaert R., 2004, *MNRAS*, 350, 517
 Springel V., 2010, *MNRAS*, 401, 791
 Springel V., White S. D. M., Tormen G., Kauffmann G., 2001, *MNRAS*, 328, 726
 Springel V. et al., 2018, *MNRAS*, 475, 676
 Szapudi I., 1998, *ApJ*, 497, 16
 Tinker J. L., Weinberg D. H., Warren M. S., 2006, *ApJ*, 647, 737
 Tinker J. L., Conroy C., Norberg P., Patri S. G., Weinberg D. H., Warren M. S., 2008, *ApJ*, 686, 53
 Vogeley M. S., Geller M. J., Park C., Huchra J. P., 1994, *AJ*, 108, 745
 White S. D. M., 1979, *MNRAS*, 186, 145
 Yang A., Saslaw W. C., 2011, *ApJ*, 729, 123

APPENDIX A: CONSISTENCY AND STABILITY OF THE MAIN CIC QUANTITIES

Following the analysis of Conroy et al. (2005), we study the sensitivity of P_0 , \bar{N} , and $\bar{\xi}$ (the main CiC statistics used in this work) to the number of test spheres N_{sph} . This test will show if our algorithm for calculating the VPF indeed is correct, and will inform on how many test spheres is necessary for a reliable estimation of the aforementioned quantities. We confirm that the estimation of the statistics stabilizes around the expected value, and see how many test spheres are necessary for this stabilization to occur.

We perform the calculations on a randomly distributed sample of points with the same density as the largest data used in this work: the TNG300-1 galaxy sample after selecting those with $10^9 \geq M/M_\odot \geq 10^{13}$. This sample size is approximately 10^7 , with a mean interparticle distance of $\simeq 0.86 h^{-1}$ Mpc. Therefore we fix our testing radius slightly smaller at $R = 0.8 h^{-1}$ Mpc so as to ensure a value of $P_0(R) \neq 0$. Results are shown in Fig. A1 with uncertainties calculated as described in Section 2.3.

As expected, for small N_{sph} , these quantities are unstable but eventually converge at the theoretical values marked with horizontal

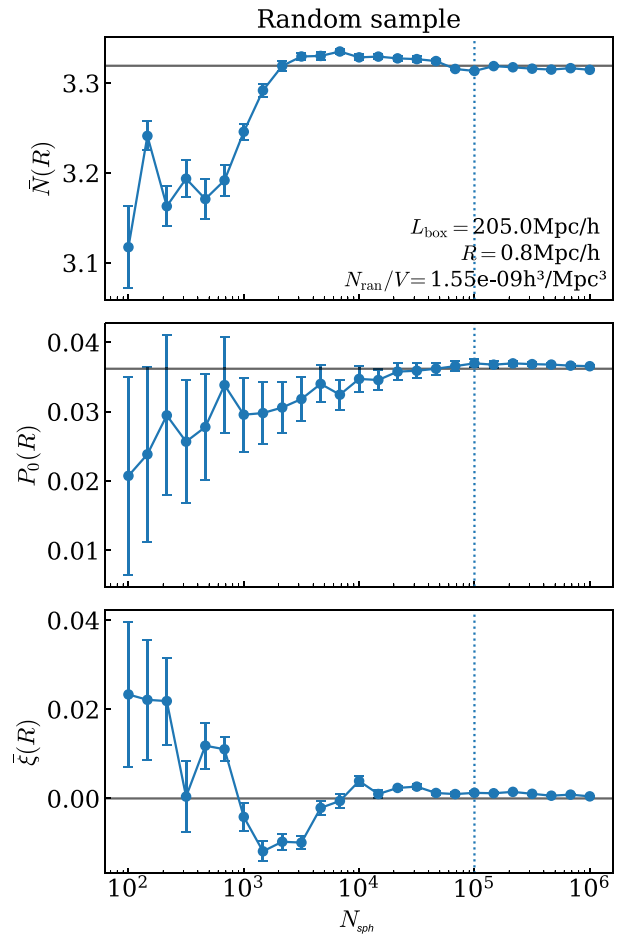


Figure A1. Consistency and stability test for the CiC statistics P_0 , \bar{N} y $\bar{\xi}$ as a function of the number of test spheres N_{sph} . The horizontal solid lines indicate the theoretical value of the statistics for a random sample with the same number density as the TNG300-1 simulation after selecting galaxies with $10^9 \geq M/M_\odot \geq 10^{13}$. The vertical dotted line shows the value of N_{sph} chosen for this work as the number of test spheres for a reliable estimation of the statistics.

grey lines. To ensure a reliable estimate of these quantities, we use $N_{\text{sph}} = 10^5$ (vertical dotted blue lines) throughout this work, although we have obtained robust results within error bars with as few as $\simeq 10^3$ test spheres.

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