Mon. Not. R. Astron. Soc. **369,** 335–348 (2006) doi:10.1111/j.1365-2966.2006.10305.x

Statistics of voids in the two-degree Field Galaxy Redshift Survey

Santiago G. Patiri,[1⋆] Juan E. Betancort-Rijo,[1,2] Francisco Prada,[3] Anatoly Klypin[4] and Stefan Gottlöber[5]

[1]*Instituto de Astrofisica de Canarias, C/Via Lactea s/n, Tenerife E38200, Spain*
[2]*Facultad de Fisica, Universidad de La Laguna, Astrofisico Francisco Sanchez, s/n, La Laguna Tenerife E38200, Spain*
[3]*Ramon y Cajal Fellow, Instituto de Astrofisica de Andalucia (CSIC), E-18008 Granada, Spain*
[4]*Astronomy Department, New Mexico State University, Department 4500, Las Cruces, NM 88003, USA*
[5]*Astrophysikalisches Institut Potsdam, An der Sternwarte 16, 14482 Potsdam, Germany*

Accepted 2006 March 8. Received 2006 March 8; in original form 2005 June 28

ABSTRACT

We present a statistical analysis of voids in the two-degree Field Galaxy Redshift Survey (2dFGRS). In order to detect the voids, we have developed two robust algorithms. We define voids as non-overlapping maximal spheres empty of haloes or galaxies with mass or luminosity above a given value. We search for voids in cosmological *N*-Body simulations to test the performance of our void finders. We obtain and analyse the void statistics for several volume-limited samples for the North Galactic Pole (NGP) and the South Galactic Pole (SGP) constructed from the 2dFGRS full data release. We find that the results obtained from the NGP and the SGP are statistically compatible. From the results of several statistical tests we conclude that voids are essentially uncorrelated, with at most a mild anticorrelation and that at the 99.5 per cent confidence level there is a dependence of the void number density on redshift. We develop a technique to correct the distortion caused by the fact that we use the redshift as the radial coordinate. We calibrate this technique with mock catalogues and find that the correction might be of some relevance to carry out accurate inferences from void statistics. We study the statistics of the galaxies inside nine nearby voids. We find that galaxies in voids are not randomly distributed: they form structures like filaments. We also obtain the galaxy number density profile in voids. This profile follow a similar but steeper trend to that followed by haloes in voids.

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

1 INTRODUCTION

There are many tests to constrain the models of structure formation, which range from the large-scale statistics such as correlation functions (Peebles 1980; Davis & Peebles 1983; Norberg et al. 2002; Zehavi et al. 2002) or the power spectrum (Peebles 1980; Tegmark et al. 2004; Cole et al. 2005), to detailed studies of the physical properties of individual galaxy clusters and voids. Although dense environments (i.e. galaxy clusters and groups) have been extensively studied, the underdense regions like giant voids attract less attention. Yet, they are not less important and can provide important information on galaxy formation (Croton et al. 2005; Goldberg et al. 2005; Hoyle et al. 2005; Rojas et al. 2005) and can give independent constraints on cosmological models (Peebles 2001; Plionis & Basilakos 2002; Croton et al. 2004; Conroy et al. 2005; Colberg et al. 2005; Solevi et al. 2006). Large voids have woken up more and more interest since their first detections 25 yr ago (Kirshner et al. 1981; Rood 1981). At present, thanks to the advent of larger redshift surveys like the two-degree Field Galaxy Redshift Survey (2dFGRS) (Colless et al. 2001) and the Sloan Digital Sky Survey (SDSS) (York et al. 2000), higher resolution of cosmological simulations and better analytical frameworks, we can extract accurate statistical information about voids. This information can be used in different ways but one of the most important is to test the models of structure formation.

Voids can be studied in different ways. One of the classical methods is the void probability function (VPF; White 1979; Fry 1986), which gives the probability that a randomly located sphere of a given radius contains no galaxies (see e.g. Einasto et al. 1991; Croton et al. 2004; Solevi et al. 2006). The number density of voids with radius greater than R is another useful statistic. This number density and also the void significance can be estimated analytically (see Patiri, Betancort-Rijo & Prada 2004). So far, it has been done only using numerical simulations or mock catalogues (Colberg et al. 2005).

⋆E-mail: spatiri@iac.es

Voids in the 2dFGRS were previously studied by Hoyle & Vogeley (2004) and Croton et al. (2004). Croton et al. (2004) have measured the VPF for volume-limited galaxy samples covering the absolute magnitude range $M_{bj} - 5\log_{10}h = -18$ to -22. Their work mainly focused on the study of the dependence of the VPF on the moments of galaxy clustering as a test to discriminate among different clustering models. They found that the VPF measured from the 2dFGRS is in excellent agreement with the paradigm of hierarchical scaling of the galaxy clustering. In addition, they showed that the negative binomial model gives a good approximation of the 2dF data over a wide range of scales. On the other hand, Hoyle & Vogeley (2004) have also calculated the VPF in the 2dFGRS obtaining similar results. They have obtained the VPF for the dark matter haloes in Lambda cold dark matter (ΛCDM) simulations and galaxy mock catalogues from semi-analytic models of galaxy formation to compare with the data. They have found that the results from the semi-analytic models that include feedback effects provide a VPF that agree with the VPF measured for the 2dFGRS and differ from that measured from the dark matter distribution.

In spite of the fact that the notion of voids is not new, there is no standard definition of what is a void. 'Voids' sometimes mean quite different objects. It all depends on used data and goals of the analysis. For example, to explain the patterns of the galaxy distribution in the Universe Van de Weygaert & Van Kampen (1993) and Sheth & Van de Weygaert (2004) define voids as irregular low-density regions in the density field. Colberg et al. (2005) use a similar definition to study void properties in a ΛCDM universe. However, as these definitions are not based on point distributions, it may be difficult to deal with the galaxy samples provided by large-scale redshift surveys. El-Ad & Piran (1997), Hoyle & Vogeley (2002) and Hoyle & Vogeley (2004) define voids as irregular regions of low number density of galaxies, which may contain bright galaxies. Thus, by construction, voids are not empty even of very luminous and likely massive galaxies. By contrast, Gottlöber et al. (2003) define voids as spherical regions which do not have massive objects (haloes in this case). Voids also can be defined in a statistical point of view as maximal spheres (Otto et al. 1986; Einasto, Einasto & Gramann 1989; Gottlöber et al. 2003; Patiri et al. 2004).

In this paper, we define voids as the maximal non-overlapping spheres empty of objects with mass (or luminosity) above a given value. For example, we could define voids as maximal spheres empty of Milky Way size galaxies. While voids are empty of these galaxies, they could have fainter galaxies inside. See Fig. 1 for a graphical representation of our definition.

The first step to follow using voids as a test for large-scale structure and galaxy formation models is to develop a robust algorithm to detect them, to calibrate it, and to obtain the statistical properties of voids for different catalogues both real and simulated ones. With this information and with the predictions made through the analytical formalism we may be able to contrast different structure formation models. To achieve this goal we develop two algorithms. They are conceptually different but are based on the same definition of void. We develop them as complementary tools. One algorithm is intended to search for all the voids in galaxy or dark matter halo samples and the other is developed to search for the rarest voids.

Once we have the tools to detect the voids, we study statistical properties that will be used to test the structure formation models. These properties go from the void correlations to the redshift dependence of voids. In the present work, we apply the tools we mentioned above in order to study the statistic of voids in the 2dFGRS. One important point that could provide clues on the galaxy formation processes is the galaxy contents of voids. In this work, we present the first results on the distribution of faint galaxies in nearby voids.

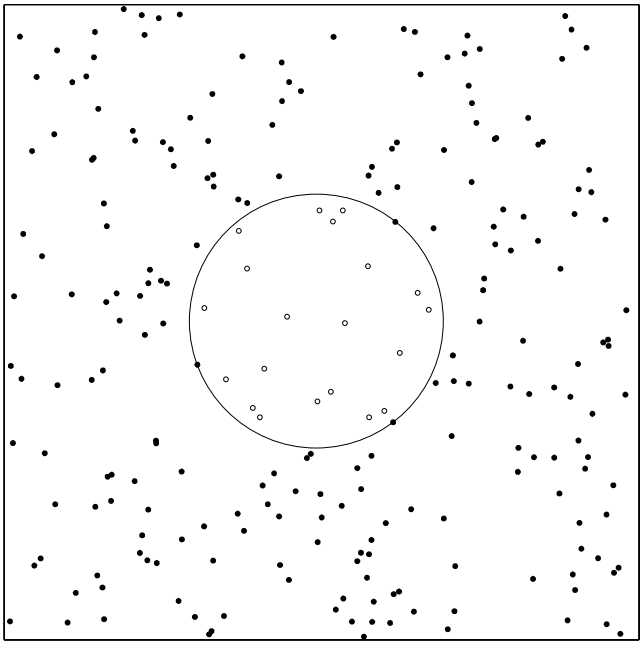


Figure 1. A graphic representation of our void definition. Voids are maximal non-overlapping spheres, in the observational domain, which are empty of objects classified by some intrinsic property. For example, in galaxy samples our voids will be defined by galaxies more luminous than a given luminosity L (filled circles). Open circles denote galaxies with luminosity fainter than L. Note that for simplicity we do not show the open circles located outside the maximal sphere.

In Section 2, we briefly describe how to detect voids. In Section 3, we present the statistics of voids found in cosmological numerical simulations and compare the performance of our algorithms. In Section 4, we present the voids that we have found in the 2dFGRS together with their statistical properties. In Section 5, we develop a method to get from the redshift space coordinates the real space ones. In Section 6, we present the statistics of the galaxies inside nearby rare voids. Finally, in Section 7 conclusions and discussions are presented. In Appendix A, we give details of our void finders presented in Section 2. Here, we also provide tests in order to check the performance of them.

Throughout this work, we adopt a ΛCDM cosmology model with parameters $\Omega_{\rm m} = 0.3$ and $\Omega_\Lambda = 0.7$.

2 VOID DETECTION ALGORITHMS

Once we have defined what is a void (as the maximal non-overlapping spheres empty of galaxies brighter than a given magnitude), the next step is to develop an algorithm to be able to find them in galaxy or dark matter halo samples. The computation structure of the algorithm naturally will depend mainly on the void definition. Following the definition we have stated in the previous section, the algorithm should try to find the maximal sphere that the void can accommodate.

There are many algorithms in the literature inspired by the different void definitions (see e.g. Einasto et al. 1989; Kauffmann & Fairall 1991; El-Ad & Piran 1997; Aikio & Mähönen 1998; Gottlöber et al. 2003). Aikio & Mähönen (1998) and Colberg

et al. (2004) have developed similar algorithms. Following their definitions of voids, they generate a field smoothing the point distribution on a cubic grid and find the local minima; then, they put spheres around those minima filling the entire underdense region. The radius of these voids are the effective ones taken from the sphere that contain the same volume as the void. Gottlöber et al. (2003) have developed an algorithm based on the minimal spanning tree for haloes selected by intrinsic properties, in this case with mass above a given value. Although they have searched for voids as local maximal spheres, they did not do void statistics in their work.

Another algorithm in literature is the so-called 'void finder' developed by El-Ad & Piran (1997) and its modifications done by Hoyle & Vogeley (2002). This algorithm searches for arbitrarily shaped regions delimited by the so-called wall galaxies in order to get the maximum volume of the void. They classify the galaxy sample in wall galaxies and field or voids galaxies by mean of a criterion that depends on the galaxy distribution itself, i.e. they define a length parameter l_n such that any galaxy that does not have n neighbours within a sphere of radius l_n is classify as field galaxy. Note that field galaxies could be, for example, bright galaxies. These field galaxies are removed before searching for voids. Once the classification is done, the algorithm searches for the maximal spheres defined by the wall galaxies based on a cubic mesh. Once all the spheres are obtained, they define an overlapping parameter to discriminate if two spheres belong to the same void. Similarly to Aikio & Mähönen (1998) and Colberg et al. (2004), they finally fill the voids with spheres to get the maximum volume. Again, they define the effective radius of a void by means of a sphere that contain the same volume as the void.

Note that, although this may be interesting for some studies (e.g. the shapes of voids), these kinds of statistics degrade in some amount the information available in the actual galaxy distribution, so they might not be particularly powerful to conduct accurate statistical inferences. This is similar to what happens with the binned data: the best statistical test using binned data is never better and usually worst than the best test using the raw data.

To identify voids, we designed two algorithms that are conceptually different but both are based in our definition of voids. The algorithms are complementary. They help us to investigate the variety of aspects present in the statistics of voids.

One of the algorithms, which we call CELLS Void Finder, was designed to search for all the voids in a galaxy or halo sample based on a computational grid. This grid defines the working resolution. To determine the void centres, the code computes the distances between each of the empty grid cells and all the galaxies or haloes in the whole observational domain, keeping the minimum distance. Once we have the list with the minimum distances, we search for the local maxima which corresponds to the void centres. Obviously, the voids radius are those maximum distances.

The other algorithm, which we denominate HB Void Finder is conceptually simple. This code searches for the maximal non-overlapping spheres with radius *larger* than a given value. First of all, we generate over the sample a large sample of random spheres of a given radius. After this, we check and keep the spheres that are empty of galaxies. We inflate these spheres until they reach the maximum radius. Finally, we eliminate the overlapping spheres keeping the maximal ones. This code is very accurate and computationally efficient to search for the biggest voids in a galaxy sample. Detailed descriptions of the two algorithms are given in Appendix A.

Table 1. Parameters of simulations.

Box (h^{-1} Mpc)	Mass resolution (h^{-1} $M_\odot$)	Number of particles	σ_8
80	3.18×10^8	512^3	0.90
120	1.07×10^9	512^3	0.90
500	7.80×10^{10}	512^3	0.90

3 VOID STATISTICS IN SIMULATIONS

3.1 Numerical simulations

We perform a series of numerical simulations with the Adaptive Refinement Tree code (ART; Kravtsov, Klypin & Khokhlov 1997) and the Tree-SPH code GADGET (Springel, Yoshida & White 2001). Dark matter haloes are identified in the simulation by the Bound Density Maxima algorithm (BDM; Klypin & Holtzman 1997; Klypin et al. 1999).

In this work, we detect and study voids in simulation boxes of $80\,h^{-1}$ Mpc, $120\,h^{-1}$ Mpc and $500\,h^{-1}$ Mpc size. The parameters of all the simulations are summarized in Table 1. With these boxes we have enough volume to study accurately the void statistics and compare the results obtained with our two void finders.

3.2 The statistics

In Fig. 2, we show the number density of voids as a function of their radius obtained with our void finders (squares for CELLS Void Finder and circles for the HB Void Finder). These results were

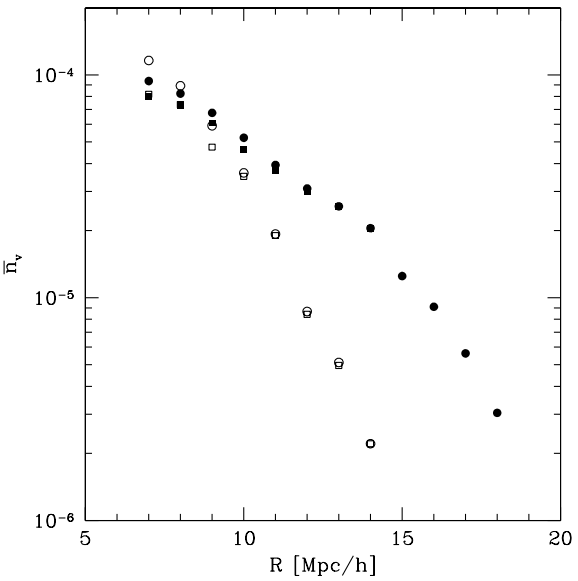


Figure 2. Number density of voids in numerical simulations obtained with our two void finders (squares for the CELLS Void Finder and circles for the HB Void Finder). These results were obtained using our 80 and $120\,h^{-1}$ Mpc boxes. The open circles and squares denotes the number density of voids defined by haloes with masses larger than $1 \times 10^{12}\,M_\odot$ while the filled circles and squares are the number density of voids defined by haloes with masses larger than $5 \times 10^{12}\,M_\odot$. Here, we can see that both algorithms get the same results as voids are bigger (symbols are overlapped). However, as we go to more common voids the differences between codes are bigger. This is due to the fact that the HB Void Finder is not so efficient detecting common voids (see text for details). $\bar{n}_v$ is in units of $\mathrm{Mpc}^3\,h^{-3}$.

obtained for voids defined by two halo masses [$1 \times 10^{12}\,\mathrm{M}_{\odot}$ (open symbols) and $5 \times 10^{12}\,\mathrm{M}_{\odot}$ (filled symbols)].

In order to search for voids with the HB Void Finder, we have generated $1 \times 10^7, 2 \times 10^7, 5 \times 10^7$ trial spheres for the $80\,h^{-1}$ Mpc, $120\,h^{-1}$ Mpc and $500\,h^{-1}$ Mpc boxes, respectively. In the case of the Cell Void Finder we defined a working resolution of $0.5\,h^{-1}$ Mpc for the 80 and $120\,h^{-1}$ Mpc boxes and $1.2\,h^{-1}$ Mpc for the $500\,h^{-1}$ Mpc box.

The obtained statistics shows that both algorithms give very similar results. These results are indeed very useful to learn about the performance of the void finders. Both algorithms give exactly the same results for the largest voids (these are rare voids), the HB Void Finder being the fastest and most precise. However, as we go to more common voids the Cell Void Finder has the best performance. The small differences between both codes for smaller voids are due to the fact that we need more realizations of the trial spheres in the HB Void Finder. As it is expected, the voids are more numerous and larger the more massive the haloes defining the void are.

4 VOID STATISTICS IN THE 2dFGRS

4.1 The galaxy samples

In the present work, we use the 2dFGRS final data release (Colless et al. 2003) to obtain the void statistics in large galaxy redshift samples. The source galaxy catalogue of the 2dFGRS is taken from the Automatic Plate Measuring (APM) galaxy catalogue (Maddox et al. 1990). The spectroscopic targets are galaxies with extinction-corrected magnitudes brighter than $b_{\mathrm{J}} = 19.45$. The median depth of the survey is $z \sim 0.11$. The final data releases contains a total of 221 414 high-quality redshifts. There are two large contiguous survey regions, one in the South Galactic Pole (SGP) and another one towards the North Galactic Pole (NGP). There are also a number of random fields which we have eliminated from our void search. Full details of the 2dFGRS can be found in Colless et al. (2001, 2003).

In order to search for the voids we have selected from 2dFGRS two rectangular regions: the region in the SGP defined by $-34^{\circ}40' < \delta < -25^{\circ}12'$ and $21^{\mathrm{h}}\,49' < \alpha < 3^{\mathrm{h}}26'$ and the region in the NGP defined by $-4^{\circ}35' < \delta < 2^{\circ}17'$ and $9^{\mathrm{h}}33' < \alpha < 14^{\mathrm{h}}54'$; that is, $\sim$690 and $\sim$550 deg^2, respectively.

The 2dFGRS is magnitude limited, i.e. the survey has been constructed by taking spectra of galaxies brighter than a fixed apparent magnitude of $b_{\mathrm{J}} = 19.45$. However, the survey is homogeneously complete up to 90 per cent at $b_{\mathrm{J}} = 19.0$ (see Norberg et al. 2002). A magnitude-limited galaxy survey is not uniform in space, since intrinsically faint galaxies have been observed only if they are relatively nearby, while at large distances only bright galaxies will be targeted. This non-uniformity of the magnitude-limited survey must be taken into account in order to make our void analysis. There are mainly two ways to deal with this; one is to use the selection functions provided by the 2dFGRS and another more simple way which we have followed here is to build volume-limited samples.

We construct four volume-limited samples, two for each survey region (SGP and NGP), one with depth $D_{\max} = 406.15\,h^{-1}$ Mpc which corresponds to $z = 0.14$. For this $D_{\max}$ we have a limiting absolute magnitude $M_{b_{\mathrm{J}}}^{\lim} = -19.32 + 5\log h$ and the other with depth $D_{\max} = 571.71\,h^{-1}$ Mpc corresponding to $z = 0.2$ and a limiting absolute magnitude $M_{b_{\mathrm{J}}}^{\lim} = -20.181 + 5\log h$. All distances are comoving ones. In Table 2, we give the properties of the volume-limited samples. Now, we have guaranteed that any galaxy brighter than $M_{b_{\mathrm{J}}}^{\lim}$ is observed in our volume. We have computed the absolute magnitudes from the apparent magnitudes assuming a ΛCDM cosmology and applying the needed corrections to model the change in the galaxy magnitudes due to the redshifted b_{J} filter bandpass (k-correction) and to account for the galaxy evolution (e-correction). These corrections for each galaxy are given in Norberg et al. (2002).

Table 2. Parameters of our volume-limited samples.

Name	$M_{\lim}$ $M_{b_{\mathrm{J}}} - 5\log h$	$z_{\max}$	$D_{\max}$ (h^{-1} Mpc)	Volume ($10^6\,h^{-3}$ Mpc^3)	N_{galaxies}
SGP1	−19.32	0.14	406.15	4.693	22 037
SGP2	−20.181	0.20	571.71	13.088	14 475
NGP1	−19.32	0.14	406.15	3.749	19 695
NGP2	−20.181	0.20	571.71	10.456	11 404

To test the spatial homogeneity of our SGP1,2 and NGP1,2 volume-limited samples we have computed the average of the cube of the radial distances, i.e. a modified version of the $V_{\max}$ test (Rowan-Robinson 1968), which for a homogeneous sample must satisfy

$$\left\langle \left(\frac{r}{r_{\max}} \right)^3 \right\rangle = \frac{1}{2} \pm \frac{1}{\sqrt{12N'}}, \qquad N' = \frac{N}{(1 + N\langle \xi(r) \rangle)}, \tag{1}$$

where r is the comoving distance to each galaxy, $r_{\max}$ is the same as $D_{\max}$, the maximum distance of the sample, N is the number of galaxies (given in Table 2) and $\langle \xi(r) \rangle$ is the average value of the correlation function over all pair of galaxy positions within the sample, defined as

$$\langle \xi \rangle = \frac{1}{V^2} \int \int \xi(|\boldsymbol{r}_1 - \boldsymbol{r}_2|)\, \mathrm{d}\boldsymbol{r}_1\, \mathrm{d}\boldsymbol{r}_2, \tag{2}$$

where V is the volume of the sample. Assuming for $\xi(r)$ (Peebles 1980):

$$\xi(r) = \left(\frac{r}{5.4\,h^{-1}\,\mathrm{Mpc}} \right)^{-1.77}, \tag{3}$$

we found $\langle \xi(r) \rangle = 6.7910 \times 10^{-3}$. It must be noted that we have used equation (3) for any value of r, while in fact, it is known that for $r \gg 10\,h^{-1}$ Mpc $\xi(r)$ must be close to the Fourier transform of the linear power spectra. This would lead to smaller values of $\langle \xi(r) \rangle$. However, this would not affect our analysis too much. With this value we find that for a homogeneous sample

$$\eta \equiv \left\langle \left(\frac{r}{r_{\max}} \right)^3 \right\rangle = \frac{1}{2} \pm 0.024. \tag{4}$$

The actual η values for the SGP1 and NGP1 are $\eta(SGP1) = 0.524$ and $\eta(NGP1) = 0.51$. For the SGP2 and NGP2 to be homogeneous (assuming that $\langle \xi(r) \rangle$ takes the same values as in the previous sample):

$$\eta \equiv \left\langle \left(\frac{r}{r_{\max}} \right)^3 \right\rangle = \frac{1}{2} \pm 0.0253, \tag{5}$$

and the actual values are $\eta(SGP2) = 0.511$ and $\eta(NGP2) = 0.475$. Therefore, we can conclude that we do not detect inhomogeneities.

4.2 The void statistics

As described in Section 3.2, once we have constructed our samples, we have to define which objects will define the voids. In the SG1 and

NGP1 volume-limited samples we present in Table 2, we search for voids defined by two types of galaxies. On the one hand, we search voids defined by galaxies brighter than $M_{b_J} = -19.32 + 5\log h$. Using these samples we also find voids defined by galaxies brighter than $M_{b_J} = -20.5 + 5\log h$. The number of galaxies brighter than $M_{b_J} = -20.5 + 5\log h$ is 2427 for the SGP1 sample and 2074 for the NGP1. In the SGP2 and NGP2 samples, we have searched for voids defined by galaxies brighter than $M_{b_J} = -20.181 + 5\log h$. In Fig. 3, we show a plot with the voids we have detected in the NGP1 sample. In Table 3, we summarize the results for the NGP1 and SGP1 samples. Note that, as we saw in the simulations, voids defined by bright galaxies are larger than the ones defined by fainter ones. In Table 4, we present the statistics for the NGP2 and SGP2 samples. We have obtained these voids applying the HB algorithm. In the radius ranges that we present here both codes have a similar performance. We generate 8×10^7 trial spheres to search for voids larger than $7.5\,h^{-1}$ Mpc which are our main interest.

We have also calculated the VPF which is the probability that a randomly located sphere of fixed radius contains no galaxies. Fig. 4 presents our VPF, which we find for the SGP1,2 and NGP1,2 samples. We calculate the *rms* of the VPF in the following way. Assuming that voids are independent, we get from the central limit theorem that

$$\mathrm{rms}^2(P_0(r)) = \frac{N(r)}{V^2}(\langle \Delta v^2\rangle - \langle \Delta v\rangle^2), \tag{6}$$

where Δv is the volume where the centre of an empty sphere of radius r may be moved so that it remains empty. V is the volume of the sample and $N(r)$ is the number of voids found with radius larger than r in that volume. The parenthesis in the right-hand side of equation (6) divided by V^2 is the contribution of each individual void to the variance of the VPF.

On the other hand, we have from Betancort-Rijo (1992) that

$$\frac{N(r)}{V} = \frac{P_0(r)}{\langle \Delta v\rangle}. \tag{7}$$

Now we square this expression and substitute it in expression (6) to obtain

$$\mathrm{rms}^2(P_0(r)) = \frac{P_0^2(r)}{N(r)}[g(r) - 1], \tag{8}$$

$$g(r) \equiv \frac{\langle \Delta v^2\rangle}{\langle \Delta v\rangle^2}, \tag{9}$$

so that, taking into account the correlations in the sample, the rms of the VPF is given by

$$\mathrm{rms}(P_0(r)) \simeq g(r)^{1/2}\frac{P_0(r)}{N(r)^{1/2}}[1 - w\bar{n}_v(R)]^{1/2}, \tag{10}$$

where $w\bar{n}_v(R)$ is as defined in equation (21). In the rare voids limit, i.e. when $P_0 \ll 1$ it is shown that (Betancort-Rijo 1992)

$$g(r) \simeq 9.20. \tag{11}$$

For voids with radius larger than $12\,h^{-1}$ Mpc in the SGP1 and NGP1 samples and larger than $17\,h^{-1}$ Mpc in the SGP2 and NGP2 samples, the asymptotic value $g(r)^{1/2} = 2.86$ is a good approximation. For smaller values of r, $g(r)^{1/2}$ is somewhat smaller (no less than $\sim$2). So in Fig. 4, where the asymptotic value of $g(r)$ have been used, the error may be slightly overestimated for small values of r. It must be noted, however, that the probability distribution of the fluctuations of the VPF around the mean is strongly non-Gaussian for small values of $N(r)$; most fluctuations being quite smaller than the rms and a few of them being very large. So, to decide the compatibility between couples of measurements, this fact has to be taken into account. In the present case, however, this problem do not arise because all results are within the error bars.

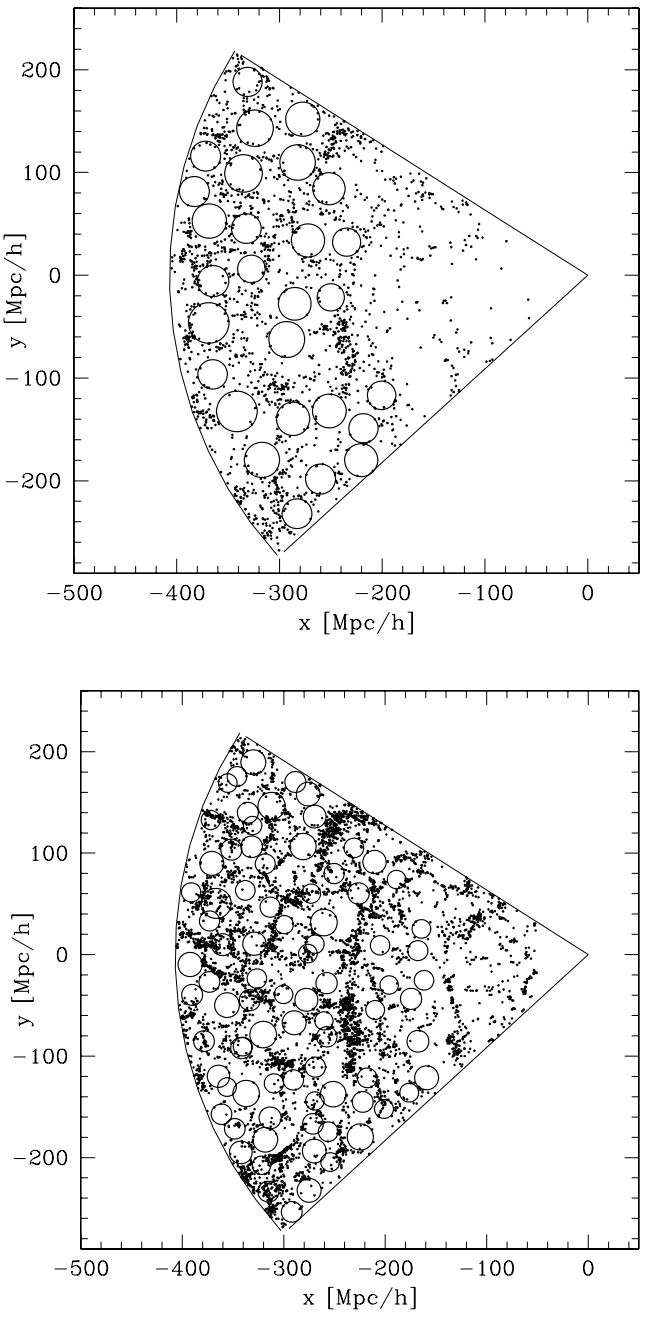


Figure 3. Voids in the 2dFGRS. We show the maximal spheres in the NGP1 volume-limited sample of galaxies (filled circles). The upper panel shows the voids with radius larger than $7.5\,h^{-1}$ Mpc defined by galaxies brighter than $M_{b_J} = -20.5 - 5\log h$. There are 2074 galaxies in this sample. In the lower panel we show the voids with radius larger than $13.0\,h^{-1}$ Mpc defined by galaxies brighter than $M_{b_J} = -19.32 - 5\log h$. The number of galaxies is 19 695. Some galaxies lie inside the maximal spheres due to projection effects.

Our results are in good agreement with the previous calculations of the VPF (Croton et al. 2004; Hoyle & Vogeley 2004). In Croton et al. (2004), the VPF was obtained for the whole 2dFGRS while

Table 3. Statistics of Voids in our NGP1 and SGP1 samples. $N_{\rm NGP1}$ and $N_{\rm SGP1}$ are the number of voids larger than the given radius. We present the void statistics for two different defining galaxies.

Radius (h^{-1} Mpc)	$N_{\rm NGP1}$	$N_{\rm SGP1}$
$M_{b_{\rm J}}^{\rm lim} = -19.32 - 5\log h$		
7.5	136	220
10.0	48	80
12.0	11	25
13.0	6	11
14.0	1	6
$M_{b_{\rm J}}^{\rm lim} = -20.5 - 5\log h$		
13.0	28	43
14.0	24	35
16.0	14	26
17.0	7	17
19.0	2	6

Table 4. Statistics of voids in our NGP2 and SGP2 samples.

Radius (h^{-1} Mpc)	$N_{\rm NGP2}$	$N_{\rm SGP2}$
$M_{b_{\rm J}}^{\rm lim} = -20.181 - 5\log h$		
13.0	68	101
15.0	36	53
16.0	25	31
17.0	9	23
18.0	5	14
21.0		2

Hoyle & Vogeley (2004) have studied the VPF for both strips (NGP and SGP). In this last result, they found that the VPF for both strips are not compatible for large spheres (see fig. 7 in Hoyle & Vogeley 2004). This is be due to the fact that they underestimated the error bars. We have shown above that using a more accurate expression for the rms of the VPF we obtain that both strips are compatible.

4.3 Voids spatial distribution

If we neglect the redshift dependence of the power spectra, the spatial distribution of the voids found in a statistically homogeneous sample must be statistically homogeneous itself, but conditioned to the fact that the maximal sphere must lay within the sample. This fact implies that the centre of the maximal spheres cannot occupy all the volume of the sample. The available volume, $V(R)$, for the maximal spheres of radius R is

$$V(R) = \int_{R/\sin(\Delta\delta/2)}^{r_{\rm max}-R} \int_{\delta_0}^{\delta_0+\Delta\delta} \int_{\alpha_0+B(r,R,\delta)}^{\alpha_0+\Delta\alpha-B(r,R,\delta)} r^2 \cos\delta \, {\rm d}\alpha \, {\rm d}\delta \, {\rm d}r$$

$$B(r, R, \delta) = \sin^{-1}\left(\frac{\sin^{-1}(R/r)}{\cos\delta}\right). \quad (12)$$

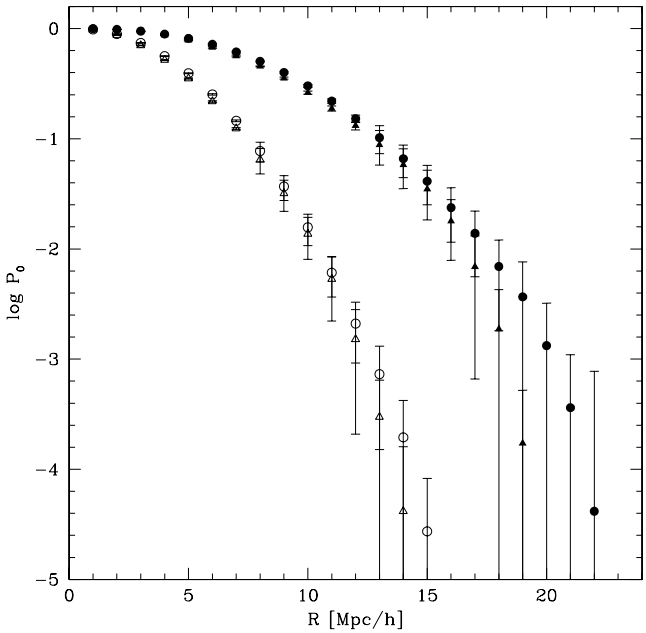


Figure 4. VPF for the 2dFGRS. We show the VPF for the SGP1,2 samples (open and filled circles, respectively) and also for the NGP1,2 samples (filled and open triangles, respectively).

Equation (12) may be considerably simplified by carrying out the integrals over α and δ keeping B fixed (using its value at $\delta_0 + \Delta\delta/2$). This is a very good approximation since $\Delta\delta \ll 1$. So, we could write

$$V(R) \simeq \int_{R/\sin(\Delta\delta/2)}^{r_{\rm max}-R} P(r, R) r^2 \, {\rm d}r, \quad (13)$$

$$P(r, R) \equiv \left[\sin\left(\delta_0 + \Delta\delta - \sin^{-1}\left(\frac{R}{r}\right)\right) - \sin(\delta_0)\right] \times \left[\Delta\alpha - 2\sin^{-1}\left(\frac{\sin^{-1}(R/r)}{\cos(\delta + \Delta\delta/2)}\right)\right], \quad (14)$$

where $\Delta\delta$, $\Delta\alpha$ are, respectively, the widths of the strips in declination and right ascension, while δ_0 is its southernmost declination. Although expression (13) for V(R) is not exact, its percentual error is completely negligible (less than 10^{-5}).

Using $P(r)$ we may obtain the mean of the cube of the distance (from the observer), r, to the centre of the maximal sphere with radius larger than R':

$$\eta(R') \equiv \left\langle \left(\frac{r}{r_{\rm max}}\right)^3 \right\rangle_{R \geqslant R'} = \frac{1}{V(\bar{R})} \int_{\bar{R}/\sin(\Delta\delta/2)}^{r_{\rm max}-\bar{R}} P(r, \bar{R}) \left(\frac{r}{r_{\rm max}}\right)^3 r^2 \, {\rm d}r, \quad (15)$$

where $\bar{R}$ is the mean radius of all maximal spheres larger than R'.

In a similar manner, we may obtain the variance for the estimate of $\eta(R/h^{-1}\,{\rm Mpc})$ from the void list we have obtained in the 2dFGRS. The values found in the SGP1,2 and NGP1,2 for voids defined by galaxies brighter than $-19.32 + 5\log h$ are listed in Table 5).

From the analysis of these results we may conclude that the voids are essentially compatible with homogeneity. There is, however, a slight, although statistically significant, trend to values lower than homogeneous that we will discuss in Section 3.5.

Table 5. The values of $\eta(R')$ found in the SGP1,2 and NGP1, 2 samples. The values within the parentheses are 1σ predictions assuming that the centre of the maximal spheres are uniformly distributed conditioned to lay entirely within the sample (see text).

Radius (h^{-1} Mpc)	η	η
	NGP1	SGP1
7.5	0.512 (0.563 ± 0.02)	0.503 (0.53 ± 0.02)
10.0	0.579 (0.579 ± 0.038)	0.508 (0.54 ± 0.03)
13.0	0.66 (0.615 ± 0.116)	0.433 (0.558 ± 0.08)
	NGP2	SGP2
18.0	0.475 (0.611 ± 0.085)	0.588 (0.556 ± 0.06)

Now, we will consider the nearest-neighbour statistics which together with the void one-point statistic we have studied above is sufficient to grant the validity of the statistical analysis in the following subsections.

The maximal spheres are chosen so that they do not overlap, in order that they very rarely corresponds to the same connected underdensity. So, by construction, for the maximal spheres with radius larger than R the two-point correlation function of their centres is -1 at least to a distance of $2R$. The main question is how is the correlation at larger distances. Note that if more than one maximal sphere were associated with the same underlying connected underdensity the correlation would be positive for distances somewhat larger than $2R$. However, as there is typically only one maximal sphere per underdense region and in the standard scenario of structure formation these regions are essentially uncorrelated (for the relevant R values) (Colberg et al. 2004), we do not expect to find correlations between the maximal spheres.

To test whether voids are correlated we compare the actual nearest-neighbour statistics with the theoretical predictions corresponding to centres which are uncorrelated for distances larger than $2\bar{R}$ and completely anticorrelated for smaller distances. For maximal spheres with radius R the theoretical predictions for the mean and quadratic mean distance to the nearest neighbour ($\langle D_{\rm oo}\rangle$ and $\langle D_{\rm oo}^2\rangle$, respectively) is given by

$$\langle D_{\rm oo}\rangle = \left[\frac{4\pi\bar{n}_v(R)}{3}\right]^{-1/3} \int_{w_0}^{\infty} {\rm e}^{-(w-w_0)} w^{1/3}\,{\rm d}w, \tag{16}$$

$$\langle D_{\rm oo}^2\rangle = \left[\frac{4\pi\bar{n}_v(R)}{3}\right]^{-2/3} \int_{w_0}^{\infty} {\rm e}^{-(w-w_0)} w^{2/3}\,{\rm d}w, \qquad w_0 \equiv \frac{4\pi}{3}(2\bar{R})^3\bar{n}_v(R), \tag{17}$$

where $\bar{n}_v(R)$ is the mean number density of the maximal spheres larger than R and $\bar{R}$ is their mean radius. The sampling error of $\langle D_{\rm oo}\rangle$ when estimated using N centres is

$${\rm rms}(\langle D_{\rm oo}\rangle) = \frac{\sqrt{5}\left(\left\langle D_{\rm oo}^2\right\rangle - \langle D_{\rm oo}\rangle^2\right)^{1/2}\right)}{(3N-2)^{1/2}}. \tag{18}$$

With these expressions we may compute the mean and the variance of q:

$$q\left(\frac{R}{h^{-1}\,{\rm Mpc}}\right) \equiv [\bar{n}_v(R)]^{1/3} E(\langle D_{\rm oo}\rangle), \tag{19}$$

where $E(\langle D_{\rm oo}\rangle)$ is the estimate of $\langle D_{\rm oo}\rangle$ using the void sample. For typical voids we find for q:

$$q_{\rm SGP1}(7.5) = 0.978\,(0.958 \pm 0.006),$$
$$q_{\rm NGP1}(7.5) = 1.029\,(0.958 \pm 0.008),$$
$$q_{\rm SGP2}(13.0) = 0.966\,(0.879 \pm 0.01),$$
$$q_{\rm NGP2}(13.0) = 1.1\,(0.879 \pm 0.013),$$

where $R = 7.5\,h^{-1}$ Mpc($\bar{R} = 9.58\,h^{-1}$ Mpc) corresponds to SGP1, NGP1 and $R = 13.0\,h^{-1}$ Mpc($\bar{R} = 15.08\,h^{-1}$ Mpc) to SGP2, NGP2. The values in the parenthesis correspond to 1σ predictions obtained using equations (16) and (17).

It is apparent that the estimated values are slightly shifted upwards. The bias is large in terms of the rms values but we cannot conclude from this that the centres are anticorrelated since the predictions do not account for border effect, which in this case is small for the SGP samples but larger than the rms.

On the other hand, for rare voids, i.e. $R = 13.0\,h^{-1}$ Mpc for the SGP1, NGP1 samples ($\bar{R}_{\rm SGP1} = 14.14\,h^{-1}$ Mpc, $\bar{R}_{\rm NGP1} = 13.42\,h^{-1}$ Mpc) and $R = 18.0\,h^{-1}$ Mpc for the SGP2, NGP2 samples ($\bar{R}_{\rm SGP2} = 19.5\,h^{-1}$ Mpc, $\bar{R}_{\rm NGP2} = 18.45\,h^{-1}$ Mpc) we find

$$q_{\rm SGP1}(13.0) = 0.959\,(0.716 \pm 0.004),$$
$$q_{\rm NGP1}(13.0) = 1.398\,(0.716 \pm 0.006),$$
$$q_{\rm SGP2}(18.0) = 0.971\,(0.724 \pm 0.046),$$
$$q_{\rm NGP2}(18.0) = 1.071\,(0.724 \pm 0.070).$$

The upward bias with respect to the values for the uncorrelated centres (without accounting for border effects) is now large even in the SGP. To assess the border effect we have simulated a random set of spheres with $R = 19.5\,h^{-1}$ Mpc (the mean radius of the maximal spheres with radius $\geqslant 18.5\,h^{-1}$ Mpc in the SGP2 sample) with the condition that they lay entirely within the catalogue and do not overlap, then we obtain the nearest-neighbour statistics. The average over several realizations of the mean nearest distance is

$$\bar{n}_v(18.0)\langle\bar{D}_{\rm oo}\rangle = 0.88 \pm (0.05 \pm \epsilon), \tag{20}$$

which is quite larger than the value predicted without border effect (0.724). This is not enough, however, to account for the measured value (0.971). The same result is found for $z_{\rm max} = 0.14$ since the border effect in this case must be very close to that in the previous case. From these results we might conclude at least at the 99 per cent confidence level (being rather conservative about the value of ϵ) that the centre of the maximal spheres show some anticorrelation. We prefer, however, to conclude simply that those centres are basically uncorrelated, which is the necessary result for the subsequent statistical inferences, and leave open the question of the possible existence of a weak anticorrelation, which shall be treated in more detail in a future work. Incidentally, it must be noted that a slight anticorrelation of the centres arise naturally in the standard scenario, being compatible with the results of Colberg et al. (2004). The reason is that the connexed underdense regions are somewhat larger than the maximal spheres they contain.

4.4 Compatibility between SGP and NGP samples

Here, we compare the void statistics found in our samples in order to assess their compatibility. To this end we must estimate the mean number density of voids, $\bar{n}_v(R)$, in each sample and obtain their sampling error. To estimate $\bar{n}_v(R)$ we simply divide the number of voids, $N(R)$, with radius larger than R, by the mean available volume $V(\bar{R})$. To estimate the sampling error we must take into account that, although the centre of the maximal spheres are basically

uncorrelated for $r > 2\bar{R}$, they are totally anticorrelated for $r < 2\bar{R}$ (we use a model in which all spheres have radius $\bar{R}$). In this case the number of spheres in a given volume do not follow a Poissonian distribution, but a negative binomial one (Betancort-Rijo 1992). For this distribution the sampling error of the estimate of $\bar{n}_v(R)$ could be obtained by means of the following expression:

$$\mathrm{rms}(\bar{n}_v(R)) = \frac{\bar{n}_v(R)}{[N(R)]^{1/2}}[1 - w\bar{n}_v(R)]^{1/2}, \tag{21}$$

$$w \equiv \frac{32\pi}{3}\bar{R}^3\left[1 + 2.73\bar{n}_v(R)\frac{32\pi}{3}\bar{R}^3 \times \left\{1 - \frac{3}{4}\frac{\bar{n}_v(R)^{-1/3}}{2\bar{R}} + \frac{1}{8}\left[\frac{\bar{n}_v(R)^{-1/3}}{2\bar{R}}\right]^2\right\}\right]^{-1}.$$

For SGP1 we find

$$\bar{n}_v(7.5) = 9.364 \pm 0.164 \times 10^{-5},$$
$$\bar{n}_v(10.0) = 3.98 \pm 0.11 \times 10^{-5},$$
$$\bar{n}_v(13.0) = 7.26 \pm 1.70 \times 10^{-6}.$$

For NGP1 we find

$$\bar{n}_v(7.5) = 9.597 \pm 0.214 \times 10^{-5},$$
$$\bar{n}_v(10.0) = 4.27 \pm 0.15 \times 10^{-5},$$
$$\bar{n}_v(13.0) = 6.56 \pm 2.10 \times 10^{-6}.$$

The corresponding $\bar{R}$ values are $\bar{R}(7.5) = 9.58$ and $\bar{R}(10.0) = 11.35$ (the mean over NGP and SGP), $\bar{R}(13.0) = 14.14$ in the SGP1 and $\bar{R}(13.0) = 13.41$ in the NGP1.

It is apparent that the three pairs of values are compatible. It is true that for the SGP1 the galaxy density is roughly a 10 per cent smaller than for the NGP1 and that the theoretical expression (Patiri et al. 2004) predicts about a 20 per cent enhancement of the density of voids with $R \geqslant 13\,h^{-1}$ Mpc in the SGP1 over that in the NGP1, and smaller differences for more common voids, but this differences do not show up above the sampling errors.

Similar results were found for the SGP2:

$$\bar{n}_v(13.0) = 1.685 \times 10^{-5},$$
$$\bar{n}_v(15.0) = 9.49 \times 10^{-6},$$
$$\bar{n}_v(18.0) = 2.978 \pm 0.58 \times 10^{-6},$$

and for the NGP2:

$$\bar{n}_v(13.0) = 2 \times 10^{-5},$$
$$\bar{n}_v(15.0) = 1.202 \times 10^{-5},$$
$$\bar{n}_v(18.0) = 2.22 \pm 0.69 \times 10^{-6}.$$

Although the implications of this results for the large-scale structure formation models is left to for a future work, it is interesting to note that they are generally in good agreement with the predictions of the standard model.

4.5 Redshift dependence of void number densities

We have checked that the galaxies in our samples are, within the sampling errors, homogeneously distributed over the sample volumes. In this situation, the only possible explanation of a statistically significant departure of the void distribution from the homogeneity must be the redshift dependence of the properties of the large-scale structure.

Using a theoretical framework (Patiri et al. 2004) to compute the number density of voids in the distribution of dark matter haloes, we found that the predicted number density of voids larger than $13.0\,h^{-1}$ Mpc for $M_{bj}^{\lim} = -19.32 + 5\log h$ decreases by 28 per cent from $z = 0$ to 0.1 and a further 28 per cent when going to $z = 0.2$. For $M_{bj}^{\lim} = -20.181 + 5\log h$ and $R \geqslant 18\,h^{-1}$ Mpc the void density decreases by 26 per cent from $z = 0$ to 0.1 and another 28 per cent from $z = 0.1$ to 0.2. Where, in order to link light with dark matter haloes, we assumed that there is one galaxy in each dark halo. These numbers suggest the possibility of a measurable effect. To this end we conduct first a test where we divide every sample into two bins (a 'near' region and a 'far' region) both with the same volume available for the voids considered. Then we compare the total number of voids larger than the chosen radius in the near regions with the total number in the far regions. If the first number is significantly larger than the latter this should be interpreted as evidence for a redshift dependence.

The problem now is that the effect we are trying to measure is strong only for the rare voids which have poorer statistics. We reach a compromise between both trends. We have chosen voids with radius ($R_{\min}$) larger than $13\,h^{-1}$ Mpc for the $z_{\max} = 0.14$ sample and $18\,h^{-1}$ Mpc for the $z_{\max} = 0.2$ sample. In order to divide a sample into two subsamples with the same volume available for the centres of the voids, we first obtain the mean radius ($\bar{R}$) of the voids larger than $R_{\min}$, and, using equation (12) with $R = \bar{R}$, we obtain the available volume of the whole sample (which has already been done in Section 4.3). Then, we use again (12) with $R = \bar{R}$ but replacing $r_{\max}$ with $r_{\lim} + \bar{R}$ and search for the value of $r_{\lim}$ giving for V half the value corresponding to the whole sample. If voids were uniformly distributed their centres would lie above and below $r_{\lim}$ with equal probability. We have explicitly checked this fact with numerical simulations. For the sample SGP1, we have chosen $R_{\min} = 13\,h^{-1}$ Mpc, so $\bar{R} = 14.14\,h^{-1}$ Mpc which implies $r_{\lim} = 371.7h^{-1}$ Mpc. Nine voids were found with their centres below this distance and two above. Proceeding in a similar way with the other samples, we found three voids closer than $349.1\,h^{-1}$ Mpc and three voids farther for the NGP1. For the SGP2 we found five voids closer than $475.0\,h^{-1}$ Mpc and eight farther, while for the NGP2 we found four voids closer than $490.2\,h^{-1}$ Mpc and one farther. So, in total we have 21 voids in near regions and 14 in the far regions.

With these numbers we can infer at least at the 88 per cent confidence level the presence of a redshift dependence. However, this is only a marginal evidence. To make more patent this evidence we use a more efficient test. This test uses the η values obtained in Section 4.3 which are shown in Table 5. We use the η values corresponding to voids larger than $13\,h^{-1}$ Mpc in the NGP1 and SGP1 samples, and voids larger than $18\,h^{-1}$ Mpc in the NGP2 and SGP2 samples. From each of those four η values, we subtract its expected value when uniformity is assumed (first number in the parenthesis in Table 5) and divide the result by the corresponding rms (second number in the parenthesis). Each of these quantities follow a Gaussian distribution with zero mean and variance 1 under the uniformity hypothesis. So, the sum of the four quantities must follow a Gaussian with zero mean and variance 4 and the probability that this sum be smaller than the actually found result is

$$1 - \frac{1}{2}\,\mathrm{erfc}\left(\frac{\sum_{i=1}^{4}\frac{\eta_i - \bar{\eta}_i}{\mathrm{rms}(\eta_i)}}{2\sqrt{2}}\right) = 0.005. \tag{22}$$

So, we may infer at the 99.5 per cent confidence level the existence of non-uniformity on the void distribution, that, given the uniformity of the galaxy distribution can only be explained by the growth of density fluctuations as redshift decreases. Note that the rms(η) given in Table 5 do not account for the anticorrelation between the

void centres (when $r < 2R$). The last factor in expression (21) approximately accounts for this fact. If we were to use it, all the rms(η) values would be smaller and, consequently, we could reject the uniformity hypothesis at a larger confidence level. However, we prefer to give the conservative value quoted above since it does not have any uncertainty.

5 REDSHIFT SPACE DISTORTIONS

Redshift galaxy maps like the 2dFGRS or SDSS are distorted by the peculiar velocities of galaxies along the line of sight. This effect, which produce deformations such as the 'finger of God', is due to the fact that the distance to the survey galaxies is obtained by means of the Hubble distance which is obtained with the total velocity (the Hubble flow plus the peculiar velocity).

We present in this section a method to correct this effect for the galaxies around voids. We first suppose that if in the distorted space we have a void of radius R, we will have the same void of radius R in the *real* space. This is a good approximation because the distortion around a void causes an elongation along the line of sight of the maximal sphere without changing the 'transversal' radius, i.e. although the distortion is not volume conservative, the maximal radius will be approximately the same in both spaces.

With this assumption, we have in the position of the void in the distorted space a mean underdensity, $\bar{\delta}_0$, within the maximal sphere with radius R (Patiri et al. 2004):

$$\bar{\delta}_0 = \int_{-1}^{\infty} \delta_0 P\left(\frac{\delta_0}{R}\right) \mathrm{d}\delta_0, \tag{23}$$

where $P(\delta_0/R)$ is the probability distribution for δ_0 within a maximal sphere of radius R.

We associate with this void a matter distribution given by the mean profile, $\delta(r/\bar{\delta}_0, R)$. In Patiri et al. (2004) we derive this profiles. So, it has a peculiar velocity profile $V(\delta(r), r)$ given by the spherical collapse model (Betancort-Rijo et al. 2005):

$$\frac{V(\delta(r), r)}{H} = \frac{-0.51}{3} \frac{r}{1+\delta} \delta_l(\delta) \left[\frac{\mathrm{d}\delta_l(\delta)}{\mathrm{d}\delta}\right]^{-1}, \tag{24}$$

where H is the Hubble constant, $\delta_l(\delta)$ is given in Sheth & Tormen (2002) and δ is the mean profile (note that for simplicity we do not show the dependence on r, δ_0 and R). We have derived an analytical approximation for $\delta(r/\bar{\delta}_0, R)$ which gives good results for $r \leqslant 1.5R$:

$$\delta(r/\bar{\delta}_0, R) \simeq \bar{\delta}_0 + (0.1645 + 0.085\bar{\delta}_0)\left(\frac{r}{1.4R}\right)^7, \tag{25}$$

so that, we correct the galaxy 'measured' distance $r(z)$:

$$r_{\mathrm{real}} \simeq r(z) - \frac{(\boldsymbol{x} - \boldsymbol{x}_c).\boldsymbol{x}}{|(\boldsymbol{x} - \boldsymbol{x}_c)||\boldsymbol{x}|} \frac{V(\delta(r/\bar{\delta}_0, R), r)}{H}, \tag{26}$$

where $\boldsymbol{x}$ is the vector to the galaxy and $\boldsymbol{x}_c$ is the vector to the centre of the void. We apply this correction to galaxies with distance to the centre of the void $\leqslant 1.5R$. Once we have the corrected catalogue, we search for voids over this new catalogue.

In order to test our technique, we have applied it to a simulated catalogue, the Millennium Run galaxy catalogue (Springel et al. 2001). The publicly available catalogue[1] contains a total of about nine million galaxies in the simulation box of $500\,h^{-1}$ Mpc. For each galaxy, it is available the position and velocity, the total and bulge galaxy magnitudes in five bands (*ugriz* SDSS bands), the total and bulge stellar mass, cold, hot and ejected gas mass, the black hole mass and the star formation rate. The dark matter haloes in the simulation were populated using semi-analytic models of galaxy formation (see Croton et al. 2005 for full details). We have constructed from the full simulated box a smaller one of $250\,h^{-1}$ Mpc. With this box we have enough volume to study large (and rare) voids. Also, as we have available in the original catalogue the coordinates in real space, we have constructed another box of $250\,h^{-1}$ Mpc with the same galaxies but with the coordinates in redshift space.

[1] It can be downloaded from http://www.mpa-garching.mpg.de/galform/agnpaper/

Table 6. The void statistics found in the three catalogues generated from the millennium galaxy catalogue: the redshift distorted, the real and the corrected. The first column show the number of voids found with radius larger than the values given in the three columns on the right-hand side corresponding to each catalogue.

Number of voids	Radius [(h^{-1} Mpc)] Distorted	Real	Corrected
1	18.17	17.48	17.58
5	16.51	16.19	15.89
10	15.70	15.09	14.93
15	15.42	14.57	14.58
20	14.93	14.36	14.15

We have applied our HB Void Finder to search for voids larger than $13.0\,h^{-1}$ Mpc in both distorted and real catalogues. With the list of voids in the distorted catalogue, we have applied the correction to those galaxies that lie within 1.4 times the radius of the voids. After we have obtained the corrected galaxy catalogue, we run the void finder over this catalogue. In Table 6, we show the statistics of voids in the three different catalogues: the distorted, the real and the corrected.

From the study of the simulated catalogues we learn that, although the correction may be larger than $3\,h^{-1}$ Mpc (for $R \geqslant 16\,h^{-1}$ Mpc) for galaxies close to the line of sight to the centre of the voids, the correction for the radius of the maximal sphere is, as expected, much smaller. Even so, the difference is not negligible for sufficiently rare voids. We find that the 20 largest voids in the simulated catalogues are on average slightly smaller ($\sim$5 per cent) in the corrected catalogue.

From the void statistics found in the corrected SGP1 and NGP1 samples we find that for the corrected SGP1, the 10 largest voids are on average $0.83\,h^{-1}$ Mpc smaller than for the uncorrected one and that the number of voids larger than $12\,h^{-1}$ Mpc in both samples is 25 for the uncorrected case and 14 for the corrected one. So, since the strongest constraint on the models comes from relatively rare voids, it seems likely that the corrected catalogues must be used in order to be able to make accurate inferences.

6 GALAXIES IN NEARBY VOIDS

In this section, we study the galaxy content of the nearby rarest voids in our 2dFGRS samples. To this end, we have selected the voids from our SGP1 volume-limited sample described in Section 5.1 (see Table 3). The voids are defined by galaxies brighter than $-19.32 + 5\log h$. We have searched for faint galaxies down to $M_{b_{\mathrm{J}}}^{\mathrm{lim}} = -18.3 + 5\log h$ inside nine voids with radius larger than $13\,h^{-1}$ Mpc in a volume-limited sample up to $z = 0.095$. These voids are uncommon (with mean radius of $14.0\,h^{-1}$ Mpc) due to the fact that a randomly placed sphere with radius equal to the mean size of these voids should contain about 50 galaxies brighter than $-19.32 + 5\log h$. In total we find 130 faint galaxies inside these nine rare

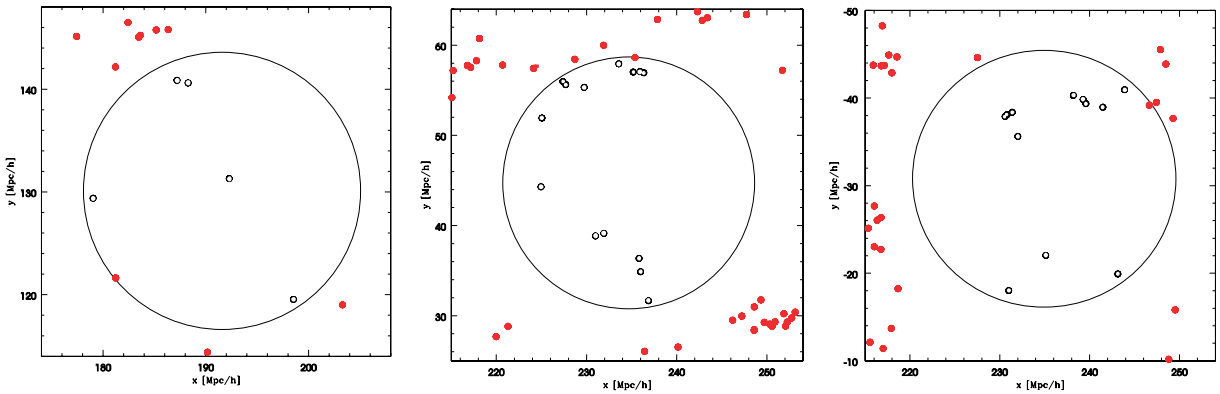


Figure 5. Three examples of faint galaxies in voids. We show for each void a $6\,h^{-1}$ Mpc central slab. In the left-hand panel is shown a $13.5\,h^{-1}$ Mpc void, in the central panel a $14.0\,h^{-1}$ Mpc void and in the right-hand panel a $14.7\,h^{-1}$ Mpc void. The filled circles are the galaxies brighter than $-19.32 + 5\log h$ that define the voids. The open circles are the fainter galaxies inside the voids down to $M_{b_J}^{\rm lim} = -18.3 + 5\log h$. Note that even though there is not a large population of faint galaxies in these voids, the galaxies are not randomly distributed, i.e. they are distributed in filamentary structures similar to those found in another scales (see central and right-hand panels).

voids, i.e. on average 14 galaxies per void. We have estimated the number density contrast of the galaxies located inside these voids by

$$\delta_{\rm gal} = \frac{\bar{n}_{\rm void} - \bar{n}}{\bar{n}}, \tag{27}$$

where $\bar{n}_{\rm void}$ is the number density of galaxies inside the void and $\bar{n}$ is the number density of galaxies in the field. For the galaxies inside the voids we obtain $\delta_{\rm gal} = -0.87$. In Fig. 5, we display a central $6\,h^{-1}$ Mpc thick slice for three of these nine voids. We can see that, despite the fact that these voids are highly empty, the faint galaxies populating them show interesting structures like filaments. Nevertheless, most galaxies are placed close to the borders of the voids being their centres much emptier. Note that these galaxy patterns are similar to those found in voids in high-resolution numerical simulations by Gottlöber et al. (2003). In Fig. 6, we show the mean number density of faint galaxies in our rare voids as a function of distance to the void centre (normalized to the void radius).

We have also studied the galaxies inside common voids, i.e. voids whose underdensities are not too big. We have selected from the SGP1 sample the voids defined by galaxies brighter than $-20.5 + 5\log h$. We found five voids larger than $15.0\,h^{-1}$ Mpc up to $z = 0.095$. For these radii, a randomly chosen sphere should contain only three galaxies brighter than $-20.5 + 5\log h$. In total, there are 666 galaxies fainter than $-20.5 + 5\log h$ down to $M_{b_J}^{\rm lim} = -18.3 + 5\log h$. These voids contain on average 10 times more galaxies than the rare voids discussed above, being the number density contrast of galaxies $\delta_{\rm gal} = -0.54$. In Fig. 7, we show an example of a $8\,h^{-1}$ Mpc thick slice of a $18.25\,h^{-1}$ Mpc void in

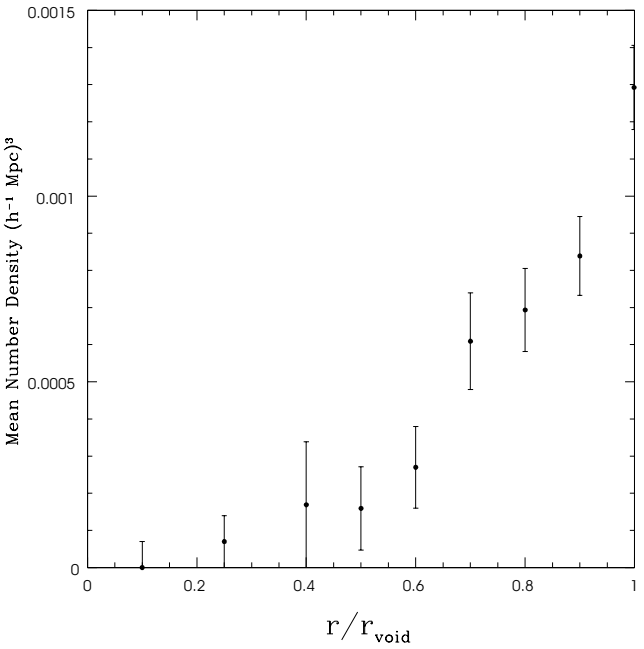


Figure 6. Averaged enclosed number density profile of faint galaxies ($-19.32 + 5\log h < M_{b_J} < -18.3 + 5\log h$) in nearby voids as a function of the distance from the centre of the void (in void radius units). The mean number density of galaxies in the field for these magnitude band is 7.0×10^{-3} $(h^{-1}\,{\rm Mpc})^{-3}$.

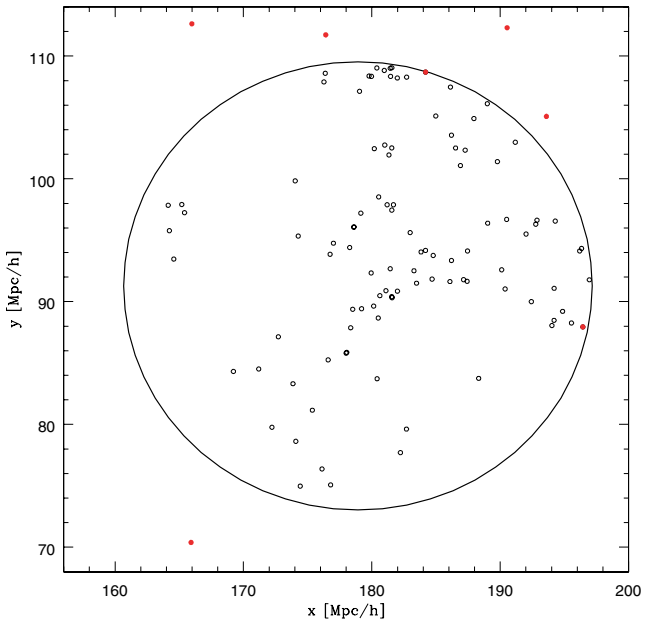


Figure 7. Galaxies inside a common void. We show a central $8\,h^{-1}$ Mpc thick slice of a $18.25\,h^{-1}$ Mpc void defined by galaxies brighter than $-20.5 + 5\log h$ (filled circles). The galaxies inside the void are fainter down to $M_{b_J}^{\rm lim} = -18.3 + 5\log h$ (open circles).

this sample. The galaxies inside this void almost fully fill the void (open circles). This is not surprising due to the fact that the density contrast of these voids is not too big.

7 DISCUSSION AND CONCLUSIONS

We have developed two robust and accurate algorithms to detect non-overlapping maximal spheres in halo or galaxy samples. We have applied them to several numerical simulations boxes in order to study the performance of the algorithms. The results found were very satisfactory.

We have then studied the void statistics in the 2dFGRS. We have detected ∼350 voids with radius larger than $7.5\,h^{-1}$ Mpc defined by galaxies with $M_{b_J} < -19.32 + 5\log h$ and ∼70 voids with radius larger than $13.0\,h^{-1}$ Mpc defined by galaxies with $M_{b_J} < -20.5 + 5\log h$ in the volume-limited samples up to $z_{\max} = 0.14$. We have also obtained the void statistics for the volume-limited samples up to $z_{\max} = 0.2$. For this case, we have detected ∼170 voids with radius larger than $13\,h^{-1}$ Mpc defined by galaxies brighter than $M_{b_J} < -20.181 + 5\log h$.

The number density of voids larger than R found in the SGP and NGP samples are in good agreement with each other for all values of R. We have obtained the VPF for both strips finding results in good agreement with previous ones (Croton et al. 2004; Hoyle & Vogeley 2004). We have shown, using an appropriate expression for the VPF sampling errors that the results found in both strips are statistically compatible.

From the results of several statistical tests we conclude that except for the anticorrelation implicit in the fact that the maximal sphere are chosen so that they do not overlap, they are essentially uncorrelated. There is, however, some evidence for a weak additional anticorrelation, which may be easily explained within the standard scenario of structure formation. We conclude at least at 99.5 per cent confidence level that there is a dependence of the void number density on redshift. We do this by means of a modified version of the $V_{\max}$ test which reveals a small trend towards small z values.

We have also obtain preliminary results on the galaxy contents of nearby voids found in volume-limited galaxy samples in the 2dFGRS. For the nine nearby voids up to $z < 0.095$ in our sample, we have found inside them on average only 15 galaxies fainter than $-19.32 + 5\log h$ (the magnitude of the galaxies which define the voids). These voids are rather empty compare to those defined by brighter galaxies. The galaxies within the voids are not randomly distributed: they are clustered forming well-defined filamentary structures as that observed in the large-scale structure of the galaxy distribution of the Universe. Moreover, this is the same pattern that show the dark matter halo distribution found inside voids in high-resolution N-body simulations. The haloes inside voids are distributed in a way that resemble a miniature of the Universe (see fig. 2 in Gottlöber et al. 2003). We have also obtained the number density profile for the galaxies inside the voids. The number density of faint galaxies falls almost a factor of 7 from the border of the voids to the inner half. Moreover, the number density of galaxies close to the borders of the voids are still too low compared with the field (almost a factor of 5). Gottlöber et al. (2003) and Patiri et al. (2004) have obtained the number density profile for haloes in voids. From the comparison of these results, we can see that even though the number density of galaxies and haloes follow similar trends, the galaxy profile is steeper than haloes (they fall just a factor of 2).

In a future work, we will use the framework developed in Patiri et al. (2004) along with the results given here for the number density of voids and their redshift dependence in order to constrain the value of σ_8. Furthermore, using a halo occupation model (e.g. Berlind et al. 2003) along the lines described in Patiri et al. (2004) we shall establish constraints in the relationship between haloes and galaxies from void statistics and the statistics of the galaxies inside the voids. In particular, we hope to be able to determine whether or not the conditional luminosity function depends on environment.

On the other hand, the study of the physical properties of the galaxies inside voids could imply constraints on the galaxy formation processes. So, we will analyse the physical properties like colours, metallicities, star formation rates, etc. of galaxies in rare voids that are available from the biggest galaxy surveys (2dFGRS and SDSS) in order to test, for example, the galaxy luminosity function in voids and the density–morphology correlations.

ACKNOWLEDGMENTS

We thank the referee for the comments and suggestions which improved the previous version of the paper. SGP would like to thank the hospitality of the Department of Astronomy of the New Mexico State University where part of this work was done. We also thank support from grant PNAYA 2005-07789.

This work was also partially supported by the *Acciones Integradas Hispano-Alemanas*. AK acknowledges support of NSF and NASA grants to NMSU. Computer simulations have been done at CESGA (Spain), NIC Jülich, LRZ Munich and NASA.

The Millennium Run simulation used in this paper was carried out by the Virgo Supercomputing Consortium at the Computing Centre of the Max-Planck Society in Garching.

REFERENCES

Aikio J., Mähönen P., 1998, ApJ, 497, 534
Berlind A. et al., 2003, ApJ, 593, 1
Betancort-Rijo J. E., 1992, Phys. Rev. A, 45, 3447
Betancort-Rijo J. E., 1999, J. Stat. Phys., 98, 917
Betancort-Rijo J. E., Sanchez-Conda M., Prada F. R., Patiri S., 2005, ApJ, submitted (astro-ph/0509897)
Colberg J. M., Sheth R. K., Diaferio A., Gao L., Yoshida N., 2005, MNRAS, 360, 216
Cole S., Hatton S., Weinberg D. H., Frenk C. S., 1998, MNRAS, 300, 945
Cole S. et al., the 2dFGRS Team, 2005, MNRAS, 362, 505
Colless M. et al., the 2dFGRS Team, 2001, MNRAS, 328, 1039
Colless M. et al., the 2dFGRS Team, 2003, preprint (astro-ph/0306581)
Conroy C. et al. 2005, ApJ, 635, 990
Croton D. J. et al., 2004, MNRAS, 352, 828
Croton D. J. et al., 2005, MNRAS, 356, 1155
Davis M., Peebles P. J. E., 1983, ApJ, 267, 465
Einasto J., Einasto M., Gramann M., 1989, MNRAS, 238, 155
Einasto J., Einasto M., Gramann M., Saar E., 1991, MNRAS, 248, 593
Einasto J., Saar E., Einasto M., Freudling W., Gramann M., 1994, ApJ, 429, 465
El-Ad H., Piran T., 1997, ApJ, 491, 421
Fry J. N., 1986, ApJ, 306, 358
Goldberg D. M., Jones T. D., Hoyle F., Rojas R. R., Vogeley M. S., Blanton M. R., 2005, ApJ, 621, 643
Gottlöber S., Lokas E., Klypin A., Hoffman Y., 2003, MNRAS, 344, 715
Hoyle F., Vogeley M. S., 2002, ApJ, 566, 641
Hoyle F., Vogeley M. S., 2004, ApJ, 607, 751
Hoyle F., Rojas R. R., Vogeley M. S., Brinkmann J., 2005, ApJ, 620, 618
Kauffmann G., Fairall A. P., 1991, MNRAS, 248, 313
Kirshner R. P., Oemler A., Schechter P. L., Shectman S. A., 1981, ApJ, 248, L57
Klypin A., Holtzman J., 1997, preprint (astro-ph/9712217)
Klypin A., Kravtsov A. V., Valenzuela O., Prada F., 1999, ApJ, 522, 82

Kravtsov A. V., Klypin A. A., Khokhlov A. M., 1997, ApJS, 111, 73
Maddox S. J., Efstathiou G., Sutherland W. J., Loveday J., 1990, MNRAS, 243, 692
Mathis H., White S. D. M., 2002, MNRAS, 337, 1193
Norberg P. et al., the 2dFGRS Team, 2002, MNRAS, 336, 907
Otto S., Politzer D. H., Preskill J., Wise M. B., 1986, ApJ, 304, 62
Patiri S. G., Betancort-Rijo J. E., Prada F., 2004, preprint (astro-ph/0407513)
Peebles P. J. E., 1980, The Large Scale Structure of the Universe. Princeton Univ. Press, Princeton, NJ
Peebles P. J. E., 2001, ApJ, 557, 495
Plionis M., Basilakos S., 2002, MNRAS, 330, 399
Rojas R. R., Vogeley M. S., Hoyle F., Brinkmann J., 2005, ApJ, 624, 571
Rood H. J., 1981, Rep. Prog. Phys., 44, 1077
Rowan-Robinson M., 1968, MNRAS, 138, 445
Sheth R. K., Van de Weygaert R., 2004, MNRAS, 350, 517
Sheth R., Tormen G., 2002, MNRAS, 329, 61
Solevi P., Mainini R., Bonometto S. A., Maccio' A. V., Klypin A., Gottloeber S., 2006, MNRAS, 366, 1346
Springel V., Yoshida N., White S. D. M., 2001, New Astron., 6, 79
Springel V. et al., 2005, Nat, 435, 629
Tegmark M. et al., 2004, ApJ, 606, 702
Van de Weygaert R., Van Kampen E., 1993, MNRAS, 263, 481
White S. D. M., 1979, MNRAS, 186, 145
York D. G. et al., 2000, AJ, 120, 1579
Zehavi I. et al., 2002, ApJ, 571, 172

APPENDIX A: VOID FINDER ALGORITHMS

A1 Algorithm 1: Cell Void Finder

The first step is to select the sample of galaxies or dark matter haloes in our redshift survey or cosmological simulation that will define the voids. We will select galaxies with luminosity greater than a given value L or haloes with virial mass greater than M.

Once the galaxy or halo sample has been selected our code generates a cubic mesh where the cell size determines the working resolution. Afterwards, the objects are assigned to cells. So we have three types of cells: the cells that contain galaxies or haloes (filled cells), the cells that are empty but are inside the observational domain (observed empty cells) and those cells which are also empty but they are located outside the observational domain or in a not observed region inside the observational domain (a 'hole') (not observed empty cells). The code searches for the voids among the observed empty cells inside the observational domain. In the case of cosmological simulations the dark matter halo samples are in 3D boxes and generally all the cells are inside the observational domain. However, in the case of galaxy samples in redshift surveys which have irregular geometry, some of the cells are located outside the observational domain. We can easily determine which cell has been observed and which has not by using the survey masks.

Once the cell classification procedure have been completed the code is then ready to search for the voids, finding their centres and radii. In principle, each observed empty cell could be a potential void centre, but it is easy to realize that the observed empty cells which are located close to filled cells that contain galaxies or haloes will not be the centre of a void. So, in order to save computational time, we mark these neighbouring cells and they will not be taken into account at the time of searching for the void centres. This is an iterative process, i.e. once we have marked the observed empty cells closer to the filled cells we can go to the next level and mark the observed empty cells that are neighbours of already marked empty cells. Note that we will stop this iteration depending on the working resolution (see Fig. A1).

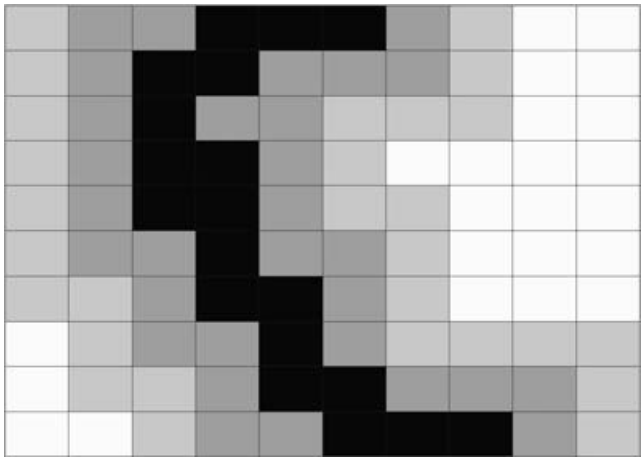


Figure A1. An example of labelling of cells for our Cell Void Finder. In order to save computational time, we mark the empty cells (dark grey cells) that are neighbours of those cells that contain objects (black cells). Note that this is an iterative process, i.e. we can now mark the empty cells that are neighbours of already marked empty cells (light grey cells). The stop of this iteration depends on the adopted resolution. The centre of the void will be searched over the empty unmarked cells (white cells).

To determine the void centres, the code computes the distances between each of the unmarked observed empty cell and all the galaxies or haloes in the whole observational domain and we retain the minimum distance. Once we have the list with the minimum distances, we search for the local maxima which correspond to the void centres. Obviously, the voids radii are those maximum distances.

Finally, the code removes the overlapping maximal spheres, keeping the biggest one, i.e. if the distance between two maximal spheres is less than the sum of their radius, then the voids overlap and we remove the smaller one.

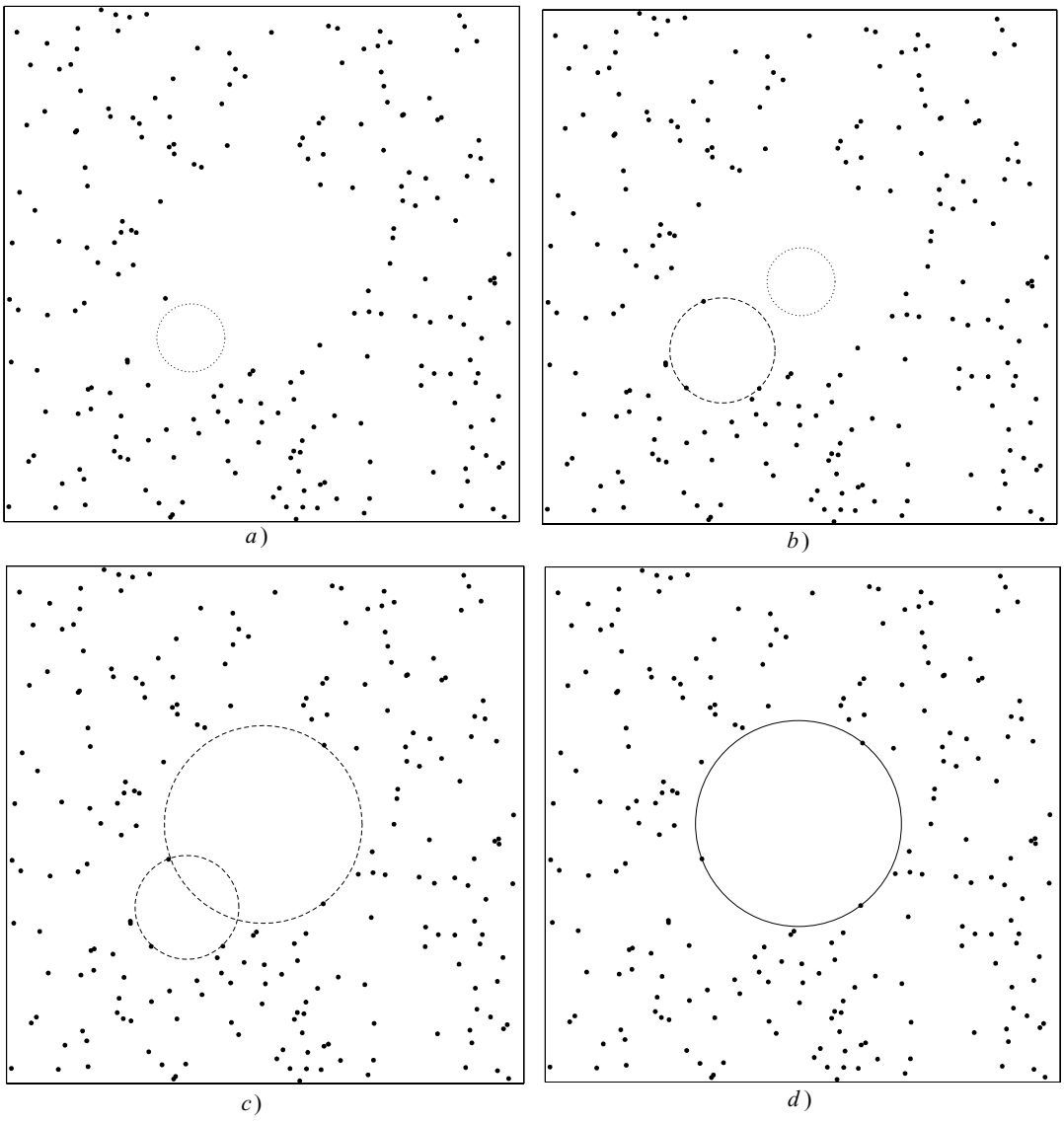
The main advantage of the Cell Void Finder algorithm is that in only one run we get all the voids in the sample. However, its main disadvantage is that it consumes quite a lot of memory, which scale with the resolution that we require. There is a similar memory problem when we have a large number of galaxies or haloes in the sample. If we are only interested in the biggest voids (rare voids) this algorithm is not the best strategy. Some studies are focused in these kind of voids, so, here we have developed a complementary algorithm which is more efficient in this respect.

A2 Algorithm 2: HB Void Finder

Here, we give the details of our second algorithm whose main task is the detection of rare voids and we will use it as a complement of the CELLS algorithm. The HB Void Finder is conceptually simple. It searches for the non-overlapping maximal spheres with radius *larger* than a given value. As we mentioned above this code is designed for statistical studies focused on the biggest voids. In these cases, the code is very accurate and computationally efficient as compared with our Cell Void Finder.

The first step in the algorithm is to generate a sample of random trial spheres with a fixed radius R. These spheres are generated directly over the observational domain with the condition that the entire sphere lies inside the observational domain. We check which spheres contains no objects and keep them.

In the next step, we find for each trial empty sphere the four nearest objects and we expand the sphere to contain them in its surface.



[t]

Figure A2. Sequence of steps in our HB Void Finder: (a) we throw a sphere of fixed radius (dotted circle) over the observed sample of objects, (b) if this sphere is empty of objects defining the void (filled dots), we then grow the sphere (dashed circle) until its surface contains the four nearest objects (three in 2D, see Appendix A). This is a potential maximal sphere. We throw another random sphere (dotted circle). Again, if this one is empty, we obtain the potential maximal sphere (dashed circle in panel c). (d) Finally, we check if the new maximal sphere overlaps with another one; if so, we keep the largest sphere. This is our void (filled circle). See text for details.

These new spheres are potentially maximal (see Section A4). Note that the new expanded spheres could in principle contain objects. If this is the case, those spheres are removed. The potential maximal spheres will be *actual* maximal spheres if they do not overlap, which is decided by means of the same criteria as for the Cell Void Finder (i.e. if the distance between two spheres is less than the sum of their radius they overlap, so we keep the largest one) and if the four nearest objects are not located in the same hemisphere (see Fig. A2). We put in both algorithms the additional constrain that the maximal spheres have to be entirely inside the observational domain. Note that, we need to do many realizations of the trial spheres in order to get the maximal spheres. Typically, four realizations are needed for each void in order to reach its maximum radius, when the maximal sphere is only slightly larger than R and an increasing number as the maximal sphere is larger with respect to R.

A3 Performance test

We construct random samples of objects in order to test the code performance and check the results of both algorithms. We generate two samples, one with 1000 points and another one with 10 000 points. Both samples are in a box of 100 Mpc. We give in Table A1 the void statistics computed from both algorithms. Note that the agreement between codes is excellent. We obtain the void statistics using $\sim 10^7$ trial spheres with the HB Void Finder. The resolution of the Cell Void Finder is 0.5 Mpc.

The CPU time of the codes mainly depends on the number of particles, the number and radius of trial spheres in the case of the HB algorithm and on the number of cells (i.e. resolution) and the levels of neighbouring cell marking for the Cell Void Finder. For example, in the case of the sample with 1000 random particles and for the voids with radius larger than 10 Mpc, the HB Void Finder

Table A1. Voids in random samples. We list the statistics of the maximal spheres found by our algorithms in the sample of 1000 and 10 000 random points in a box of 100 Mpc . $N_{\rm HB}$ is the number of voids found with the HB algorithm with radius larger than the value given in the first column. $N_{\rm CELLS}$ is the same but for the voids found with the Cells Void Finder. See text for details.

Radius (Mpc)	$N_{\rm HB}$	$N_{\rm CELLS}$
Sample 1000		
10.0	23	25
11.0	15	15
12.0	7	7
Sample 10 000		
5.5	69	73
6.0	20	21
6.5	3	3
7.0	1	1

takes ∼1 h for 10^7 trials, while the Cell Void Finder takes ∼3 h with a resolution of 0.5 Mpc. Notice that the voids with radius larger than 10 Mpc are common in this box, so the running time differences are not so big between both algorithms. However, if we search for voids with radius larger than 12 Mpc, the HB Void Finder takes only 20 min, while the Cell Void Finder last the same ∼3 h. These tests were done in a Pentium IV processor (3.06-GHz clock and 2 GB RAM) and in a Itanium-2 processor (1.5-GHz clock, 2 GB RAM) giving both similar performances.

A4 How to grow the trial spheres

To determine the sphere passing through the four nearest objects to an empty trial sphere we proceed as follows. We first take the two nearest objects (whose coordinates we denote as $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$) and calculate the middle point $(\boldsymbol{x}_1 + \boldsymbol{x}_2)/2$. From this point, we move along a vector in the plane containing the three nearest objects and perpendicular to $\boldsymbol{x}_2 - \boldsymbol{x}_1$ until we reach the point, q, where the distances to the third nearest object ($\boldsymbol{x}_3$) is the same as that to object 1, then the distance between object 2 and $\boldsymbol{q}$ is also the same that the previous two (See Fig. A3). Then we need to solve

$$|\boldsymbol{x}_1 - \boldsymbol{q}(w_0)| = |\boldsymbol{x}_3 - \boldsymbol{q}(w_0)|, \tag{A1}$$

where

$$\boldsymbol{q}(w) = \frac{\boldsymbol{x}_1 + \boldsymbol{x}_2}{2} + w\boldsymbol{j} \tag{A2}$$

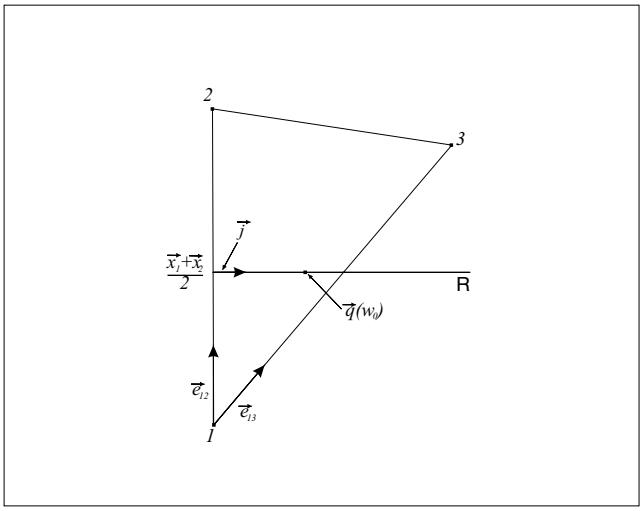


Figure A3. How we grow the spheres.

and

$$\boldsymbol{j} = \frac{\boldsymbol{e}_{13} - (\boldsymbol{e}_{13}.\boldsymbol{e}_{12})\boldsymbol{e}_{12}}{|\boldsymbol{e}_{13} - (\boldsymbol{e}_{13}.\boldsymbol{e}_{12})\boldsymbol{e}_{12}|}, \tag{A3}$$

where

$$\boldsymbol{e}_{12} \equiv \frac{\boldsymbol{x}_1 - \boldsymbol{x}_2}{|\boldsymbol{x}_1 - \boldsymbol{x}_2|} \quad \text{and} \quad \boldsymbol{e}_{13} \equiv \frac{\boldsymbol{x}_1 - \boldsymbol{x}_3}{|\boldsymbol{x}_1 - \boldsymbol{x}_3|} \tag{A4}$$

with $w_0 \in (-2R_0, 2R_0)$, R_0 is the radius of the trial sphere.

Now, we repeat the same procedure described above but taking into account the fourth object, i.e. we move from $\boldsymbol{q}\,(w_0)$ perpendicularly to the plane of Fig. A3 until we reach the point $\boldsymbol{P}\,(t)$ where the distance between $\boldsymbol{x}_4$ and $\boldsymbol{P}\,(t)$ is the same as the distance from $\boldsymbol{x}_1$ to $\boldsymbol{P}\,(t)$. So,

$$\boldsymbol{P}(t) = \boldsymbol{q}(w_0) + t\boldsymbol{n}, \tag{A5}$$

where

$$\boldsymbol{n} = \boldsymbol{e}_{12} \wedge \boldsymbol{e}_{13}. \tag{A6}$$

Solving

$$|\boldsymbol{x}_1 - \boldsymbol{P}(t)| = |\boldsymbol{x}_4 - \boldsymbol{P}(t)| \tag{A7}$$

with $t \in (-2R_0, 2R_0)$, we finally obtain the coordinates of the centre of the maximal sphere, $\boldsymbol{P}\,(t)$, and its radius R, which is simply given by $|\boldsymbol{x}_1 - \boldsymbol{P}\,(t)|$.

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