8

Statistical Study of the Galaxy Distribution

Antoine Labatie, Jean-Luc Starck and Marc Lachièze-Rey
Laboratoire AIM (UMR 7158), CEA/DSM-CNRS-Université Paris Diderot, IRFU, SEDI-SAP, Service d'Astrophysique, Centre de Saclay, F-91191 Gif-Sur-Yvette cedex France

1. Introduction

Large-scale structures in the Universe provide crucial information about formation of structures and can be used to test cosmological models. The good agreement between large-scale observations and the now-standard Lambda-Cold Dark Matter (ΛCDM) model gives hope for this model to be a lasting foundation. Going a step further, these observations have been used for precision cosmology to constrain cosmological parameters inside the model.

Large-scale structures are mainly studied through galaxy surveys, while on the other hand, cosmological models give predictions in terms of the global matter density field. Thus we need to understand the relationship between the matter density field and galaxy field, which is still a subject of research. Yet galaxy surveys have the advantage to map very large volumes. Recent galaxy surveys such as the Sloan Digital Sky Survey (SDSS, York et al. (2000)) and the 2 degree Field (2dF, Colless et al. (2001)) map unprecedented 3D volumes in the Universe. They have notably confirmed the view of large-scale structures as a collection of giant bubble-like voids separated by sheets and filaments of galaxies. This pattern has become known as the "Cosmic Web" (Bond et al. (1996), Springel et al. (2005)). The SDSS survey also enabled the first convincing detection of Baryon Acoustic Oscillations (BAOs) in large-scale structures (Eisenstein et al. (2005)), which opened a new field in precision cosmology.

Future galaxy surveys will map always bigger volumes, with more galaxies and continue to improve our knowledge of large-scale structures and cosmological models. Though data sets are of crucial importance, statistical methods extracting the information also have an important role. They should be optimized in order to exploit the full potential of these surveys. The purpose of this chapter is to review the different methods that can be used.

A first class of methods that we present is based on Fourier analysis, using the correlation function and the power spectrum. It originates from the simplicity of the Gaussian field, fully described by its second order moments. The Gaussian model gives a simple approximation of the matter density field, that works well on large scales. Another reason for the study of correlation function and power spectrum is that they are well understood and can be predicted in ΛCDM models.

Recently, Fourier analysis has been used to study BAOs. These structures are remnants of acoustic waves, traveling in the hot plasma before recombination, and that have stayed frozen in large-scale structures. The detection of BAOs in the SDSS has been a strong support for the ΛCDM model. Besides, they provide a statistical standard ruler since their absolute

size is known with small uncertainty. Thus their measurement in galaxy catalogues gives information on distances, and enables to constrain cosmological parameters.
A second class of methods use geometrical and topological tools to describe large-scale-structures. Among this class of methods, we present the Minkowski functionals and the genus statistic, providing a rigorous mathematical framework for the analysis. Among other advantages, Minkowski functionals are known analytically for gaussian random fields. Historically, the main application of those statistics was to determine the scale at which the distribution is gaussian, i.e. approximately the scale of linear evolution. Among other applications, we show how they can be used for model discrimination, or to provide a standard ruler in galaxy catalogues.
Finally we present the approach based on fractal analysis to characterize the galaxy distribution. It is motivated by the well-established scale-invariance of the galaxy clustering at small scales ($r < 10h^{-1}$ Mpc). Yet fractality on all scales would put into question current cosmological models, which assume large-scale homogeneity. It would also call into question the correlation function analysis, which assumes a well-defined mean density of the field. We discuss the question of the large-scale homogeneity. We show how the extension to multifractal analysis can represent more general distributions, in particular the distribution of galaxies, and enable to study the scale of homogeneity.

2. Fourier analysis of clustering

2.1 Power spectrum in the ΛCDM model

The Fourier analysis of clustering mainly originates from a simple approximation of the large-scale distribution of matter being a Gaussian field. The justification is that, in current cosmological models involving inflation, the matter distribution originates from tiny gaussian random fluctuations that gravitationally evolved. Let us note δ the fluctuation field, from the matter density field ρ and its mean value $\bar{\rho}$:

$$\delta(\mathbf{x}) = \frac{\rho(\mathbf{x}) - \bar{\rho}}{\bar{\rho}} \tag{1}$$

In the linear regime of evolution, the fluctuation is supposed to be small ($\delta \ll 1$), and the equation of evolution is linear with δ:

$$\frac{\partial^2 \delta}{\partial t^2} + 2\frac{\dot{a}}{a}\frac{\partial \delta}{\partial t} = 4\pi G\bar{\rho}\delta \tag{2}$$

with a the expansion factor of the Universe, $\dot{a}$ its time derivative (i.e. $H = \frac{\dot{a}}{a}$), G Newton's gravitational constant.
In this regime, the Fourier modes $\tilde{\delta}(\mathbf{k})$ all evolve proportionally to an evolution factor. There are two solutions for this equation of evolution, with one growing mode and one decaying mode. The growing mode, which necessarily governs the evolution, has the following expression in a ΛCDM model:

$$G(z) = \frac{5\Omega_m E(z)}{2}\int_z^\infty \frac{(1+z')}{E(z')^3}dz' \tag{3}$$

with Ω_m the dimensionless matter density and $E(z) = \frac{H(z)}{H_0}$ the dimensionless Hubble constant at redshift z. In the linear regime, the gaussianity of the field is preserved, as well as

the gaussianity of the Fourier coefficients $\tilde{\delta}(\mathbf{k})$ of the fluctuation. Thus a complete description of the field can be given in this regime in terms of the power spectrum $P(\mathbf{k})$ of the fluctuation field

$$\mathbb{E}[\tilde{\delta}(\mathbf{k})\tilde{\delta}^{\star}(\mathbf{k}')] = (2\pi)^3\delta_D(\mathbf{k}-\mathbf{k}')P(\mathbf{k}) \tag{4}$$

with δ_D the 3-dimensional Dirac function. The power spectrum quantifies the expected strength of the oscillations at wavelength $\mathbf{k}$. With the usual hypothesis of isotropy, the power spectrum only depends on the norm k of the vector $\mathbf{k}$. However in galaxy catalogues, the galaxy are studied in redshift space which slightly breaks the isotropy property.
The linear approximation works very well at large scale, where the strength of the fluctuation is small. On smaller scale, the matter field is in a much more advanced evolution phase and has already overtaken non-linear evolution, which makes it much harder to understand. To take into account non-linear evolution, one has to use N-body simulations and empirically infer formulas.
Overall, the dark matter power spectrum can be written as

$$P_{DM}(k,z) = G^2(z)T^2(k,z)P_l(k) \tag{5}$$

where $T(k,z)$ is the transfer function at redshift z including non-linear evolution and $P_l(k)$ is the primordial dark matter power spectrum.

2.2 Two-point correlation function in the ΛCDM model

Just as the power spectrum, the two-point correlation function is a second order statistic that describes the clustering of a field or a point process. It is simply the inverse Fourier transform of the power spectrum, i.e. when P and ξ are isotropic:

$$\xi(r) = \frac{1}{2\pi^2}\int k^2 P(k)\frac{\sin(kr)}{kr}dk \tag{6}$$

The correlation function $\xi(r)$ at distance r also has a statistical interpretation. It measures the excess of probability to find a pair of points in two volumes dV_1 and dV_2 at distance r compared to a random distribution.

$$dP_{12} = \bar{\rho}^2[1+\xi(r)]dV_1dV_2 \tag{7}$$

The correlation function is directly linked to the fluctuation field $\delta(\mathbf{x}) = \frac{\rho(\mathbf{x})-\bar{\rho}}{\bar{\rho}}$:

$$\xi(r) = \langle\delta(\mathbf{x})\delta(\mathbf{x}+\mathbf{r})\rangle \tag{8}$$

with $\|\mathbf{r}\| = r$. As for the power spectrum analysis, the standard ΛCDM model predicts a certain shape for $\xi(r)$ with a dependence on the cosmic parameters. In this model the correlation is rapidly decreasing at small scale as a power law as found in the distribution of galaxies (Norberg & Baugh (2001), Zehavi et al. (2005)):

$$\xi(r) = \left(\frac{r}{r_0}\right)^{-\gamma} \tag{9}$$

with $r_0 \approx 5h^{-1}$Mpc and $\gamma \approx 1.7$.
At large scale the correlation presents a very interesting feature caused by BAOs. It is a relic of sound waves traveling in the early Universe when baryon and photons were coupled in a

relativistic plasma before recombination which caused the wave propagation to end Eisenstein & Hu (1998). It has imprinted the large scale distribution and can be seen as a small peak in the correlation function at a scale r_s, corresponding to the comoving distance of the sound horizon.

2.3 Modeling clustering bias in the galaxy distribution

As described in the previous section, theoretical predictions about clustering are given for the global dark matter density field. On the other hand, current large-scale surveys map the galaxy field. Thus, if we want to relate cosmological predictions to these observations, we have to understand the link between dark matter and galaxies.

The distribution of galaxies n_g is usually modeled by a Cox process, i.e. a Poisson process with an intensity given by a continuous field $\rho_g(\mathbf{x})$, which itself is a statistical process. Knowing $\rho_g(\mathbf{x})$, the number of galaxies in a volume dV around $\mathbf{x}$ is a Poisson variable with intensity $\rho_g(\mathbf{x})dV$. The field ρ_g is linked to the underlying dark matter density field ρ_m (at least stochastically) since galaxies tend to form in over-densities, but is not supposed to be identical. As a consequence, there are differences in the clustering of the two fields, and thus in their correlation functions $\xi_g(r)$ and $\xi_m(r)$. The relationship between ρ_g and ρ_m depends on the way that galaxies are grafted onto the matter density field. This is a critical question that remains an active research subject.

Historically, it has been observed that clusters of galaxies cluster more strongly than galaxy themselves. In more recent surveys, more luminous galaxies have been found to cluster more strongly. This shows a non trivial relation between the dark matter and the visible matter, which depends also on the type of objects. This effect is referred as the bias of visible matter with respect to the dark matter:

$$\xi_g(r) = b(r)^2\, \xi_m(r) \tag{10}$$

with $b(r)$ the bias of the galaxy field with respect to dark matter. The bias b modulates the galaxy correlation with respect to the dark matter correlation and should depend on the scale. This concept of bias appeared in the 80's to reconcile the CDM model with observational constraints. A first attempt to understand this bias was made by Kaiser in 1984 with the high peak model, which considers the formation of objects at local maximums of the smoothed density field.

Recently, the halo model (Cooray & Sheth (2002)) has been a fruitful step in order to connect the galaxy and dark matter distributions. A given halo model predicts the expected number of galaxies above given luminosity thresholds inside halos of different mass. Halo models can be parametrized to match very well current observations and naturally explains the biasing phenomenon. The halo model has become particularly useful in cosmological N-body simulations, where it is used to populate dark matter halos with galaxies, providing realistic simulations for current and future galaxy surveys.

2.4 Correlation function estimation from galaxy surveys

Both the correlation function and power spectrum have been extensively measured on galaxy surveys. The power spectrum has the advantage that the different bands are weakly correlated, which is not the case for the correlation function. On the other hand the correlation function is more intuitively linked to 3D space. As an illustration, BAOs from the early plasma stayed frozen at the sound horizon scale r_s, which produces an excess of clustering at this

particular distance. Thus BAOs manifest as a localized bump at the sound horizon scale in the correlation function, whereas they manifest as an infinite number of wiggles in the power spectrum.
In this section we decide to focus on the correlation function, explaining details about its estimation and how it can be used for BAO study.

2.4.1 Correlation function estimation

When estimating the correlation function, there are two types of statistical uncertainties. The first one comes from cosmic fluctuations due to limited sample volume, and the other one from the finite number of galaxies which do not trace exactly the underlying field (i.e. shot noise). In general, estimators of the correlation are subject both to bias and variance. In practice there is no way to evaluate the bias of the estimator if it exists, and it must be considered itself as a source of uncertainty, in addition to the variance.
There are various estimators of the correlation function, most using random catalogues with identical geometry to measure this excess of probability. Let N_D and N_R be the number of galaxies respectively in the data and random catalogues. We define $DD(r)$, $RR(r)$ and $DR(r)$ as the number of pairs at a distance in $[r \pm dr/2]$ of respectively data-data, random-random and data-random galaxies. We also define N_{DD}, N_{RR} and N_{DR} as the total number of corresponding pairs in the (real or random) catalog. With the convention of counting pairs only once we have:

$$N_{DD} = \frac{N_D(N_D-1)}{2} \tag{11}$$

$$N_{RR} = \frac{N_R(N_R-1)}{2} \tag{12}$$

$$N_{DR} = N_R N_D \tag{13}$$

$$\tag{14}$$

Let us give the expression of 4 of the most commonly used estimators used, from Peebles & Hauser (1974), Davis & Peebles (1983), Hamilton (1993) and Landy & Szalay (1993)

$$\hat{\xi}_{PH}(r) = \frac{N_{RR}}{N_{DD}}\frac{DD(r)}{RR(r)} - 1$$

$$\hat{\xi}_{DP}(r) = \frac{N_{DR}}{N_{DD}}\frac{DD(r)}{DR(r)} - 1$$

$$\hat{\xi}_{HAM}(r) = \frac{{N_{DR}}^2}{N_{DD}N_{RR}}\frac{DD(r)RR(r)}{[DR(r)]^2} - 1$$

$$\hat{\xi}_{LS}(r) = 1 + \frac{N_{RR}}{N_{DD}}\frac{DD(r)}{RR(r)} - 2\frac{N_{RR}}{N_{DR}}\frac{DR(r)}{RR(r)}$$

In general, the analysis is complicated by selection effects in the catalogues, with a density decreasing farther away. One has to consider a mean density varying with the position in the catalogue $\bar{\rho}(\mathbf{x})$, i.e. fluctuations $\delta(\mathbf{x}) = \frac{\rho(\mathbf{x})-\bar{\rho}(\mathbf{x})}{\bar{\rho}(\mathbf{x})}$. This is taken into account in the random catalogues which have the same selection effect as the catalogue. Finally an additional weight

can be given to the galaxies at different position of the survey (both in the data and random catalogues) in order to minimize the estimator's variances (Kaiser (1986)).
Estimating ξ would be easier knowing the exact number of points in the volume expected from the distribution. In practice we can only estimate it with the empirical quantities N_D and N_{DD}. Let us show that Hamilton and Landy-Szalay only depend on the second order on this uncertainty in the mean density, and thus perform better. We repeat the calculations given in Hamilton (1993) in a simple case where the sample is volume-limited (i.e. with a constant expected density $\bar{\rho}$ in the sample), so that the optimal strategy is to weight all galaxies equally. The empirical density in the catalogue ρ is a sum of Dirac functions on the galaxies of the catalogue. For a volume-limited survey with constant expected density $\bar{\rho}$, the relative fluctuation in the sample is given by

$$\delta(\mathbf{x}) = \frac{\rho(\mathbf{x}) - \bar{\rho}}{\bar{\rho}} \tag{15}$$

We write W the indicator function of the sample volume and $\langle . \rangle$ the integration on the volume. For example $\langle W(\mathbf{x})\,\rho(\mathbf{x})\rangle$ is the integration of the empirical density and thus equals the number of points in the sample. We introduce the following quantities that have statistical expectations of 0 for $\bar{\delta}$ and Ψ, and ξ for $\hat{\xi}$:

$$\bar{\delta} = \frac{\langle W(\mathbf{x})\,\delta(\mathbf{x})\rangle}{\langle W(\mathbf{x})\rangle} \tag{16}$$

$$\Psi(r) = \frac{\langle \delta(\mathbf{x})\,W(\mathbf{x})W(\mathbf{y})\rangle_r}{\langle W(\mathbf{x})W(\mathbf{y})\rangle_r} \tag{17}$$

$$\hat{\xi}(r) = \frac{\langle \delta(\mathbf{x})\delta(\mathbf{y})\,W(\mathbf{x})W(\mathbf{y})\rangle_r}{\langle W(\mathbf{x})W(\mathbf{y})\rangle_r} \tag{18}$$

where $\langle . \rangle_r$ means a double integration in the volume, restricted to $\mathbf{x}$ and $\mathbf{y}$ separated by a distance in $[r \pm dr/2]$. $\hat{\xi}$ is an unbiased estimator of the real ξ but we cannot calculate it since we do not know $\bar{\rho}$ and δ.
With short calculations it is possible to express the different estimators with the quantities $\hat{\xi}$, $\bar{\delta}$ and Ψ (Hamilton (1993)):

$$\hat{\xi}_{PH}(r) = \frac{\hat{\xi}(r) + 2\Psi(r) - 2\bar{\delta} - \bar{\delta}^2}{[1+\bar{\delta}]^2} \tag{19}$$

$$\hat{\xi}_{DP}(r) = \frac{\hat{\xi}(r) + \Psi(r) - \bar{\delta} - \Psi(r)\,\bar{\delta}}{[1+\bar{\delta}]\,[1+\Psi(r)]} \tag{20}$$

$$\hat{\xi}_{H}(r) = \frac{\hat{\xi}(r) - \Psi(r)^2}{[1+\Psi(r)]^2} \tag{21}$$

$$\hat{\xi}_{LS}(r) = \frac{\hat{\xi}(r) - 2\bar{\delta}\Psi(r) + \bar{\delta}^2}{[1+\bar{\delta}]^2} \tag{22}$$

These formulas explain the superiority of Hamilton and Landy-Szalay estimators, with Ψ and $\bar{\delta}$ terms at the second order in the numerator. Terms in the denominator are not important since they generate a small relative error, whereas terms in the numerator can generate a high relative error when their values become non negligible compared to $\hat{\xi}$. For Hamilton and

Landy-Szalay estimators, the error is dominated by the one of $\hat{\xi}$ and not really affected by Ψ and $\bar{\delta}$, which are linked to the uncertainty in $\bar{n}$.
The superiority of Hamilton and Landy-Szalay has been observed empirically in different studies (Labatie et al. (2010), Kerscher et al. (2000), Pons-Bordería et al. (1999)). In Labatie et al. (2010) it is also shown that these estimators are nearly unbiased for current galaxy surveys. In particular the integral constraint (caused by the uncertainty in the mean density $\bar{\rho}$) has nearly no effect on the estimation. We show in figure 1 results with different estimators when estimating the correlation function on simulations of the SDSS Luminous Red Galaxies (LRG) sample. Estimators are nearly unbiased for this survey size and Hamilton and Landy Szalay estimators show much less variance than the other estimators.

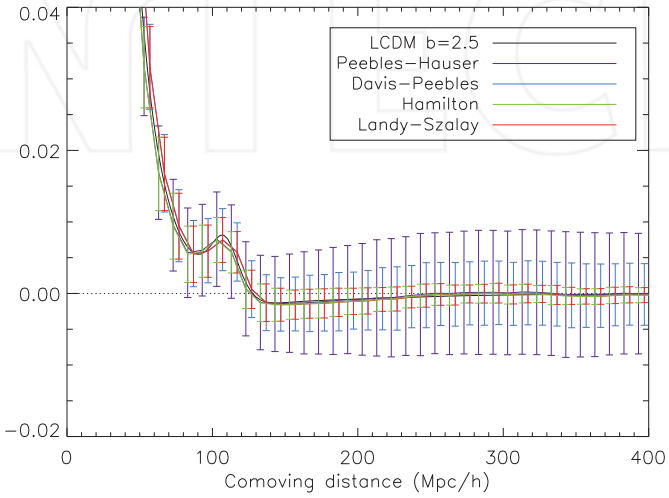


Fig. 1. Estimation of the correlation function on the SDSS LRG catalogue. The different colored curves show the result for each estimator with error bars showing the standard deviations of the estimation. They were estimated using 2000 simulations. The real correlation function is given by the black curve. The small bump in the correlation at scale $r \approx 105h^{-1}$Mpc corresponds to the BAO feature. The bias is negligible for all estimators but there are great differences in their quality. Bias and uncertainty are important to know the error when fitting the estimated correlation function to a model correlation function (as in BAO studies).

As a result, recent studies usually focus on the Landy-Szalay estimator to measure the galaxy correlation function.

2.4.2 Correlation function and BAO study

Recently a lot of studies using the power spectrum or the correlation function have been motivated by BAO study. The reason is that BAOs provide a statistical standard ruler, i.e. a powerful tool to study the geometry of the Universe. Since their absolute size is known with small uncertainty from measurements in the Cosmic Microwave Background (CMB), their detection in galaxy surveys gives access to real distances in the catalogue.

The primary informations obtained on galaxies are the angular positions of the objects on the sky, and their redshift. In order to obtain 3D positions, one has to assume a fiducial cosmological model and convert redshifts into distances. We show in figure 2 how the size of an object is related to its angular extent θ and its redshift extent dz, using the real cosmological quantities $d_A(z)$ (angular diameter distance) and $H(z)$. The interest of the standard ruler is that its size is already known, and thus from the angular and redshift extents, it naturally provides the quantities $d_A(z)$ and $H(z)$ at the redshift of the survey.

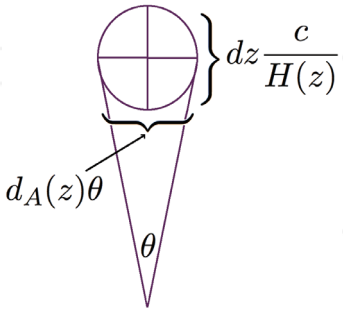


Fig. 2. Size of a spherical object as a function of its angular extent θ and redshift extent dz, using the real angular diameter distance $d_A(z)$ and Hubble constant $H(z)$ at redshift z of the survey. Figure adapted from Bassett & Hlozek (2010).

BAOs constrain cosmological parameters through $d_A(z)$ and $H(z)$ in a complementary way with CMB and type Ia supernovae, which explains the recent interest in this field. BAOs were first convincingly detected in the correlation function of the SDSS survey (Eisenstein et al. (2005)).
BAO study through correlation can be divided in two different parts. The first one consists in detecting the BAO feature in the correlation function, which can be done by rejecting the null hypothesis that there are no BAOs. When the detection is made significantly, the second step is to determine distances in the survey, which provides the best fit in the BAO peak between the measured and the model correlation functions.

2.4.2.1 BAO detection and significance

Let us describe more precisely the procedure for the detection. In most BAO studies (for example Eisenstein et al. (2005), Hütsi (2006), Percival et al. (2010), Blake et al. (2011)), the χ^2 statistic is used. Here the observations are the estimated correlation function values in different bins $(\hat{\xi}_i)_{1\leq i\leq n} = (\hat{\xi}(r_i))_{1\leq i\leq n}$. Given the covariance matrix $(C_{i,j})$ of this measurement vector, and the expected value given by the model $\xi = (\xi_i)$, the χ^2 statistic is defined as

$$\begin{aligned}\chi^2 &= (\hat{\xi}-\xi)^T C^{-1}(\hat{\xi}-\xi)\\ &= \sum_{i,j}(\hat{\xi}_i-\xi_i)C^{-1}_{i,j}(\hat{\xi}_j-\xi_j)\end{aligned}$$

The covariance matrix is usually estimated empirically using mock catalogues, and then inverted. Assuming that the measurement vector is gaussian, the χ^2 statistic follows a

$\chi^2(n)$ law, i.e. a chi-square law with n degrees of freedom. The situation is a bit more complex here because the model correlation function $\xi(\theta_1, ..., \theta_k)$ depends on cosmological parameters $(\theta_1, ..., \theta_k)$, and so the $\chi^2(\theta_1, ..., \theta_k)$ statistic also depends on a choice of model. The method used consists in finding the best fit model $(\theta_1^0, ..., \theta_k^0)$, i.e. the model with $\xi(\theta_1, ..., \theta_k)$ minimizing the χ^2 statistic

$$(\theta_1^0, ..., \theta_k^0) = \underset{(\theta_1,...,\theta_k)}{\operatorname{argmin}} \chi^2(\theta_1, ..., \theta_k) \tag{23}$$

With k parameters used for the fitting, it can be shown that the χ^2 at the best fit value follows a $\chi^2(n-k)$ law (under the assumption that one of the model is correct), i.e. a chi-square law with $n-k$ degrees of freedom. Even though this result is widely used, it is worth noting that it is subject to regularity assumptions The best fit χ^2 statistic is computed with the 2 different models (with and without BAOs) which leads to a difference $\Delta\chi^2$ in the best fit χ^2 for the two different models.

This difference $\Delta\chi^2$ can be used to test between 2 competing model that are nested, i.e. one model can be obtained by fixing/eliminating parameters in the other model. If we note 1 the model with fewer parameters, i.e. a higher value of χ^2, and 2 the other model, then it can be shown that the difference $\chi_{\text{diff}} = \chi_1^2 - \chi_2^2$ follows a $\chi^2(d_{diff})$ law (under the assumption that the restrained model is also correct). Here, $d_{diff} = d_1 - d_2$ is the difference in the number of parameters of the two models. In the case we are interested in, an artificial parameter is introduced to account for the presence or not of BAOs in the correlation function. The global model then becomes an union of the 2 precedent models and the restrained model contains only the model without BAOs. Then the χ_{diff} corresponds to the difference $\Delta\chi^2$, with a value that can be interpreted in terms of significance.

This methodology has been extensively used, for example in Eisenstein et al. (2005) for the first BAO detection at the 3.4 σ level (the convention is that a number of σ's is converted from a p-value using the relation for a Gaussian distribution). In Cabré & Gaztañaga (2011) it is used to show with mock catalogues that a detection higher than 3σ is not likely in the SDSS.

2.4.2.2 Constraining cosmological parameters through BAO scale

A later step consists in using the BAO feature in the correlation function to constrain cosmological parameters. We remind that 3D catalogues are created assuming a fiducial cosmology to obtain the distance from the observer to the objects. If we change the fiducial cosmology to the real cosmology, then the objects move and dilate. Spherical objects have a size which approximately scales as the dilation scale $D_V(z) = (1+z)^2 d_A(z) cz/H(z)$ (Eisenstein et al. (2005)) . Therefore, the correlation function in the true cosmology is obtained by dilating it with the ratio of the true to the fiducial dilation scales. When doing the χ^2 test this dilation scale is introduced as a parameter, and an estimation of the true dilation scale is obtained for the best fit. Other cosmological parameters can also be introduced during the fitting procedure. The dilation scale itself provides constraints since it depends on cosmological parameters.

In Eisenstein et al. (2005), there are 3 fitted parameters: an overall amplitude A, $\Omega_m h^2$ and $D_V(0.35)$, where $z = 0.35$ is the mean redshift of the SDSS LRG sample used. The other parameters intervening in the correlation function are kept fixed because they are well constrained by other measurements, e.g. on the CMB. The posterior probability on those parameters is simply equal to the likelihood (assuming a flat prior). This is also directly linked

to the $\chi^2(A, \Omega_m h^2, D_V(0.35))$ statistic with the gaussian hypothesis. So the χ^2 statistic directly gives the confidence regions for these parameters as shown in figure 3, where the predictions of different models are overplotted.

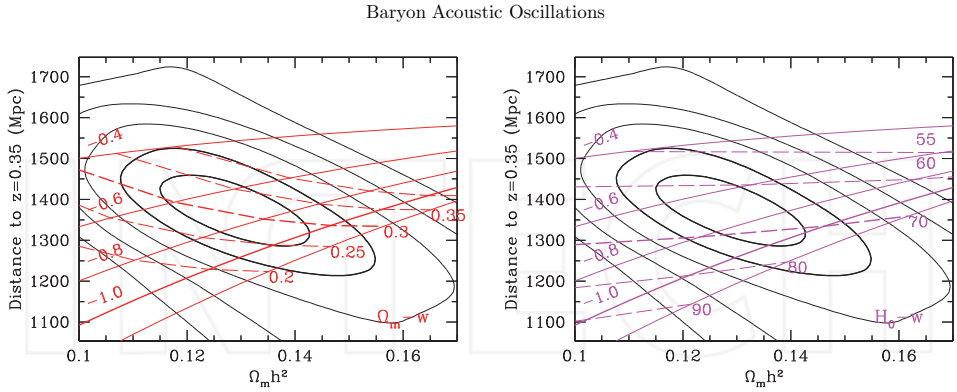


Fig. 3. The likelihood contours of CDM models as a function of $\Omega_m h^2$ and $D_V(0.35)$. The likelihood has been taken to be proportional to $\exp(-\chi^2/2)$, and contours corresponding to 1σ to 5σ for a 2D Gaussian have been plotted. Overplotted are the model predictions for fixed Ω_m, w and H_0. Figure courtesy of Eisenstein et al. (2005).

3. Geometrical analysis of clustering

In the first section we presented the Fourier analysis of clustering based on the second order statistics which are the power spectrum and its real space equivalent, the 2-point correlation function. In general a random field is characterized by all its moments and N-point correlation functions. The 2-point correlation function necessarily presents limitations for model discrimination and may not be the best tool for some applications. One example of such application is the geometrical analysis of large-scale structures. Obviously the 2-point correlation function does not provide much geometrical information since it provides spherically averaged quantities.

In recent galaxy surveys, the Universe appears as a collection of giant bubble-like voids separated by sheets and filaments of galaxies, with the superclusters appearing as occasional relatively dense nodes. This network, usually called the "Cosmic Web", is clearly visible in the 2dF Galaxy Redshift Survey and the SDSS survey.

The study of the geometry and topology of large-scale structures can carry useful information and complement the study of second order clustering. We review different geometrical methods used for this task and results obtained.

3.1 Minkowski functionals

Minkowski functionals are a very useful tool to describe hypersurfaces of dimension $d-1$ in a space of dimension d. Hadwiger theorem states that any function defined on subsets A of the space that satisfies the properties of translation and rotation invariance, additivity and continuity can be decomposed as a sum of the Minkowski functionals. In a space of dimension

d there are $d+1$ Minkowski functionals. In the case of the 3D distribution of galaxies, where $d=3$, Minkowski functionals are

$$
\begin{aligned}
V_0(A) &= \int_A d^3\mathbf{x} \\
V_1(A) &= \frac{1}{6}\int_{\delta A} dS(\mathbf{x}) \\
V_2(A) &= \frac{1}{6\pi}\int_{\delta A}\left(\frac{1}{R_1(\mathbf{x})}+\frac{1}{R_2(\mathbf{x})}\right)dS(\mathbf{x}) \\
V_3(A) &= \frac{1}{4\pi}\int_{\delta A}\frac{1}{R_1(\mathbf{x})R_2(\mathbf{x})}dS(\mathbf{x})
\end{aligned}
$$

where R_1 and R_2 are the principal curvatures of the surface δA.
These functions have intuitive interpretations in the case $d=3$. V_0 corresponds to the volume of A, V_1 is proportional to the surface of the boundary δA, V_2 corresponds to the mean curvature on the surface δA and V_3 is half the Euler characteristic χ of the subset, linked to the genus G

$$\chi = 2(1-G) \tag{24}$$

V_3 is a topological measure, which equals 1 for a ball and any analogue convex set. It can be interpreted as the surface drawn on the unit sphere by the vector normal to the hypersurface, when covering δA. The surface is counted positively when the product of the curvature radii is positive and negatively otherwise.
The genus also has a simple interpretation. It corresponds to the number of torus holes, minus the number of isolated regions, plus 1.

3.1.1 Characterization of random fields

Minkowski functionals can be calculated on scalar fields, for example on realizations of random fields or on the distribution of galaxies Gott et al. (1986). In this context Minkowski functionals are calculated on the excursion sets of these fields, i.e. on the isodensity surfaces. The excursion set of a function for a threshold u is given by the set $F_u = \{\mathbf{x}, Z(\mathbf{x}) \geq u\}$. It is more complicated for a point distribution like the galaxy distribution since excursion sets are not well defined. Thus, the point distribution is usually smoothed at different scales giving a continuous 3D field, where the Minkowski functionals can be calculated.
An example of isodensity surfaces is given in figure 4 for a Gaussian random field with a power spectrum $P(k) \sim k$ smoothed with kernel width $s=3$ in a cube of size 128. The left column shows the lower density regions and the right column the higher density regions with respective fractions of the volume of 7%, 50% and 93%. The symmetry between high and low density is clearly seen. Another important aspect of Gaussian field is the sponginess at median threshold, quantified by a positive genus, which can be seen here and also in the matter distribution (Gott et al. (1986)).
For random realizations, Minkowski functionals have statistical variations as the field itself. However if the field is homogeneous above a certain scale (i.e. if it presents an ergodicity property), Minkowski functional densities (i.e. Minkowski functionals divided by the volume) converge to their expected value.

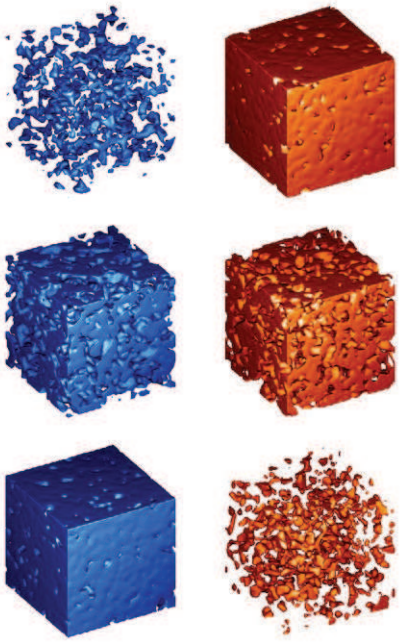


Fig. 4. Isodensity surface of a Gaussian random field smoothed with kernel width $s = 3$. The left column shows the lower density regions and the right column the higher density regions with respective volume fractions of 7%, 50% and 93%.The symmetry between high and low density is clearly seen, as well as the sponginess of the distribution at the median density threshold. Figure courtesy of Martínez et al. (2005).

A change of variable is often made to replace the threshold u by a standard variable, $\nu = \frac{u-\mu}{\sigma}$, where μ and σ are respectively the expected value and standard deviation of the field. Finally the Minkowski functionals are divided by the total volume studied, which gives functional densities which can be compared between different volumes. A great advantage of Minkowski functionals is that their values are known for Gaussian fields. The expected value of the first Minkowski functional $V_0(u)$ is given by the cumulative density function of a standard Gaussian distribution.

$$\mathbb{E}[V_0(u)] = V_{tot} \frac{1}{\sqrt{2\pi}\sigma} \int_u^\infty e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt \tag{25}$$

with $\nu = \frac{u-\mu}{\sigma}$ this gives

$$\mathbb{E}[V_0(\nu)] = V_{tot} \frac{1}{\sqrt{2\pi}} \int_\nu^\infty e^{-\frac{t^2}{2}} dt \tag{26}$$

So the functional density is given by

$$\mathbb{E}[v_0(\nu)] = \frac{1}{\sqrt{2\pi}} \int_\nu^\infty e^{-\frac{t^2}{2}} dt \tag{27}$$

In Tomita (1990) the theoretical values of the other Minkowski functional densities are given in expected value

$$\mathbb{E}[v_1(\nu)] = \frac{2}{3} \frac{\lambda}{\sqrt{2\pi}} \exp\left(\frac{-\nu^2}{2}\right) \tag{28}$$

$$\mathbb{E}[v_2(\nu)] = \frac{2}{3} \frac{\lambda^2}{\sqrt{2\pi}} \nu \exp\left(\frac{-\nu^2}{2}\right) \tag{29}$$

$$\mathbb{E}[v_3(\nu)] = \frac{\lambda^3}{\sqrt{2\pi}} (\nu^2 - 1) \exp\left(\frac{-\nu^2}{2}\right) \tag{30}$$

where $\lambda = \sqrt{\frac{\sigma_1^2}{6\pi\sigma^2}}$, $\sigma = \mathbb{E}\left[\delta^2\right]^{1/2}$, $\sigma_1 = \mathbb{E}\left[|\nabla\delta|^2\right]^{1/2}$, and δ is the fluctuation field.

Note that for gaussian fields, Minkowski functionals are unique up to a multiplicative factor. This is very useful to test for gaussianity, which was the primary use of Minkowski functionals (in particular the genus). Since the matter distribution originates from a Gaussian field, it should also remain gaussian when smoothed at the linear scales of evolution. This overall picture can be tested by smoothing the field at the linear scales and comparing Minkowski functionals with the ones expected for a Gaussian field.

3.1.2 Results obtained on galaxy surveys with Minkowski functionals

3.1.2.1 Test of gaussianity

Let us first illustrate how to test the gaussianity of the field using the third Minkowski functional v_3, which is a topology measure. We perform the calculation for a catalogue described in Martínez et al. (2005). It is a volume limited sample extracted from the 2dF Galaxy Redshift Survey, which contains 8487 galaxies and has been cut to obtain a rectangular volume.

We show in figure 5 the $v_3(\nu)$ curve with a smoothing of the field by gaussian windows with different sizes $s = 1, 2, 4$ and $8\,h^{-1}$Mpc. The smoothing consists in applying a low-pass filter on the field, and only keeping the low Fourier modes. For $s = 4, 8\,h^{-1}$Mpc the remaining Fourier modes present a v_3 statistic very similar to a gaussian field. This suggests that non-linear evolution has not much changed the field topology at these scales. On the contrary, for $s = 1, 2\,h^{-1}$Mpc, there are significant deviations from a gaussian topology.

In Martínez et al. (2005) the galaxy field is first denoised using a multiscale wavelet decomposition. The idea is to look at the topology of the real continuous field, instead of the discrete galaxy distribution smoothed at different scales. Results show that the continuous field has clearly a non-gaussian topology, whereas the galaxy field becomes rapidly gaussian when increasing the smoothing size. Thus, even though the matter distribution is gaussian at linear scales, its real global topology is highly non-gaussian.

3.1.2.2 Model discrimination

A second application of Minkowski functionals, and in particular the genus, is the power of discrimination between models. In Sheth et al. (2003), Minkowski functionals are calculated

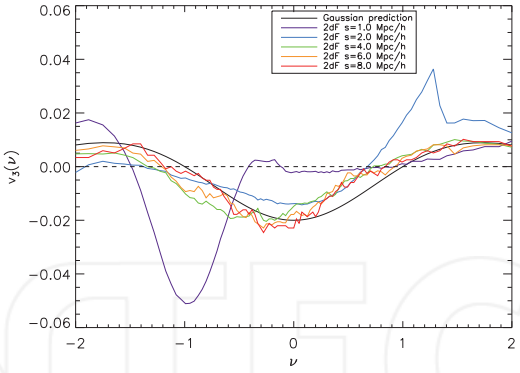


Fig. 5. $v_3(\nu)$ statistic on the 2dF survey. The different curves show $v_3(\nu)$ obtained when smoothing the galaxy field at different scales $s = 1, 2, 4, 6$ and $8\,h^{-1}$Mpc (color curves). We also show the $v_3(\nu)$ expected for a Gaussian field from equation (30) (black curve). Each curve has been globally rescaled to make the comparison possible. We see the curves approaching the gaussian result when increasing the smoothing radius. This supports the fact that the galaxy field becomes gaussian at large scale.

on cosmological simulations of the Virgo consortium, using SURFGEN. Calculations are made on 3 rival cosmological models - a ΛCDM model with $\Omega_m = 0.3$, a Standard-CDM (SCDM) model of a flat Universe with $\Omega_m = 1$, and a τCDM model also flat and with decaying massive τ neutrino as the dark matter constituent. The models show very distinct morphologies, which can be easily distinguished through their Minkowski functionals.
In Hikage et al. (2003) Minkowski functionals are calculated on the SDSS Early Data Release showing an agreement (up to statistical uncertainties) between estimated Minkowski functionals in galaxy catalogues and Minkowski functionals obtained from ΛCDM N-body simulations. Results also show a clear incompatibility with simulations in the SCDM model.

3.1.2.3 Constraining cosmological parameters

Minkowski functionals can also be used to infer real distances of the survey. As for BAOs, the topology of large-scale structures (quantified by the genus statistic) can provide a standard ruler. Up to statistical uncertainties, survey distances can be adjusted to obtain a measured genus that matches the one expected for a given cosmological model. This approach has been first proposed in Park & Kim (2010), and applied in Zunckel et al. (2011) to show how it could constrain the equation of state of the dark energy for the upcoming Baryon Oscillation Spectroscopic Survey (BOSS, www.sdss3.org). They show that BOSS could constrain a constant parameter w in the equation of state with a 5% uncertainty, using this approach.
The quality of a standard ruler depends on the precision in the knowledge of its real size, and also on the uncertainty of the measure. Here, the knowledge of the expected genus may not be as good as the knowledge of the baryon acoustic scale. Yet the genus has the advantage that it can be applied at different scales in the distribution, thus decreasing the measurement uncertainty. Therefore it could be very useful in order to cross-check results obtained using BAOs.

4. Fractal analysis

4.1 The concept of fractal

The idea of a fractal Universe is very old, and already in the eighteenth century Immanuel Kant and Johann Lambert imagined the Universe as an infinite hierarchy of worlds. Later this type of Universe was also invoked to explain Olbert's paradox that the sky is dark at night, in contradiction with the assumption of an infinite number of stars. The mathematical description of fractals only came with Mandelbrot (Mandelbrot (1982)), that reinterpreted them with the concepts of self-similarity and scale-invariance. The denomination of fractal by Mandelbrot comes from the latin word *fractus* which means break. An exact definition of a fractal is difficult since it would exclude possible interesting cases. In Falconer (1990) a loose definition is given for a fractal F:

" 1. F has a fine structure, i.e. details on arbitrary small scales.
2. F is too irregular to be described in traditional geometrical language, both locally and globally.
3. Often F has some form of self-similarity, perhaps approximate or statistical.
4. Usually, the 'fractal dimension' of F (defined in some way) is greater than its topological dimension.
5. In most cases of interest F is defined in a very simple way, perhaps recursively. "

Fractals can be either deterministic, i.e. constructed by an iterative process or recursion, or they can be random, i.e. generated by a stochastic process. Usually, only fractals constructed using an iterative process, rigorously present a self-similarity property. Random fractals are the most used in practice, and can describe numerous irregular objects in the real world, for example clouds, mountains, coastlines or turbulent fluids.

4.1.1 Fractal dimensions

Fractal dimension are extensions of traditional notions of dimensions, that exist for simple usual objects as lines, surfaces, and volumes. A first dimension is the box-counting dimension (also called Minkowski-Bouligand or Minkowski dimension). It is calculated by filling the space with small cubes of size δ, and counting the number of cubes that intersect the object (figure 6). For a line of length l, the number of such cubes $N(\delta)$ will be roughly $N(\delta) \approx \frac{l}{\delta}$, and for a surface of area A the number will be roughly $N(\delta) \approx \frac{A}{\delta^2}$.

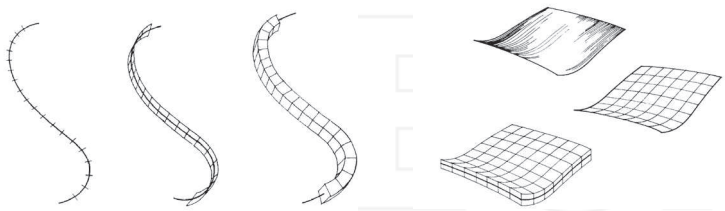


Fig. 6. Measuring the box-counting dimension of a line (left panel) and a surface (right panel). For a line the number of intersecting cubes of size δ scales as $\frac{1}{\delta}$, and for a surface it scales as $\frac{1}{\delta^2}$. Figure courtesy of Feder (1988).

The box-counting dimension of a set F is defined as

$$\dim_{\text{box}}(F) = \lim_{\delta \to 0} \frac{\log(N(\delta))}{\log(1/\delta)} \tag{31}$$

If the limit does not exist, one must talk about the upper box dimension ($\dim_{\text{upperbox}}$) and lower box dimensions ($\dim_{\text{lowerbox}}$) as the upper and lower limits.
Another important dimension is the Hausdorff dimension. A r-cover of F is a countable collection of sets $\mathcal{A} = \bigcup_i \mathcal{A}_i$ which all have a diameter less than r, and such that F is totally covered by this collection. The d-dimensional Hausdorff measure of S is defined as

$$H^d(F) = \lim_{r \to 0} \inf_{\mathcal{A}} \sum_{i=0}^{\infty} diam(\mathcal{A}_i)^d \tag{32}$$

The Hausdorff dimension d_H is defined such that

$$H^d(F) \to \begin{cases} 0 \text{ for } d > d_H(F) \\ \infty \text{ for } d < d_H(F) \end{cases} \tag{33}$$

It can be shown that the Hausdorff dimension is always less than the box-counting dimensions

$$d_H(F) \leq \dim_{\text{lowerbox}} \leq \dim_{\text{upperbox}} \tag{34}$$

Another fractal dimension, which is well adapted to describe random point patterns, is the correlation dimension. In particular, it is very useful for measuring the fractality of the galaxy distribution. Noting $N(< r)$ the number of neighbors that a point has in average within a distance r, the point pattern is said to be fractal with the correlation dimension D_2 when

$$N(< r) \propto r^{D_2} \tag{35}$$

In other terms, D_2 is defined as

$$D_2 = \frac{d(\log N(< r))}{d(\log r)} \tag{36}$$

Such a measure is locally self-similar, i.e. the number of points varies a power law around every point. Intuitively, D_2 gives a local description of the measure linked to the point process. If $D_2 = 1$ then the points will be located on structures that are locally filamentary, and if $D_2 = 2$ these local structures will be planar.
For many well behaved fractals, all these dimensions are equal. For usual geometrical objects (lines, surfaces, volumes), they are integers, thus representing very particular cases. For purely fractal sets however, the Hausdorff dimension can be non integer. Let us give an example of such a case. Figure 7 shows the construction of the von Koch curve F, with segments of the curve each time subdivided into 4 new sub-segments. For this set, the fractal dimensions agree and are equal to $\frac{\ln 3}{\ln 2}$. Concerning the length of the curve E_k at iteration k, it is equal to $\left(\frac{4}{3}\right)^k$, and thus the limit curve has infinite length. On the other hand, the curve occupies zero area in the plane, so neither its length nor its area are useful descriptors.

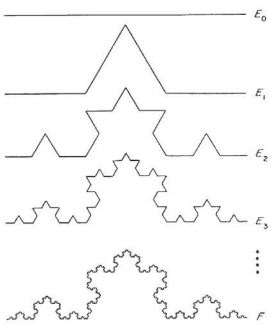


Fig. 7. Successive iterations in the construction of the von Koch curve F. Figure courtesy of Falconer (1990)

4.2 Multifractal measures

A multifractal measure μ on a set A is a generalization of a fractal. It is used when a single exponent (the fractal dimension) is not enough to describe the measure. It is everywhere self-similar, i.e. it varies locally as a power law at every point of its support.

4.2.0.1 Hölder exponent

The Hölder exponent or singularity exponent at a point x_0 is defined

$$\alpha(x_0) = \lim_{\delta \to 0} \frac{\log \mu(B_{x_o}(\delta))}{\log \delta} \tag{37}$$

where $B_{x_0}(\delta)$ is a ball centered on x_0 of radius δ. We have

$$\mu(B_{x_o}(\delta)) \propto \delta^{\alpha(x_0)} \tag{38}$$

The smaller the value of $\alpha(x_0)$, the less the measure is regular around x_0.

4.2.0.2 Singularity Spectrum

As previously mentioned, a multifractal measure is not characterized by a single exponent. For each value of α, a quantity of interest is the fractal dimension $f(\alpha)$ of the set of points that share this singularity exponent α

$$f(\alpha) = d_F\left(\{x_0 \in A \mid \alpha(x_0) = \alpha\}\right) \tag{39}$$

The function $f(\alpha)$ is denoted as the singularity spectrum of μ. The singularity spectrum describes the fraction of points with a given exponent. So if we cover A with boxes of size δ, the number of boxes containing a point with exponent α will scale as $N_\alpha(\delta) \propto \delta^{-f(\alpha)}$, for small δ. The function $f(\alpha)$ is usually a single-humped function with its maximum at $\max_\alpha f(\alpha) = D$, where D is the dimension of the measure support. In the simple case of a monofractal, the function $f(\alpha)$ is reduced to a single point.
In order to calculate the singularity spectrum, the box-counting method consists in calculating the partition function

$$Z(q,\delta) = \sum_{i=0}^{N(\delta)} (\mu_i(\delta))^q \tag{40}$$

where $N(\delta)$ is the number of boxes of size δ intersecting the measure support, and we denoted $\mu_i(\delta) = \mu(B_i(\delta))$. If the measure μ is multifractal, $Z(q,\delta)$ follows a power law in the limit $\delta \to 0$ with

$$Z(q,\delta) \propto \delta^{\tau(q)} \tag{41}$$

The function $\tau(q)$ is related to the singularity spectrum by a Legendre transform

$$f(\alpha) = \min_q(q.\alpha - \tau(q)) \tag{42}$$

In the case of a monofractal, one has $\alpha = f(\alpha) = D$, with D the fractal dimension of μ (because the set of points with exponent α is the support of the measure itself). In terms of the Legendre transform this corresponds to $\tau(q) = D(q-1)$, i.e. the behavior of $\tau(q)$ versus q is a straight line with a slope given by the fractal dimension.
Some years ago, a new method based on wavelets, called Wavelet Transform Modulus Maxima (Muzy et al. (1994)), has shown to provide a more precise method for calculating the singularity spectrum.

4.2.0.3 Rényi dimension

The generalized fractal dimension also called Rényi dimension of order q is given by

$$D_q = \frac{\tau(q)}{q-1} \tag{43}$$

D_0 corresponds to the Hausdorff dimension, D_1 is called the information dimension and D_2 is the correlation dimension, already defined in equation (36).
In practice, because of the limits in resolution and noise, the exponents can only be computed for a limited range of scales. Thus Rényi dimensions depend on the range of scales considered. In the example of the distribution of galaxies, the partition function can be written in a slightly different way

$$Z(q,r) = \frac{1}{N}\sum_{i=0}^{N} n_i(r)^{q-1} \tag{44}$$

where $n_i(r)$ is the numbers of neighbours of point labeled by i within a sphere of radius r. Scale-dependent Rényi dimensions $D_q(r)$ are defined as

$$D_q(r) = \frac{1}{q-1}.\tau(q,r) = \frac{1}{q-1}\frac{d\log Z(q,r)}{d\log r} \tag{45}$$

We show section 4.3.2 how these dimensions can be used in order to test the homogeneity of the galaxy distribution at different scales.

4.3 Fractality of large-scale structures

While it is clear that the distribution of galaxies is fairly inhomogeneous at small scales, the homogenization at larger scales has long been an debated question. This was due to

the lack of data to provide a definitive answer. For a detailed discussion on this matter at that period of time, we refer to Davis (1997). This is an important question, since the Friedmann-Robertson-Walker metric presumes large scale homogeneity and isotropy of the Universe (the cosmological principle). This is the simplest cosmological model and the cornerstone of modern cosmology, so it is a fair question to ask the degree to which it is supported by the observational evidence.

Due to the fractality over a wide range of scales, early redshift surveys did not show a trend to homogenization, and revealed always larger coherent structures (e.g. the CfA stickman, da Costa et al. (1994)). As we show section 4.3.1, the correlation function is not a well-defined quantity for a fractal distribution. Indeed it presupposes the existence of a mean density $\bar{\rho}$, which is not the case for a fractal distribution. So when estimating the correlation function on a fractal distribution, the expected result will have a systematic change with the sample size. In particular, the correlation length r_0 (defined such that $\xi(r_0) = 1$) increases linearly with the sample size. This systematic change in r_0 was observed in early galaxy surveys, which supported the fractality of the distribution. Another argument in favor of a fractal Universe was given by more luminous galaxies and clusters. They have been found to be more correlated than standard galaxies, and therefore to form larger coherent structures.

While all these arguments are valid, they are also explained in the standard model, that assumes an homogenization at large scale. The unstable behavior of r_0 is explained by the small scale fractality of the distribution, whereas the difference of clustering of different object populations is explained by the 'bias' of rare events, luminosity bias, and environmental effects. On the other hand, there are solid arguments in favor of an homogenization of the distribution. This phenomenon is referred as the End of Greatness, i.e. scales where no coherent structures can be observed anymore (thus one does not have to find new superlatives to describe structures).

Let us briefly present the numerous observations in favor of the standard picture, with a large-scale homogeneity of the matter distribution. The first type of observations is given by 2-dimensional data sets on the celestial sphere, including X-ray counts, radio sources counts, γ-ray bursts and CMB, which all present a remarkable degree of isotropy. If we do not assume our position to be special in the Universe, this 2D isotropy is a strong indication of homogeneity. Indeed, it is not clear at all, how random fractal distributions could have such isotropic 2D projections.

The second type of observations is given by 3-dimensional data, in particular the most recent redshift surveys, the 2dF and SDSS, which present a high degree of homogeneity at large scales. As an example of test, in Hogg et al. (2005) the mean number of points at distance R of LRG galaxies is studied and found to vary as $N(R) \propto R^3$ for scales $R > 70h^{-1}$Mpc. This is not compatible with a strictly fractal model, where this number would scale as $N(R) \propto R^D$, with $D < 3$ the fractal dimension. Another type of 3D data is given by the Lyman-α absorption lines, observed in the spectra of quasars, due to intervening clouds of neutral hydrogen. These clouds are found to be ubiquitous and nearly uniformly distributed in space.

4.3.1 Fractility and correlation function

There is a basic assumption behind the notion of correlation function, that the distribution has a well defined mean density $\bar{\rho}$. Indeed the correlation function measures the excess of probability to find a point at distance r of another given point, compared to an unclustered random distribution. The probability dP_r to find a galaxy in a volume dV which is separated by r of a given galaxy is

$$dP_r = \bar{\rho}(1 + \xi(r))dV \tag{46}$$

For a fractal field, the mean density $\bar{\rho}$ is not a well defined concept, thus the correlation function is also not well defined. As we show, if we still try to estimate the 'correlation function' of a fractal field, the expected result will depend on the volume size. Indeed, it can be proven that for a fractal of dimension D, the estimated 'correlation function' in a sphere of size R_s is given by (Sylos Labini et al. (1998))

$$\xi(r) = \frac{D}{3}\left(\frac{r}{R_s}\right)^{D-3} - 1 \tag{47}$$

This behavior is observed, but only at small scales, when measuring the correlation function on redshift surveys with $\xi(r) \propto r^{-\gamma}$ and $\gamma \approx 1.7$.
The correlation length r_0 is defined as the distance where $\xi(r_0) = 1$, i.e. the transition from strong to weak clustering. When estimating ξ on a fractal, the correlation length depends on the size of the volume (Pietronero (1987)). Using formula (47) we get

$$r_0 = \left(\frac{6}{D}\right)^{1/D-3} R_s \tag{48}$$

The correlation length is thus proportional to the volume size R_s. This behavior was observed in early galaxy surveys but more recent redshift surveys, from the Infrared Astronomical Satellite (IRAS, Neugebauer et al. (1984)) to the 2dF and SDSS, have contradicted this trend. The correlation length has become very stable for a given galaxy population ($r_0 \approx 5 - 6h^{-1}$ Mpc).

4.3.2 Multifractality and transition to homogeneity

The multifractal approach has been applied to the distribution of galaxies for this first time in Jones et al. (1988). In this study made on the CfA survey, the distribution of galaxies is found to be better described by a multifractal than by a monofractal. Thus, even for a small survey size, where the galaxy distribution is very inhomogeneous, it did not appear to be purely fractal.
More recent works have applied the multifractal formalism on the SDSS survey and SDSS mock catalogues. The method of analysis consists in studying the Rényi dimensions $D_q(r)$ as defined equation (45) for different scales r. If the exponents are found to be constant and equal to the euclidian dimension, this means that the galaxy distribution is homogeneous for the scales considered.
In Yadav et al. (2005) the exponents $D_q(r)$ are calculated on 2D slices of the SDSS Data Release 1 as well as ΛCDM mock galaxy catalogues, with different mass-luminosity bias (see section 2.3). Calculated in the range of scales from $60 - 70\,h^{-1}$Mpc to $150\,h^{-1}$Mpc, exponents $D_q(r)$ are found to vary between 1.7 and 2.2, i.e. close to a constant value equal to the dimension of the 2D slices. Furthermore, these results have been found to be consistent with the biased mock galaxy catalogues.
In Sarkar et al. (2009) the same analysis is performed on SDSS Data Release 6, but this time using the full 3D survey. This enables to reduce uncertainties in the estimation of the Rényi dimensions $D_q(r)$. Results at two different scales $r = 60\,h^{-1}$Mpc and $r = 70\,h^{-1}$Mpc are shown in figure 8. At scale $r = 60\,h^{-1}$Mpc, Rényi dimensions show small deviations from the constant value $D = 3$ equal to the dimension of the space. They are not consistent with

random catalogues (i.e. with randomly located points, homogeneous by construction) at the 1σ level. However at scale $r = 70\,h^{-1}$Mpc, the Rényi dimensions are nearly constant with q and close to the value $D = 3$. They also become consistent with results obtained on random catalogues. The conclusion of these studies is that the galaxy distribution is homogeneous at large scale, with a transition to homogeneity at around $70\,h^{-1}$Mpc.

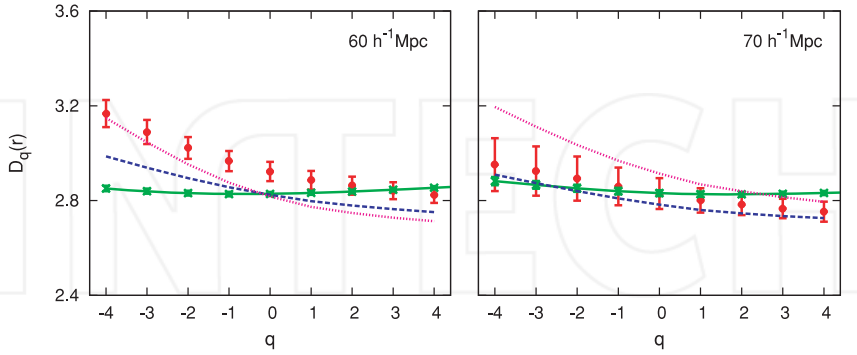


Fig. 8. The Rényi dimensions $D_q(r)$ as a function of q for $r = 60\,h^{-1}$Mpc (left panel) and $r = 70\,h^{-1}$Mpc (right panel). The figure shows the SDSS data (red points), random data (green solid line), N-body (blue dashed) and Millennium N-body simulation (pink dotted). Five independent N-body realizations are used to estimate the 1σ error bars shown on the SDSS data. Figure courtesy of Sarkar et al. (2009)

We conclude this section by quoting Davis (1997), which sums up the current accepted view of the galaxy distribution having a fractal behavior at small scale, and at the same time getting homogeneous at large scales:
"The observed galaxy distribution, being a real physical system rather than a mathematical idealization, is a beautiful example of a limited-scale fractal joined onto sensible, dynamically evolving outer boundary conditions."

5. Conclusion

We have presented several methods for studying the galaxy distribution, which are sensitive to different aspects of it. They should not be directly compared, but rather be considered as complementary. Their goal is to test cosmological models, and constrain parameters inside them. Ideally, the statistics are given by analytic formulas for each model, which can be compared to data results. N-body simulations can also be used for this task, but one has to trust these simulations, which are usually done for a single set of cosmological parameters. The most popular method is Fourier analysis, based on the second order statistics, which are the correlation function and power spectrum. One advantage is that they can be analytically predicted at linear scales, for usual ΛCDM and wCDM models. Another advantage is that BAOs have their signature directly imprinted in these statistics, as they manifest by an excess of clustering at the sound horizon scale. BAOs in the correlation function (or power spectrum) confirms the usual cosmological paradigm of linear gravitational evolution from redshift $z \approx 1000$, and the existence of dark energy. They also provide a standard ruler to quantify the evolution of distances, i.e. constrain cosmological parameters.

A second type of methods studies the global morphology and topology of the galaxy field, through Minkowski functionals. They can also be analytically predicted at linear scales where the matter distribution is gaussian. They were first used to determine the transition from linear to non-linear regimes, where the field becomes non gaussian. Recently they have been used as a standard ruler to constrain cosmological parameters, since their value per unit volume is predicted analytically. They have also been used to distinguish between models with different morphologies, that are well differentiated by Minkowski functionals.
Finally we presented the fractal analysis of the galaxy distribution, also probing the global morphology of the field. Fractality of the galaxy distribution is well-established at small scales. On the other hand, observations confirm large-scale homogeneity, as required by the cosmological principle which is the cornerstone of modern cosmology. One has to turn to multifractal analysis to model the galaxy distribution and study this transition to homogeneity. This is done by studying the scaling exponents of the distribution, which are found to approach the dimension of the space at scales around $70h^{-1}$Mpc, indicating large-scale homogeneity.

6. References

Bassett, B. & Hlozek, R. (2010). *Baryon acoustic oscillations*, Cambridge, UK.

Blake, C. et al. (2011). The WiggleZ Dark Energy Survey: testing the cosmological model with baryon acoustic oscillations at z=0.6, *ArXiv e-prints* .

Bond, J. R., Kofman, L. & Pogosyan, D. (1996). How filaments of galaxies are woven into the cosmic web, *Nature* 380: 603–606.

Cabré, A. & Gaztañaga, E. (2011). Have baryonic acoustic oscillations in the galaxy distribution really been measured ?, *Monthly Notices of the Royal Astronomical Society* 412: L98–L102.

Colless, M. et al. (2001). The 2dF Galaxy Redshift Survey: spectra and redshifts, *Monthly Notices of the Royal Astronomical Society* 328: 1039–1063.

Cooray, A. & Sheth, R. (2002). Halo models of large scale structure, *Physics Reports* 372: 1–129.

da Costa, L. N. et al. (1994). A complete southern sky redshift survey, *Astrophysical Journal Letters* 424: L1–L4.

Davis, M. (1997). Is the Universe Homogeneous on Large Scales?, *Critical Dialogues in Cosmology*, pp. 13–+.

Davis, M. & Peebles, P. J. E. (1983). A survey of galaxy redshifts. V – The two-point position and velocity correlations, *Astrophysical Journal* 267: 465–482.

Eisenstein, D. J. & Hu, W. (1998). Baryonic Features in the Matter Transfer Function, *Astrophysical Journal* 496(2): 605.

Eisenstein, D. J. et al. (2005). Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies, *Astrophysical Journal* 633: 560–574.

Falconer, K. (1990). *Fractal Geometry*, Mathematical Foundations and Applications, John Wiley & Sons Ltd., New-York.

Feder, J. (1988). *Fractals*, Plenum Press, New York.

Gott, J., Dickinson, M. & Melott, A. (1986). The sponge-like topology of large-scale structure in the universe, *Astrophysical Journal* 306: 341–357.

Hamilton, A. J. S. (1993). Toward better ways to measure the galaxy correlation function, *Astrophysical Journal* 417: 19.

Hikage, C. et al. (2003). Minkowski Functionals of SDSS Galaxies I : Analysis of Excursion Sets, *Publications of the Astronomical Society of Japan* 55: 911–931.

Hogg, D. W. et al. (2005). Cosmic homogeneity demonstrated with luminous red galaxies, *Astrophysical Journal* 624: 54–58.

Hütsi, G. (2006). Acoustic oscillations in the SDSS DR4 luminous red galaxy sample power spectrum, *Astronomy and Astrophysics* 449: 891–902.

Jones, B. J. T., Martinez, V. J., Saar, E. & Einasto, J. (1988). Multifractal description of the large-scale structure of the universe, *Astrophysical Journal Letters* 332: L1–L5.

Kaiser, N. (1986). A sparse-sampling strategy for the estimation of large-scale clustering from redshift surveys, *Monthly Notices of the Royal Astronomical Society* 219: 785–790.

Kerscher, M., Szapudi, I. & Szalay, A. S. (2000). A Comparison of Estimators for the Two-Point Correlation Function, *Astrophysical Journal* 535: L13–L16.

Labatie, A., Starck, J., Lachièze-Rey, M. & Arnalte-Mur, P. (2010). Uncertainty in 2-point correlation function estimators and BAO detection in SDSS DR7, *ArXiv e-prints* .

Landy, S. D. & Szalay, A. S. (1993). Bias and variance of angular correlation functions, *Astrophysical Journal* 412: 64–71.

Mandelbrot, B. (1982). *The Fractal Geometry of Nature*, Times Books.

Martínez, V., Starck, J., Saar, E., Donoho, D., Reynolds, S., de la Cruz, P. & S., P. (2005). Morphology of the Galaxy Distribution from Wavelet Denoising, *Astrophysical Journal* 634: 744–755.

Muzy, J. F., Bacry, E. & Arnéodo, A. (1994). The multifractal formalism revisited with wavelets, *International Journal of Bifurcation and Chaos* 4: 245–302.

Neugebauer, G. et al. (1984). The Infrared Astronomical Satellite (IRAS) mission, *Astrophysical Journal Letters* 278: L1–L6.

Norberg, P. & Baugh, C. M. (2001). The 2df galaxy redshift survey: The dependence of galaxy clustering on luminosity and spectral type, *Monthly Notices of the Royal Astronomical Society* 332: 827. 13 p.

Park, C. & Kim, Y. (2010). Large-scale Structure of the Universe as a Cosmic Standard Ruler, *Astrophysical Journal Letters* 715: L185–L188.

Peebles, P. J. E. & Hauser, M. G. (1974). Statistical Analysis of Catalogs of Extragalactic Objects. III. The Shane-Wirtanen and Zwicky Catalogs, *Astrophysical Journal Supplement Series* 28: 19–+.

Percival, W. J. et al. (2010). Baryon acoustic oscillations in the Sloan Digital Sky Survey Data Release 7 galaxy sample, *Monthly Notices of the Royal Astronomical Society* 401: 2148–2168.

Pietronero, L. (1987). The fractal structure of the universe: Correlations of galaxies and clusters and the average mass density, *Physica A Statistical Mechanics and its Applications* 144: 257–284.

Pons-Bordería, M., Martínez, V. J., Stoyan, D., Stoyan, H. & Saar, E. (1999). Comparing estimators of the galaxy correlation function, *Astrophysical Journal* 523: 480–491.

Sarkar, P., Yadav, J., Pandey, B. & Bharadwaj, S. (2009). The scale of homogeneity of the galaxy distribution in SDSS DR6, *Monthly Notices of the Royal Astronomical Society* 399: L128–L131.

Sheth, J. V., Sahni, V., Shandarin, S. F. & Sathyaprakash, B. S. (2003). Measuring the geometry and topology of large-scale structure using SURFGEN: methodology and preliminary results, *Monthly Notices of the Royal Astronomical Society* 343: 22–46.

Springel et al. (2005). Simulations of the formation, evolution and clustering of galaxies and quasars, *Nature* 435: 629–636.

Sylos Labini, F., Montuori, M. & Pietronero, L. (1998). Scale-invariance of galaxy clustering, *Physics Reports* 293: 61–226.

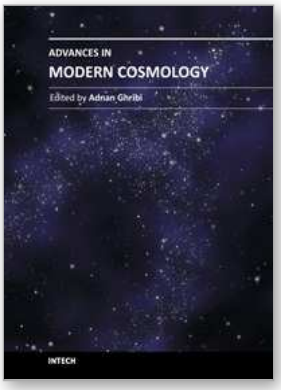
Tomita, H. (1990). *Formation, Dynamics and Statistics of Patterns, Vol.1, (World Scientific), 113*, Vol. 1, World Scientific, pp. 113–157.

Yadav, J., Bharadwaj, S., Pandey, B. & Seshadri, T. R. (2005). Testing homogeneity on large scales in the Sloan Digital Sky Survey Data Release One, *Monthly Notices of the Royal Astronomical Society* 364: 601–606.

York, D. G. et al. (2000). The Sloan Digital Sky Survey: Technical Summary, *Astrophysical Journal* 120: 1579–1587.

Zehavi, I. et al. (2005). The Luminosity and Color Dependence of the Galaxy Correlation Function, *Astrophysical Journal* 630: 1–27.

Zunckel, C., Gott, J. R. & Lunnan, R. (2011). Using the topology of large-scale structure to constrain dark energy, *Monthly Notices of the Royal Astronomical Society* 412: 1401–1408.



Advances in Modern Cosmology
Edited by Dr. Adnan Ghribi

ISBN 978-953-307-423-8
Hard cover, 198 pages
Publisher InTech
Published online 29, August, 2011
Published in print edition August, 2011

The twentieth century elevated our understanding of the Universe from its early stages to what it is today and what is to become of it. Cosmology is the weapon that utilizes all the scientific tools that we have created to feel less lost in the immensity of our Universe. The standard model is the theory that explains the best what we observe. Even with all the successes that this theory had, two main questions are still to be answered: What is the nature of dark matter and dark energy? This book attempts to understand these questions while giving some of the most promising advances in modern cosmology.

How to reference
In order to correctly reference this scholarly work, feel free to copy and paste the following:

Antoine Labatie, Jean-Luc Starck and Marc Lachièze-Rey (2011). Statistical Study of the Galaxy Distribution, Advances in Modern Cosmology, Dr. Adnan Ghribi (Ed.), ISBN: 978-953-307-423-8, InTech, Available from: http://www.intechopen.com/books/advances-in-modern-cosmology/statistical-study-of-the-galaxy-distribution

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821