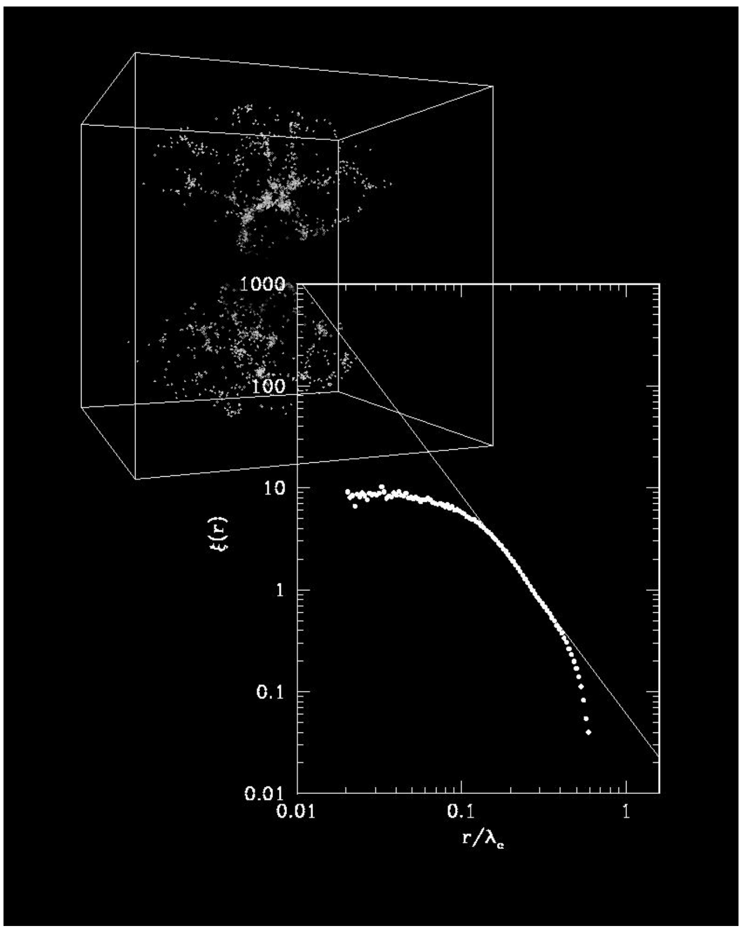
Selection Effects on the Two-Point Correlation Function of Voronoi Distributions of Galaxies

Characterizing the complete from the incomplete

Jonathan Heiner

Under supervision of Rien van de Weygaert



Contents

1 Introduction

The search for unravelling the large-scale structure of our universe is ongoing, but will we be able to see all there is? We have, for example, the so-called zone of avoidance, the part of our sky that is obscured by the silhouets of the galaxy we live in. Of course, concerning the largest scales we are able to see, the universe appears homogeneous. It may not be necessary to see all objects in all directions. Just an appreciably sized sample should do. Or does it?

It is customary to assume that the properties of the complete distribution of galaxies can be characterized by the limited volume, especially concerning the shape obtained by galaxy redshift surveys. Corrections are made to compensate for the different shapes. Often this involves a weighting scheme.

Simulations can be performed to imitate the observed structures. Numerical simulations, in particular kinematic models using Voronoi tessellations, are used in trying to mimic the distribution of galaxies in the real universe. The geometrical limits of sky surveys are incorporated by creating structure simulations with the same shape as the observed galaxies in the surveys. The properties of this simulations can then be compared to the complete simulation, which has a cubic shape.

The effects of this shape selection on this simulations of the large-scale structure of the universe are examined, with the aim of verifying the assumed invariance of large-scale clustering properties. This work focuses on the effects of not seeing the whole sky and not

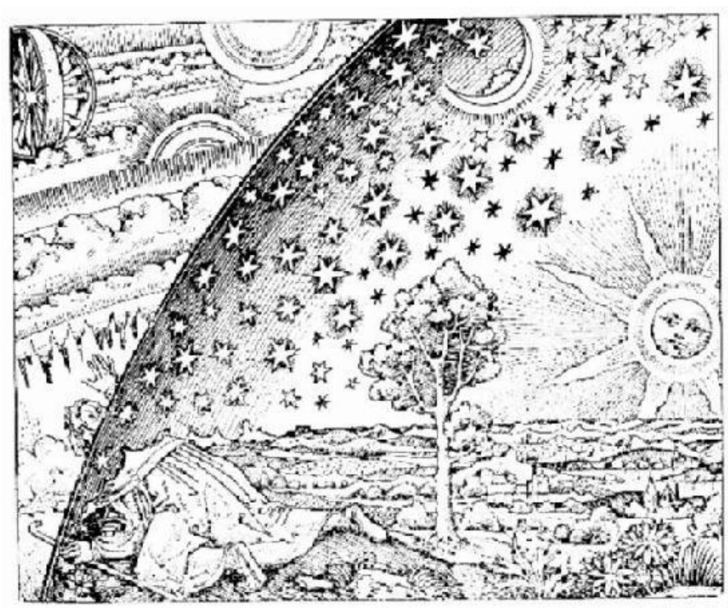


Figure 1: A man looking beyond the edge of space into the realm of God. We might not ever be able to travel to all places in the universe, but we can sure try and get to see as much as we can. This image is commonly thought to be originating from the 16th-century.

seeing all galaxies within range by considering the two-point correlation function of the sample of galaxies and its derived subsamples simulating observational constraints. The two-point correlation function is used to quantify in a statistic manner the possible effects resulting from seletions.
Some basic cosmological concepts will be treated first. The large-scale structure as is being observed nowadays is treated next, followed by a description of the basic concepts underlying the formation of structure in the universe.
Having gone through these basic descriptions, we proceed to the more specific focus of this work: the two-point correlation function and its possible change when observing parts of a complete galaxy distribution.

The 'Groot Onderzoek', of which this report is the culmination, has been supervised by Prof. Dr. Rien van de Weygaert.

Jonathan Heiner,
April 7, 2004.

2 Cosmological Background

2.1 The Great Outdoors

Extending our physical theories to encompass the vastness of the universe requires some adaptation of the standard physical methods. The forces that dominate on the scale of electrons are not the same one dominating our macroscopic world. In the vastness of space dominates - as is generally assumed - gravity, described by the general theory of relativity. Before the realm outside of our own galaxy could really started to be observed, the nature of the objects outside of it had to be determined. Remember that galaxies outside our own were still called nebulae only four decades ago. (Hubble in 1938 noted that the nearby nebulae were already resolved into individual stars, but the distance to them remained unknown. The shape of the nebulae had already been observed for decades, e.g. Fig. 2) Relativistic theory was not as firmly established as it is nowadays.
The universe can be approached bottom-up or top-down (see e.g. the introduction of a published lecture of Milne, 1940). Cosmological Theory is concerned with the distribution of matter and motion in the universe as a whole, Milne poses. Bottom-up in this context means starting from all available empirical and theoretical laws and theories and each step would have to be thoroughly verified in order to proceed. This would exclude the use of too many assumptions or the use of approximations to further develop theories on top of those approximations. This is why, as Milne continues to describe, the top-down approach in general is more favored in cosmology. This strategy treats the universe as a large experiment, observing it without much assumptions. It is a justification to infer empirical

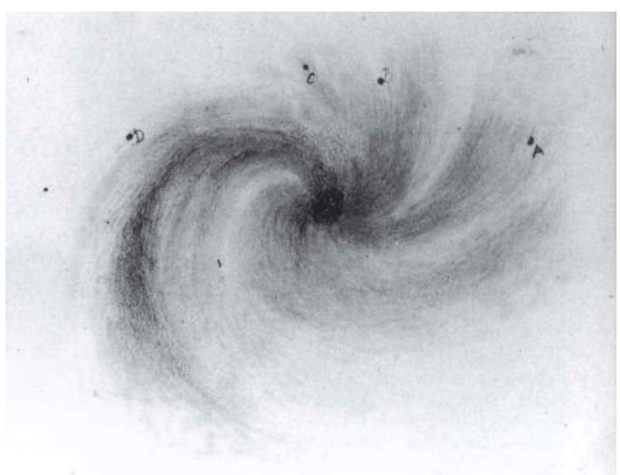


Figure 2: Sir Rosse found object M99 to have a spiral shape in 1848, shortly after he discovered the shape of M51, the 'Whirlpool Galaxy'. The debate about what these nebulae might be, however, had only just begun. (Source: *The Cambridge Illustrated History of Astronomy*, Cambridge University Press 1997)

principles and build upon them without having to work ones way up from established principles verified to work on small scales. It also allows the simplification of neglecting small-scale effects that do not appear to be influencing evolution on the larger scales.
The practice to mimick the observed galaxy distribution in an effort to identify the underlying processes that created it is widespread. The universe is like an experiment on its own, just one possible realization of many, allowing an observer to deduce principles from it in an empirical fashion. Physical principles and laws could then be determined from the resulting empirical laws and deductions. Cosmology has since its early days progressed to incorporate a wide array of disciplines of the physical sciences to describe the universe, such as quantum mechanics, thermodynamics and the already mentioned general relativity.

2.2 The Cosmological Principle As Cornerstone Of Reasoning

Simplifying assumptions and generalizations have to be made to get a first handle on the universe as a whole. The Cosmological Principle and Weyl's postulate are each generalizations for the (absence of) structure on the largest scales to be able to describe the universe in general with continuous approximations. A compact overview can be found in Reid et al., 2002.
As initial assumption about the universe as a whole, we have the cosmological principle. It is based on non-uniqueness and simple observation: the universe is homogeneous and isotropic.
The Copernican principle is a part of this: the notion of our living at a location of space that is not particular special is in a sense an extrapolation of the findings of man through history: it was first thought that the earth was flat, later that the sun was circling around the earth and that the stars were holes in some kind of ceiling around the 'universe'. Our Sun was found to inhabit a particularly non-special part of the galaxy. Our own galaxy was, though fairly large as they come, not unique. And so we apparently find ourselves in a universe that really appears homogeneous and isotropic.
In addition to the universe being homogeneous and isotropic, it is observed to be uniformly expanding. The uniform expansion of the universe assumes that every part of the universe expands in the same way. This is an essential requirement in light of the other two assumptions because a non-uniformly expanding universe would instantaneously cease being homogeneous as one part expands differently than another part. The cosmological principle is instrumental in deriving basic properties of the universe and provides a starting point for interpretation.

2.3 Weyl's Postulate

Considering each galaxy represented by a line depicting the three spatial coordinates and the time coordinate (a so-called worldline), Weyl's postulate states that these lines do not intersect, expect for the point were these lines have originated.
This allows for the construction of a coherent coordinate system, with space-like smooth hypersurfaces with a constant time coordinate. It is then possible to identify different

epochs in time in a consistent way as measured from the point of origin..
The ability to construct these hypersurfaces implies each galaxy moving at a non-relativistic velocity and that the assignment of a convenient coordinate frame is possible.
The postulate also allows the derivation of a constraint on the form of the geometrical description of space-time. In general, $ds^2 = g_{\mu\nu}dx^\mu dx^\nu$ $(\mu, \nu = 0, 1, 2, 3)$ per definition. Now Weyle's postulate allows us to obtain $g_{00} = 1$ (a regular time axis) and the equation for the geodesic ds becomes

$$ds^2 = (dx^0)^2 + g_{ij}dx^i dx^j = c^2dt^2 + g_{ij}dx^i dx^j (i, j = 1, 2, 3). \tag{1}$$

t is then called the proper time local to each galaxy, signifying a homogeneous passing of time at large scales.

2.4 Establishing the Metric

A general description of the metric of the universe needs to be homogeneous, isotropic and allow for possible curvature. The curvature is described by κ, which has value -1 for open, 0 for flat, 1 for closed space (respectively hyperbolic, euclidean or spherical, see Fig. 3) .

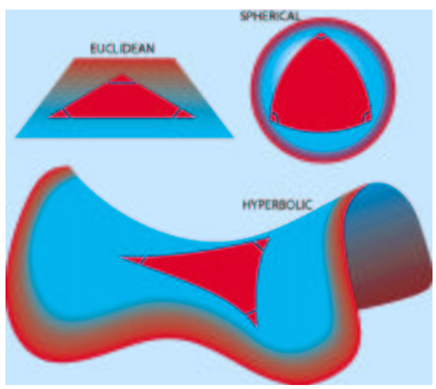


Figure 3: Two-dimensional analogy of the open, closed or flat spaces, from Luminet et al., 1999.

These are the only values of κ which allow for an isotropic universe. κ is related to the radius of curvature by $\kappa = \frac{1}{R_c^2}$.
Einstein gave a solution for a closed universe with matter, which he had to get static and closed by introducing the now famous cosmological constant Λ; a parameter representing a repelling force to counter gravitation. De Sitter gave a solution which was expanding but had no matter content (only mass-less particles). Robertson and Walker derived the metric that could describe the cosmological principle conforming universe in a satisfying fashion. The Robertson-Walker metric is:

$$ds^2 = c^2dt^2 - a^2(t)\left\{dr^2 + R_c^2 S_k^2\left(\frac{r}{R_c}\right)\left[d\theta^2 + \sin^2(\theta)d\phi^2\right]\right\}, \tag{2}$$

where $a(t) \equiv \frac{R(t)}{R_c}$ and

$$S_k\left(\frac{r}{R_c}\right) = \begin{cases} \sin\frac{r}{R_c} & k = 1 \\ \frac{r}{R_c} & k = 0 \\ \sinh\frac{r}{R_c} & k = -1 \end{cases} \tag{3}$$

The introduction of a scale factor $a(t)$ is convenient because it incorporates the rate of expansion of the universe independent from its physical laws.

2.5 Hubble's Constant

A so-called co-moving coordinate system is defined which takes the expansion out of the regular space coordinates,

$$s(t) = a(t)x, \tag{4}$$

Note that this can only hold for a uniformly expanding universe, as the equation must be valid at any position in space and because the scale factor is independent of physical coordinates.

From this definition the time derivative can be taken:

$$\dot{s(t)} = v(t) = \frac{\dot{a(t)}}{a(t)}a(t)x = \frac{\dot{a(t)}}{a(t)}s(t) = H(t)s(t) \tag{5}$$

This parameter H(t) is the Hubble parameter measuring the expansion rate of the universe at time t. The value of H at the present epoch is defined as H_0 and is known as the Hubble Constant. Convenient units to give its value are $kms^{-1}Mpc^{-1}$. Often H_0 is written in units of $100hkms^{-1}Mpc^{-1}$, which allows distances to be given in units of $h^{-1}Mpc$.

A lot of effort has gone into determining the value of H_0 with various methods. Chen et al., 2003, present a thorough analysis of the more than 400 measurements of H_0 collected by Huchra[1]; Fig. 4. They highlight both a $H_0 = 71kms^{-1}Mpc^{-1}$ as determined by the WMAP experiment (the Wilkinson Microwave Anisotropy Probe[2]) and a $H_0 = 67kms^{-1}Mpc^{-1}$ as inferred from median statistics analysis. Uncertainty in these value is of the order of $10kms^{-1}Mpc^{-1}$.

2.6 Friedmann Models

The energy momentum tensor $T_{\mu\nu}$, considering isotropic pressure in a perfect fluid,is given by

$$T_{\mu\nu} = \left(\rho + \frac{p}{c^2}\right) u_\mu u_\nu - p g_{\mu\nu}, \tag{6}$$

with u_μ, u_ν being worldline velocities of fluid particles.

We take $g_{\mu\nu}$ from the Robertson-Walker metric (flat space) as a constraint due to homogeneity and isotropy.

[1]http://cfa-www.harvard.edu/∼huchra

[2]http://map.gsfc.nasa.gov/

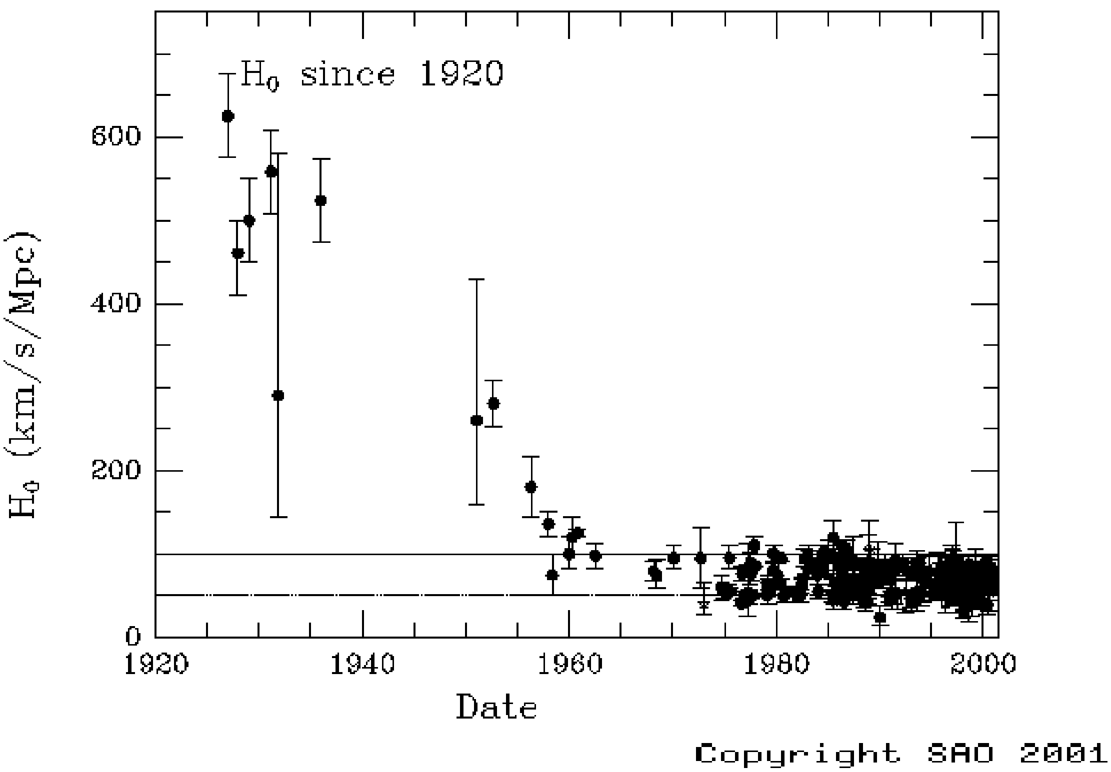


Figure 4: Determinations of the Hubble constant, as collected by Huchra.

The Einstein tensor is defined as

$$G^{\mu\nu} = R^{\mu\nu} - \frac{1}{2} g^{\mu\nu} R, \tag{7}$$

where the Ricci tensor $R^{\mu\nu}$ en the curvature scalar R are contractions of the Riemann tensor.
The Einstein Field Equations are then:

$$G^{\mu\nu} = -\frac{8\pi G}{c^4} T^{\mu\nu}, \tag{8}$$

as both the Einstein tensor and the energy-momentum tensor have zero co-variant divergence. The proportionality constant follows from requiring Einstein's theory to become Newtonian gravity in the low mass limit.
Adding the before mentioned cosmological constant this equation simply gains another term, $G^{\mu\nu} + \Lambda g^{\mu\nu} = -\frac{8\pi G}{c^4} T^{\mu\nu}$.
Taking the perfect fluid energy momentum tensor, this finally gives the Friedmann equations for a universe with a cosmological constant Λ, expressing energy conservation and the relation between expansion and energy content.

$$\ddot{a} = -\frac{4\pi G}{3}\left(\rho + \frac{3p}{c^2}\right)a + \frac{\Lambda}{3}a \tag{9}$$

$$\frac{\dot{a}^2}{a^2} = \frac{8\pi G}{3}\rho - \frac{kc^2/R_0^2}{a^2} + \frac{\Lambda}{3}a^2 \tag{10}$$

These equations are the mainstay of the standard model describing the dynamics of the universe.
Friedmann published his equations in 1922, being the first to derive a solution containing only matter from the Einstein equations. Lemaître independently discovered the same cosmological model, so the model is sometimes referred to as the Friedmann-Lemaître Cosmological Model. However, the combination the name Friedmann-Robertson-Walker (FRW) equations is used to signify the crucial element of the Robertson-Walker metric in the model.

2.7 Ω Critical to a Flat Universe

The density of the universe, ρ, is directly coupled to the geometry, described by k, the curvature. The matter content of the universe is therefore responsible for a (spatially) closed, flat or open universe. Requiring $k = 0$, the flat universe, means the density has to be

$$\rho_{crit} = \frac{3H^2}{8\pi G}, \tag{11}$$

the critical density. For convenience, the mass density parameter Ω is defined as

$$\Omega = \frac{\rho}{\rho_{crit}}. \tag{12}$$

Its current value is denoted as Ω_0. Different subscripts are added to Ω to distinguish matter, radiation or vacuum components if required: $\Omega_m, \Omega_r, \Omega_{vac}$.
The components of the density ρ are matter (non-relativistic) and radiation (relativistic), whereas the Λ term carries the vacuum density.
Recent observations suggest a flat universe, or one very close to it. WMAP results (from Bennett et al. 2003, Sperger et al. 2003) are $\Omega_{tot} = 1.02 \pm 0.02$. This is why the Einstein-de Sitter (EdS) universe (flat, $\Omega = 1$, dominated by matter) carries so much importance. The slightest perturbation from a flat universe leads to a definitive departure from flatness. The universe being so close to flatness at present means it must have been extremely close to flatness in the beginning (this is termed the flatness problem). The EdS flat universe has specific solutions to the FRW-equations, making it a good starting point for further deductions.
The matter component can be derived to evolve proportional to R^{-3}, the radiation component proportional to R^{-4}. Finally, the vacuum density is assumed to be constant, with $\frac{\Lambda}{3} \equiv \frac{8\pi G}{3}\rho_{v,0}$.
The different constituents of Ω to the density, and therefore also the geometry, are dominant to the dynamics at different values of the expansion parameter, as the universe shifts from

radiation-dominated via matter-dominated to vacuum-dominated. Our universe is believed to be currently matter-dominated and possible close to a transition to vacuum-dominated. The Hubble parameter evolves like

$$H^2(a) = H_0^2 \left\{ \Omega_{m,0} a^{-3} + \Omega_{rad,0} a^{-4} + \Omega_{vac,0} - (\Omega_0 - 1) a^{-2} \right\} . \tag{13}$$

in terms of the different contributions to Ω, reflecting the forementioned proportionalities.

3 Large-Scale Structure on Display

3.1 Observe the Universe, Look Up to the Sky

One of the observations with the most public impact in recent years have been the Hubble Deep Field pictures. Speckled with lots and lots of galaxies, these colorful images awaken the curiosity of mankind to see what is beyond. (See Fig. 5 for an example.) The Hubble space telescope peered through the universe, able to see even the earliest galaxies.
A rough description of the large-scale structure of the universe starts with a short overview of what is known from observations.
Distances in our expanding universe can be inferred from the redshift of observed objects through the well-known formula $d = \frac{cz}{H}$, at least for small redshifts. At larger redshifts one has to account for the geometrical structure of space characterized by the different

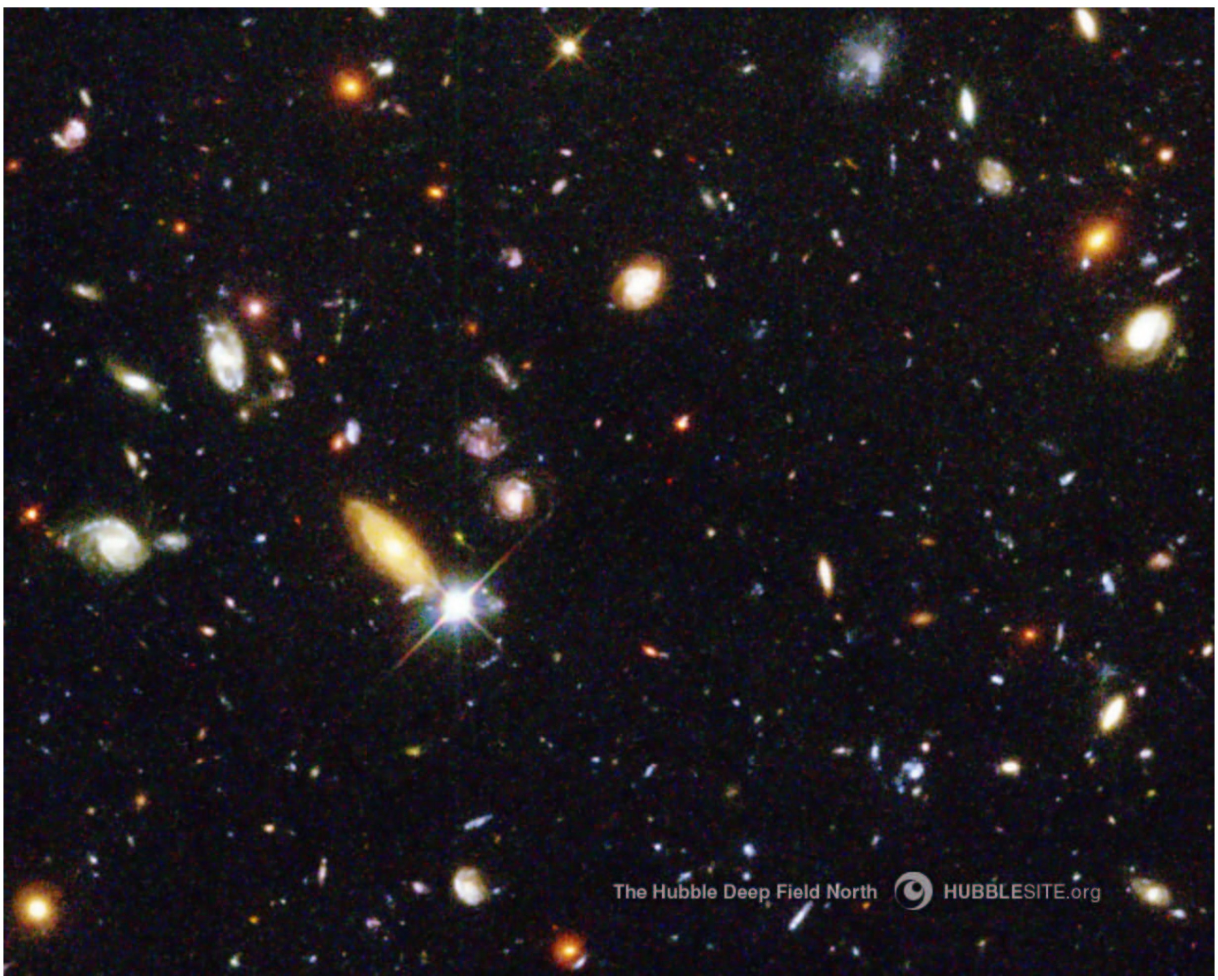


Figure 5: Hubble Deep Field North picture showing a wealth of galaxies (from hubblesite.org; Robert Williams and the Hubble Deep Field Team)

constituents of Ω, but essentially the coupling of distance and redshift remains. This makes redshift determatination the prime tool for measuring distances at the largest scales. By giving a rough estimation of distance, redshift surveys have been instrumental in mapping the universe. As more sophisticated methods for determining distances from these redshifts emerged, taking more and more correction factors into account, this became even more true. The surveys provide the positional and velocity information needed to get a clearer picture of the distribution of galaxies and other gravitating matter. Together with the information from the cosmic microwave background radiation, N-body simulations can be done to mimic the dynamics of the universe.

This purely experimental approach of simulating the dynamics, as discussed by e.g. Jones et al. (unpublished), has as its eventual goal to understand the underlying principles behind the parameters of the model. With the data on how the universe started out and the resulting structure in sight, one hopes to be able to determine how structure formation works and how the universe became the way it looks like today. Last but not least is the anticipation of the ongoing evolution of the universe.

Of course, sky redshift surveys wouldn't have been possible if the universe was indeed, as was expected from simplicity arguments (Occam's Razor), static and without structure. We turn our attention to the discovery of the rich structure at the larger scales.

3.2 Discovering the Foam-Like Structure of the Universe

While the idea of a universe with homogeneously distributed galaxies was a popular one and hard to refute, the large-scale structure of voids, filaments, walls and clusters slowly became clear during the 1980's. The CfA redhift survey (Davis et al., 1982, see Fig. 6) provided an indication of a structure with clusters - with filaments between them - and 'empty holes'.

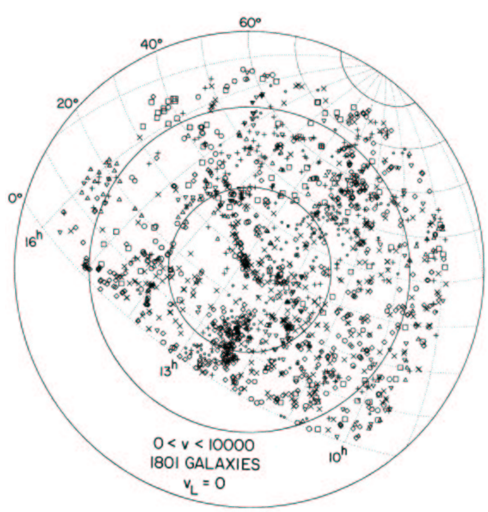


Figure 6: Hints of structure from the CfA survey, northern hemisphere, Davis et al., 1982

The volume, however, was too limited to say anything with certainty about larger scales. Structure at larger scale could not be ruled out yet. The CfA survey was extended, as discussed in 1986 by De Lapparent et al., who tried to explain the observed large bubble-like structures with diameters in the order of $20h^{-1}$ Mpc. The observed structures were still comparable with the sample depth in size, so more accurate observations were needed to truly show the structures involved. Slices in multiple directions however, implicated the structure (see e.g. Huchra et al., 1988).
The Las Campanas Deep Redshift Survey (Oemler Jr. et al. 1993), with more than 20,000 galaxies out to a depth of $600h^{-1}$ Mpc convincingly showed the cellular structure in which the galaxies are organized. It also showed large-scale homogeneity at a scale above $100h^{-1}$ Mpc. Galaxy redshift surveys, like the recently completed 2dF survey (Taylor et al., 1997) and the Sloan Digital Sky Survey (see e.g. Einasto et al., 2003) continue to reach further into the universe, confirming the by now well known spunge-like structure and confirming homogeneity on the largest scales. More about these surveys follows below.
On the different scales of increasing magnitude it could be observed that, while the local group galaxies are completely virialized, the superclusters of galaxies are believed to be still during the linear stages of collapse. These conglomerations are actively evolving, and, because they are not yet virialized, are probed for the matter distribution that gave rise to the observed structure.

3.3 Recent Large Sky Surveys

The recent large sky surveys deserve special attention as they have mapped, or are still in the process of mapping, an unprecedented amount of galaxy redshifts and provide the most complete view of the spatial distribution of galaxies at the largest scales so far.

The 2dF Galaxy Redshift Survey
The name 2dF is a reference to the 2 degree Field spectrometer on the 3.9m Anglo-Australian Telescope that was used to collect redshift of over 200,000 galaxies. The survey recently finished in 2003 and has measured galaxy clustering on scales up to a redshift of 0.2. Of course the full results of the survey will require years of analysis, but a map with the observed galaxy distribution has been released; see Fig. 7. Among the results are an accurate measurement of the power spectrum of galaxy clustering, measurements of the distortion of the clustering pattern in redshift space, measurements of the galaxy luminosity function, the Hubble constant and the cosmological parameter. (A complete description is available on the 2dFGRS homepage.) To cite some figures: 2dF data suggests a power spectrum corresponding to an $\Omega_m h = 0.20$ model, with $\Omega_b/\Omega_m = 0.10$ (Percival et al., 2001; b is for baryon). Zandivarez et al., 2003, determine from the 2dF data for the power law regime of the two-point correlation function (explained later on) a slope of $\gamma = 1.6 \pm 0.1$ on scales smaller than $20h^{-1}Mpc$.

The Sloan Digital Sky Survey
The SDSS began in 1998 and will map one-quarter of the sky, by measuring positions and

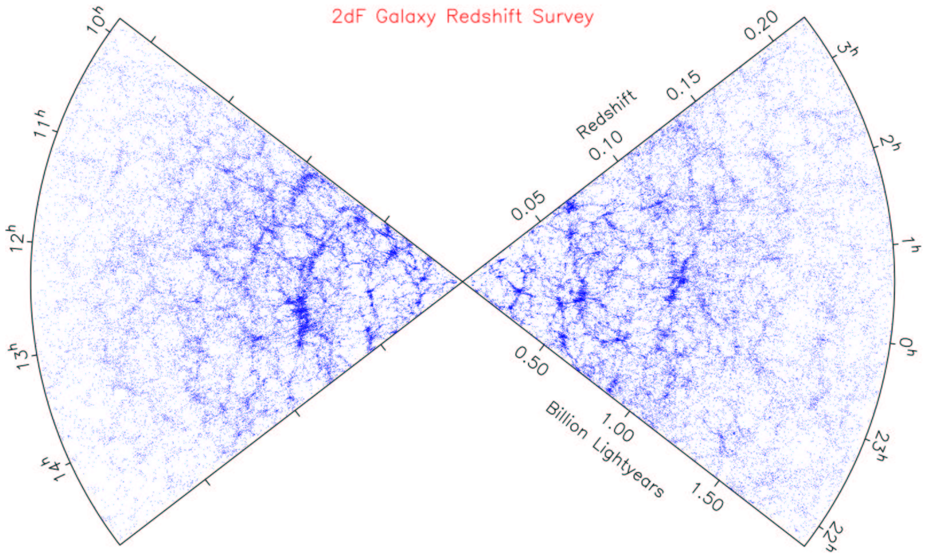


Figure 7: 2dF release showing large-scale distribution of galaxies.

brightnesses of more than 100 million of objects on the sky. Also, the distances to more than a million galaxies and quasars will be obtained. The SDSS obtains skypictures in five different wavelengths. The map will give a 3-dimensional representation of a large fraction of the visible universe. It will be used to study large-scale structures throughout the universe; the voids, filaments, walls and clusters. Knowledge of this structure will enable the validation of various theories about the way this structure was formed. It will also provide more information about the nature of the pervasive 'invisible' dark matter. Because of the huge range of the survey, it is expected that the largest structures of the universe can be observed. Or, differently said, it will show the scale at which the universe appears homogeneous. The results so far have been impressive to say the least. Fig. 8 features a Delaunay representation of the observed structure by the SDSS, showing detailed structure. The power law approximated real-space two-point correlation function from the SDSS data yields a value of $\gamma = 1.75 \pm 0.03$ according to Zehavi et al., 2001.

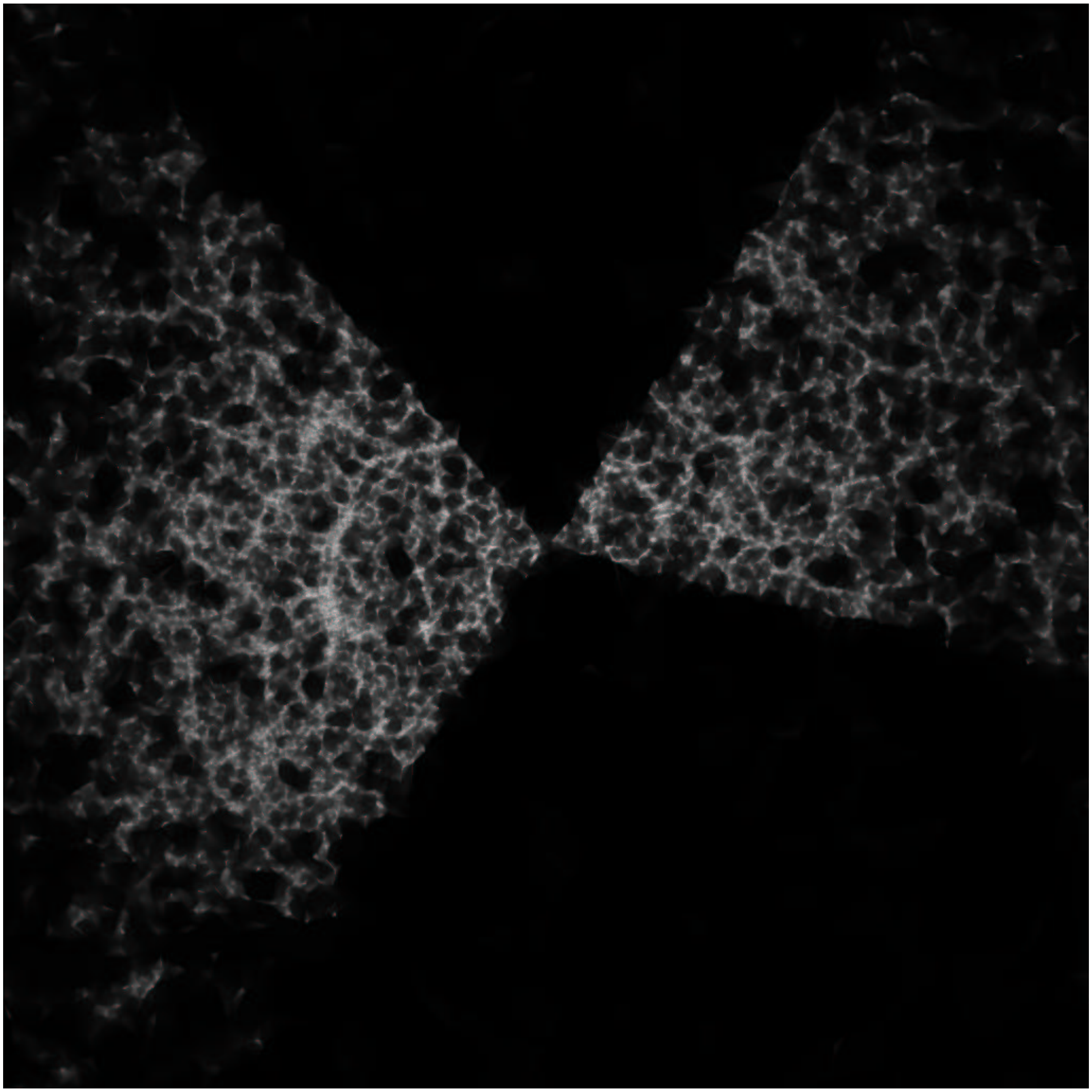


Figure 8: SDSS preliminary release showing the extreme range that has been covered by this survey. This figure was made using a Delaunay tessellation, depicting the density distributions rather than individual galaxies. (Courtesy: Willem Schaap; Rien van de Weygaert, private communication)

4 Structure Formation Walkthrough

4.1 Commence at the Beginning

The well-known Big Bang theory describing the origin of Everything is relatively young. As the existence of galaxies besides our own used to be uncertain, the finite age of the universe was equally unestablished. And indeed, why should the universe have had a beginning? When the idea of a static universe seemingly had to be discarded because of the observation that all galaxies recede from ours (the expanding universe), the theories about a violent beginning, a kind of Big Bang, gained more ground. The term Big Bang, by the way, was made up by one of the opponents of the theory, Sir Fred Hoyle. The name was however readily adopted by supporters of the theory of a violent beginning..

The Big Bang refers to the presumed infinite density and temperature from which the universe started to expand, the intersection of all worldlines. In 1965, Penzias and Wilson discovered the Cosmic Microwave Background, the thermal radiation originating from the surface of last scattering at the epoch of recombination (proton and electrons combined to form hydrogen atoms). This CMB was a clear echo of the beginning of our universe and proved to be extremely isotropic, up to the μK scale. It fitted a perfect blackbody of $T = 2.725 \pm 0.002K$ (see Fig. 9 from HyperPhysics[3]).

Although homogeneous on the largest scales, the CMB shows anisotropies on scales expected to correspond to structure in our present universe. The anisotropy on smaller scales is being probed for hints of structure formation and the value of cosmological constants. The term 'anisotropic collapse', dealing with the collapse of matter into galaxies and clusters, is related to CMB anisotropy.

A much more detailed treatment of the content of the following sections in this chapter can be found in Van de Weygaert (2002) and references therein.

[3]http://hyperphysics.phy-astr.gsu.edu/hbase/bkg3k.html

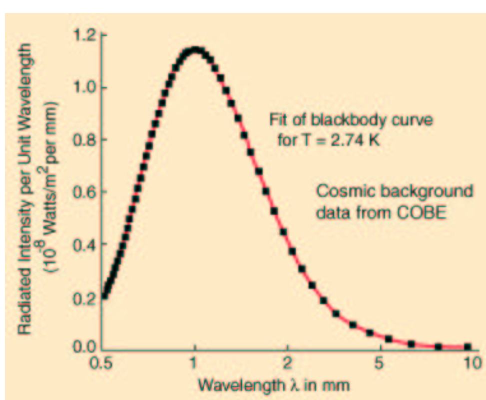


Figure 9: The COBE satellite measured the CMB up to the μK scale, obtaining a perfect blackbody spectrum. (From HyperPhysics)

4.2 Growth of Fluctuations

The radiation dominated epoch ended when the universe became transparent and the CMB radiation was emitted. As has been mentioned above, it was nearly homogeneous, but contained noise. The noise is supposed to be gaussian in nature, a reasonable assumption from the central limit theorem. Gaussianity is also a believed to be the result of the assumed inflationary fase when quantum fluctuations blew up to small curvature fluctuations and the universe grew exponentially.

The reason for the growing structure is sought in the fluctuations in the primordial matter (be it baryonic or the yet to be observed dark matter) distribution. The noise in the gravitational field is described by its power spectrum, the definition of which is, according to the wikipedia[4], 'a plot of the portion of a signal's power (energy per unit time) falling within given frequency bins'. This means that the power spectrum is a representation of the strength of different frequency contributions to the initial density fluctuations when related to cosmology.

The power spectrum is related to the density fluctuations by

$$(2\pi)^3 P(\mathbf{k_1})\delta_D(\mathbf{k_1}-\mathbf{k_2}) = \left\langle \hat{f}(\mathbf{k_1})\hat{f}^\star(\mathbf{k_2})\right\rangle, \tag{14}$$

but is also often (less acurate) denoted by $P(k) = \left\langle \left|\hat{f}(\mathbf{k})\right|^2 \right\rangle$. $\lambda_k = \frac{2\pi}{k}$ as is customary; δ_D is the Dirac delta function.

The density field is coupled to the gravitational potential through the Poisson equation. The density fluctuations grew because matter began to move. The fluctuations with the highest amplitudes are generally believed to have collapsed first, though this depends on the shape of the power spectrum. (See below for more.)

The corresponding wavelength indicates what structure will form: galaxies at small wavelengths, (super-)clusters from larger wavelengths. So a power spectrum with high values at high wavenumbers means structure formation at small wavelengths first: galaxies.

The use of the continuity equation, describing mass conservation, and the Euler equation, describing the flow of matter, enables the prediction of the complete velocity field, at least in the linear regime. Fluctuations can be expected to be small for a certain limited amount of time initually, so equations can be approximated linearly.

When the density fluctuations grow larger, the linear approximations break down and non-linear effects start to kick in. Matter moves no longer only as a consequence of the global gravitational potential but also starts to influence other matter leading to chaotic effects. No exact mathematical methods have been developed yet to describe these processes, so the study of the non-linear is dependent on numerical simulations and extrapolations of linear theory into the non-linear realm as far as possible.

The noisy nature of the initial density fluctuations result in a departure from perfect homogeneity. Randomly spread small density fluctuations will grow and result in anisotropic collapse. The infall of matter will eventually be halted through the formation of a gravi-

[4] http://www.wikipedia.org

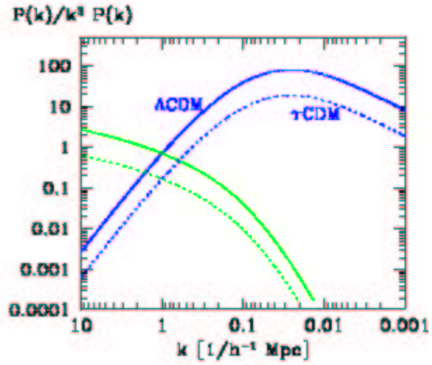


Figure 10: Shape of the power spectrum for two common varieties of the cold dark matter model. The blue lines represent $P(k)$; the green lines $k^3P(k)$. (Rien van de Weygaert, private communication.)

tationally bound and virialized object, stabilized against the gravitational pressure by the kinetic energy of its constituents.

4.3 Dark Matter Short Detour

In the review so far no reference has been made to the presumed collisionless gravitationally bound non-baryonic matter commonly known as dark matter. The expected baryonic fraction from nucleosynthesis is too low to explain the total amount of matter in the density paramater Ω_m, prompting the need for dark matter.
The two most important models that have been developed are Hot Dark Matter (HDM) and Cold Dark Matter (CDM). Hot Dark Matter depends on neutrino masses with large quasi-thermal velocities.
The most popular model is the one in which dark matter consists of arbitrarily massive particles which interact only weakly and have a thermal velocity which is effectively zero. This Cold Dark Matter model (e.g. Bond and Szalay, 1983) is integrated in the established standard cosmological model (Spergel et al., 2003). The CDM model predicts hierarchical collapse in which the smallest structures form first. Cole et al., 1998 have constructed mock redshift surveys based on the 2dF and SDSS redshift surveys. Their power spectrum follows the CDM model, which features a simpel power law at lower wavenumbers and a logarithmic fall-off at higher wavenumbers, corresponding to the fluctuations the size of galaxies. Different variations to the CDM models have been developed with varying characteristics, (see Fig. 10), like the shape of the power spectrum, different values for different constituents of Ω etc. The standard CDM model for example utilizes a flat $\Omega = 1$ universe, where the ΛCDM model has a cosmological constant which together $\Omega_0 = 0.3$ produces a flat geometry for the universe.
Enough assumptions can be made to use dark matter in numerical simulations, or just to use them as gravitating particles without discerning between conventional matter.

4.4 Simple Symmetrical Prelude to Structure

A useful first step in assessing general principles governing the evolution of density fluctuations and the forming of structure is looking at the most simple symmetrical form: the sphere. A spherical mass distribution can be described as divided into a distinct number of shells. The individual shells are characterized by their motion under the central mass distribution and their energy:

$$\ddot{r} = -\frac{GM(r)}{r^2} \tag{15}$$

$$\frac{1}{2}(\dot{r})^2 - \frac{GM}{r} = E \tag{16}$$

The assumption is made that the different shells are homogeneous, do not cross and experience no outside influences.
As the universe expands, an individual shell, depending on its mass, initially expands and then keeps expanding or collapses. The density parameter equation determining this possible collapse can be derived to be

$$\Delta(r,t) = \frac{3}{r^3}\int_0^r \delta(y,t) y^2 dy \tag{17}$$

where $r(t) = a(t)x$.
The density can be expressed in an excess density with respect to a flat universe and the kinetic energy of a shell can be assigned in a dimensionless fashion:

$$1 + \Delta_{ci} = \Omega_i \left[1 + \Delta(t_i, r_i)\right] \tag{18}$$

$$\alpha_i = \left(\frac{v_i}{H_i r_i}\right)^2 - 1 \tag{19}$$

(i for initial). An open shell has $\alpha_i > \Delta_{ci}$, a critical shell (expands along with the universe) has $\alpha_i = \Delta_{ci}$ and a closed shell has $\alpha < \Delta_{ci}$.
For such a shell a number of interesting characteristics can be determined. Following the linear extrapolated density excess $\Delta_{lin}(t)$, while keeping in mind that the shell itself may also be in the non-linear regime, it can be derived that the original tophat overdensity stops expanding and turns into collapse at $\delta_{c,turn} = 1.08$. The shell reaches its virial radius at $\delta_{c,vir} = 1.59$ and has hypothetically collapsed into a point source at $\delta_{c,coll} = 1.69$. $\Delta_{lin}(t)$ of the perturbation is related to the real perturbation through

$$\Delta_{lin}(t) = \frac{D(t)}{D(t_i)}\Delta_i, \tag{20}$$

where D(t) is the growth factor related to linear gravitational instability theory.
Of course in reality, the shells do not collapse completely as the contributions of individual particles start to become significant. The collapse is halted when the system becomes

virialized (when the kinetic energy amounts to twice the potential energy). Additionally, spherical collapse is also highly improbable. An elliptical model is slightly more realistic. In the approximation treated above, more symmetrical models can be analysed with increasing complexity. The picture that should remain from this simple model is the notion of various initial density perturbations growing and collapsing at different epochs which can be determined for each perturbation by its initial density contrast.

4.5 Growth in the Linear Regime

As has been mentioned before, the equations governing the displacement of particles treated as fluid elements are the Poisson equation, the Euler equation and the continuity equation. The linearized continuity equation is

$$\dot{\delta} = -\frac{1}{a}\nabla \cdot \mathbf{v}, \tag{21}$$

the linearized Euler equation is

$$\ddot{\boldsymbol{x}} + \frac{\dot{a}}{a}\dot{\boldsymbol{x}} = -\frac{1}{a}\nabla\phi, \tag{22}$$

for a pressureless fluid; the linearized Poisson equations is

$$\nabla^2\phi = 4\pi G a^2 \rho_0 \delta. \tag{23}$$

Combining these equations yields a differential equation for δ:

$$\ddot{\delta} + 2\frac{\dot{a}}{a}\dot{\delta} - 4\pi G\bar{\rho}\delta = 0 \tag{24}$$

This equation solves into an independent growing and decaying mode,

$$\delta(\mathbf{x}, t) = D_{\pm}(t)\delta(\mathbf{x}). \tag{25}$$

The dependence of the growing and decaying modes depend on the applied cosmological model (e.g. Einstein-de Sitter). The decaying mode is often ignored in further approximations.

The linear nature of the mass flow means that structures in the linear regime reflect the original density fluctuations giving rise to those structures. As has been mentioned before, the observed nodes in the network of filaments, the superclusters, are still in the process of accreting matter and are believed to be in the linear regime. They are therefore viable candidates for tracing the primordial mass distribution.

4.6 Zel'dovich Approximation

The description by Zel'dovich (1970) of growing density fluctuations in a linear regime takes an initial mass distribution with its associated gravitational potential and determines the

distribution at a later time directly. The method is extrapolated slightly outside of the linear regime while estimating deviations resulting from neglecting non-linear terms. This gives an accurate qualitative description, describing anisotropic collapse.
Following the Lagrangian formalism of following individual particles instead of looking at physical information at certain fixed coordinates (Eulerian approach), each particle is denominated by $\boldsymbol{q} = \boldsymbol{x}(t = 0)$ and its position is tracked with $\boldsymbol{x}(\boldsymbol{q}, t)$. This means $\boldsymbol{x}(t, \boldsymbol{q}) = \boldsymbol{q} + t \cdot \boldsymbol{v}(\boldsymbol{q})$, neglecting higher order terms. Extending this approach to a fluid element means monitoring the shape and movement of such an element together with its position and density.
The fluid elements move under the influence of the growth of the linear density perturbations and the potential Ψ.

$$\boldsymbol{x}(\boldsymbol{q}, t) = \boldsymbol{q} - D(t)\nabla\Psi(\boldsymbol{q}); \tag{26}$$

Ψ is proportional to the linearly extrapolated gravitational potential ϕ, $\Psi = \frac{2}{3Da^2\Omega H^2}\phi$.
The extrapolated gravitational potential is a crucial element of the approximation, as it allows the determination of the gravitational field directly from the initial potential and through it the evolution of the density field.
Mass-conservation requires

$$\frac{\rho(\boldsymbol{x}, t)}{\rho_0} = det \left\| \frac{\partial \boldsymbol{x}}{\partial \boldsymbol{q}} \right\|^{-1}, \tag{27}$$

which, using the eigenvalues λ stemming from the Jacobian, solves to

$$\frac{\rho(\boldsymbol{x}, t)}{\rho_0} = \frac{1}{(1 - a(t)\lambda_1)(1 - a(t)\lambda_2)(1 - a(t)\lambda_3)}. \tag{28}$$

This can be approximated to $1 + a(t)(\lambda_1 + \lambda_2 + \lambda_3)$. The eigenvalues can be ordered by magnitude. They represent the different dimensions of the fluid element. One or more positieve eigenvalues guarantees collapse at one instance or another.
With growing fluctuations b, each dimension of the element in turn collapses. This is the collapse into pancake shapes and further into filamentary shapes.
At that stage, though, the paths of the individual elements start to cross and the approximation is no longer valid, as the individual elements start to influence eachother. The result of ignoring the other elements is a structure that is less compact than it should be.
The linearly extrapolated density fluctuations evolve like

$$\delta(t) = a(t)\Sigma_m \lambda_m. \tag{29}$$

An example of the effects of the Zel'dovich approximation and the resulting structure from a random primordial matter distribution can be seen in Fig. 11.

4.7 Venturing Into the Non-Linear Realms

After extrapolating the linear approximations as far as possible into the range where they are technically no longer applicable, the problem of the dynamical evolution of large-scale

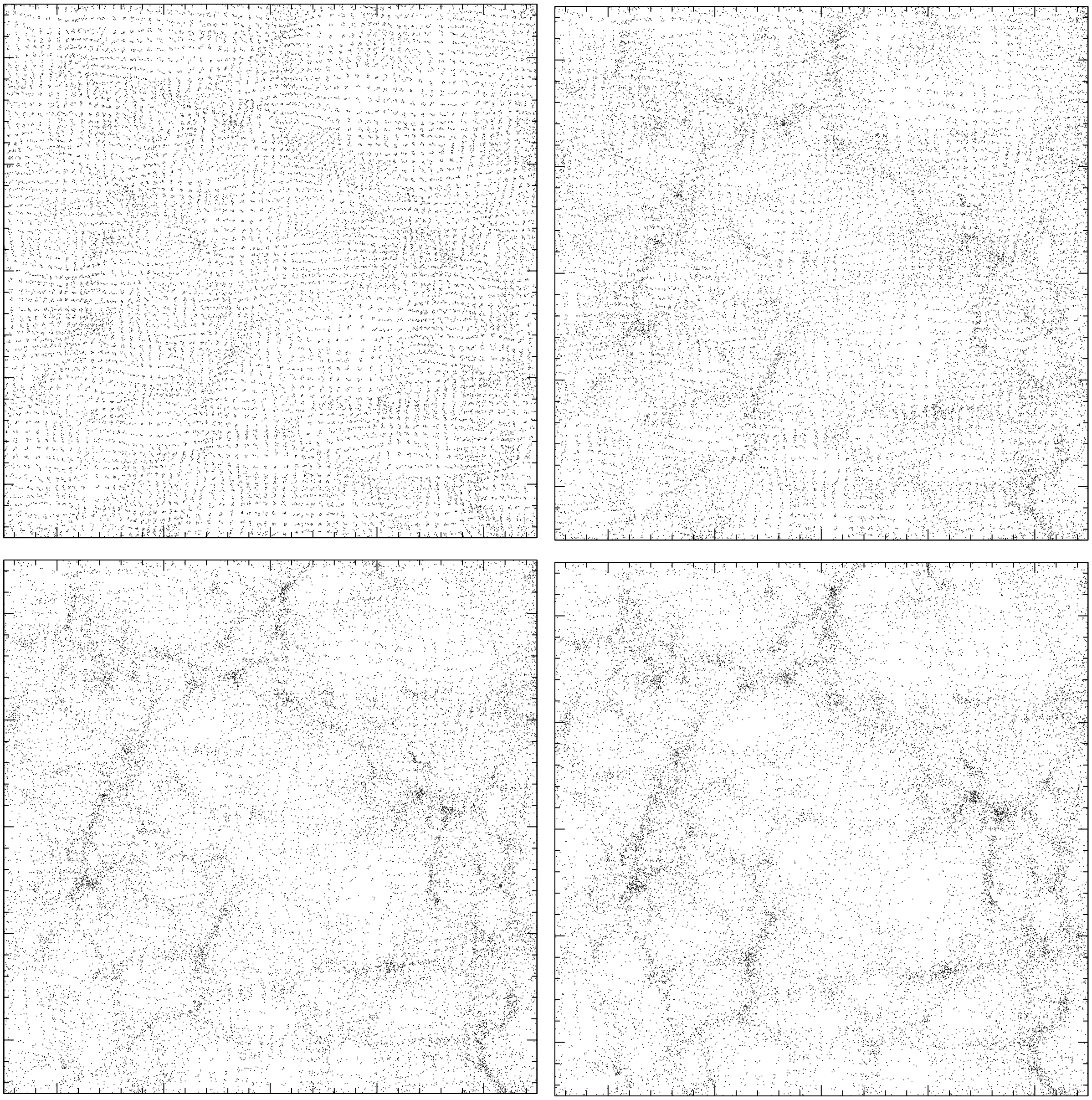


Figure 11: Zel'dovich formalism at work from top left to bottom right, showing the resulting structure from a random gaussian matter distribution. From Van de Weygaert (2002).

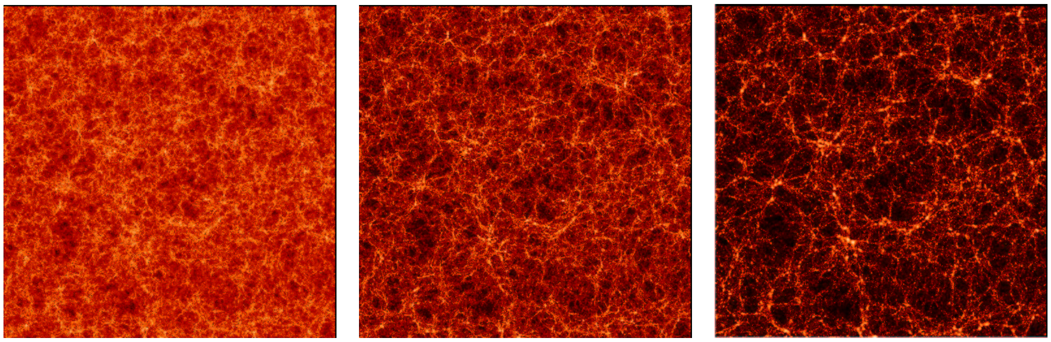


Figure 12: N-body simulation by the Virgo Collaboration: sequence of evolutions of a primordial density field in a standard cold dark matter scheme. The depicted epochs are z=3 (left panel), z=1 and z=0 (right panel). The developing structure from small initial fluctuations is clearly visible. (Image by Joerg Colberg; see Jenkins et al., 1998)

structure seems to be tractable only by doing numerical calculations. The collapse into the structures that can be observed is readily simulated using N-body tree models or kinematic models. See e.g. simulations carried out by the Virgo Collaboration[5]: Fig. 12. These simulations confirm the general notions of hierarchic clustering expected by the evolution of the power spectrum and e.g. the Zel'dovich approximation.

4.8 Formation of the Cosmic Foam Overview

Summarizing the above sections, it can be asserted that once a initial density fluctuation field has been decided on, the formation of structure proceeds described by linear fluctuation theory. Further evolution can initially be approximated by the Zel'dovich formalism, but becomes completely nonlinear eventually. The shape of the power spectrum, related to the initial density fluctuations, determines at which scales collapse starts first (hierarchic clustering). Collapse will be anisotropic, because asymmetrisch among the different axes will grow during collapse. A linear model like the Zel'dovich approximation predicts collapse into a pancake shape first, then into a filament and finally into a cluster. Underdense regions, known as voids, will get increasingly spherical during their expansion, as opposed to overdense regions. The resulting foamy structure can be described by a Voronoi tessellation once the different void 'bubbles' start to overlap, as will be dealt with in the next chapter.

[5]all simulation data available at http://www.mpa-garching.mpg.de/NumCos

5 Reproducing Large-Scale Structure

5.1 Modelling of the Galaxy Distribution through Voids

Simultaneous with the advent of the deep redshift surveys came the efforts to simulate the observed structures. Although larger scale clustering had already been suggested early on, e.g. in Icke (1973), in Aarseth et al. (1978), supported by numerical simulations and in Zel'dovich et al. (1982), it took the results of the redshift surveys to jumpstart the development of real large scale structure modelling.
Without analytical solutions to the complicated equations governing the motions of individual galaxies making up the large scale structure, the use of numerical simulations are an essential tool in analyzing the principles behind structure formation.
N-body simulations proved to be a valuable tool for realistic modelling of the large-scale structure. The discovery in the early 1980s (e.g. Davis et al., 1982) of the 10-50 Mpc regions practically devoid of galaxies, was quickly explained by models concentrating on the evolution of underdense regions. The Boötes void (Kirshner et al., 1981) was the

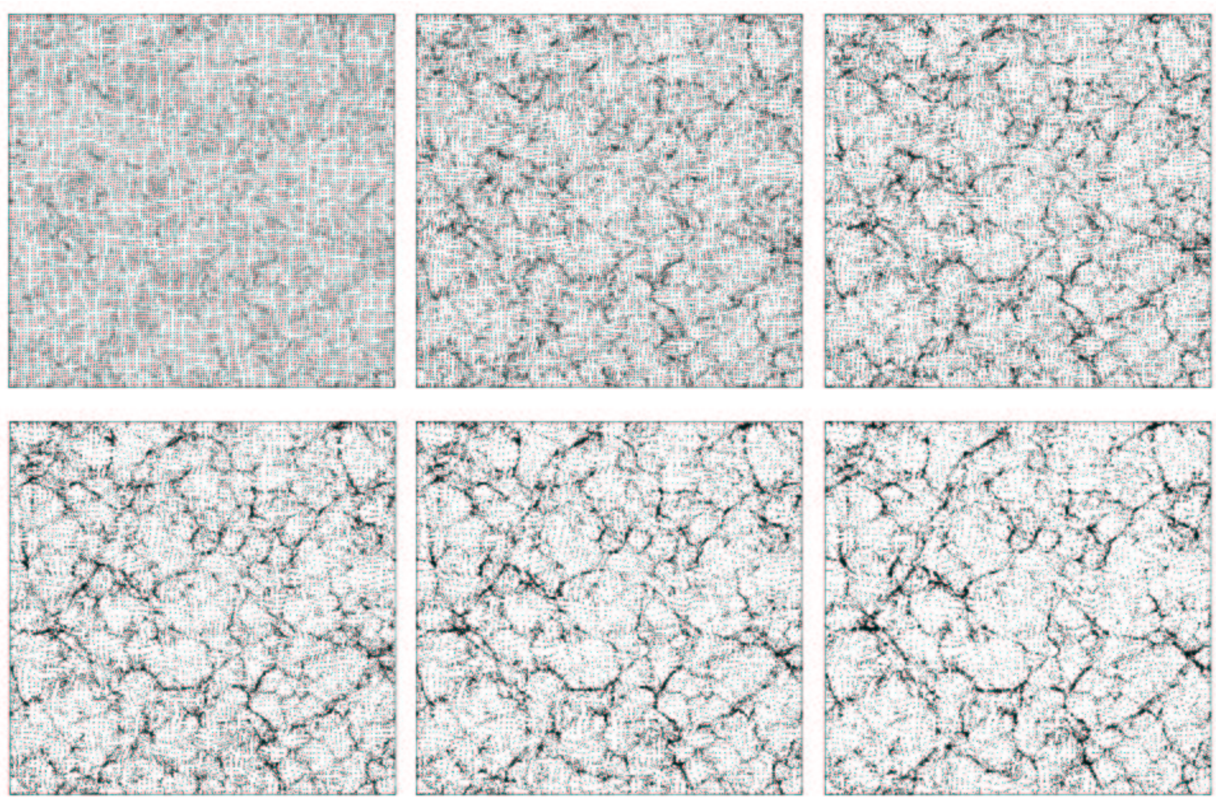


Figure 13: Sequence of N-body CDM simulations showing the evolution of voids and the formation of walls, filaments and clusters. The expansion factors are from top-left to bottom-right: $a_{exp} = 0.05, 0.1, 0.15, 0.2, 0.25, 0.3$. (1 is the present epoch.) Courtesy: Erwin Platen

first void to be discovered. The title of their article seems to sound dramatic as well as uncertain: “A Million Cubic Megaparsec Void In Boötes?”
Indeed, N-body simulations of the evolution of the primordial density field readily produce the rich structure with voids as can be seen from the slices of a three-dimensional simulation in Fig. 13. From an initial random noise field, the mass is displaced from which voids can be discerned early on. As mass streams out of the void, the filamentary structure can be seen to grow. The 'knots' in the filaments are the sites where clusters are in the process of being formed.

5.2 Void Properties

While the high-density regions will collapse into filamentary shapes, as has been illustrated with the Zel'dovich approximation, Icke (1984) showed analytically that low-density regions will become increasingly spherical. A void may be considered as a negative-density region with respect to the background universe. Voids are growing entities, effectively pusing matter outwards[6]. Matter is expected to be streaming out of voids, enhancing their negative density contrast.
From Icke (1984, and references therein) we have the equations for the evolution of an ellipsoid, which he compares with the evolution of a void:

$$-\frac{\ddot{X}}{X} = 2\pi G\rho\alpha; \quad -\frac{\ddot{Y}}{Y} = 2\pi G\rho\beta; \quad -\frac{\ddot{Z}}{Z} = 2\pi G\rho\gamma. \tag{30}$$

α, β and γ are related to the axes of an ellipsoid in a way that if $a > b > c$ then $\alpha < \beta < \gamma$, which means

$$-\frac{\ddot{X}}{X} < -\frac{\ddot{Y}}{Y} < -\frac{\ddot{Z}}{Z}. \tag{31}$$

The axial ratios $a : b : c$ increase with time and aspherities of an overdensity are continuously magnified.
The signs in the equations above reverse for an underdense region and the aspherities of the void decrease with time. The void therefore becomes increasingly spherical.
Qualitatively, when the voids have grown to such a size that adjacent voids begin to get close to each other, the overdense regions have collapsed, or are in the process of collapsing to pancake shapes, filamentary shapes and clusters. The analogy of voids pushing matter outwards is useful again, when it is considered what happens when voids start to get close to each other.
At the point where two voids are pressing matter outwards against each other, matter is being 'pressed' into the pancake shapes. These structures are short-lived, as the matter in them is being pushed outwards toward the points where multiple voids touch. The location where three voids touch, the production of filaments takes place. Finally, the spot where more voids are pressing against each other is the place where clusters will be formed.

[6] While matter appears to be pushed out of voids, they are actually regions with a lower gravitational potential, so matter is being pulled out towards the regions with a higher potential.

The focus on the importance of voids has lead to knowledge of void properties, including (see Sheth and Van de Weygaert, 2004, in preparation) expansion, evacuation and spherical shape. (Void edges will be of more irregular shape in general because of the influences of surrounding structures, Van de Weygaert and Van Kampen, 1993.) This has been mentioned, but it is worth imagining voids pointedly as expanding spheres with decreasing density. The 'disappearing' matter is a non-linear effect taking place at the void boundaries. In addition, the more technical aspects of voids are their (negative) tophat density profile resulting from the continuing outflow of matter, a constant velocity divergence comparable to the uniform expansion of the universe itself ("Super Hubble Bubbles") and suppressed structure growth. The latter would be expected intuitively, because as the void starts to resemble an underdense universe, structure freezes out. The boundary of a void becomes the place where matter accumulates, a ridge where matter is pushed into. The spherical nature of voids finally can be described by a shells. The quasi-linear stages of the evolution of voids ends with shell-crossing, when inner shells surpass outer shells and the void enters the non-linear stage.
If the voids are distributed homogeneously through space, the resulting structure would be a honeycomb structure, as studied by Hoffman et al., 1983. Voids are, however, distributed in a more random fashion and their growth into spherical objects leads eventually to a structure resembling a Voronoi tessellation (Matsuda and Shima, 1984). This is discussed in more detail in the next paragraph.

5.3 From Void to Voronoi

The advantages of using a kinematic Voronoi model over N-body simulations lies in its flexibility. N-body simulations start with a density field that evolves through the predefined physics of (but not limited to) gravitational influences. The end product can be adjusted through the initial parameters and the initial density field, but the simulations remain restricted to the known physical process influencing the evolution of the simulated structure. This poses difficulties when trying to analyze the observed structure, from which information is often available from the present epoch only. The kinematic Voronoi model simply mimicks the observed structure with a realistic framework on which galaxies can move and evolve in position. Various paramaters can be adjusted to yield different kinds of structures: different void sizes, different amounts of clustering etc. The characterization of the Voronoi tessellation will be given below.
The validity of the use of a Voronoi tessellation to characterize the observed large-scale structure can be found, as has been mentioned above, in the evolution of voids. The Voronoi model relies on voids of approximately equal sizes. This is not necessarily the case, as the initial density field is thought to consist of gaussian fluctuations, with no preferred size. Recent investigations by Sheth and Van de Weygaert (2004, in preparation) show, besides noting that voids indeed fill the universe by a certain fraction, that small voids tend to vanish in overdensities. The distribution of void sizes tends to be peaked around a characteristic void size. The characteristic void size increases through time, as can be seen in Fig. 14. The movement of matter is also clearly visible in these simulations,

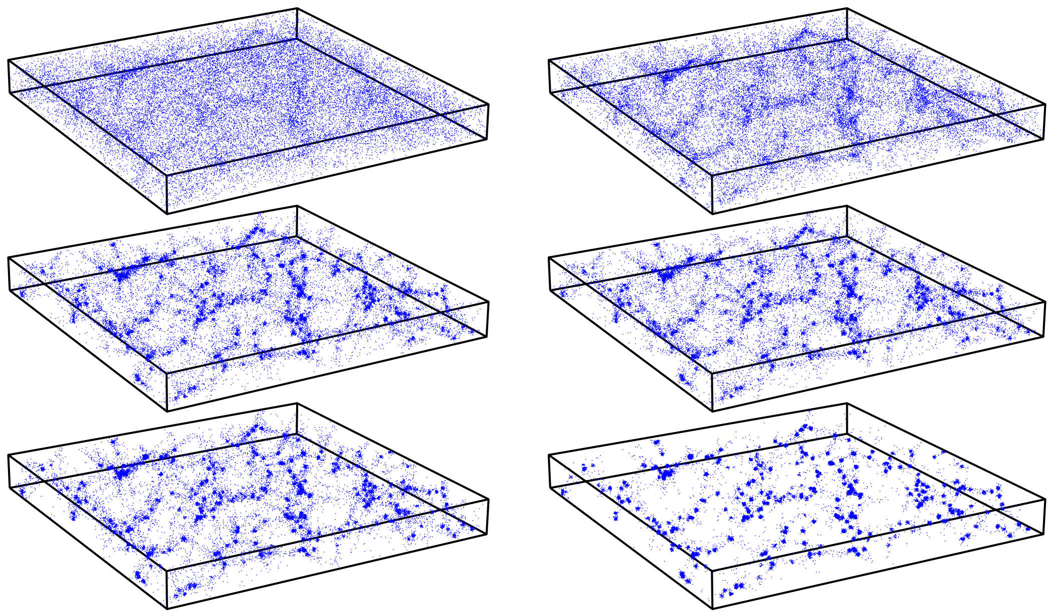


Figure 14: Slices from a kinematic Voronoi model simulation (from Van de Weygaert, 2002), showing the formation of structure in this model.

into walls, filaments and clusters.
In an extreme approximation, the peaked void size distribution can be replaced by a spiked void size distribution, in which all voids are approximately the same size. Walls and filaments will then be located exactly between the different cells and the Voronoi model is recovered (Icke and Van de Weygaert, 1987).
The justification of the idea of a Voronoi tessellation as an approximation of the observed large-scale structure combined with the flexibility associated with the building of this Voronoi framework makes it an excellent model to study for properties of the large-scale structure.
The concept of Voronoi tessellations in large- scale structure simulations has been explored in much more detail ever since; see e.g. Icke and Van de Weygaert, 1987; Van de Weygaert and Icke, 1989; Yoshikoka and Ikeuchi, 1989 and more recently Ramella et al., 2000, El-Ad and Piran, 1997, Goldwirth et al. 1995 and Zanetti, 1995.

5.4 Voronoi Foam

The Voronoi tessellation is named after the Russian mathematician G.F. Voronoi (1868 - 1908), although the ideas behind his tessellation have much deeper roots. A Voronoi tessellation is a division of an N-dimensional space into a cellular structure. A quick look at http://www.voronoi.com, dedicated to be a meeting point for people using this tessellation, reveals its usage in many disciplines. Among these are biology, chemistry, computational mechanics, forestry (modelling forest dynamics!), geophysics, hydrodynamics, medicine

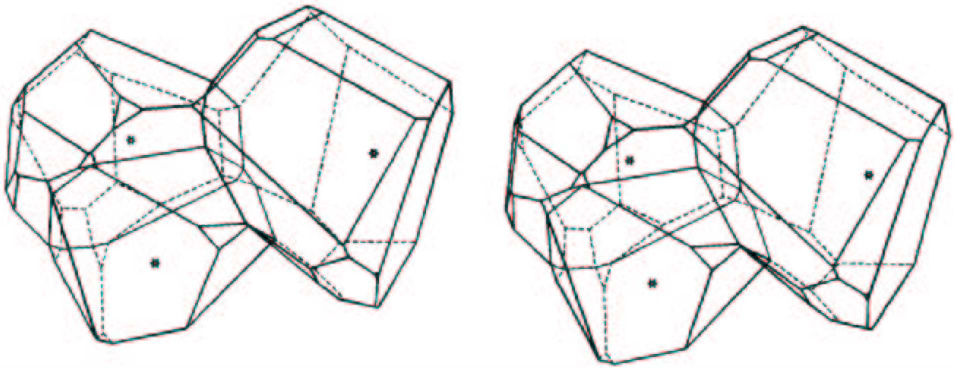


Figure 16: Stereoscopic Voronoi cell example (Van de Weygaert & Icke 1989). The inner space within the cell represents a void, the planes enclosing the cell represent walls. The lines represent filaments and the intersections of lines represent clusters when distributing galaxies in a Voronoi tessellation.

and route planning.
A Voronoi tessellation is completely determined by a discrete set of points, which can be seen as the nuclei in the cellular structure. Each cell consists of all points closer to its nucleus than any other nuclei. The borders of the cells are formed by the points that are equidistant to two or more nuclei. The resulting structure consists of convex cells, Voronoi Polygons, of which the exact definition according to Wolfram Mathworld[7] reads:

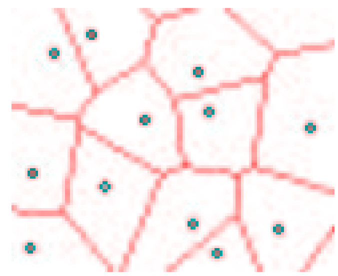


Figure 15: Simple example in two dimensions of a Voronoi tessellation. (Picture from manifold.net

"A polygon whose interior consists of all points in the plane which are closer to a particular lattice point than to any other. The generalization to n dimensions is called a Dirichlet region, Thiessen polytope, or Voronoi cell."

In the formal form of a formula, this amounts to

$$\Pi_i = \{\mathbf{x} | d(\mathbf{x}, \mathbf{x_i}) < d(\mathbf{x}, \mathbf{x_j}), \forall j \neq i\}\,, \tag{32}$$

where Π_i is the i-th Voronoi region associated with a set Φ of nuclei $\{x_i\}$ in an n-dimensional space.
In two dimensions this looks like depicted in Fig. 15.
Voronoi walls Σ_{ij} consist of all points in Π_i and Π_j that are located exactly halfway between both nuclei x_i and x_j of Pi_i and Π_j, which means

$$\Sigma_{ij} = \{\mathbf{x} | d(\mathbf{x}, \mathbf{x_i}) = d(\mathbf{x}, \mathbf{x_j}), \forall j \neq i\}\,. \tag{33}$$

We can proceed in the same way to define Voronoi edges (filaments) Λ_{ijk} and Voronoi vertices V_{ijkl}.

[7]http://mathworld.wolfram.com

A Voronoi tessellation is made from a number of nuclei. These nuclei form the center of the cells. See fig. 16. The nuclei represent the original underdense regions in the primordial density fluctuations that have expanded into voids. Galaxies are positioned in this structure as desired; the inner parts of the cells are voids. The cell-boundaries are walls (between two cells), or filaments (intersections along more than two cell), or clusters (where filaments come together). The resulting spunge-like structure is periodic in nature. In fig. 18 a slice through a Voronoi distribution is shown, which shows filaments or walls (not discernable that well in 2D) clearly, as well as the voids and clusters. Galaxies can be distributed by various techniques using increasingly realistic methods. In this work the galaxies have been distributed by defining fractions of the total amount of galaxies that are being put into the different structure parts: void, wall, filament or cluster. These fractions are chosen to produce a realistic simulation of the galaxy distribution.
Evolution of the galaxy distribution happens through the principle, that galaxies stream out of voids, into walls, then into filaments and finally into clusters. See Fig. 17.
The two-point correlation function (treated in more detail further on) is used to compare the resulting distribution with what is observed in the real universe. The Voronoi tessellation two-point correlation function has the same power-law shape on moderate radius as is observed in the universe. Software from Rien van de Weygaert has been used to create suitable Voronoi distributions.

5.5 Voronoi Kinematic Model

A kinematic Voronoi model was used in this work (Van de Weygaert and Icke, 1988), which differs from N-body simulations in the method through which a realistic structure of galaxies is generated, as has been discussed above. The simulation of a Voronoi galaxy distribution is customizable in parameters like: the number of cells in the total volume, galaxy luminosity distribution through the Schechter function parameters (see e.g. Turner

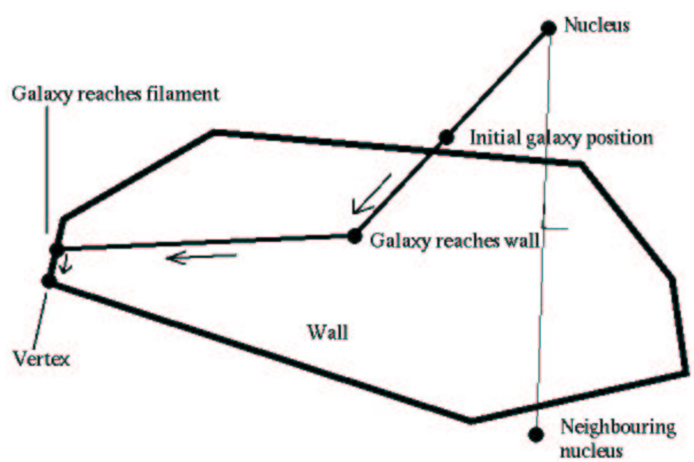


Figure 17: Movement of a galaxy through a Voronoi cell in the kinematic Voronoi model. Courtesy: Jacco Dankers.

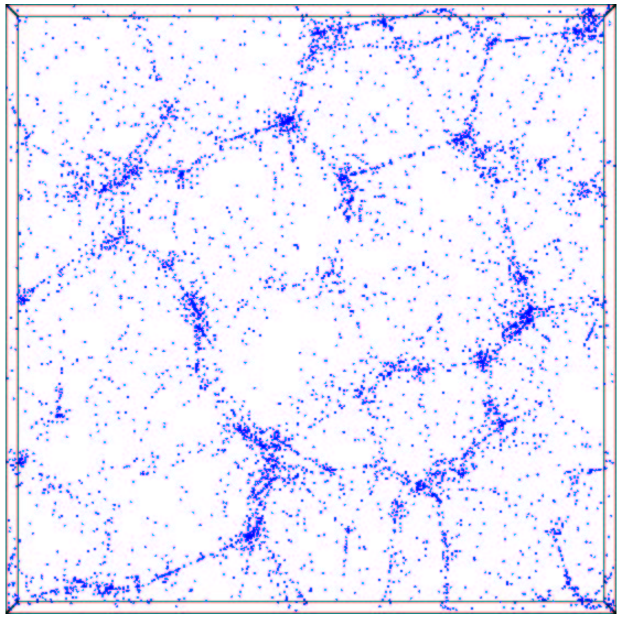


Figure 18: Slice of 100x100x20 Mpc through an example Voronoi distribution with distinct filaments and clusters. The cellular structure is also discernable very well.

and Gott, 1976 about the schecter luminosity function), number of galaxies through the Schechter function and population proportions.
Repeating the corresponding of a Voronoi cell to a void region, the cell surfaces to walls, the surface edges to filaments and the edge vertices to clusters, we have a convenient and easy to produce skeleton that can be used to stick realistic galaxy distributions on.
An approach resembling dynamical simulations, but somewhat idealized, sees a galaxy, which is positioned inside a Voronoi cell, moving to the wall of the cell, to the edge of the wall and finally into a vertex where multiple cells touch (Fig. 17). This process yields similar results as the N-body simulations, in which galaxies stream out of voids and into walls and filaments in the same way.
In slightly more detail, this works as follows. Galaxies in voids stream outwards and the mean distance between them increases uniformly (the Super-Hubble property described above). The reaching of the void boundary slows a galaxy down through dissipational processes and the gravity of the (self-gravitating) wall. The void edge is a stiff border to the approaching galaxy and it consecutively loses its velocity component perpendicular to the wall. The galaxy then proceeds along the wall towards the edge of the cell border, where it becomes a member of a filament. Eventually, galaxies aggregate in the Voronoi vertices, representing clusters. These vertices are identified with rich Abell clusters. The process is strikingly illustrated by Fig. 14, where the initially featureless distribution of 'galaxies' quickly starts to show the resulting cellular structure, all the way up to the extreme clustering stage that our own universe hasn't reached yet.
The finishing touch of this model to produce realistic galaxy distribution in a flexible way is attaching proper luminosities to the galaxies.

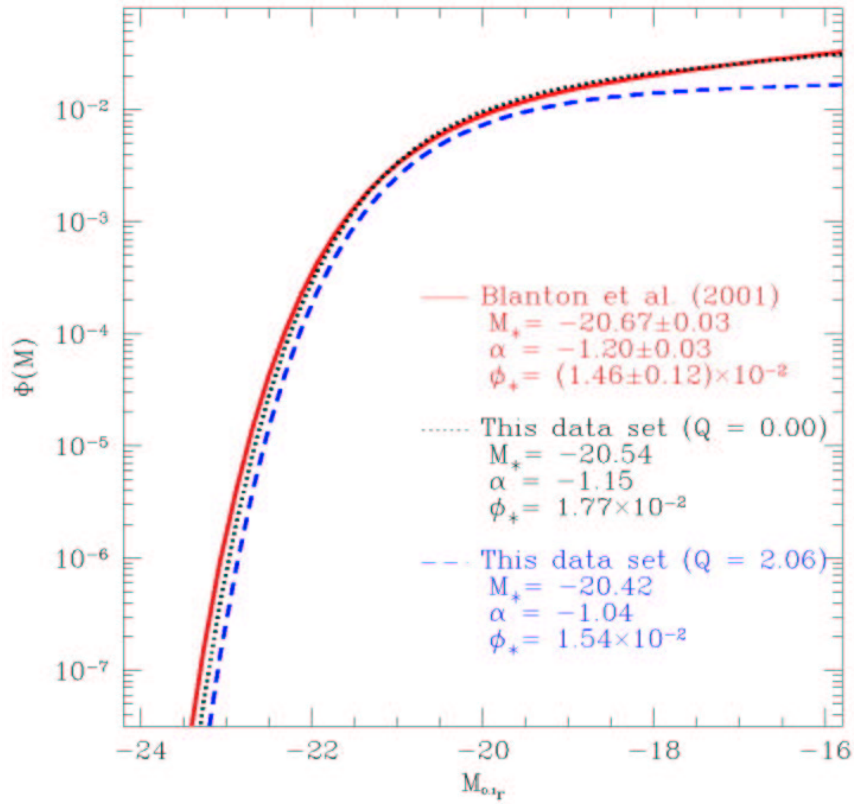


Figure 19: Galaxy luminosity function fit by a Schechter function (from Blanton et al., 2003)

To obtain realistic galaxy luminosities, the Schechter function is utilized. The Schechter function is

$$\phi(L)dL = n_\star (L/L_\star)^\alpha \exp(-L/L_\star) d(L/L_\star), \tag{34}$$

which can be rewritten to

$$\phi(M)dM = 0.4\phi_\star \ln(10)(\alpha + 1)^{10^{0.4(M_\star - M)}} \exp(-10^{0.4*(M_\star - M)}). \tag{35}$$

The Schechter function is used to fit the observed galaxy luminosity function, see Fig. 19. This function is used to obtain magnitudes for simulated galaxies and to deterimine the number of galaxies in each magnitude range. It has a steep part (the exponential cut-off) and a power-law part. The Schechter function parameters from Efstathiou, Ellis and Peterson (1988) have been used, with $H_0 = 100 kms^{-1}Mpc^{-1}$. The values are $\alpha = -1.07$, $M_\star = -19.68$ and $phi_\star = 1.5610^{-2}(h^{-1}Mpc)^{-3}$. They derive the values from redshift surveys. Galaxies are assumed to have a minimum absolute magnitude of -17 for the simulations. For example, in a cubic volume with a boxlength of 100 Mpc the number of galaxies would equal 33529, a number that has been frequently used in the simulations for this work.

The quantitative analysis of the model and the galaxy distribution can be found in correlation functions. Peebles (1978) states that for analyzing the galaxy distribution the n-point correlation functions are the natural choice. The dynamics of groups and clusters of galaxies is compared to non-ideal gases. On small scales, the two-point correlation function is approximated by a simple power law,

$$\xi = \left(\frac{r}{r_0}\right)^{-\gamma}, r_0 \approx 5h^{-1}Mpc, \tag{36}$$

for which γ is approximated to be 1.75 (Zehavi et al., 2001). Voronoi tessellations are statistical in nature and can also be described quantitatively by the two-point correlation function. This characterizes the strength of the clustering in the tessellation and provides a way to compare it with the observations.
The next chapter will explain the nature and characteristics of the two-point correlation function.

6 Two-Point Correlation Function Connects Observation and Simulation

It is generally believed that observing an arbitrary volume of the universe instead of the complete volume has no repercussions on the statistical properties that can be derived from the observations. The redshift sky surveys, for example, focus at a strip on the sky and produce a slice with galaxies. The real-space two-point correlation function that is extracted from this limited volume is assumed to represent the complete sky up to the (magnitude or volume) limit. Apparently, this has not been confirmed before. Although deviations are not expected, it is not entirely trivial to assume that no effects occur.
Simulations of large-scale structure are usually performed in a cubic (periodic) volume. Clustering properties can be analyzed using n-point correlation functions, of which the two-point correlation function is the first clustering measure. The two-point correlation function can be determined comparatively easy.
When consecutively the simulations are compared to the observations, care is being taken to minimize the effects of edge effects when estimating the two-point correlation function of the observed distribution. The estimators that are being used might be insensitve to edge effects by using random distributions to compare the observations with or employ a weighting scheme in an effort to minimize effects at the borders of the distribution.
The aim of this work has been to create a realistic simulation of a distribution of galaxies in the shape and with the characteristics that are obtained using galaxy redshift surveys. The two-point correlation function is then determined and compared to the correlation function that is obtained from the complete cubic distribution.

6.1 Intuitive Meaning of the Two-Point Correlation Function

The two-point correlation function is useful to determine strengths of patterns in distributions. As has been mentioned by e.g. Kerscher et al. (2000), the two-point correlation function has become the standard tool to study clustering in the large scale structure of the universe. The two-point correlation function is believed to be directly coupled to the initial mass power spectrum. It is a useful tool to study the results of recent galaxy redshift surveys. Large scale structure can be conveniently characterized statistically by the two-point correlation function.
The two-point correlation function $\xi(r)$ is a measure of the amount of clustering of points in a certain volume. $\xi(r)$ is an indication of the chance of finding a different amount points at a distance r than the amount that will be found in a homogeneous point distribution. A value larger than zero means more particles (overdensity) and smaller than zero means less particles than expected (underdensity) in a homogeneous point distribution.
The two-point correlation function gives an overall measure of the amount of clustering in a point distribution throughout a certain volume. These points could represent individual galaxies or mass clumps having masses in the order of galaxy mass. In this work the points in the simulations are each assumed to represent individual galaxies.

The definition of the (cosmological) two-point correlation function $\xi(r)$ is

$$dP_{12} = \bar{n}^2 \left[1 + \xi(r_{12})\right] dV_1 dV_2, \tag{37}$$

where dP_{12} is the probability of finding a particle in volume dV_1 as well as one in volume dV_2 which are at positions r and $r + dr$. $\bar{n}$ is the particle density. ξ, according to the formula, is an excess probability to finding two particles at a certain distance apart.

6.2 Estimating the Two-Point Correlation Function

A multitude of methods are available to estimate the two-point correlation function. The development of different estimators is a work in progress of its own. Pons-Bordería et al. (1999) and Kerscher et al. (2000) conveniently summarize the different estimators for ξ, with each estimator's merits and drawbacks. Often, a homogeneous random sample is used in the estimation process to compare with the dataset from which the correlation function is to be determined. This allows datasets of arbitrary spatial shapes, which is important to this work.

The estimator that is most commonly used is the one by Davis and Peebles (1983), using a random distribution in addition to the distribution that is used to estimate the correlation function from. Its widespread use is the reason why this estimator is known as the standard estimator. It is defined by

$$\hat{\xi}_{DP} = \frac{N_{rd}}{N}\frac{DD(r)}{DR(r)} - 1. \tag{38}$$

$DD(r)$ is defined as the number of particles in the data set that are within a distance $r \pm \delta$ of each other. $DR(r)$ is defined in the same way, only now pairs of particles are being considerd that has one particle from the dataset and one from the random set. N and N_{rd} are the number of particles in the dataset and the random set.

Another estimator is the one by Hamilton (1993):

$$\hat{\xi}_{Ham} = \frac{DD(r) \cdot RR(r)}{DR(r)^2} - 1 \tag{39}$$

with a contribution from the number of pairs $RR(r)$ in the random sample. The Hamilton estimator has a second order uncertainty in the mean density, whereas the standard estimator has first order uncertainties.

In the same year that the Hamilton estimator was proposed, Landy and Szalay (1993) proposed an alternative one:

$$\hat{\xi}_{LS} = 1 + \frac{DD(r)}{RR(r)}\left(\frac{N_{rd}}{N}\right)^2 - 2\frac{DR(r)}{RR(r)}\frac{N_{rd}}{N}. \tag{40}$$

This estimator is said by its conceivers to perform similar to the Hamilton estimator except for a small bias.

The estimator that has been used in this work is the standard estimator, for its simplicity and adaptability, especially with respect to the selection of shapes and magnitudes. The usefulness of this estimator lies in its reduced sensitivity to border effects.(See e.g. Martinez and Saar, 2001.) At the borders the distribution suddenly stops, which, if unaccounted for, would result in large measured underdensity. This results in unrealistic two-point correlation functions.

The number of points in the homogeneous random sample is taken to be twice the size of the dataset to be examined. Martinez and Saar (2001) recommend using ten times the number of points of the dataset, but using less points did not influence the results qualitatively.

The homogeneous samples have been checked for having indeed a (near-)zero two-point correlation function. Although the correlation function of the produced samples in general appear to be lying slightly below the axis, the error bars estimated from the uncertainties in the correlation function reach out over the axis, making the deviation insignificant. An example of the two-point correlation functions resulting from the homogeneous galaxy samples can be found in Fig. 30(a), later on.

The code for creating the homogeneous samples can be found in Appendix D.

For low radii the correlation function behaves like a power law (e.g. Totsuji and Kihara, 1969; Peebles, 1978). At a radius comparable to the size of a typical cell, the correlation function will drop below this power law. At a radius near zero the power law approximation also breaks down because of influences from distinct clusters. The two-point correlation function behaves like

$$\xi(r) \sim \left(\frac{r}{r_0}\right)^{-\gamma} \tag{41}$$

in this range. A lot of effort has been made over the years to determine the value of this γ for the large-scale structure of galaxies in the universe. Peebles (1978) for example concludes $\gamma = 1.77 \pm 0.04$ for scales from $100h^{-1}kpc$ up to $10h^{-1}Mpc$.

Hawkins et al. (2003), from the 2dF survey, find a value of 1.67 ± 0.03.

The application of this power law fit to clustering at different scales (galaxies and clusters) can be expected from the initial power spectrum: the gaussian noise has no preferred scale (e.g. Colombi et al., 1996).

6.3 Averaged Two-Point Correlation Function

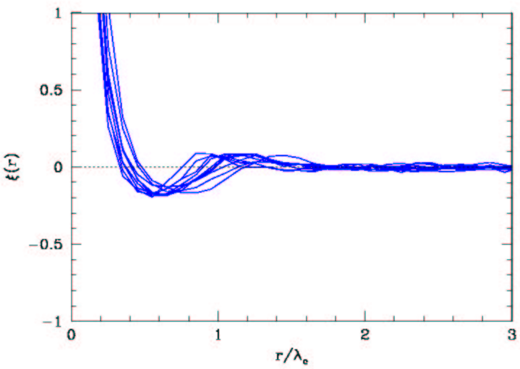
The resulting correlation functions are averaged to estimate uncertainties involved in the determination of the two-point correlation function .

Averaging occurs in an unweighted fashion,

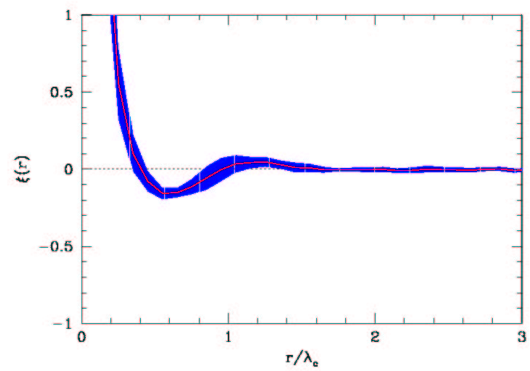
$$\bar{\xi} = \frac{1}{N}\sum_{i=1}^{N} \xi_i. \tag{42}$$

The sigma-deviation, to estimate the uncertainty in the determination, is obtained in the usual way:

$$\sigma_i = \sqrt{\frac{\left(\xi_i - \bar{\xi}_i\right)^2}{N-1}}. \tag{43}$$



(a) Various two-point correlation functions



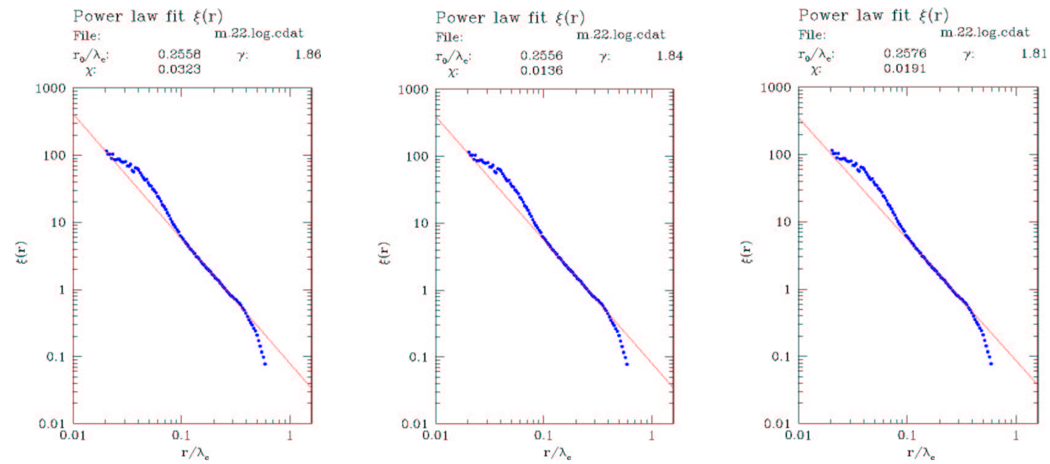
(b) Averaged two-point correlation function with one-sigma region

Figure 20: On the left are some two-point correlation functions, on the right is the averaged function. Note that the averaged function seems to imply smooth and identical behaviour of al the constituents, which is not necessarily the case.

Using the average correlation function and its sigma-deviation the uncertainties in the correlation function estimage are presented in a more insightful manner.
As an example, see Fig. 20, where the correlation functions making up an averaged function and its one-sigma envelope can be seen.
One has to be careful not to loose characteristics by averaging them out. However, the one-sigma envelope does characterizes the range of values for the different two-point correlation functions well.

6.4 Estimation Accuracy

The making of different realisations, the producing of simulations with the same parameters (but different point locations because of the random distribution of particles), yields a generic precision in the estimation of the two-point correlation function.
As the part of the two point correlation function which is conforming to a power law needs to be determined by eye, the value of gamma is prone to extra uncertainty. The expected error in gamma is about 0.02 for a normal fit. To illustrate the additional error (which is not additive though) some fits have been produced that show how much gamma can vary while at the same time produce equally acceptable fits (Fig. 21). As can be seen in the figures the value of gamma in this example ranges from 1.81 to 1.86, which is slightly larger then the 0.02 expected error.
In Davis and Peebles (1983) the error in γ is estimated by eye to be about 0.04 (on the basis of maximum and minimum slopes) and estimated to be about two standard deviations.
The value of χ^2 has been calculated between the two points through which the fitting line



(a) (b) (c)

Figure 21: Different power law fits to the two-point correlation function illustrating the fit uncertainties: γ ranges over about 0.05 with the different fits.

has been made. A lower χ^2 therefore does not automatically means a better fit.
The value of r_0/λ_c is less sensitive to the fitting process. An uncertainty in the order of 0.005 seems justified. With a cellradius of 4.64 (usually, with 100 nuclei, 100Mpc box), the error in r_0 would amount to 0.02 h^{-1} Mpc. Davis and Peebles (1983) conclude an error 0.3, which is one magnitude higher.
Deviations larger then the mentioned approximate value of 0.04 in γ will be expected to be due to uncertainties in the quantity that is being observed.
Uncertainties in the linear range can be estimated from translating all galaxies, in addition to using the averaged two-point correlation function. This should be possible because the periodic Voronoi distribution does not loose any intrinsic information when translating all its particles into another coordinate system.

6.5 Measures of the Two-Point Correlation Function

The two-point correlation function can be characterized by a clustering length r_0 and a correlation length r_a.
The clustering length is defined as the radius at which the correlation function for the first time is equal to 1 (taking an increasing radius). It is useful for the logarithmic plot of the correlation function.
The correlation length is defined as the radius at which the correlation function is 0 for the first time, again as seen from the origin. This is useful to use with the linear plot of the correlation function.

7 Preparations

7.1 Creating the Simulated Datasets

Different attempts at producing a realistic Voronoi tessellation have been made, using a variety of simulation options. As the aim of this work is to investigate the effects of shape selection, producing Voronoi tessellations as nearly as possible matching the observed galaxy distributions was given a lower priority. See Fig. 22 for a few different realizations of a Voronoi tessellations with identical parameters for the amount of galaxies and the distribution into voids, filaments, walls and clusters. Many more were created for use with the different shape and magnitude selections.
A suitable amount of nuclei for the Voronoi tessellation was chosen, together with standard values for the galaxy luminosity function. An amount of galaxies could be chosen by hand or determined from the galaxy luminosity function. To generate the desired structure, a scheme was chosen to specify the relative amount of galaxies in each structural type: field, wall, filament or cluster.
See Appendix A for an extended description of the program used to create these simulations. Appendix D shows the code that has been used to carry out the manipulations outlined below.

7.2 Translation and Rotation

The basis of flexible shape selection in simulations lies in the ability to choose a suitable center of observation and the possibility to choose the direction of observation. All manipulations are are trivial mathematical issues, but require some additional programming to implement. Both translation and rotation are applied to the coordinates if required to achieve the desired effects.
Translation is used to put the origin of the sample in different locations within the sam-

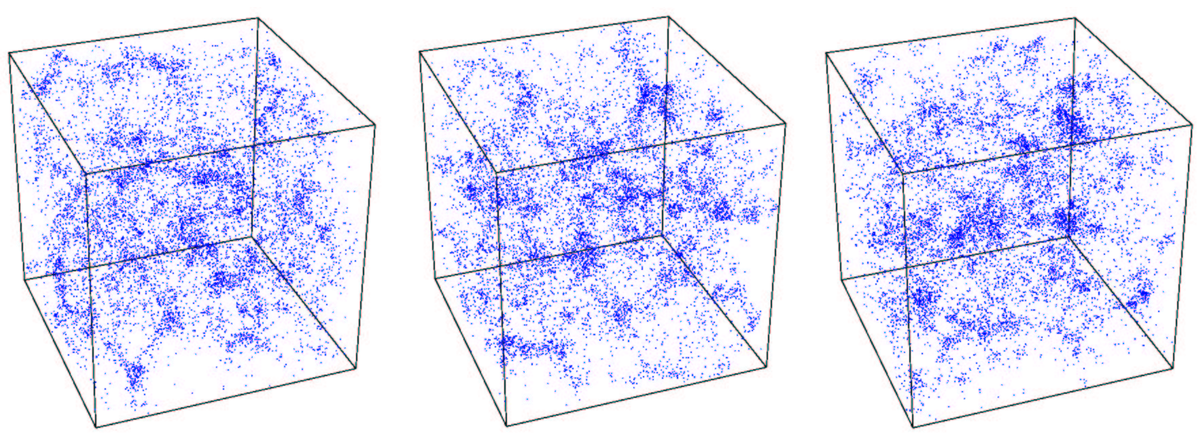


Figure 22: Examples of Voronoi distributions of galaxies with the same model parameters.

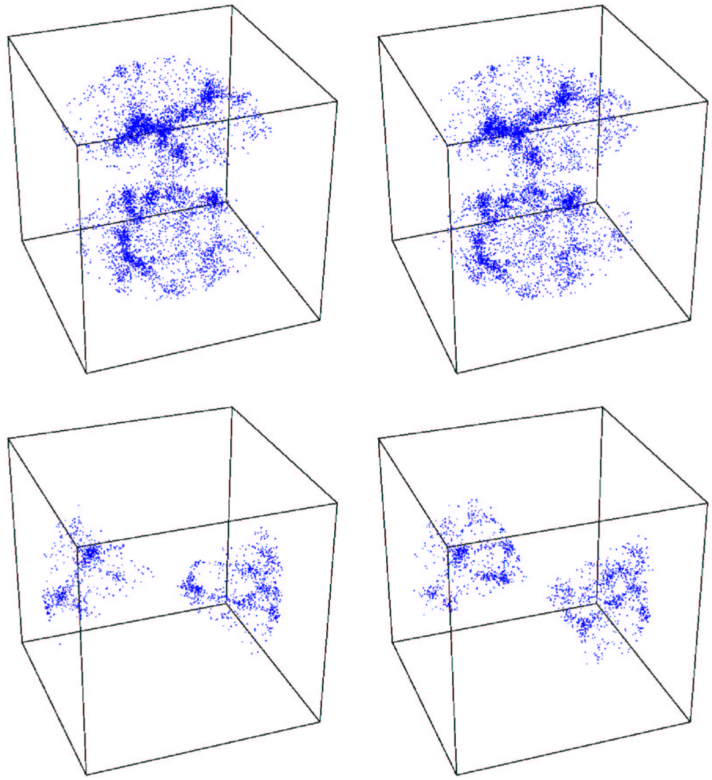


Figure 23: Examples of translation of the origin and rotation of the shapes used for selection. Upper left panel: double cone. Upper right panel: double cone shifted 7 Mpc in the x-direction. Lower left panel: double slice rotated 30 degrees (ψ). Lower right panel: double slice rotated 50 degrees (ψ).

pleset. In this case the coordinates of the samplepoints are all translated to define a new origin. The translation process assumes the already mentioned periodicity of the constructed Voronoi tessellation, so translated sample particles do not leave the dataset.
To observe in different direction one can either rotate all particles before selection or incorporate rotation into the selection formula. The effect is in both cases the same and both methods have been used depending on which method was more convenient in programming. Fig. 23 shows examples of translation and rotation, with values chosen to keep the simulated structures recognizable. (Distances are given in Mpc, assuming $h = 1$ for convenience. The value of h is of no importance to this work as all simulated structures remain static.)

7.3 Selection

Sphere selection
The sphere selection (Fig. 24, left panel) is used to set a limiting depth to the simulated survey, as the simulated survey is always made in cubical, periodical form.

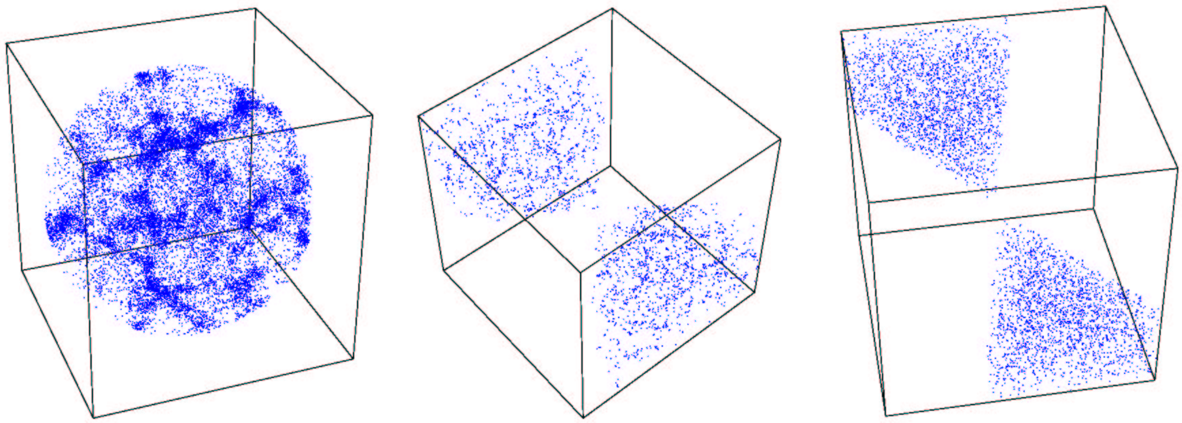


Figure 24: Examples of shape selection from the cubic dataset. From left to right: sphere selection, double cone selection and double slice selection.

Sphere selection is done by the commonly known formula

$$x^2 + y^2 + z^2 < r^2 \tag{44}$$

to select a sphere with radius r from the sample. No orientation is required for a sphere.

Cone selection
The cone selection is used to simulate looking at a spherical patch of the sky. A sphere selection can be applied first to complete this analogy.
Cone selection is basically done with the formula

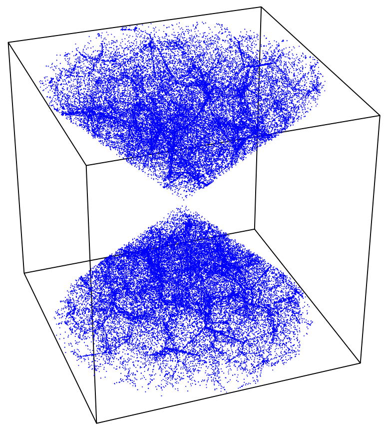
$$x^2 + y^2 = C \cdot z^2, \tag{45}$$

where C is a constant equaling

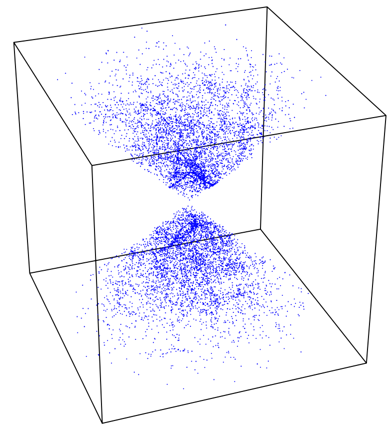
$$C = \tan \frac{\alpha}{2}^2. \tag{46}$$

α is the opening angle. Particles are in the cone if $x^2 + y^2 < C \cdot z^2$.
The orientation of the cone is controlled by prerotating the particles in theta and phi direction.
A double cone is selected to simulate looking in two directions (e.g. northern and southern hemisphere), see the middle panel of Fig. 24.

Slice selection
The selection of the slice shape is accomplished by projecting each particle onto four planes. The four planes together enclose the slice, leaving the slice open in the plane opposite to the point where the four planes intersect. This is equivalent to looking at a rectangular patch of the sky, for example with square photographic plates. See the right panel of Fig. 24.



(a) Cone selection, no magnitude selection



(b) Cone selection with limiting magnitude 15.

Figure 25: Effect of magnitude selection double cone selection. The amount of galaxies decreases from 70079 to 9354 in this selection with a limiting magnitude of 15.

The most comprehensive way to project the particles onto each plane is to rotate the x-y plane to one of the four enclosing planes. The slice has its tip on the origin, and points along the x-axis.
The first plane is rotated by 90 degrees clockwise around the x-axis and then rotated around the original z-axis by half the requested first opening angle counterclockwise.
The second plane is rotated the same, except by half the first opening angle clockwise.
The third plane is rotated around the x-axis by half the second opening angle clockwise and then rotated around the original z-axis by 90 degrees counterclockwise.
The fourth plane is rotated in the same way, except that it is rotated by half the second opening angle counterclockwise.
Now it can be checked whether each particle is within this shape. To accomplish the projection, each particle is inverse-rotated by the mentioned angles for each of the four plane. Then it can simply be seen from the z-coordinate (being smaller or greater than zero) whether the particle is above or below the plane. Combining the projections for the four planes yields whether the particle is within or not within the slice.
Rotation of the complete slice shape is done by inverse rotating each particle through euler angles theta, phi and psi. It is like moving the particles in front of the lens, the lens being the slice, instead of moving the lens to view the particles. This is being done before it is being determined whether each particle is within the slice or not.
This way the slice can be oriented in any direction.

To test viewing in many different directions, a regular grid of 100 Mpc boxlength has been used to select double slices in random directions. The directions used can be found below in the specific survey simulation results: Observing in Arbitrary Directions. The amount of particles that were selected fluctuated only by 0.7 percent.

Magnitude Selection

Magnitude selection is crucial in the simulation process, as all raw survey data is magnitude limited. Each galaxy in the simulated survey is assigned a realistic absolute magnitude using a Schechter function. The relative magnitude is determined with distance information only. Other (systematic) effects are ignored.

Default survey magnitude limit has been chosen to cope with the comparatively slow falloff of galaxies within the standard simulation volume ($100Mpc$). Shown in Fig. 25 however, is a sample with a size of $200Mpc$ to show the fall-off more clearly.

In Fig. 26 the amount of galaxies per shell is shown after a magnitude selection, compared to a theoretical Schechter function. A standard Kolmogorov-Smirnov test confirmed the functions indeed to be originating form a similar parent pouplation, but the fit can also qualitatively seen to be correct. The standard values parameters for the schechter function have been taken: $\phi_* = 1.5610^{-2}(h^{-1}Mpc)^{-3}$, $\alpha = -1.07$, $M_* = -19.68 + 5\log h$ and $M_{min} = -17.00 + 5\log h$. The logs are taken to be zero since h is taken to be 1 for convenience.

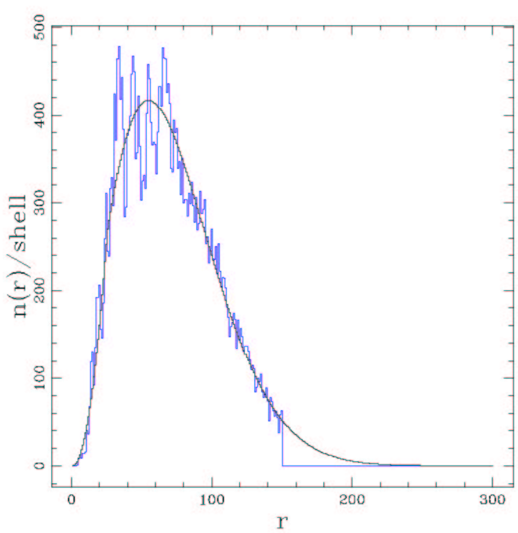


Figure 26: The theoretical galaxy luminosity function (Schechter) with the result a magnitude limited homogeneous sample with realistic luminosities. The 150 Mpc cut-off is the result of a spherical selection with that radius. The number of galaxies per shell is plotted against the radius from the center in Mpc.

8 Results

8.1 Manipulation Effects on the Two-Point Correlation Function

The effects of manipulation on a Voronoi distribution of galaxy will be illustrated using a cubic sample of 33529 galaxies with 100 cells. The galaxy distribution is shown in Fig. 28. 10% of the galaxies were put in voids, 20% in filaments, 30% in walls and 40% in clusters. It has a power law γ of 2.10. In general, it can be said that r_a and r_0 vary slightly: 0.05 for both. Additional results are shown for the selection of slices and the selection with limiting magnitude.

8.1.1 Translation

Attention is focussed first on the effects of translation. The Voronoi foam is periodic in nature. In principle, this means that any translation in an arbitrary direction should have no repercussions on the two-point correlation function. Indeed, the differences are very small, as can be seen from the black and red lines in Fig. 27(a). The two-point correlation function varies slightly in amplitude and the point where it cuts through the axis (r_a). The differences are even smaller on the logarithmic scale (Fig. 27(b)).
The chosen translation for this plot is representative for the effect on the two-point correlation function after translation. The dataset was translated over (20,30,40).

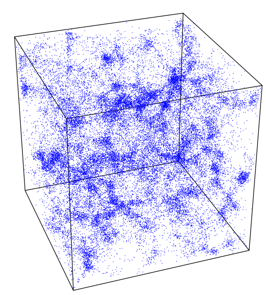


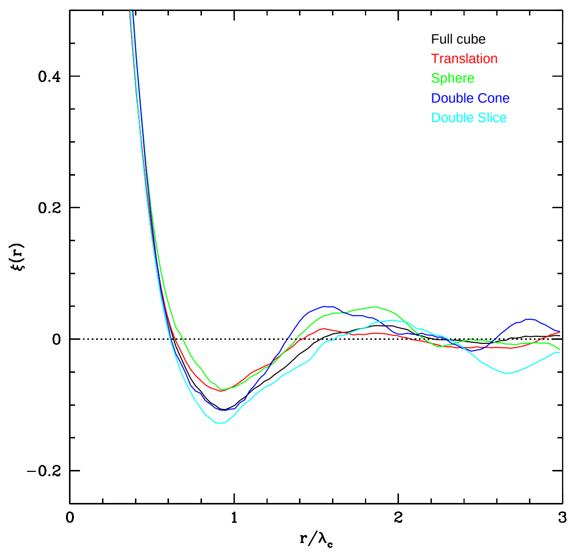
Figure 28: Section default simulation.

8.1.2 Spherical Shape Selection

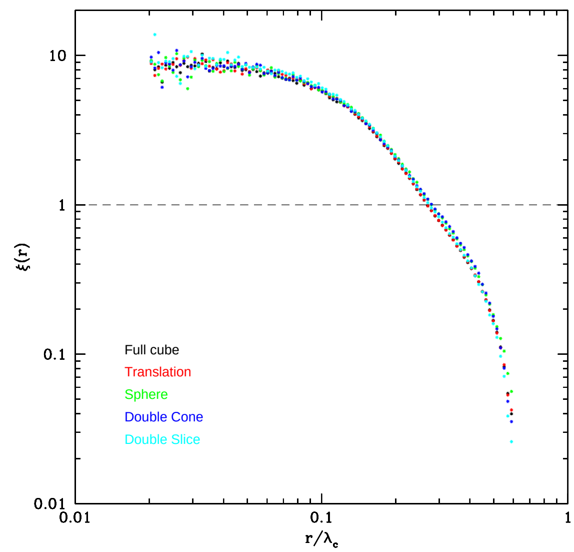
The possible effects of shape selection are considered next.
Fluctuations in the estimate of the two-point correlation function with the selection of a spherical region are comparable to fluctuations due to translations An example of the result of a sphere selection with a radius equal to half the boxlength of the original cube is plotted in green in Fig. 27(a). After selection 17412 out of the original 33529 galaxies remained.
On the logarithmic scale the differences are also small. An estimation of the uncertainties showed the two depicted curves in 27(b) shows the curve of the full cubic distribution and the curve from the sphere selection overlap within their estimated uncertainties. The power law γ estimate of the Voronoi cube is 2.10; the sphere selection has $\gamma = 2.22$, but is easily matched with the γ of the cube with a slight increase in the χ^2 goodness of fit.

8.1.3 Cone Selection

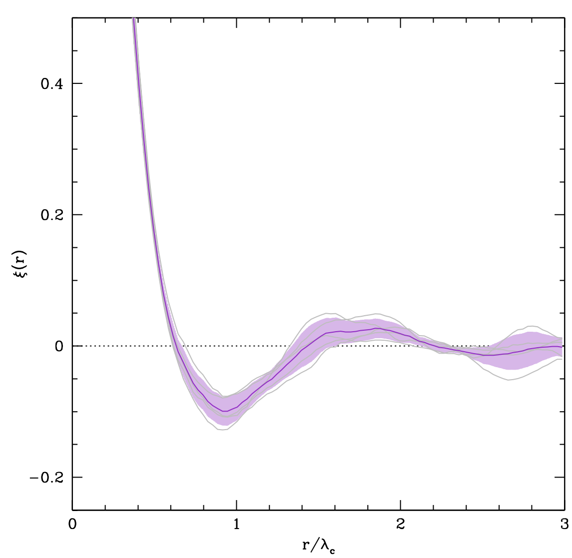
The effect of the selection of a cone out of a cube can be seen by a representative double cone selection (dark blue) in Fig. 27(a). Its opening angle is 120 degrees; which leaves 18388 particles of the cube's original 33529. The resulting two-point correlation function seems



(a) Linear range two-point correlation function manipulation effects.



(b) Logarthmic range two-point correlation function manipulation effects.



(c) Averaged two-point correlation function over the different manipulations, showing an indication of the variation in the two-point corelation function resulting from different selections.

Figure 27: Two-point correlation function manipulation effects. The different manipulation effects are indicated in the figures by color.

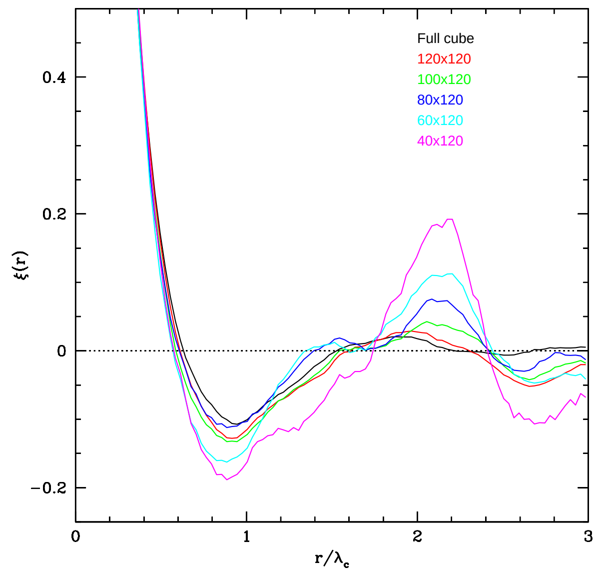


Figure 29: Effects of a double slice selection with decreasing opening angle (flattening wide shape) in the linear range.

to fluctuate more at the end of the plotted range, but the uncertainties in the estimations are ever increasing towards larger radii. Deviations from the two-point correlation function of the original dataset are no more than 0.05.
The power law γ's have a value of 2.14 and 2.17 for the cube and the cone respectively. The cone selection in the logarithmic range has been plotted in Fig. 27(b).

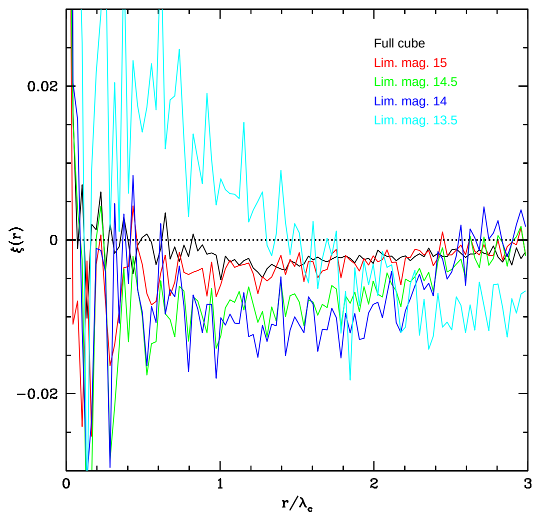
8.1.4 Slice Selection

The light blue color in Fig. 27(a) is the two-point correlation function of a representative double slice with an opening angle of 120 by 120 degrees. Differences are comparable to the other selections. This is also true for the logarithmic range.
The power-law γ's are close again, respectively 2.08 and 2.13, as can bee seen in Fig. 27(b)

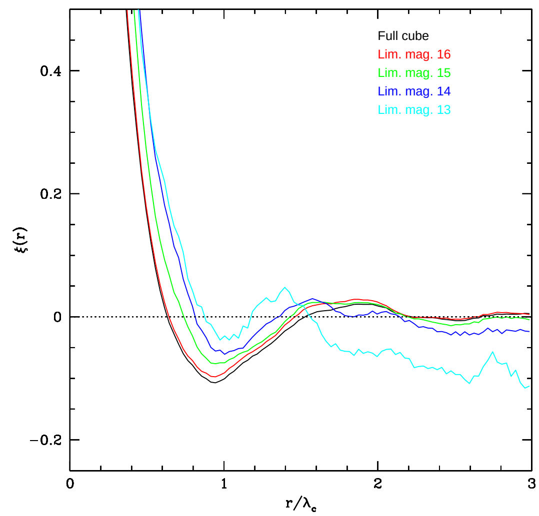
In addition to the double slice mentioned above, Fig. 29 shows the two-point correlation function of a double slice with opening angles 120 by 120 degrees initially, which gets increasingly flat until an opening angle of 40 degrees has been reached. The smallest double slice has 4895 galaxies.
Up to an opening angle of about 80 degrees little dispersion can be seen, especially in the correlation length r_a. The amplitude of the two-point correlation function increases noticably with decreasing opening angle, which indicates stronger clustering. The logarithmic range (not shown) features no changes with opening angle; not in clustering length r_0 and not in power law γ.

8.1.5 Magnitude Selection

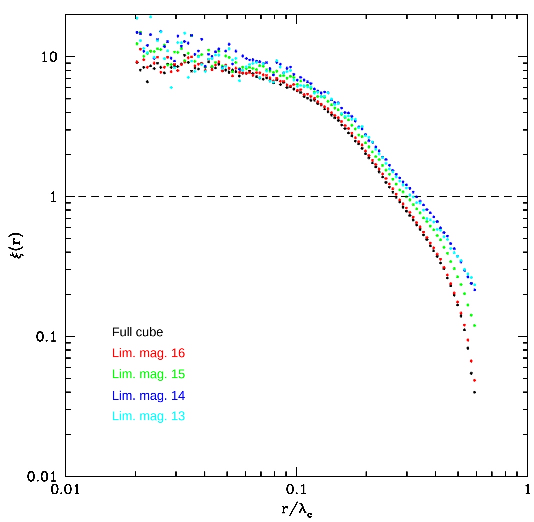
After shape selection, the last of applied selections is the magnitude selection. The selection on a limiting apparant magnitude seems to influence the two-point correlation function, which is why a separate plot has been made for the magnitude selection effects.
The effects on the two-point correlation function for a homogeneous dataset have been plotted in Fig. 30(a). The two-point correlation function for the full cube (black) is near zero, as it should be, though it is positioned 0.002 below the axis on average. The change in the two-point correlation function due to increasingly strict magnitude selections is shown by magnitude selections of 15, 14.5, 14 and 13.5. 3948 out of the original 30000 galaxies homogeneous sample remained for the magnitude selection of 13.5. At first, the correlation function hardly changes (red), but it seems to drop in the middle for selection with limiting magnitudes of 14.5 (blue) and 14 (light blue). Finally, the selection with a limiting magnitude of 13.5 shows a completely different two-point correlation function. Although the amplitude is still comparably small, this one starts above the axis and declines until it ends below it.
The plot of the magnitude selections on the Voronoi distribution (Fig. 28) is shown in Fig. 30(b). The effects are visible in this plot as well, although slightly different from the effects on the homogeneous sample. Instead of assuming lower values with increasinly strict magnitude selection, the values of the two-point correlation function are rising, together with a clearly visible increase in correlation length r_a. The logarithmic range is plotted in Fig. 30(c). The clustering length r_0 increases about 0.05 over the magnitude range, but the power law γ does not change.
The plotted sequence of magnitude selections seems to imply a departure from the regular shape of the two-point correlation function with a strong magnitude selection. This means that many galaxies are rejected because they are to faint. This should be countered with a similar rejection rate in the homogeneous sample, so it is not quite clear why any effect should appear. Strong magnitude selection can be compared to a very narrow shape selection and one would have expected extra noise in the estimate of the two-point correlation function, which was found, but the shape at the strictest magnitude selection was not anticipated.



(a) Linear range two-point correlation function for a homogeneous sample with magnitude selections.



(b) Linear range two-point correlation function for a Voronoi sample.



(c) Logarithmic range two-point correlation function.

Figure 30: The effects on the two-point correlation of magnitude selection has been plotted in these figures: for a homogeneous sample and the Voronoi sample that has been used before (Fig. 28).

8.2 Specific Survey Simulation Results

To investigate the influence of limiting effects on surveys, simulations have been carried out with different 'realistic' selections: varying opening angles, different viewpoints, arbitrary lines of sight.

8.2.1 Increasing Opening Angle in a Double Slice Shape

From a simulation cube of 268235 galaxies with a boxlength of 200 h^{-1} Mpc, a double slice was selected with the first opening angle of 10 degrees and the second opening angle varying. The opening angle varied from 80 degrees to 160 degrees, with steps of 10 degrees. In addition, a magnitude limit of 15 was imposed together with an artificial distance limit of 100 h^{-1} (a spherical selection). 32618 galaxies remained after magnitude selection on the cube. 28972 remained after the sphere selection, so only about 10 percent of the were lost due to the artificial distance limit. The narrow double slices contained about 1000 to 1500 galaxies. An example double slice is shown in Fig. 31.

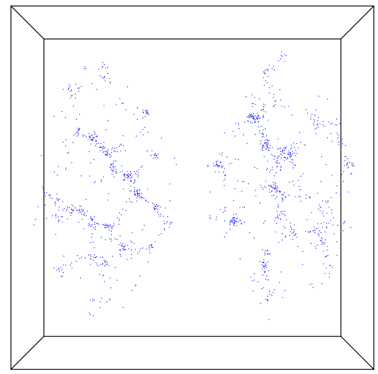


Figure 31: Double slice simulated survey.

The results are shown in Fig. 32, in which the averaged correlation function is plotted over all opening angles. As can be seen from this averaged function and its error estimate, the varying of opening angle does not really influence the correlation function.

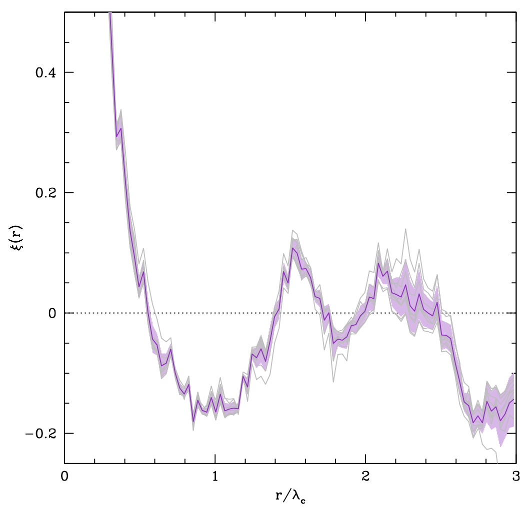
Determining the clustering length and correlation length of the different samples reveals little variation in its values at all, as is shown clearly in Fig. 32(b). The clustering length r_0 hovers at about 0.23 with an uncertainty of roughly 0.05. The correlation length at approximately 0.535 with an uncertainty in the order of 0.1.

Comparing this outcome with the effect of plain slice selection indicates an increased amplitude for a flat slice, but varying the second opening angle of such a flat slice does not seem to influence the correlation function any further.

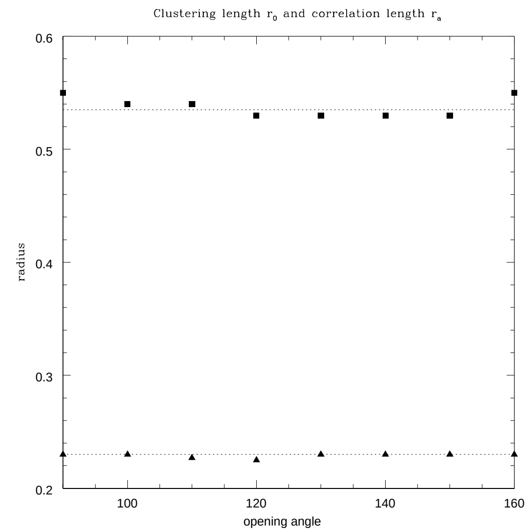
8.2.2 Observing Within Different Structures

Combining magnitude selection with sphere selection and slice or cone selection produces a simulated survey with which, combined with translating the origin to different positions in the sample, observing different viewpoints can be mimicked. Only magnitude and sphere selection have been applied to the sample of Fig. 28 here for simplicity. Different points within a Voronoi distribution have been used to deterimine the correlation function: from a field point and from a cluster point, each ten times.

Four example Aitoff-projections have been made in Figures 33 and 34 to show what the sky looks like from within a cluster and from within a void, out to different radii. It can be seen clearly that the galaxy density is lower in a void and that little structure can be seen in the distribution of galaxies. On the other hand, from within a cluster, the galaxy



(a) Averaged correlation function for the varying opening angles in the linear range.



(b) Plot of the clustering length r_0 (triangles) and the correlation length r_a (squares) due to a changing opening angle of a double slice.

Figure 32: Two-point correlation functions resulting from varying the opening angle of a flat double slice.

density is larger and some clusterings can be seen. The cluster in the middle-south part of the plot can be seen on both 10 and 20 h^{-1} Mpc distance, so this structure is close. The one in the upper-left corner is hardly visible in the plot up to 10 h^{-1} Mpc, so this one is further away.
The linear correlation functions (not shown) from clusters and fields turn out to be similar. Looking at the logarithmic results shows they are identical in γ (2.36) and clustering length r_0 (0.26). Shown is the averaged two-point correlation functio of ten different viewpoints (cluster) and ten different viewpoints (field) in the logarithmic range, with a magnitude limit of 15 and a sphere selection of radius 50 (Fig. 35). This means that each averaged correlation function contains ten independent correlation functions.

8.2.3 Observing in Arbitrary Directions

In the following galaxy dataset it has been simulated to be observing random rectangular patches on the sky, in two opposed directions. The set is magnitude limited and spherically selected. The actual selections that have been performed are: magnitude limit 15, spherical radius $100h^-1Mpc$ and a double slice with opening angles of 10 by 120 degrees. 20 random directions, or pointings, have been used. The original cube contained 268235 particles and is identical to the one used in the previous section Increasing Opening Angle in a Double Slice Shape.

(a) Observing from a void, out to 10 h^{-1} Mpc

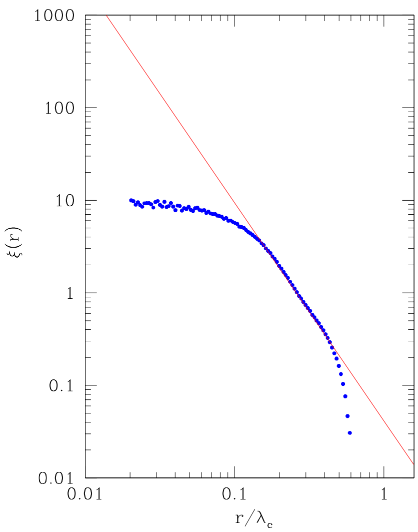
(b) Observing from a void, out to 20 h^{-1} Mpc

Figure 33: Looking at the sky: Aitoff projections as seen from two different viewpoints: within a void and within a cluster (next figure). The viewpoints show little structure and less stars from within a void and observable structure and more stars from within a cluster.

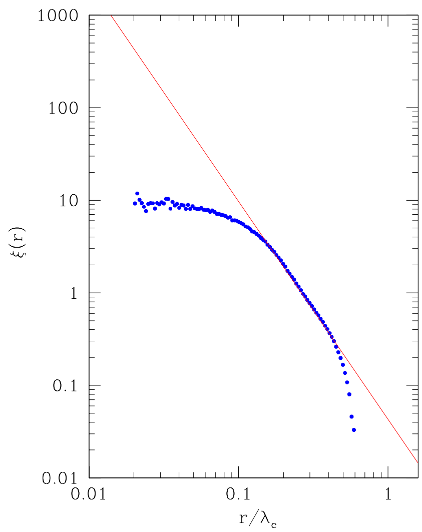
(a) …from a cluster, out to 10 h^{-1} Mpc

(b) …from a cluster, out to 20 h^{-1} Mpc

Figure 34: Looking at the sky: Aitoff projections as seen from two different viewpoints: within a void (previous figure) and within a cluster. The viewpoints show little structure and less stars from within a void and observable structure and more stars from within a cluster.



(a) Field averaged logarithmic two point correlation functions



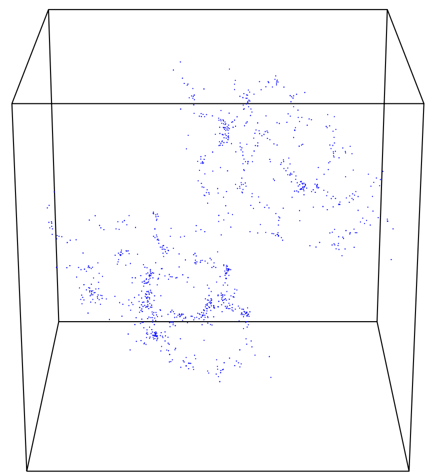
(b) Cluster averaged logarithmic two point correlation functions

Figure 35: (No) differences in average logaritmic two point correlation functions as seen from the field or from a cluster

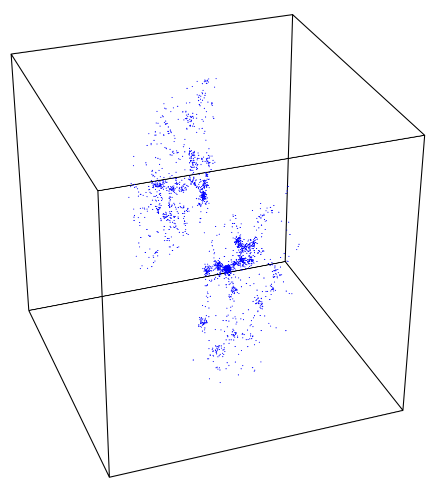
By looking in different directions it is expected that some beams encounter many clusterings and some beams might not. The amount of particles in the beam may therefore vary appreciable. The actual variation ranges from about 1,000 particles to about 2,000 particles. Looking at the examples given in Fig. 36, it can be seen that the slice with 2062 contains some thick clusters, while the one with 1052 only contains some clusters that are less rich.

The averaged two-point correlation function over 20 random directions can be seen in Fig. 37(a). The averaged correlation function over the first half of correlation functions (sorted by amount of particles) and the second half are compared. Notice that the first half correlation function has a narrower range in estimated uncertainty, but otherwise it has the same form. There is no reason to see any difference between the samples with less particles as compared to those with more particles (looking in the direction of thick clusters). See Figures 37(b) and 37(c).

The 1σ spread is considerable in all these plots, which can be thought to be due to the different pointings representing different parts of the complete structure. This is different from increasing the opening angle of a cone or slice selection, because in the latter case the structure inside the selection grows instead of changing.

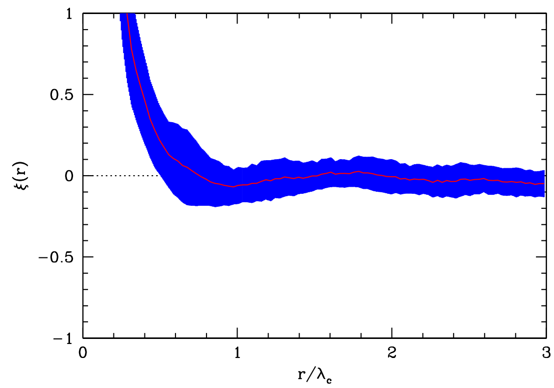


(a) Slice with 1052 particles

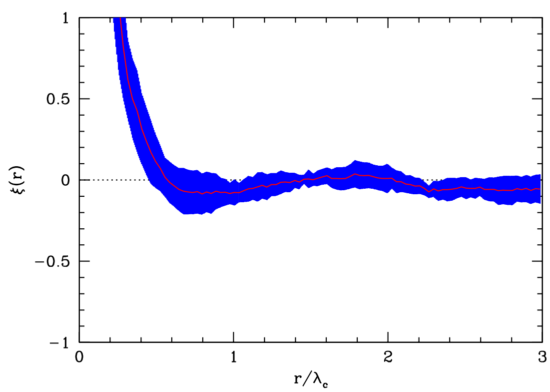


(b) Slice with 2062 particles

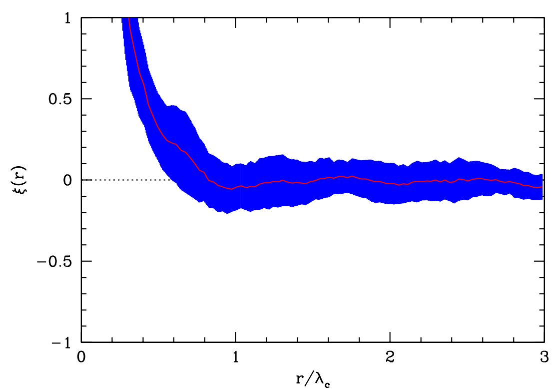
Figure 36: Two slices pointing in different directions; one with thick clusters in it, one without.



(a) Complete averaged correlation function over the random directions



(b) Averaged correlation function: first half in number of particles



(c) Averaged correlation function: second half in number of particles

Figure 37: Averaged correlation functions of two slices, one resulting from pointing in directions with thick clusters (bottom left panel) and one resulting from pointing in directions without (bottom right panel). The top panel shows the averaged function over all random directions.

9 Conclusions

After looking at the different selection effects and the associated uncertainties, it can be confirmed that there is no reason to believe that shape selection influences the two-point correlation function, save for the introduction of extra statistical noise as the selected volume becomes less representative of the complete distribution. Selection with a magnitude limit has not been convincingly shown to have no effect, though the effect seems to occur only at very strict magnitude limits leaving a small fraction of galaxies.

10 Future work

A comparison of different estimators have been published in e.g. Pons-Bordería et al. (1999). Accuracy of the results could be obtained by using the Hamilton estimator, like Goldwirth et al. (1995) did to complement their study into the size of voids. However, the results are expected to remain the same qualitatively.
A lot more selections could be tried. Especially, a detailed probe into the effect of the inclusion or exclusion of clusters in the more narrow beams could be interesting. This would be a variation of observing from voids or clusters.
The effects of magnitude selection remain unresolved for now, allthough these effects are arising at very strict magnitude limits. Present day redshift surveys operate at such sensitivity that any hypothetical statistical effect should be near undetectable. Recovering the cause of the effects of magnitude selection will still be insightful, though.

Acknowledgements

I wish to thank in no preferred order:
Rien for never ceasing to push me in the right direction(s).
Liesbeth for stimulating discussions about everything directly or remotely related to (Voronoi) simulations.
Chris for helping out with loads of editing tricks in VIM and LaTeX features.
The other roommates of ZG193 for tolerating my often non-standard break regime.
My housemates for accepting my working hours, which got more exotic with every passing year.
My parents for motivating me when completion of this work seemed really far away .
All my friends that I forgot to mention for forcing me to relax and enjoy my spare time every once in a while.

References

- Computer Simulations of Galaxy Clustering, Aarseth, S.J., Gott III, J.R., Turner, E.L., IAUS, 1978

- First Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Preliminary Maps and Basic Results, Bennett, C.L. and 20 coauthors, ApJS, 2003

- The galaxy luminosity function and luminosity density at redshift $z = 0.1$, Blanton, M.R. and 19 coauthors, ApJ, 2003

- The Collisionless Damping of Density Fluctuations in an Expanding Universe, Bond, J.R., Szalay, A.S., ApJ, 1983

- Non-Gaussian Error Distribution of Hubble Constant Measurements, Chen, G., Gott III, J.R., Ratra, B., PASP, 2003

- Self-Similarity and Scaling Behavior of Scale-Free Gravitational Clustering, Colombi, S., Bouchet, F.R., Hernquist, L., ApJ, 1996

- A Survey of Galaxy Redhifts. II. The Large Scale Space Distribution, Davis, M., Huchra, J., David W.L., Tonry, J., ApJ, 1982

- A survey of galaxy redshifts. V. The two-point position and velocity correlation, Davis, M., Peebles, P.J.E., ApJ, 1983

- A Slice of the Universe, De Lapparent, V., Geller, M.J., Huchra, J.P., ApJ, 1986

- Analysis of a complete galaxy redshift survey - II. The field-galaxy luminosity function, Efstathiou, G., Ellis, R.S., Peterson, B.A., MNRAS, 1988

- Clusters and superclusters in the Sloan Digital Sky Survey, Einasto, J., Hütsi, G., Einasto, M., Saar, E., Tucker, D.L., Müller, V., Heinämäki, P., Allam, S.S., A&A, 2003

- Voids in the Large-Scale Structure, El-Ad, H., Piran, T., ApJ, 1997

- The two-point correlation function and the size of voids, Goldwirth, D.S., Da Costa, L.N., Van de Weygaert, R., MNRAS, 1995

- Formation of Galaxies inside Clusters, Icke, V., A&A, 1973

- Voids and filaments, Icke, V., MNRAS, 1984

- Fragmenting the universe. I - Statistics of two-dimensional Voronoi foams, Icke, V., Van de Weygaert, R., A&A, 1987

- The Galaxy Distribution as a Voronoi Foam, Icke, V., Van de Weygaert, R., QJRAS, 1991

- Toward better ways to measure the galaxy correlation function, Hamilton, A.J.S., ApJ, 1993

- The 2dF Galaxy Redshift Survey: correlation functions, peculiar velocities and the matter density of the Unverse, Hawkins, E. and 29 coauthors, MNRAS, 2003
- Spherical Simulations of Holes and Honeycombs in Friedmann Universes, Hoffman, G.L., Salpeter, E.E., Wasserman, I., ApJ, 1983
- The CfA Redshift Survey, Huchra, J.P., Geller, M.J., De Lapparent, V., Burg, R., IAUS, 1988
- Evolution of Structure in Cold Dark Matter Universes, Jenkins, A., Frenk, C.S., Pearce, F.R., Thomas, P.A., Colberg, J.M., White, S.D.M., Couchman, H.M.P., Peacock, J.A., Efstatiou, G., Nelson, A.H., ApJ, 1998
- Scaling Laws in the Distribution of Galaxies, Jones, B.J.T., Martínez, V.J., Saar, E., Trimble, V., in preparation
- A comparison of estimators for the two-point correlation function, Kerscher, M., Szapudi, I., Szalay, A.S., ApJ, 2000
- A Million Cubic Megaparsec Void in Boötes?, Kirshner, R.P., Oemler Jr., A., Schechter, P.L., Shectman, S.A., ApJ, 1981
- Topology of Supercluster-Void Structure, Matsuda, T., Shima, E., PThPh, 1984
- Cosmological Theories, Milne, E.A., ApJ, 1940
- The Las Campanas Deep Redshift Survey, Oemler Jr., A., Tucker, D.L., Kirshner, R.P., Lin H., Shectman, S.A., Schechter, P.L., Obs. Cosm. Astron. Soc. of the Pacific Conf. Series Vol. 51, 1993
- Comparing Estimators of the Galaxy Correlation Function, Pons-Bordería, M.-J., Martínez, V.J., Stoyan D. and Stoyan H., Saar, E., ApJ, 1999
- Finding galaxy clusters using Voronoi tessellations, Ramella, M., Boschin, W., Fadda, D., Nonino, M., A&A, 2001
- A hierarchy of voids: Much ado about nothing, Sheth, R.K., Van de Weygaert, R., astro-ph/0311260
- Estimating the power-law of the two-point correlation function, Snethlage, M., A&A, 2001
- First Year Wilkinson Microwave Anisotroyp Probe (WMAP) Observations: Determination of Cosmological Parameters, Spergel, D.N. and 16 coauthors, ApJS, 2003
- A New Class of Estimators for the N-Point Correlations, Szapudi, I., Szalay, A.S., ApJ, 1998
- Anglo-Australian Telescope's 2dF Facility, Taylor, K., Cannon, R.D., Watson, F.G., Proc. SPIE Vol. 2871, 1997
- The Correlation Function for the Distribution of Galaxies, Totsuji, H., Kihara, T., PASJ, 1969

- Groups of Galaxies. II. The Luminosity Function, Turner, E. L., Gott III, J.R., ApJ, 1976
- Fragmenting the universe. II - Voronoi vertices as Abell clusters, Van de Weygaert, R., Icke, V., A&A, 1989
- Quasi-periodicity in deep redshift surveys, Van de Weygaert, R., MNRAS, 1991
- Fragmenting the Universe III. The construction and statistics of 3-D Voronoi tessellations, Van de Weygaert, A&A, 1994
- Explorations of a Foamy World, Voronoi patterns as templates for the spatial cosmic galaxy distribution, Van de Weygaert, R., in preparation
- Voids in gravitational instability scenarios - I. Global density and velocity fields in an Einstein-de Sitter universe, Van de Weygaert, R., Van Kampen, E., MNRAS, 1993
- The large-scale structure of the universe and the division of space, Yoshioka, S., Ikeuchi, S., ApJ, 1989
- Galaxy groups in the 2dF Galaxy Redshift Survey: large-scale structure with groups, Zandivarez, A., Merchán, M.E., Padilla, N.D., MNRAS, 2003
- Dynamical Voronoi tessellation. V. Thickness and incompleteness, Zaninetti, L., A&AS, 1995
- Galaxy clustering in early Sloan Digital Sky Survey redshift data, Zehavi, I. and 68 coauthors, ApJ, 2002
- Cosmological fluctuations produced near a singularity, Zel'dovich, Ya. B., MNRAS, 1980
- Gravitational Instability: An Approximate Theory for Large Density Perturbations, Zel'dovich, Ya. B., A&A, 1970
- Giant Voids in the Universe, Zel'dovich, Ya. B., Einasto, J., Shandarin, S.F., 1982

Consulted literature:

- Large-Scale Structure, Theory and Statistics, Coles, P., astro-ph/0103017
- A Short Course in General Relativity, Foster, J., Nightingale, J.D., second edition, 1995
- Statistics of the Galaxy Distribution, Martinez, V.J., Saar, E., 2001
- An Introduction to Cosmology, Narlikar, J.V., third edition, 2002
- Large scale clustering in the universe, Peebles, P.J.E., IAUS, 1978
- The picture of our universe: A view from modern cosmology, Reid, D.D., Kittel, D.W., Arsznov, E.E., Thompson, G.B., astro-ph/0209504
- Froth Across the Universe: Dynamics and Stochastic Geometry of the Cosmic Foam, Van de Weygaert, R., 2002

- Lecture notes structure formation and cosmology, Van de Weygaert, R.
- Course notes Graduate Cosmology: VIII. Cosmological inhomogeneities, Weinberg, D.H., http://www-astronomy.mps.ohio-state.edu/∼dhw/A873/notes8.ps
- Introductory Astronomy and Astrophysics, Zeilik, M., Gregory, S.A., Smith, E v.P., third edition, 1992
- The 2dF Galaxy Redshift Survey, http://msowww.anu.edu.au/2dFGRS/

Appendices

A Geomsur

Geomsur is a program by Rien van de Weygaert used to produce Voronoi tessellations. It has a wide range of simulation options, but the ones that have been used in this work are the option to generate a galaxy distribution around a Voronoi framework, with information about galaxy membership (voids, filaments, clusters, walls). Galaxy magnitudes are assigned through the Schechter luminosity function. Clustering schemes include the Void Superhubble Scheme, the Void Explosion Scheme, but a simple Four Element Mixture Scheme has been used here. With this scheme, one has to specifiy the percentages of galaxies in fields, walls, filaments and clusters. The thickness of these different elements are given through a radial density profile. A Gaussian profile has been used in this work. An example Voronoi distribution of galaxies is shown below in Fig. 38.

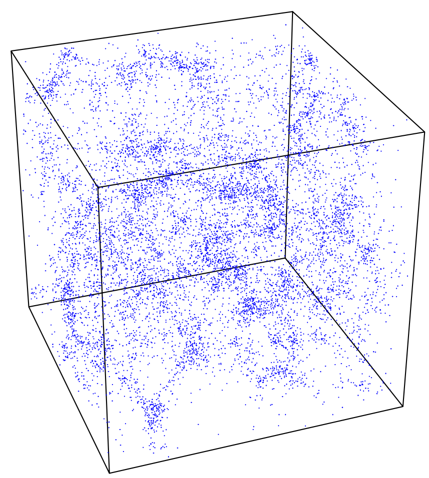


Figure 38: Sample plot of a galaxy distribution produced by geomsur.

B Cubeplotsm and P3D

Cubeplotsm and P3D are used to plot the simulated galaxy distribution in a three-dimensional fashion.
Cubeplotsm is a utility by Rien van de Weygaert. It uses the Supermongo libraries to create a three-dimensional cubical representation of a galaxy distribution file in the .dpm format common to the output of N-body simulation codes. It can also be used to create customizable slices of the cube with galaxies, or a sequence of slices, along the different axes and from different viewpoints. See e.g. Fig. 18, depicting a slice P3D uses PGPLOT to create a similar three-dimensional representation in its own fileformat, with the possibility to assign colors to different galaxy brightness. The program P3D was written by Andrey V. Kravtsov, 1998 and is available under the GNU-General Public License. It can be used to create animations from changing viewpoints or time evolution. See Fig. 39.

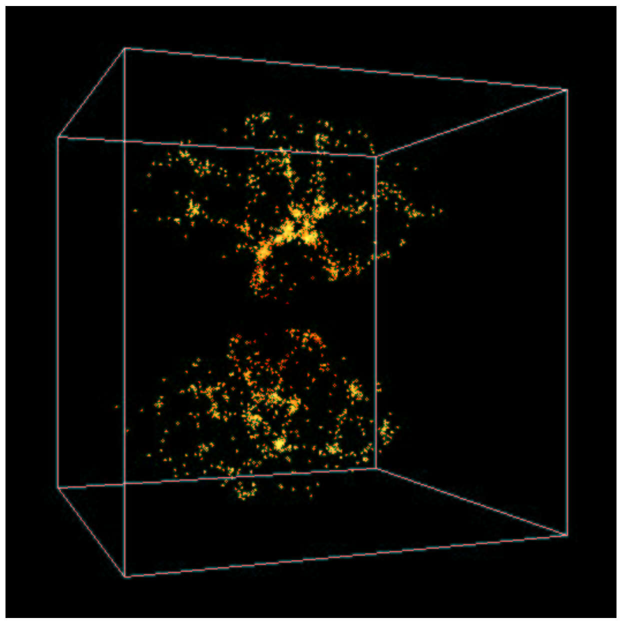


Figure 39: Sample P3D plot of a galaxy distribution: a double slice selection combined with sphere and magnitude selection.

C Sampleselect

The program sampleselect has been written for this work to accomodate shape and magnitude selections on datasets. It has a modular design as to accomodate the usage of many different filetypes. The main functionality of sampleselect lies in its ability to read a batch file with instructions for carrying out selections. The design features allow for an easy adaptation for complete commandline control. See Fig. 40 for the structure of the program sampleselect.

The batch file contains information about the datafile on which the selections are to be carried out. A file prefix for the automatic generation of suitable files after selection is given. The first line of the batch file contains a number giving the file type of the filename in the next line. After the filename comes the file prefix.

Then come the command lines; each containing instructions to perform e.g. magnitude selection, cone selection with rotation etc. A typical command line goes like

```
ta10_0_0 m15 sph50 ds120_120_30_0_0
```

Sampleselect recognizes this as a command to:

- translate 10 Mpc in the x-direction,
- do a magnitude selection, excluding every galaxy fainter than 15 mag,
- perform a spherical selection of radius 50 Mpc,
- select a double slice of 120 by 120 degrees, oriented 30 degrees in theta direction.

Files from the resulting data are being written using a nomenclature that shows which selections have been performed. Different files are suitable for viewing and analysis.

As an added bonus, the program uses the interface to different filetypes used by the simulation and plotting routines to convert datafiles from one type to another.

The supported file types are:

- .DAT unformatted datafile. Plain text file with columns for spatial coordinates
- .VDAT unformatted datafile (read only), produced by the geomsur program. Contains spatial coordinates, information about void, filament, wall or cluster type of galaxy and optional each galaxy's absolute magnitude.
- .DPM formatted datafile, which is usually output from the N-body simulation codes and can be read by cubeplotsm.
- .TPCF formatted datafile (write only), used as input in the program estimating the two-point correlation function. Sampleselect attaches extra selection history information to this file.

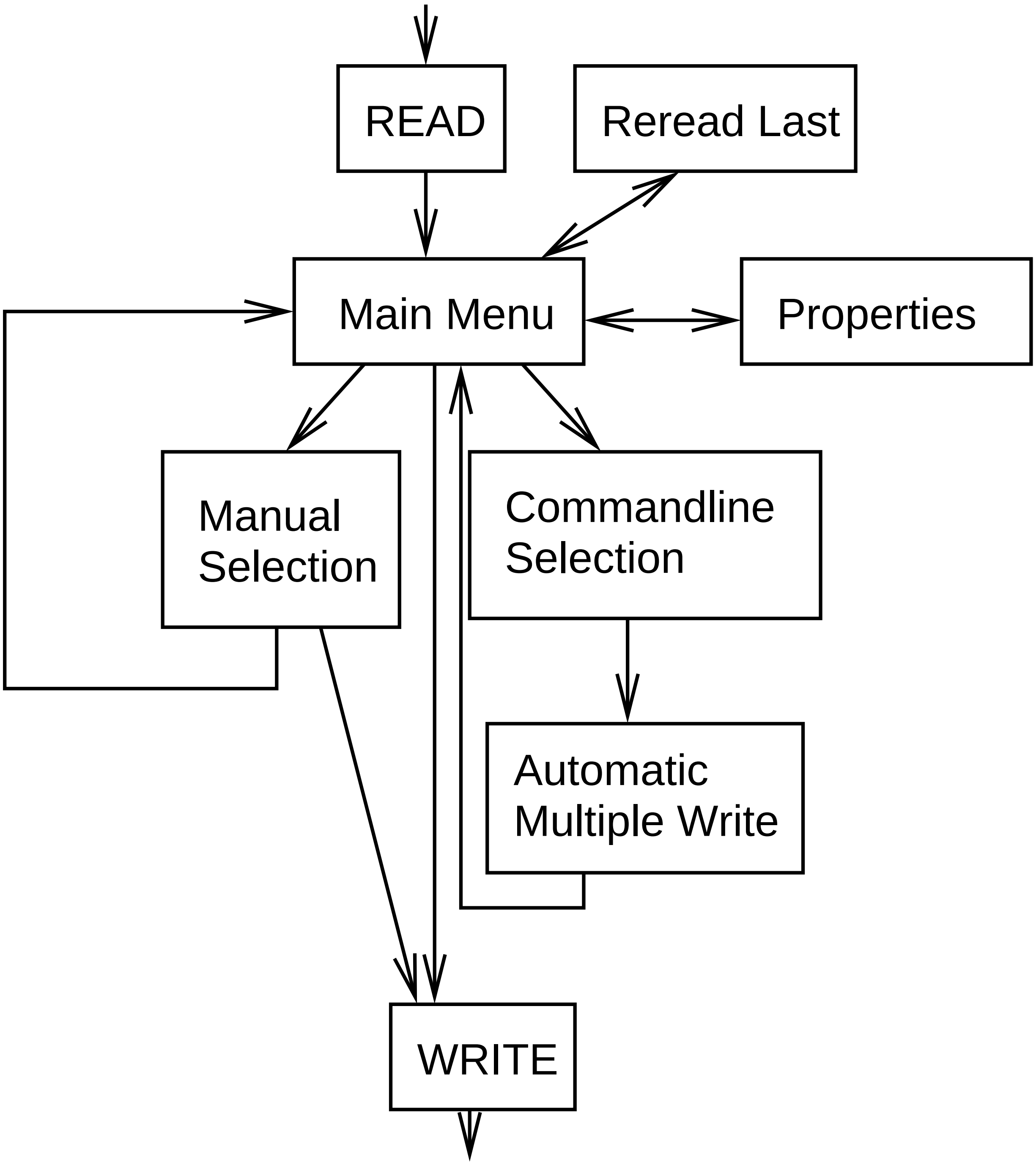


Figure 40: Sampleselect program structure flowchart.

- (.log).CDAT unformatted datafile, which is output of the estimator program. It can be used to plot the (logarithmic range of the) two-point correlation function
- .P3D unformatted datafile (write only), which can be read by the P3D plotting program.

The file routines are called from a wrapper read or write routine.

D Selected Fortran Code Routines

```
SUBROUTINE SelectCone(d, o_angle, theta, phi, singlecone)
  !x^2 + y^2 = cz^2 <=> x^2 + y^2 = z^2 tan^2(theta)
  !using abbreviated names for readability

  TYPE(dataobject), INTENT(INOUT) :: d !datapoints

  REAL, INTENT(IN) :: o_angle !opening angle (deg)
  REAL, INTENT(IN) :: theta, phi !orientation of axis system (deg)

  LOGICAL, INTENT(IN) :: singlecone

  INTEGER :: n, m, tmp_n_o_p

  REAL :: constant !the constant c in the formula mentioned above
  REAL :: rad_angle, rad_theta, rad_phi, rad_psi
  REAL :: cosphi, sinphi, costheta, sintheta

  !auxiliary for readability and comprehensibility
  REAL                 :: x, y, z, u, v, w, uu, vv, ww
  !REAL, DIMENSION(dim) :: rot
  !Decrease variables after problem has been solved

  LOGICAL :: below_o

  IF (info) PRINT *, "Subroutine SelectCone, single cone: ", singlecone

  rad_angle = o_angle * pi / 180.
  rad_theta = theta   * pi / 180.
  rad_phi   = phi     * pi / 180.
  rad_psi   = 0         !Angle not required with this symmetry
  constant  = (TAN((rad_angle/2.))) ** 2.

  !efficiency suggestion:
  !cphi = cos(rad_phi); sphi = sin(rad_phi); ctheta etc.
  !Should reduce calculating time
  cosphi   = cos(rad_phi)
  sinphi   = sin(rad_phi)
  costheta = cos(rad_theta)
  sintheta = sin(rad_theta)

  tmp_n_o_p = d%number_of_particles  !gets decreased, so preserve it
  n = 1 !need to count independently
  DO m = 1, d%number_of_particles
```

```
    !The number of thrown out points cannot exceed the original n_o_p

    !Need a back-rotation to get points on standard simple cone
    !So, adjust relation for incoming angles.

    !x, y, z translation
    x =  (d%point(n)%x - d%origin%x)
    y =  (d%point(n)%y - d%origin%y)
    z =  (d%point(n)%z - d%origin%z)

    !x, y, z transformed into u, v, w
    u =              cosphi * x &
      & +            sinphi * y
    v = - costheta * sinphi * x &
      & + costheta * cosphi * y &
      & + sintheta *          z
    w =   sintheta * sinphi * x &
      & - sintheta * cosphi * y &
      & + costheta *          z

    uu = u**2
    vv = v**2
    ww = w**2

    IF ( (uu + vv) < (constant * ww) ) THEN !point is within a double cone
      below_o = (w < 0)
      IF ( (singlecone) .AND. (below_o) ) THEN !point outside single c
        CALL TrashPoint(d,n)
        n = n - 1 !make sure the new point is considered as well
      END IF
    ELSE !point is outside a double cone
      CALL TrashPoint(d,n)
      n = n - 1
    END IF

    n = n + 1
    IF (n > d%number_of_particles) EXIT !Otherwise we do many points twice.
  END DO

END SUBROUTINE SelectCone

SUBROUTINE SelectSlice(data, theta, phi, psi, o_a1, o_a2, single)
  !use 4 planes, rotate each through the right angles, see if points below
  !or above eache plane
  !for double slice adjust angles and calculate separately
```

```
TYPE(dataobject), INTENT(INOUT) :: data

REAL, INTENT(IN) :: theta, phi, psi, o_a1, o_a2

LOGICAL, INTENT(IN) :: single

REAL, DIMENSION(4) :: w
REAL, DIMENSION(5) :: rad_theta, rad_phi, rad_psi
REAL :: rad_o_a1, rad_o_a2
REAL :: x, y, z, xr, yr, zr !x, y, z tempororary translated coord.

INTEGER :: n, m, t, tmp_n_o_p

LOGICAL :: is_in

rad_o_a1 = o_a1 * pi / 180.
rad_o_a2 = o_a2 * pi / 180.

!Convert the orientation angles
rad_theta(5) = theta * (pi/180.)
rad_phi(5)   = phi   * (pi/180.)
rad_psi(5)   = psi   * (pi/180.)

!Now back-rotate angles for the particles
!using Irotate
!Plane 1
rad_theta(1) = +(pi/2.)
rad_psi(1)   = +(rad_o_a1 / 2.)

!Plane 2
rad_theta(2) = +(pi/2.)
rad_psi(2)   = -(rad_o_a1 / 2.)

!Plane 3
rad_theta(3) = +(rad_o_a2 / 2.)
rad_psi(3)   = -(pi/2.)

!Plane 4
rad_theta(4) = -(rad_o_a2 / 2.)
rad_psi(4)   = -(pi/2.)

!Back-rotate every particle after pre-rotation.
tmp_n_o_p = data%number_of_particles
n = 1
DO m = 1, tmp_n_o_p
```

```
    !x, y, z translation to origin, rotate.
    x = data%point(n)%x - data%origin%x
    y = data%point(n)%y - data%origin%y
    z = data%point(n)%z - data%origin%z
    !Rotate all particles prior to selection
    xr = IRotateX(x, y, z, rad_theta(5), rad_phi(5), rad_psi(5))
    yr = IRotateY(x, y, z, rad_theta(5), rad_phi(5), rad_psi(5))
    zr = IRotateZ(x, y, z, rad_theta(5), rad_phi(5), rad_psi(5))

    DO t = 1, 4 !Project points at all four planes
      w(t) = IRotateZ(xr, yr, zr, rad_theta(t), rad_phi(t), rad_psi(t))
    END DO

    !Plane 1 in: w(1) > 0
    is_in =             (w(1) .GT. 0)

    !Plane 2 in: w(2) < 0
    is_in = is_in .AND. (w(2) .LT. 0)

    !Plane 3 in: w(3) < 0
    is_in = is_in .AND. (w(3) .LT. 0)

    !Plane 4 in: w(4) > 0
    is_in = is_in .AND. (w(4) .GT. 0)

    !Reverse criteria
    IF (.NOT. single) THEN
      is_in = is_in .OR. &
        & ((w(1) < 0) .AND. (w(2) > 0) .AND. (w(3) > 0) .AND. (w(4) < 0))
    END IF

    IF (.NOT. is_in) THEN
      CALL TrashPoint(data, n)
      n = n - 1
    END IF
    n = n + 1
    IF (n > data%number_of_particles) EXIT

  END DO

CONTAINS

  REAL FUNCTION IRotateX(x, y, z, rtheta, rphi, rpsi)
  REAL FUNCTION IRotateY(x, y, z, rtheta, rphi, rpsi)
  REAL FUNCTION IRotateZ(x, y, z, rtheta, rphi, rpsi)
```

```
END SUBROUTINE SelectSlice

 SUBROUTINE SelectMag(d, lim_mag)
   !Select on magnitude, so magnitude must be available or otherwise
   !no galaxies will be selected
   !m = M - 5 + 5log(r) - Kz    --> Kz assumed to be neglegible
   !Datapoint coordinates are in Mpc

   TYPE(dataobject), INTENT(INOUT) :: d

   REAL, INTENT(IN) :: lim_mag

   REAL :: r, x, y, z, mag !radius from point to center

   INTEGER :: n, m, tmp_n_o_p

   LOGICAL :: rflag

   tmp_n_o_p = d%number_of_particles
   n = 1
   DO m = 1, tmp_n_o_p
     !Calculate apparent magnitude from the d%origin
     x = d%point(n)%x - d%origin%x
     y = d%point(n)%y - d%origin%y
     z = d%point(n)%z - d%origin%z
     r = ((x**2. + y**2. + z**2.)**(1/2.)) * 10.**6.
     mag = d%point(n)%mag - 5. + (5. * log10(r))
     IF (mag > lim_mag) THEN
       CALL TrashPoint(d, n) !magnitude is too faint
       n = n - 1
     END IF
     n = n + 1
     IF (n > d%number_of_particles) EXIT
   END DO

 END SUBROUTINE SelectMag
```

The creation of the homogeneous random point samples involves the use of simply rejecting random particle positions that are outside the desired shape. The routines magapproved and shapeapproved establish whether a particle is inside the desired shape or magnitude range in a similar fashion as the routines above.

```
      SUBROUTINE POISSAMPLE(INPT,PNTRAN)
```

```
      INTEGER    DIM,INMPOS
      PARAMETER  (DIM=3,INMPOS=125000)

      integer maxmag
      parameter(maxmag=10000)

      INTEGER    INPT
      REAL       PNTRAN(DIM+1,INMPOS)
      double precision XRAND

      REAL       BND(2,DIM), SIZE(DIM)
      COMMON     /BOUND/BND,SIZE

      REAL       DISDIM
      COMMON     /DISDIM/DISDIM

      integer shape
      real    orienttheta, orientphi, orientpsi
      real    o_angle1, o_angle2, sradius, boxlength
      real    originx, originy, originz
      real    r_a1, r_a2, r_theta, r_phi, r_psi
      real    coneconst, cosphi, sinphi, costheta, sintheta
      real    a_r_theta(4), a_r_phi(4), a_r_psi(4), magsel

      common     /shapeparams/shape,orienttheta, orientphi, orientpsi,
     *            o_angle1, o_angle2, sradius, boxlength,
     *            originx, originy, originz,
     *            r_a1, r_a2, r_theta, r_phi, r_psi,
     *            coneconst, cosphi, sinphi, costheta, sintheta,
     *            a_r_theta, a_r_phi, a_r_psi, magsel

      real        phistar, alpha, mstar, mmin, mmax

      common      /schechter/phistar, alpha, mstar, mmin, mmax

      integer    inmlum
      real     lumcum(maxmag), magn(maxmag)

      common     /lum/lumcum, magn, inmlum

      real     gallum

      logical    is_true, shapeapproved, magapproved
```

```
      integer p

C       ========================================================================

C11   Initialise schechter stuff (expect default values if not available)
      CALL SCHINI

C2    routine generates points in range [0,1) * size(:) + bnd(1,:)
        DO 130 N=1,INPT

C4        INPNT iterations should be enough to give at least 1 point inside desired shape.
          do 145 p=1,INPT

            DO 140 M=1,DIM
              CALL RANDA(XRAND)
              PNTRAN(M,N)=XRAND*SIZE(M)+BND(1,M)
140         CONTINUE
            CALL RANDA(XRAND)
            PNTRAN(4,N) = GALLUM(XRAND)

C4          Check whether point within shape before moving to next N.
C12         Also check magnitude (if only that and no shape)
            is_true = magapproved(pntran(1,n),pntran(2,n),
     3                           pntran(3,n),pntran(4,n),magsel)
            if ((shape .ne. 0) .or. (boxlength .ne. 0.)) then
              is_true=is_true .and. (shapeapproved(pntran(1,n),
     1                              pntran(2,n),pntran(3,n)))
            endif
            if (is_true) then
              goto 130
C4            goto: leaves the do-loop. Sorry.
C4            shape params are globally passed; if point is in specified shape exit loop
            endif

145       continue

130     CONTINUE

C       ========================================================================

       RETURN
      END
```