Galaxy surveys

Somnath Bharadwaj

Department of Physics and Meteorology, and Centre for Theoretical Studies, Indian Institute of Technology Kharagpur, Kharagpur 721 302, India

Galaxy surveys provide us with a map of the local universe. They also contain detailed information about galaxy properties. Both these inputs, coming from galaxy surveys, have played a very important role in shaping our present picture of the universe. This article provides an introduction to galaxy surveys. Some of the large optical galaxy surveys carried out till date are described, and a few of the scientific findings of these surveys are highlighted.

Keywords: Galaxy surveys, Milk Way, redshift.

Introduction

The German philosopher Immanuel Kant proposed, in 1775, that the sun is a part of an isolated collection of stars distributed in a plane and bound together by gravity. This disk-like stellar system, he argued, is consistent with the Milky Way, a luminous band that is observed to run across the entire sky. Kant also pointed out that the Milky Way might not be the only such stellar system. Some of the faint, fuzzy, elliptical patches of light seen in the sky and referred to as nebulae might be ‘island universes’, stellar systems similar to the Milky Way viewed from a distance. The word ‘galaxy’ comes from the Greek word for milk, and we refer to each island universe as a galaxy.

The effort of several astronomers over a period of many decades, and the advent of large telescopes (eg. 100 inch at Mt Wilson) were required. It was only in the 1920s that the nature and the spatial extent of our own galaxy the Milky Way, also referred to as the galaxy, could be determined. It was also in the 1920s that the distances to some of the nebulae could be measured revealing them to be located beyond the extent of our own galaxy. This established the fact that some of the nebulae were indeed island universes, galaxies viewed from a distance. Our Galaxy is a collection of $10^{11}$ stars, distributed in a disk of diameter $D$ 23 kpc (kpc = kilo-parsec; 1 pc = $3.08 \times 10^{16}$ m) and thickness $H$ between 260 pc and 1350 kpc. The entire disk rotates around its centre with a speed $\sim 220$ km/s which is roughly independent of the distances from the centre. The Andromeda Nebulae (M31), the nearest full fledged disk galaxy like our own, is at a distance of $0.785 \pm 0.025$ Mpc (Mpc = Mega-parsec) well beyond the extent of our own galaxy. It was also evident that our Galaxy and the Andromeda galaxy were not unique, there were several other galaxies already identified in the 1920s. The universe around us appears to be filled with galaxies. These findings threw open a new field in astronomy, the study of galaxies. Figure 1 shows an image of NGC 628, a galaxy similar to our Galaxy and M31.

What are the properties that characterize a galaxy? Are all galaxies similar, or are there different types of galaxies? How are the galaxies distributed in space? These are just a few of the questions that come up. A systematic study of galaxies is needed in order to address these questions. The scope of such a study could, in general, be two-fold:

1. To determine the properties of individual galaxies.
2. To determine how the galaxies are distributed in the sky, or in full three-dimensional space if the distances can be estimated.

We refer to such a systematic study of galaxies as a galaxy survey, the topic of this article.

Preliminaries

The reader is referred to Binney and Merrifield for a detailed discussion of the contents of first three

Figure 1. Blue-band image of the spiral galaxy NGC 628 from the Digital Sky Survey.
subsections and to any textbook on cosmology for the last.

Angular coordinates

I will briefly discuss the equatorial system of angular coordinates on the sky. The right ascension (RA) $\alpha$ and declination (DEC) $\delta$ define an angular coordinate system on the sky or the celestial sphere. The celestial sphere is a sphere of infinite radius centred on the Earth. The celestial equator is the great circle where the Earth’s equatorial plane intersects the celestial sphere, and the north and south celestial poles (NCP and SCP) respectively refer to the two points where the Earth’s rotation axis intersects the celestial sphere. The coordinate system is fixed to the celestial sphere as it appears to rotate from east to west due to the Earth’s rotation.

The right ascension and declination are similar to the latitude and longitude, while the latter are circles drawn on the Earth’s surface, the former are circles drawn on the celestial sphere. The declination $\delta$ like the latitude, is measured in degrees from the equator and the NCP and SCP are at $\delta = 90^\circ$ and $\delta = -90^\circ$ respectively. The right ascension differs from the longitudes in that it is measured in hours, minutes and seconds ($24 \text{ h} = 360^\circ$) increasing from west to east. Further, the origin $\alpha = 0$ is not located at the Greenwich meridian, it is located at the meridian through the Sun’s position at the Vernal Equinox. We use ($\alpha, \delta$) to denote the position of an astronomical source on the celestial sphere.

Colour and magnitudes

We use $f_\lambda$, the energy flux per unit frequency interval [W m$^{-2}$ Hz$^{-1}$] or equivalently $f_\lambda$, the energy flux per unit wavelength interval [W m$^{-2}$ nm$^{-1}$] to quantify the spectrum of energy flux received from any astronomical source. The quantity that is usually measured in either a CCD or an electronic photometer is the energy flux incident in a wavelength band, typically defined by a filter of width $\Delta \lambda$ centred on a wavelength $\lambda_{\text{eff}}$

$$f = \int d\lambda \, f_\lambda S_\lambda. \tag{1}$$

Here, $S_\lambda$ is the combined sensitivity of the atmosphere, telescope and filter. The procedure of measuring the energy flux $f$ incident from an astronomical source is called astronomical photometry.

The apparent brightness of any astronomical source is quantified in terms of the apparent magnitude $m$ defined as

$$m = -2.5 \log f + C. \tag{2}$$

The larger the magnitude, the fainter the source. A magnitude difference of 5 corresponds to a source that is 100 times fainter.

Astronomical photometry is usually carried out in specific, well-defined wavelength bands. The bands are defined by filters which restrict the wavelength range to which the observations are sensitive. This is quantified by the sensitivity $S_\lambda$ (eq. (1)), and the parameters $\lambda_{\text{eff}}$ and the full width at half maxima (FWHM) of $S_\lambda$ respectively parametrize the effective wavelength and width in wavelength of a band. The standard photometric system is based on the ultraviolet-blue-visual system (UBV) bands whose parameters are listed in Table 1. In a particular band, say the V band, the apparent magnitude of a source is denoted using either $V$ or $m_V$. As we shall see later, there now are a variety of bands extending into the infrared in which photometric observations are carried out.

The difference in the magnitudes measured in two different bands is referred to as colour index or colour. This is essentially a measure of the ratio of the flux at two different wavelengths, a quantity that matches with our notion of colour. We could, for example, have the U−B and the U−V colours. Unlike the apparent magnitude which depends on the distance to the source, the colour does not depend on distance. We expect the colour to be related to the physical properties of the source. An larger U−B colour implies an increase in the B band flux relative to U indicating a redder source. For stars, the colour serves as an indicator of its temperature, the cooler stars being redder.

The light from astronomical sources undergoes extinction through scattering in the interstellar medium enroute to us. This scattering is more efficient at small wavelengths than at large wavelengths, and as a consequence the sources are not only dimmed but also reddened. As a consequence it is possible to devise combinations of colours to determine the amount of reddening introduced by the interstellar extinction, and thereby correct for both the extinction and the reddening.

The flux, and consequently the apparent magnitude, do not reflect how luminous the source is, rather it reflects only the apparent brightness which also depends on the distance. Consider a source at a distance $d$ from which the observed flux is $f$, the observed flux $F$ would be

$$F = \left( \frac{d}{D} \right) f. \tag{3}$$

<table>
<thead>
<tr>
<th>Band</th>
<th>$\lambda_{\text{eff}}$ (nm)</th>
<th>FWHM (nm)</th>
<th>$L_\odot / 10^{25}$ W</th>
<th>$M_\odot$</th>
</tr>
</thead>
<tbody>
<tr>
<td>U</td>
<td>365</td>
<td>66</td>
<td>1.86</td>
<td>5.61</td>
</tr>
<tr>
<td>B</td>
<td>445</td>
<td>94</td>
<td>4.67</td>
<td>5.48</td>
</tr>
<tr>
<td>V</td>
<td>551</td>
<td>88</td>
<td>4.64</td>
<td>4.83</td>
</tr>
</tbody>
</table>

Table 1. The definition of the UBV bands. The solar luminosity and corresponding absolute magnitude are also shown for each band. From Binney and Merrifield.
SPECIAL SECTION: ASTRONOMY

if the same source were at a distance \( D \) instead. The absolute magnitude \( M \) of a source is defined as the apparent magnitude that would be measured if it were at a standard distance of \( D = 10 \) pc. The absolute magnitude quantifies the luminosity of the source in the band of observations. We use \( M_U, M_B, \ldots \) to denote the absolute magnitudes of a source in the U, B, \ldots bands respectively.

The apparent magnitude \( m \) and absolute magnitude \( M \) of a source at a distance \( d \) are related as

\[
m - M = 5 \log d - 5.
\]

Here \( m - M \) is referred to as the distance modulus. Interstellar extinction causes the source to appear fainter than it should. This is accounted for by introducing an extinction correction \( A \). The extinction is highly wavelength dependent and the value of \( A \) differs considerably from band to band. There is another correction that is required for sources which are moving away or towards the observer. In this case because of Doppler shift, a particular band corresponds to a different set of wavelengths in the rest frame of the source. The absolute magnitude refers to the flux that would be observed if the source were at rest at a distance of \( 10 \) pc. The \( K \) correction accounts for this, and it is possible to predict \( K \) if the slope of the source spectrum is known. Including both these corrections eq. (4) is modified to

\[
m - M = 5 \log d - 5 + A + K.
\]

I finally discuss the calibration of the magnitude scale. This is usually specified using the bright star Vega. The zero point of the magnitude scale is chosen so that the magnitude is zero in each band \( (U = V = B = \ldots = 0) \) for Vega. Table 1 provides the solar luminosity \( L_\odot \) and the corresponding absolute magnitude \( M_\odot \) for the UBV bands, which can be used to convert the UBV magnitudes to the corresponding flux in \( \text{W m}^{-2} \). We also note that the total luminosity of the Sun (including all wavelengths), referred to as the bolometric luminosity, is \( 3.8 \times 10^{26} \text{ W} \).

Classification of galaxies

The most widely accepted classification of galaxies, based on their morphologies, was introduced by Hubble in 1936. Hubble’s classification scheme is illustrated by the tuning-fork diagram shown in Figure 2, the galaxies that appear smooth and featureless are places at the left of the tuning-fork. The elliptical galaxies have shapes that vary from round (E0) to highly elongated (E7). In addition to the luminous ellipticals shown in the Hubble tuning-fork diagram, there also are the less luminous dwarf elliptical (dE) and dwarf spheroidal (dSph) galaxies which are of even lower luminosity. Elliptical galaxies largely contain cool, red stars characteristic of an old stellar population. Spiral galaxies are located at the right of the tuning-fork diagram. The normal spiral galaxies (Sa, Sb, Sc) contain a central bright region that resembles an elliptical galaxy embedded in a thin disk containing spiral arms. The barred spiral galaxies (SBa, SBB, SBc) have a bar at the centre. The spiral galaxies, in addition to the cool, red stars also contain luminous, hot, blue stars which are predominantly located along the spiral arms. As a consequence, the spirals are bluer than the elliptical galaxies. The spiral galaxies contain inter-stellar gases which are mainly hydrogen and helium. Further, the inter-stellar gas fuels star formation which is observed mainly in spiral galaxies. A spiral galaxy’s mass content in inter-stellar gas, stars and dark matter are typically of the order of \( 10^{10} M_\odot, 10^{11} M_\odot \) and \( 10^{12} M_\odot \) respectively, where \( M_\odot = 2 \times 10^{30} \text{ kg} \) refers to the Solar mass.

The lenticular galaxies are located at the intersection of the two arms of the tuning fork. They have a central concentration and a disk, similar to a spiral but these are structureless. They do not pose spiral arms, are red-like ellipticals and usually do not have interstellar gas.

Galaxies of asymmetric shapes are classified as irregular galaxies (Irr I and Irr II).

Hubble suggested that galaxies evolved from the left to right on the tuning-fork, and thus ellipticals and spirals are referred to as early and late type galaxies respectively. The current belief about galaxy formation is exactly the opposite, though the early–late nomenclature introduced by Hubble is still in use.

The expanding universe

Spectral lines identified in spectroscopic studies of galaxies are typically observed to have a wavelength \( \lambda_\circ \) that
differs from $\lambda_c$, the characteristic wavelength of the transition. We use the redshift $z = (\lambda_c - \lambda_0)/\lambda_c$ to quantify this. Hubble, in 1929, made the remarkable discovery that the redshift $z$ was proportional to $r$, the distance to the galaxy

$$z = \frac{H_0}{c} r,$$  \hfill (6)

where the Hubble constant $H_0 = 100$ km s$^{-1}$ Mpc$^{-1}$ and $c$ is the speed of light. Interpreting this relation, known as Hubble’s Law in terms of the Doppler shift – the galaxies are all receding from us with a radial velocity proportional to the distance. Although Hubble had initially determined a large value for $H_0$, it is currently accepted that, $0.5 \leq h \leq 1$.

The Cosmological Principle which states that the universe, on sufficiently large scales, is homogeneous and isotropic, provides a theoretical framework for picturing the universe. Combined with the Hubble’s Law this gives a picture where the whole universe is at present filled with galaxies whose number density, when averaged over a suitably large length-scale, is the same everywhere. These galaxies move apart from one another whereby we have an expanding universe. This expansion has no preferred centre, and observers on each galaxy see exactly the same Hubble’s Law. The Hubble time $t_0 = H_0^{-1} = 9.78 h^{-1} \times 10^9$ yr and the Hubble radius $cH_0^{-1} = 300 h^{-1}$ Mpc respectively, determine the present age and the furthest distance visible at present.

It is convenient to introduce comoving coordinates $\bar{x}$ which are fixed to the galaxies as they move apart due to the expansion of the universe. The comoving separation $\bar{r}$ between a pair of galaxies does not change because of the expansion of the universe, but the physical separation $r = a(t)\bar{r}$ increases through the scale factor $a(t)$ which is an increasing function of time. The expansion of the universe is contained in the scale factor $a(t)$ whose dynamics is governed by the Friedmann equation. Here, the different constituents of the universe contribute through the ratio of their density to the critical density whose present value is

$$\rho_{\text{crit}} = \frac{3H_0^2}{8\pi G} = 1.88h^2 \times 10^{-26} \text{ kg m}^{-3}.$$  \hfill (7)

The $\Lambda$CDM model with two constituents suffices here. The first constituent, referred to as matter, is pressureless (cold dark matter and baryons), and its present density $\rho_{\text{mat}}$ is parameterized through the matter density parameter $\Omega_{\text{mat}} = \rho_{\text{mat}}/\rho_{\text{crit}}$. The second constituent is the dark energy which has negative pressure ($P_{\Lambda} = -\rho_{\Lambda}c^2$), and the dark energy density parameter is $\Omega_{\Lambda} = \rho_{\Lambda}/\rho_{\text{crit}}$. The cosmological constant $\Lambda$ is one of the most viable dark energy candidates. The difference $\Omega_{\Lambda} = 1 - \Omega_{\text{mat}} - \Omega_{\Lambda}$ is a measure of the spatial curvature. In this model the evolution of $a(t)$ is governed by

$$H^2(a) = \left(\frac{\dot{a}(t)}{a(t)}\right)^2 = H_0^2\left[\Omega_{\text{mat}}a^{-3} + \Omega_{\text{r}}a^{-2} + \Omega_{\Lambda}\right].$$  \hfill (8)

where $\dot{a} = (da/dt)$. Unless stated otherwise, we shall use the $\Lambda$CDM model with $\Omega_{\text{mat}} = 0.3, \Omega_{\Lambda} = 0.7$ and $h = 0.7$.

The comoving distance $x(z)$ to a source at redshift $z$ is given by

$$x(z) = c \int_{(1+z)}^{1} \frac{da}{H(a)a^2},$$  \hfill (9)

which goes over to eq. (6) for $z \ll 1$. The distance-redshift relation is different when $z \sim 1$ or larger, and it depends on the values of $\Omega_{\text{mat}}$ and $\Omega_{\Lambda}$.

The physical length $\Delta$ of a source that subtends and angle $\Delta\theta$ on the sky and has a redshift $z$ is given by $\Delta = d_L(z) \Delta\theta$ where $d_L(z) = (1+z)x(z)$ is the angular diameter distance. The luminosity $L$ and flux $f$ are related as $f = L/(4\pi d_L^2(z))$ where $d_L(z) = (1+z)x(z)$ is the luminosity distance. Note that for sources at cosmological distances $d_L(z)$ should be used to calculate the absolute magnitude from the measured apparent magnitude (eq. (4)).

The measured redshift $z$ provides a relatively easy method of determining the distance to a galaxy, provided that atomic transition lines can be identified in the spectrum. Distances estimated in this way depend on $H_0$, and it is common practice to take this into account by using units $h^{-1}$ Mpc for these distances. The redshift–distance relation also depends on the parameters ($\Omega_{\text{mat}}, \Omega_{\Lambda}$) at high $z$, and it is necessary to specify these values along with the distance to a source. Further, this distance estimate rests on the assumption that the entire measured redshift is due to the expansion of the universe. It is now well established that, in addition to the Hubble expansion, galaxies also have peculiar velocities which contribute to the measured redshifts. The matter distribution is inhomogeneous on small scales, with both, over-dense and under-dense regions being present. The over-dense regions attract galaxies (inflow) while the under-dense regions have the opposite effect (outflow). This introduces coherent, local peculiar velocity patterns of magnitude of the order of 200–300 km s$^{-1}$. The galaxies in the most over-dense regions usually belong to gravitationally bound, vialized systems – the clusters of galaxies. While the whole cluster participates in the Hubble expansion, the galaxies inside the cluster have random motions that can be of the order of 1000 km s$^{-1}$. Both of these, namely the coherent and the random peculiar velocities cause errors in distances estimated from the measured redshift. Fortunately, the relative error reduces as the distance to the source increases. Finally, we note that there are several direct
distance estimators which do not depend on the redshift, but these are all not so straightforward to determine and hence cannot be applied to large samples of galaxies.

Catalogues and surveys

During the years 1758–1782, the French astronomer Charles Messier compiled a catalogue of nearly 100 diffuse, ‘deep sky’ objects which could be mistaken for a comet. Many of these objects were discovered by either Messier himself using a 4-inch telescope or by another French astronomer Pierre Mecham. The Messier catalogue was one of the first comprehensive and reliable lists of astronomical sources, many of whose members were later found to be galaxies. The objects in the Messier catalogue are referred to as M1, M2, ..., and this nomenclature still continues to be used for the objects that were originally included in this catalogue. For example, M31 refers to the Andromeda galaxy.

Dreyer, in 1887, published the New General Catalogue (NGC). This contained 7840 objects which were identified as not being stars, mostly from observations by William Herschel and his son John Herschel using a number of telescopes, the largest being of 49.5-inch diameter. This catalogue was subsequently enlarged with the Index Catalogue (IC1 1995, IC2II 1908) which added 5320 new objects. These catalogues together contained all objects visually identified as non-stellar objects prior to 1908. These catalogues, compiled prior to the advent of photography, were homogeneous across the sky and had several other errors. There have been several efforts to revise the NGC, the NGC2000.0 by Sinnott being one of the recent ones. Approximately 75% of the data in the NGC and IC have been later identified to be galaxies, and the NGC and IC nomenclature still continues to be used for the galaxies originally in the respective catalogues.

Many of the photographic plates observed by Hubble were later compiled and published by Sandage as the Hubble Atlas of Galaxies.

The above-mentioned collections of galaxy data suffer from the major drawback that the members were not selected systematically based on some quantitative criteria, and it therefore is very difficult to draw statistical inferences regarding the galaxy properties or regarding the distribution of the galaxies either on the sky or in space. A galaxy sample which is suited for this purpose is constructed by first carrying out a sky survey. This makes an image of a part of the sky, or possibly the entire sky, with a particular limiting apparent magnitude in a well-defined band. The image contains all sources brighter than the limiting apparent magnitude. The First Palomar Observatory Sky Survey (POSS-I) was one of the first major, sensitive sky surveys. It surveyed the northern skies with a limiting apparent magnitude of 21 and 20 in the B and R bands respectively. Table 2 lists a few of the existing sky surveys at optical wavelengths. The digitized images of these surveys, originally carried out using photographic plates, are now available on line as the STScI Digital Sky Survey (DSS).

The sources in a sky survey will, in general, be a combination of stars and galaxies. Once a sky survey is available, the next step is to identify the galaxies and separate them out. This gives us a galaxy catalogue, or a galaxy survey, typically a list of the angular coordinates and the apparent magnitudes of all the galaxies brighter than the limiting magnitude. The Upsala General Catalogue of Galaxies (UGC), by Peter Nilson, is based on POSS-I. It is complete to the limiting angular diameter of 1.0′ on the blue prints of the Palomar sky survey. The catalogue is also designed to include all galaxies brighter than $m = 14.5$. Galaxies brighter than $m = 14.5$ but smaller than 1.0′ were included from the Catalogue of Galaxies and of Clusters of Galaxies (CGCG, Zwicky et al.).

Abell published a list of 2712 rich clusters of galaxies (or just clusters) visually identified from the POSS. The clusters were large concentrations of galaxies, the number of galaxies in a small region being significantly in excess of the average. To be included in the Abell catalogue, a cluster had to have at least 30 galaxies with magnitude in the range $m_1$ to $m_1 + 2$, where $m_1$ is the magnitude of the third brightest member with a distance of 1.5 Mpc. The Abell catalogue initially covered only the Northern sky ($\delta > -27$). The UK Schmidt Survey was later used to identify rich clusters in the southern sky, and an all sky catalogue (ACO catalogue) containing 4073 rich clusters was published in 1989. This continues to serve as one of the most comprehensive catalogues of rich clusters. Figure 3 shows an image of the rich cluster A2147, a member of the Abell catalogue. Studies have later shown that clusters predominantly contain elliptical galaxies, whereas field galaxies (not members of clusters) are a mixture of ellipticals and spirals.

Table 2. Some sky surveys. From Binney and Merrifield

<table>
<thead>
<tr>
<th>Survey</th>
<th>Telescope</th>
<th>Sky coverage</th>
<th>Limiting magnitude (and band)</th>
</tr>
</thead>
<tbody>
<tr>
<td>POSS-I, 1960</td>
<td>Polar 48-inch Schmidt</td>
<td>$-33^\circ \leq \delta \leq 90^\circ$</td>
<td>21 (B) 20 (R)</td>
</tr>
<tr>
<td>ESO/ESERC Southern Sky Survey, 1980</td>
<td>UK 1.2 m Schmidt and ESO 1.0 m Schmidt</td>
<td>$-90^\circ \leq \delta \leq -17^\circ$</td>
<td>23 (B) 22 (R)</td>
</tr>
<tr>
<td>POSS-II, 19</td>
<td>Polar 48-inch Schmidt</td>
<td>$-3^\circ \leq \delta \leq 90^\circ$</td>
<td>22.5 (B) 21 (R)</td>
</tr>
<tr>
<td>Second Southern Sky Survey</td>
<td>UK 1.2 m Schmidt</td>
<td>$-90^\circ \leq \delta \leq -17^\circ$</td>
<td>21.5 (R)</td>
</tr>
</tbody>
</table>
The Abell cluster A 2147, each of the extended object in the image is a galaxy.

The APM galaxy survey

The early galaxy surveys were constructed by visually inspecting the photographic plates and identifying the galaxies. The APM Galaxy Survey\(^{18}\) presents a major advancement in that the galaxy identification was automated. Photographic plates of the southern sky taken by the UK Schmidt Telescope Unit in Australia were scanned using the Automatic Plate Measuring (APM) machine. This automatic machine locates images on photographic plates, and measures the brightness, position and shape parameters for each image. The galaxy-star separation is based on the fact that a star is a point source, and hence the profile of the image of a star is decided by instrumental effects and the seeing whereas a galaxy is an extended object with an intrinsic brightness profile. The image parameters corresponding to a star are expected to lie in a well-defined region of the parameter space which allows them to be identified and separated. The limiting magnitude of the APM Galaxy Survey, in the wide blue passband \(b_j\), is \(b_j = 20\), and the galaxies have magnitudes in the range \(17 < b_j < 20.5\). The initial APM Galaxy Survey\(^{18}\) contains 2 million galaxies brighter than the limiting magnitude (Figure 4). The angular coverage is restricted to \(b \leq -40^\circ\) and \(\delta < -20^\circ\). Here \(b\) refers to the Galactic latitude, which is the angle measured from the plane of our Galaxy’s disk. The Galactic disk contains dust which causes severe extinction for sources located in the Galactic plane, and hence the plane of the Galactic disk is avoided in optical surveys. The APM survey has later been extended to cover the equatorial South Galactic Pole (SGP) area between \(\delta = +2.5^\circ\) and \(-17.5^\circ\), and also the equatorial North Galactic Pole (NGP) area between \(\delta = +2.5^\circ\) and \(-7.5^\circ\). It now contains around 3 million galaxies. The total area covered is now over 7000 deg\(^2\), roughly a quarter of the whole sky (Figure 5)\(^{19}\).

Finally, we note that there are several other optical surveys, and surveys at other frequencies (e.g. infrared, radio) which we have not touched upon in the discussion.

Redshift surveys

A galaxy survey, as described above, gives a picture of only the distribution of galaxies on the sky (e.g., Figure 4). A three-dimensional view of the galaxy distribution requires the distances to the galaxies to be known. This is achieved by carrying out spectroscopic observations of each galaxy. A redshift measurement is possible if some spectral line can be identified, and the redshift gives an estimate of the comoving distance to the galaxy (eq. 9). This gives what is known as a galaxy redshift survey. In addition to the angular coordinates and apparent magnitude, a galaxy redshift survey also gives the redshift of each galaxy.

Redshift space distortions

The measured redshift can be used (eq. (9)) to determine the comoving distance to a galaxy, it is common notation
to use $s$ to denote this distance. The observed redshift, in addition to the expansion of the universe, also has a contribution from peculiar velocities (section on 'preliminaries') and as a consequence is, in general, different from the galaxy's actual comoving distance $x$, the two being related as:

$$s = x + \frac{v}{aH(a)}$$  \hspace{1cm} (10)

where $v$ is the radial component of the galaxy's peculiar velocity away from us. Redshift surveys provide us with a three-dimensional view of the galaxy distribution, this view is somewhat distorted due to the peculiar velocities which introduce errors in the radial distances. This effect, known as redshift space distortion, has interesting consequences that we shall encounter later.

**The CfA redshift survey**

The Center for Astrophysics (CfA) redshift survey\textsuperscript{21,22} was the first attempt to map the large scale structure of the universe. The main motivation behind the survey is to assemble complete redshift information for well defined samples of galaxies to use in the statistical analysis. The CfA survey is based on a magnitude limited sample of objects from the merge of the Zwicky catalogue\textsuperscript{13} and the UGC\textsuperscript{12}. The CfA started in 1977 and completed in 1982. This survey contains 2401 galaxies (1845 in the northern galactic hemisphere and 556 in the southern Galactic hemisphere) with magnitude brighter than or equal to $m_r = 14.5$ and with galactic latitude $b \geq 40^\circ$ and $|\delta| \geq 0^\circ$ or $b \leq -30^\circ$ and $|\delta| \geq -2^\circ$. Figure 6 shows the galaxy distribution in a part of the CfA survey.

The CfA\textsuperscript{23,24} survey started in 1984/5 and completed in 1995. This extension of the earlier survey includes galaxies brighter than or equal to $m_r = 15.5$. The northern Galactic cap and the southern Galactic cap is bounded by $8^\circ \leq |\alpha| \leq 17^\circ$ in the RA and $0^\circ \leq \delta \leq 90^\circ$ in the declination and $20^\circ \leq |\alpha| \leq 4^\circ$ in the RA and $-2.5^\circ \leq \delta \leq 90^\circ$ in the declination respectively. This survey measures the redshift of about 18,000 galaxies.

**The LCRS**

The Las Campanas Redshift Survey\textsuperscript{25} (LCRS) has been carried out at the Carnegie Institution’s Las Campanas Observatory in Chile over the years 1988–1994. The survey contains over 26,000 galaxies whose spectra were measured and redshifts determined. The survey used a telescope with 2.1$^\circ$ diameter field of view and a multiobject fibre-optic spectrograph system which permits simultaneous observations of 112 galaxy spectra. The survey covers six long, thin strips each 1.5$^\circ$ in $\delta$ and 80$^\circ$ in $\alpha$, with three strips in the northern Galactic cap and three in the southern Galactic cap. The galaxies have an average redshift $z = 0.1$.

**The 2dF-GRS**

The 2dF Galaxy Redshift Survey (2dFGRS)\textsuperscript{26} is a spectroscopic survey of the galaxies identified by the APM Galaxy Survey. The survey was conducted using an instrument that can simultaneously record 400 spectra over a $2^\circ$ field of view of the sky. The 2dFGRS obtained spectra for 245,591 objects. These are mainly galaxies brighter than a nominal extinction-corrected magnitude limit of
Table 3. Characteristics of the SDSS photometric system. From Fukugita et al.31

<table>
<thead>
<tr>
<th>Band</th>
<th>$\lambda_{\text{eff}}$ Å</th>
<th>FWHM/Å</th>
<th>$Q$ (flux sensitivity)</th>
</tr>
</thead>
<tbody>
<tr>
<td>u</td>
<td>3557</td>
<td>599</td>
<td>0.0181</td>
</tr>
<tr>
<td>g</td>
<td>4825</td>
<td>1379</td>
<td>0.110</td>
</tr>
<tr>
<td>r</td>
<td>6261</td>
<td>1382</td>
<td>0.101</td>
</tr>
<tr>
<td>i</td>
<td>7672</td>
<td>1535</td>
<td>0.0766</td>
</tr>
<tr>
<td>z</td>
<td>9097</td>
<td>1370</td>
<td>0.0354</td>
</tr>
</tbody>
</table>

$b_j = 19.45$. Reliable (quality ≥3) redshifts were obtained for 221,414 galaxies. The galaxies cover an area of approximately 1500 deg$^2$ selected from the extended APM Galaxy Survey in three regions: an NGP strip, an SGP strip and random fields scattered around the SGP strip. The sky coverage is shown in Figure 5. The survey is now complete, and the Final Data Release$^{19}$ is dated 30 June 2003. Figure 7 shows the galaxy distribution in the 2dFGRS. We see that the 2dFGRS traces out the galaxy distribution to a redshift of $z \sim 0.20$ or a distance of $\sim 600 \, h^{-1}$ Mpc.

The Sloan Digital Sky Survey

The Sloan Digital Sky Survey (SDSS)$^{27,28}$ is the largest galaxy redshift survey to date. It is a wide-field photometric and spectroscopic survey of the high galactic latitude sky visible from the northern hemisphere. It employs a specially designed 2.5 m telescope with a 3$^\circ$ field of view, a mosaic CCD camera, and dual fibre-fed spectrograph, at Apache Point Observatory, New Mexico to obtain five band (u, g, r, i, z) digital photometry and spectroscopy over effectively the full range of optical wavelengths.

Thus the SDSS consists of two surveys: one is a photometric survey mapping the sky across the optical region of the electromagnetic spectrum and a follow-up spectroscopic survey over that same region.

Photometric survey: The photometric survey is done simultaneously across five optical bands: u, g, r, i and z, in order of increasing wavelength. The band definitions are given in Table 3. The imaging array consists of 30, 2048 x 2048 Tektronix CCDs, placed in an array of six columns and five rows. The telescope scanning is aligned with the columns and each row observes the sky through a different filter. The imaging is done using drift scanning mode, i.e. the camera continually sweeps the sky in great circles. Since the camera contains 6 columns, the result is a long strip of six scan lines, containing almost simultaneously observed five band data for each of the six CCD columns. The scanline columns are separated by roughly 0.2$^\circ$. When subsequent scans are interlaced to fill in their respective gaps, they form a imaging region roughly 2.5$^\circ$ wide. This associated pair of scans is referred to as a stripe and typically run between 90$^\circ$ and 120$^\circ$ in length.

Spectroscopic survey: After a region of the sky has been imaged by the photometric survey, objects are detected,
Figure 7. Galaxy distribution in the 2dFGRS. Each point in the figure represents a galaxy. We are located at the centre, and the redshift gives the radial distance to any galaxy. The galaxy positions are plotted as a function of the radial distance $r$ and the right ascension $a$. The third coordinate $\delta$, where the extent is smaller (Figure 5), has been collapsed. From the 2dFGRS website.39

Figure 8. Sky coverage of the data included in DR7 shown in an Aitoff equal area projection in J2000 equatorial coordinates. The centre of each panel is at $a = 120^\circ = 3^\circ$ and the plots cut-off at $\delta = -25^\circ$, below which the SDSS do not extend. The contiguous region covered in NGP and three strips in SGP are shown in shaded black and grey. In addition, several other strips are shown in red, blue and green which are for auxiliary imaging data, SEGUE (Sloan Extension for Galactic Understanding and Exploration) imaging scan and additional runs respectively. From Abazajian et al.40

classified as point source and extended, and measured by the image analysis software. These imaging data are used to select in a uniform way different classes of objects whose spectra will be taken. A number of selection algorithms are applied to the associated data set to produce a list of targets for follow-up spectroscopy. The spectroscopy is done using two fibre-fed optical spectrographs mounted on the same 2.5 m telescope. These spectrographs can observe 640 objects simultaneously. 640 holes are drilled in an aluminum plate, with each hole corresponding to the position of a selected star, galaxy, or quasar. Then the holes are plugged with optical fibre cables. The fibres capture light from the 640 objects simultaneously and send it into the two spectrographs. The spectrographs split the light form each object into composite colors, and the resulting spectra are recorded using CCDs. Each spectrum is measured from 3800 Å (blue) to 9200 Å (near infrared) on 2048 × 2048 CCDs.

The sky coverage of the photometric survey and the spectroscopic survey are respectively shown in Figure 8.

Current status: The main spectroscopic galaxy sample of the SDSS includes objects having Petrosian magnitude of $r < 17.77$ after correction for Galactic extinction. It is designed to measure $10^6$ galaxy redshifts over ~104 deg$^2$ of sky. The latest and final release of the SDSS data, the SDSS DR730 has been publicly released now. As of now, the SDSS DR7 has covered 11,663 deg$^2$ of imaging data, and spectroscopy is now complete over a contiguous area of 7500 deg$^2$ in the northern Galactic cap, closing the gap that was present in previous data releases. The SDSS DR7 data release includes 1.6 million spectra in total, including 930,000 galaxies, 120,000 quasars and 460,000 stars.

The galaxy distribution: The main galaxy sample of SDSS, whose spatial distribution is shown in Figure 9, has a median redshift of 0.1. The luminous red galaxy (LRG) sample extends to as large as redshift of 0.5 (Fig-
Figure 9. Equatorial distribution of right ascension and redshift for main galaxy sample within 6° of the equator in a LSS sample of the SDSS. From Blanton et al.33.

The LRGs are very large, very old galaxies that typically reside in the centres of massive galaxy clusters. They typically have little interstellar dust and no active star formation. Because they are more luminous than normal field galaxies, they can be observed at greater distances for a given magnitude limit and their stable colours make them relatively easy to pick out from the rest of the galaxies32 using the SDSS multi-band photometry. The redshift obtained for these galaxies are the photometric redshift.

The SDSS also has a quasar survey which has redshifts for quasars as far as $z = 5$ and the imaging survey has been involved in the detection of quasars beyond a redshift 6.

Data products: Following data products are available from the SDSS database:

- Image parameters: positions, fluxes and shapes of all detected objects.
- Spectroscopic parameters: redshift, spectral classification and detected lines of each spectrum.
- Colour images: JPEG images constructed from the imaging data. A GIF image of each spectrum with features identified.
- Images: FITS image files of the corrected frames in five bands.
- Spectra: The flux- and wavelength-calibrated sky subtracted spectra with error and mask arrays.
- Other data products: Summaries of observing conditions for imaging fields and for spectroscopic plates.

Fits images, spectra and catalogue tables are served by the Data Archive Server (DAS) of SDSS. Catalogues and JPEG images can be accessed through the Catalog Archive Server (CAS) database of SDSS. The data is now available online34,35.

The luminosity functions

Galaxies span a wide range in luminosity, and it is interesting to ask 'How are the galaxy luminosities distri-
and the galaxy density is low contain mainly spiral galaxies, and the luminosity function shows a deficit in the number of more luminous galaxies compared to a region of average galaxy density. In contrast, regions of high galaxy density (clusters) contain very luminous elliptical galaxies.
mainly focus on quantifying the large-scale structures seen in the galaxy distribution. Presumably, the observed clustering dies away on sufficiently large-scales where the universe is homogeneous.

**Correlation functions**

The probability $dP_1$ of finding a galaxy in a volume element $dV_1$ located at $\bar{x}_1$ is $dP_1 = \bar{n} dV_1$, where $\bar{n}$ is the mean galaxy density. Let us now ask the question: What is the joint probability $dP_{12}$ of finding two galaxies, one in $dV_1$ at $\bar{x}_1$ and another in $dV_2$ at $\bar{x}_2$? In case the galaxies are randomly distributed, the answer would be $dP_{12} = \bar{n} dV_1 dV_2$, which is the product $dP_{12} = dP_1 \times dP_2$ of the individual probabilities. As already mentioned, it is quite evident from the pictures of the galaxy distribution that the galaxies have a clustered distribution, there are some regions which are densely filled with many galaxies and there are other regions which are largely devoid of galaxies. In such a situation, the joint probability is not the product of the individual probabilities, and we express $dP_{12}$ as

$$dP_{12} = \bar{n}^2 [1 + \xi'(\bar{x}_1, \bar{x}_2)] dV_1 dV_2.$$  

Here $\xi'(\bar{x}_1, \bar{x}_2)$ quantifies the degree of clustering or the correlation in the galaxy distribution, and it is referred to as the two-point correlation function. Further, we assume that the clustering pattern is statistically homogeneous and isotropic, whereby

$$\xi'(\bar{x}_1, \bar{x}_2) = \xi(x) = |\bar{x}_1 - \bar{x}_2|.$$  

The function $\xi(x)$ refers to pairs of galaxies whose members are separated by a distance $x$. A positive or negative $\xi(x)$ respectively imply an excess or decrement in the number of such pairs as compared to a situation where the galaxies are randomly distributed.

Extending the discussion to triplets of galaxies, the joint probability $dP_{123}$ of finding a galaxy each in $dV_1$ at $\bar{x}_1$, $dV_2$ at $\bar{x}_2$ and $dV_3$ at $\bar{x}_3$ is expressed as

$$dP_{123} = \bar{n}^3 [1 + \xi'(\bar{x}_1, \bar{x}_2) + \xi'(\bar{x}_2, \bar{x}_3) + \xi'(\bar{x}_3, \bar{x}_1) + \xi(\bar{x}_1, \bar{x}_2, \bar{x}_3)] dV_1 dV_2 dV_3,$$  

where $\xi(\bar{x}_1, \bar{x}_2, \bar{x}_3)$ is the three-point correlation function. Statistical homogeneity and isotropy imply that $\xi'(\bar{x}_1, \bar{x}_2, \bar{x}_3)$ depends only on the shape and the size of the triangle formed by the three points $\bar{x}_1$, $\bar{x}_2$ and $\bar{x}_3$, and it does not depend on the origin or the orientation of the triangle. The length of the three sides of the triangle $x_{12} = |\bar{x}_1 - \bar{x}_2|$, $x_{23} = |\bar{x}_2 - \bar{x}_3|$ and $x_{31} = |\bar{x}_3 - \bar{x}_1|$ are adequate for this purpose and we have $\xi(\bar{x}_{12}, \bar{x}_{23}, \bar{x}_{31})$.

A positive value of the correlation functions $\xi$ and $\xi'$ indicates galaxy clustering at the length-scale in question, whereas a negative value indicates anti-clustering. Higher order correlation functions, like the four-point correlation function, can be defined in a similar fashion, but these are difficult to estimate and we do not consider this here.

In one of the first studies, Groth and Peebles measured the angular two-point correlation function $w(\theta)$ of the galaxies in the Lick catalogue. The angular two-point correlation function is similar to the spatial two-point correlation function $\xi$ discussed earlier, the former is used when only the galaxy angular coordinates are available. They found that a power law

$$w(\theta) = A \theta^\gamma$$

with $A = 0.0684 \pm 0.0057$,  

$$\gamma = 1.741 \pm 0.035,$$  

where $\theta$ is in degrees, gives a good fit to the observed data at $\theta < 2^\circ$. The observed correlation falls rapidly below the power law at $\theta > 2^\circ$. Interpreting these observations in terms of the spatial two point correlation function, they find that this is a power law

$$\xi(r) = (r/r_0)^{-1.17}$$

with $r_0 = 4.7 \ Mpc$,  

over the range of length-scales $0.05 \ Mpc \leq r \leq 9 \ Mpc$. Here the length-scale $r_0$ is referred to as the correlation length. They also measured the angular three-point correlation function. Interpreting this in terms of the spatial three-point correlation function, they find that this can be expressed as

$$\xi(x_{12}, x_{23}, x_{31}) = Q(\xi(x_{12})\xi(x_{31}) + \xi(x_{12})\xi(x_{23}) + \xi(x_{23})\xi(x_{31})).$$
where \( Q = 1.29 \pm 0.21 \) is a constant independent of the shape or size of the triangle. Equation (17) is often referred to as the hierarchical form for the three-point correlation. Subsequent measurements of the angular-two-point correlation function and the inferred \( \xi \) of the APM galaxy survey\(^{48,49} \) are roughly consistent with the earlier findings.

Redshift surveys allow the two-point correlation function to be measured in redshift space. As noted earlier, redshift space is somewhat different from the actual physical space because of peculiar velocities. For the LCRS, the observed two-point correlation function\(^{30} \) is found to be well fit by a power law

\[
\xi(s) = (s/s_0)^{\gamma} \quad \text{with} \quad s_0 = 6.28 \pm 0.27 \, h^{-1} \, \text{Mpc}, \\
\gamma = 1.52 \pm 0.03,
\]  

(18)

over the range of length-scales \( 2.0 < s < 16.4 \, h^{-1} \, \text{Mpc} \). They find that \( \xi(s) \) has a zero crossing between 30–40 \( h^{-1} \, \text{Mpc} \) beyond which it is consistent with zero. The LCRS real-space correlation function\(^{31} \) is well fitted by the power law

\[
\xi(r) = (r/r_0)^{\gamma} \quad \text{with} \quad r_0 = 5.06 \pm 0.12 \, h^{-1} \, \text{Mpc}, \\
\gamma = 1.862 \pm 0.034.
\]  

(19)

The LCRS three-point correlation function\(^{31} \), both in redshift space and in real space is found to be roughly consistent with the hierarchical form (eq. (17)). There is some evidence for a small, but significant, variation of \( Q \) with the shape and size of the triangle.

Figure 12 (Peacock et al.)\(^{52} \) shows the redshift space two-point correlation function \( \xi(\sigma, \pi) \) for the 2dFGRS.

Here, \( \sigma \) and \( \pi \) respectively refer to the transverse and radial components of the pair separation \( s \). Although the real space two-point correlation function \( \xi(s) \) is an isotropic function of the pair separation \( s \), the redshift space two-point correlation function is not isotropic because of the redshift space distortion caused by peculiar velocities which affect only the radial component \( \pi \) and not the transverse component \( \sigma \). The elongation along \( \pi \) seen at small scales is referred to as the ‘fingers of God’ effect. This arises from the random motions of the galaxies that are located in clusters, and this anisotropy can be modelled through a single parameter, the pairwise velocity dispersion \( \sigma_v \). The flattening along \( \pi \), seen at large scales, is known as the Kaiser effect\(^{20} \). It arises because of the coherent flow of galaxies into overdense regions and out of under-dense regions. This anisotropy can be modeled using the linear distortion parameter\(^{33} \) defined as \( \beta = \Omega_m/b \). Here \( \Omega_m \) is the matter density parameter at the epoch being probed, and \( b \) is the galaxy linear bias parameter. The galaxy bias parameter takes into account the possibility, that the galaxies may not perfectly trace the total matter distribution. The latter is believed to be largely dark, and therefore not amenable to direct observations. A bias \( b > 1 \) implies that galaxies are more clustered relative to the dark matter, and \( b < 1 \) implies just the opposite. Galaxies exactly trace the dark matter if \( b = 1 \). The best fit value for \( \beta \) from the anisotropy in \( \xi(\sigma, \pi) \) measured for the 2dFGRS is \( \beta = 0.43 \pm 0.07 \). This was used, in combination with results from the measurements of the Cosmic Microwave Background anisotropy, to show that \( \Omega_m \approx 0.3 \).

Hawkins et al.\(^{24} \) present a detailed analysis of the two-point correlation function, \( \xi(\sigma, \pi) \) from the 2dFGRS. The effective redshift at which their estimates are made is \( z_s \approx 0.15 \), and similarly the effective luminosity, \( L_s \approx 1.4L^* \). They estimate the redshift-space correlation function, \( \xi(s) \) from which they measure the redshift space clustering length, \( s_0 = 6.82 \pm 0.28 \, h^{-1} \, \text{Mpc} \). They also estimate the projected correlation function, \( \Xi(\sigma) \), and the real-space correlation function, \( \xi(r) \), which can be fit by a power-law \( (r/r_0)^{\gamma} \), with \( r_0 = 5.05 \pm 0.26 \, h^{-1} \, \text{Mpc}, \gamma = 1.67 \pm 0.03 \). For \( r > 20 \, h^{-1} \, \text{Mpc}, \xi \) drops below a power law as, for instance, is expected in the popular \( \Lambda \)CDM model. The ratio of amplitudes of the real and redshift-space correlation functions on scales of \( 8–30 \, h^{-1} \, \text{Mpc} \) gives an estimate of the redshift-space distortion parameter \( \beta \). They conclude that \( \beta L = L^* \), \( z = 0 \) is \( 0.47 \pm 0.08 \), and that the present day matter density of the Universe, \( \Omega_m \approx 0.3 \), consistent with other 2dFGRS estimates and independent analysis. Figure 13 shows the 2dFGRS \( \xi(s) \) along with results from some of the other surveys.

In a series of papers, Zehavi et al.\(^{55–58} \) have analysed the two-point correlation function of the SDSS galaxies. An early study\(^{55} \) shows the inferred real-space correlation function to be well described by a power law with \( r_0 = 6.1 \pm 0.2 \, h^{-1} \, \text{Mpc} \) and \( \gamma = 1.75 \pm 0.03 \) for \( 0.1 \, h^{-1} \, \text{Mpc} \leq s \leq 5.0 \, h^{-1} \, \text{Mpc} \).
$r \leq 16 \, h^{-1} \, \text{Mpc}$, in reasonable agreement with the results from earlier surveys. The large number of galaxies available in the 2dFGRS and the SDSS makes it possible to study if the clustering depends on the type of galaxy. These studies\textsuperscript{55,59} show that (1) red galaxies are more strongly clustered than blue galaxies, (2) more luminous galaxies are more strongly clustered than the less luminous ones, (3) elliptical galaxies are more strongly clustered than the spirals. It should be noted that ellipticals are red and more luminous in comparison so spirals which are blue, so the three points mentioned earlier are possibly correlated and not independent.

While the SDSS main galaxy sample, and the earlier surveys largely probe the galaxy clustering at $z \approx 0.1$, it is possible to probe galaxy clustering at length-scales 0.3–40 $h^{-1} \, \text{Mpc}$ at redshifts 0.16–0.44 using the SDSS luminous red galaxy (LRG) sample (sub-section titled ‘The galaxy distribution’). Zehavi \textit{et al.}\textsuperscript{58} study the redshift-space two-point correlation function $\xi(s)$, the projected correlation function $w_p(r_p)$, and the deprojected real-space correlation function $\xi(r)$, for approximately volume limited samples. They find that the galaxies are highly clustered, with the correlation length varying from $9.8 \pm 0.2$ to $11.2 \pm 0.2 \, h^{-1} \, \text{Mpc}$, dependent on the specific luminosity range. For the $-23.2 < M_r < -21.2$ sample, the inferred bias relative to that of $L*$ galaxies is $1.84 \pm 0.11$ for $1 \, h^{-1} \, \text{Mpc} < r < 10 \, h^{-1} \, \text{Mpc}$, with yet stronger clustering on smaller scales. They detect luminosity dependent bias within the sample but see no evidence for redshift evolution between $z = 0.2$ and $z = 0.4$. They find a clear indication for deviations from a power-law in the real-space correlation function, with a dip at $\sim 2 \, h^{-1} \, \text{Mpc}$ scales and an upturn on smaller scales. The precision measurements of these clustering trends offer new avenues for the study of the formation and evolution of these massive galaxies.

Gaztañaga \textit{et al.}\textsuperscript{60} have measured the 2dFGRS galaxy three-point correlation function parameterized using $Q$ (eq. (17)). They use the observed dependence of $Q$ on the shape and size of the triangle to infer that the galaxy clustering is consistent with a picture where it grows through the process of gravitational instability\textsuperscript{61} from some initially small density fluctuations, these fluctuations being a Gaussian random field. Further, they use these observations to determine the galaxy bias. Kayo \textit{et al.}\textsuperscript{62} have studied carried out similar studies using the SDSS. They find that $Q$ is nearly scale-independent. They also study if $Q$ depends on the galaxy colour, morphology and luminosity, and do not detect any statistically significant dependence. Nichol \textit{et al.}\textsuperscript{52} have measured the three-point correlation function in the SDSS, and show that the estimate of $Q$ is considerably influenced by the presence of a single, very large structure the ‘Sloan Great Wall’ (at $z \sim 0.08$). They propose that a larger volume needs to be surveyed for it to be a ‘fair sample’ of the universe for the estimation of $Q$.

\textbf{The power spectrum}

The power spectrum $P(k)$ is the Fourier transform of the two-point correlation function

$$P(k) = \int d^3 x \xi(x)e^{i k \cdot x},$$

(20)

here $k$ is the comoving wavenumber in $h \, \text{Mpc}^{-1}$. While $\xi(x)$ is usually measured by counting pairs of galaxies, $P(k)$ is estimated using a different approach\textsuperscript{63,64} where the galaxy distribution is treated as a density field which is decomposed into modes. The power spectrum has several advantages\textsuperscript{53} over the two-point correlation function at large scales where the clustering signal is extremely small.

Much effort has gone towards estimating the galaxy power spectrum from different surveys. We mention here some of the work carried out during 1992–96: CfA2\textsuperscript{55}, IRAS\textsuperscript{65,66}, Southern Sky Redshift Survey (SSRS)\textsuperscript{68}, APM\textsuperscript{69,70}, ICRS\textsuperscript{71,72}.

In the currently accepted cold dark matter (CDM) scenario\textsuperscript{53,74}, initially small fluctuations grow by the process.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{fig13.png}
\caption{Comparison of 2dFGRS $\xi(s)$ with (a) other $b$-band selected surveys and (b) R-band selected surveys. From Hawkins \textit{et al.}\textsuperscript{57}.}
\end{figure}
of gravitational instability to give the large-scale structures observed at present. The initial density fluctuations are assumed to be a Gaussian random field, whose properties are completely specified by the power spectrum (or equivalently the two-point correlation) which is predicted to be of the form

$$P(k) = Ak^n T^2(k),$$  \hspace{1cm} (21)

where $n_s \approx 1$, $T(k)$ is the transfer function, and $A$ the normalization coefficient. On length-scales where the density fluctuations are small, and linear theory can be applied, the shape of the power spectrum does not change as it grows through the process of gravitational instability. The shape of the power spectrum changes from eq. (21) at length-scales where the density fluctuations become large (nonlinear). We expect the power spectrum at present to be well described by eq. (21) at large scales (small $k$) where the density fluctuations are observed to be small, whereas there would be deviations at small scale which are nonlinear. The present value of the normalization coefficient $A$ is determined through observations of the CMBR anisotropies$^{53}$. The transfer function $T(k)$ is unity for $k \ll k_{eq} = 0.2\Omega_{m0} h^2$ Mpc$^{-1}$, and $T(k) \sim (k/k_{eq})^{-2}$ for $k \gg k_{eq}$. The accurate form of $T(k)$ can be calculated numerically for any cosmological model. Thus, the power spectrum is expected to be $P(k) = Ak$ at $k \ll k_{eq}$, have a maximum around $k \approx k_{eq}$ and then decline at larger $k$. A point to note is that eq. (21) refers to the power spectrum of the fluctuations in the underlying dark matter distribution, it is necessary to allow for a possible bias when comparing this to the observed galaxy power spectrum $P_g(k)$, and $P_g(k) = h^2 P(k)$.

Tegmark et al.$^{76}$ measure the large-scale real-space power spectrum $P(k)$ using the SDSS galaxies with a mean redshift $z = 0.1$ (Figure 14). Their results suggest that the bias is independent of length-scale to better than a few percent for $k < 0.1$ h/Mpc. They find that the power spectrum is not well characterized by a single power law, but unambiguously shows curvature. As a simple characterization of the data, their measurements are well fit by a flat scale-invariant adiabatic cosmological model with $h\Omega_m = 0.213 \pm 0.023$ and $\sigma_8 = 0.89 \pm 0.02$ for L* galaxies, when fixing the baryon fraction $\Omega_b/\Omega_m = 0.17$ and the Hubble parameter $h = 0.72$. Note that $\sigma_8$ here is the rms mass fluctuation in a sphere of comoving radius 8 h$^{-1}$ Mpc, this can be calculated if the power spectrum is known. Figure 15 shows the SDSS $P(k)$ in comparison to various other estimates of $P(k)$ and the CDM power spectrum.

Cole et al.$^{77}$ measure the power spectrum in the final 2dFGRS. They are confident that the 2dFGRS power spectrum can be used to infer the matter content of the universe. On large scales, their estimated power spectrum shows evidence for the ‘baryon oscillations’ that are predicted in CDM models. Fitting to a CDM model, assuming a primordial $n_s = 1$ spectrum, $h = 0.72$ and negligible neutrino mass, the preferred parameters are $\Omega_m h = 0.168 \pm 0.016$ and a baryon fraction $\Omega_b/\Omega_m = 0.185 \pm 0.046$ ($1\sigma$ errors). This analysis implies a density significantly lower than the best-fit value for CDM.
below the standard $\Omega_m = 0.3$: in combination with CMB data from WMAP, they infer $\Omega_m = 0.231 \pm 0.021$.

Tegmark et al.\textsuperscript{74} measure the large-scale real-space power spectrum $P(k)$ using the SDSS LRG sample and use this measurement to sharpen constraints on cosmological parameters from the Wilkinson Microwave Anisotropy Probe (WMAP). Results from the LRG and main galaxy samples are consistent, with the former providing higher signal-to-noise. Combining only SDSS LRG and WMAP data places robust constraints on many cosmological parameters. They also detected a clear signature of baryon oscillations.

Percival et al.\textsuperscript{79} measure power spectrum of galaxies in the combined Main galaxy and LRG SDSS Data Release 5 (DR5) sample. The aim of their analysis is to consider how well they can measure the cosmological matter density using the signature of $k_{\text{eq}}$ embedded in the large-scale power spectrum. The new data constrains the power spectrum on scales 100–600 h$^{-1}$ Mpc with significantly higher precision than previous analysis of just the SDSS Main galaxies, due to their larger sample and the inclusion of the LRGs. This improvement means that they can now reveal a discrepancy between the shape of the measured power and linear CDM models on scales $0.01 < k < 0.15$ h Mpc$^{-1}$, with linear model fits favouring a lower matter density ($\Omega_M = 0.22 \pm 0.04$) on scales $0.01 < k < 0.06$ h Mpc$^{-1}$ and a higher matter density ($\Omega_M = 0.32 \pm 0.01$) when smaller scales are included, assuming a flat $\Lambda$CDM model with $h = 0.73$ and $n_s = 0.96$. The lower matter density favoured by fitting their SDSS data for $0.01 < k < 0.06$ h Mpc$^{-1}$ is a better match to the best-fit WMAP 3-year cosmological model, and to results from the positions of the baryon oscillations observed in the SDSS DR5 power spectrum.

Fractals and homogeneity

Modern cosmology is built on the assumption that the universe is homogeneous and isotropic on sufficiently large scales. This assumption, known as the Cosmological Principle, can be tested using galaxy surveys. The correlation analysis and the power spectrum both rely on this assumption, as they both assume that it is possible to define a mean density. Coleman and Pietronero\textsuperscript{80} propose that the universe has a fractal structure. If true, this implies that the Cosmological Principle does not hold and it is not possible to define a mean density. While some of the subsequent analysis of galaxy surveys shows evidence for homogeneity\textsuperscript{80-83}, there are other who claim\textsuperscript{84} a fractal structure on the largest scales probed and some studies which are unable to make a strong statement either way\textsuperscript{85,86}.

In a recent study Hogg et al.\textsuperscript{87} have analysed the SDSS LRG sample. They measure the number of galaxies $N(R)$ in spheres of different comoving radii $R$, this is expected to scale as $N(R) \propto R^D$ for a fractal of dimension $D$. They find $D = 3$ at $R \sim 70$ h$^{-1}$ Mpc, indicating that the universe is homogeneous at this length-scale. Sylos Labin et al.\textsuperscript{88} have analysed the galaxy distribution in the SDSS Data Release 4, and do not find homogeneity even at the largest scales (~100 h$^{-1}$ Mpc).

Bharadwaj et al.\textsuperscript{89} and Yadav et al.\textsuperscript{90} have carried out a multi-fractal analysis of the galaxy distribution in the LCRS and SDSS respectively. The data analysed in these works is nearly two-dimensional. The multi-fractal analysis was carried out by considering the scaling of the number of galaxies $N(R)$ inside circles of comoving radius $R$ centred on galaxies. These studies find clear evidence for a transition to homogeneity at ~70 h$^{-1}$ Mpc.

Filaments in the galaxy distribution

A visual inspection of the galaxy surveys (Figures 7, 9 and 11) shows that the galaxies are distributed along filaments. These filaments encircle nearly empty regions, voids. The filaments are interconnected and form a complex network known as the ‘Cosmic Web’. Quantifying this filamentary pattern is an issue of considerable interest.

The analysis of filaments and voids in the galaxy distribution has a long history dating back to papers\textsuperscript{91-93} by Zel’dovich and collaborators in the 1980s. There have been several studies of the filaments and voids in the LCRS\textsuperscript{94-96}, PSCz\textsuperscript{27}, Abell/ACO cluster catalogue\textsuperscript{98}, 2dFGRS\textsuperscript{99} and SDSS\textsuperscript{98,100,101}. The topology of iso-density surfaces of the galaxy distribution has also been used to quantify the Cosmic Web\textsuperscript{102-104}. Here we shall restrict the subsequent discussion to an approach in which the author has been involved.

The question is ‘How to quantify the shapes of the patterns seen in the galaxy distribution’? Sahni et al.\textsuperscript{105} have proposed the Shapefinder statistics to quantify the shape of a three-dimensional surface (e.g. a sheet or a filament). The LCRS and the early versions of the SDSS are nearly two dimensional (2D), i.e. they provide a picture of the galaxy distribution in regions, one of whose dimensions is much smaller than the other two, thus, a 2D version of the Shapefinder statistics suffices\textsuperscript{106}. In the analysis, the galaxy distribution is successively coarse-grained. The galaxies occupy an increasingly larger fraction of the area of the survey as the coarse-graining proceeds. This is quantified through the Filling Factor (FF) whose value increases from ~0.01 to 1 as the coarse-graining proceeds. After each iteration of coarse-graining, a Friend-of-Friend (FOF) algorithm is used to identify connected patterns in the galaxy distribution. Each distinct interconnected set of galaxies is referred to as a cluster. We have a large number of isolated galaxies (many small clusters) at the beginning of the coarse-graining procedure. The galaxies connect up as coarse-graining proceeds, and finally all the galaxies are connected into a single cluster.
Figure 16. Largest cluster statistics and the average filamentarity ($F_2$) for two of the SDSS strips together with the values for their random counterparts. Nine realizations are used to determine the mean values and the 1–σ error-bars shown for the random data. From Pandey and Bharadwaj\textsuperscript{109}.

Figure 17. Galaxy distribution in the SDSS NGP strip. The clusters identified using FOF are shown at the coarse-graining corresponding to the percolation transition. The filamentary nature of these clusters and the interconnected network (Cosmic Web) are quite evident. From Pandey and Bharadwaj\textsuperscript{109}.

at the end of coarse-graining. The propensity of the galaxies to connect up into a single structure is quantified through the Largest Cluster Statistics (LCS) which is the ratio of area occupied by the largest cluster to the total area occupied by galaxies. The value of LCS increases as the coarse-graining proceeds and is finally LCS = 1 when all the galaxies have connected up into a single, interconnected structure the Cosmic Web. Figure 16 shows LCS as a function of FF at different stages of coarse-graining. Note the sharp rise in LCS at FF ≈ 0.5, this corresponds to the percolation transition when a substantial fraction of the galaxies connect up to the largest cluster. Also note that percolation occurs at a smaller FF in the SDSS as compared to a random point distribution, indicating a network like topology for the SDSS. For a strip of the SDSS, Figure 17 shows the clusters at the percolation transition. The filamentary nature of the clusters, and the interconnected network known, the Cosmic Web, are quite evident.

The Shapefinder $\mathcal{F}$ is used to determine the shape of all the clusters after every iteration of coarse-graining. $\mathcal{F}$ has a value 0 if the cluster is a disk, it is 1 if the cluster is a filament and the value of $\mathcal{F}$ changes from 0 to 1 as the shape of the cluster is deformed from a disk to a filament. The average filamentarity $F_2$, defined as the area...
SPECIAL SECTION: ASTRONOMY

weighted average of $F$ over all the clusters, gives an estimate of the overall degree of filamentarity after each iteration of coarse-graining. Figure 16 shows $F_2$ as a function of $F$ at different levels of coarse-graining. The Average Filamentarity increases as the coarse-graining proceeds and finally saturates at $F_2 \sim 1$ when the entire galaxy distribution connects up into the Cosmic Web. Note that the average filamentarity of the SDSS exceeds that of a random point distribution.

Initial studies, based on the LCRS, showed that (1) there was excess filamentarity\footnote{1} in the galaxy distribution compared to a random point distribution, (2) the filaments were statistically significant\footnote{2} to scales as large as 70–80 $h^{-1}$ Mpc but not beyond, (3) the observed filamentarity was consistent\footnote{3} with the LCDM model of structure formation with the need for a bias somewhat in excess of $b = 1$. The degree of filamentarity was also found to be sensitive to the galaxy bias. A subsequent study\footnote{4} using the SDSS confirmed the earlier findings. The larger SDSS data made it possible to test if the filamentarity depends on the galaxy properties\footnote{5}. It was found that at the same $F$, the red, more luminous, elliptical galaxies had a less filamentary distribution as compared to the blue, less luminous, spiral galaxies. The filamentarity of galaxy samples in different luminosity bins was compared to LCDM N-body simulations varying the galaxy bias, this was used to determine a luminosity-bias relation\footnote{6}. The filamentarity of star-forming galaxies was found to be higher for the galaxies with a low star formation rate as compared to those with a high rate\footnote{7}.

---


7. \url{http://www.messier.obspm.fr/Messier.html}


11. \url{http://stldata.stsci.edu/cgi-bin/dss_form}


19. \url{http://www.mso.anu.edu.au/2DFGRS/}


34. \url{http://www.sdss.org/dr7/products/index.html}

35. \url{http://cas.sdss.org/dr7/en/}


ACKNOWLEDGEMENTS. I am grateful to Biswajit Pandey without whose help this work could not have been completed. I thank Prasun Dutta and Prakash Sarkar for their help.