Investigating the large scale distributions of galaxies using the archival data resources

Chuan-Chin Lai

Department of Physics, National Kaohsiung Normal University

Supervisors:

Primary supervisor: Yas Hashimoto
Department of Earth Sciences, National Taiwan Normal University

Secondary supervisor: Seb Foucaud
Department of Earth Sciences, National Taiwan Normal University

ABSTRACT

Large scale structure of the universe is one of the most important research fields in the modern Astronomy because it provides us with vital information towards understating of such fundamental problems as cosmology, big bang theory, and formation and evolution of galaxies. I study the large scale structure of the universe by investigating the distribution of galaxies taken from the archival resources available through Internet. Three dimensional information, right ascension, declination, and redshift, of more than 10,000 galaxies are assembled from the Internet databases, such as NED (NASA/IPAC Extra-galactic Database) and SDSS (Sloan Digital Sky Survey). To handle such large datasets, I have extensively used various computer languages/tools, such as C, Fortran, shell script, and PGPLOT. Using these tools, I study the distribution of galaxies in the structures, such as walls, filaments, bubbles, and voids. Statistical nature of the galaxy distribution is also studied using statistical measures, such as two dimensional auto correlation function.

1 DATA

1.1 NED

NED (NASA/IPAC EXTRAGALACTIC DATABASE) is a database contains ~1,040,000 astronomical objects. NED provides different ways to search the data like 'by position', 'by near position' or 'by parameter'. NED also supports a search by batch job. I choose a range of red-shift between 0.1 and 0.105 and RA between 0 hour to 6 hour. Example of the NED data is presented in Table 1.1

#	Object Name	Type	Position	
3677	SDSS J025142.72-07125	G	02h51m42.73s	-07d12m51.0s
3678	SDSS J025143.40-07465	G	02h51m43.41s	-07d46m52.7s
3679	2dFGRS S315Z133	G	02h51m45.10s	-29d11m07.0s
3680	2dFGRS S467Z131	G	02h51m49.81s	-31d59m19.5s
3681	2dFGRS S161Z106	G	02h51m57.99s	-25d38m30.5s

Table1.1 I cut the part of table, we only choose the position like RA and Dec.

We are interested in distribution, so I use shell scripts to remove "h,m,s,d" from position, and convert the unit of RA from hour to degree to make plotting easier. Then I project the sky from 3-D to 2-D (using Aitoff Projection), then plot the all sky structure like Figure 1.1

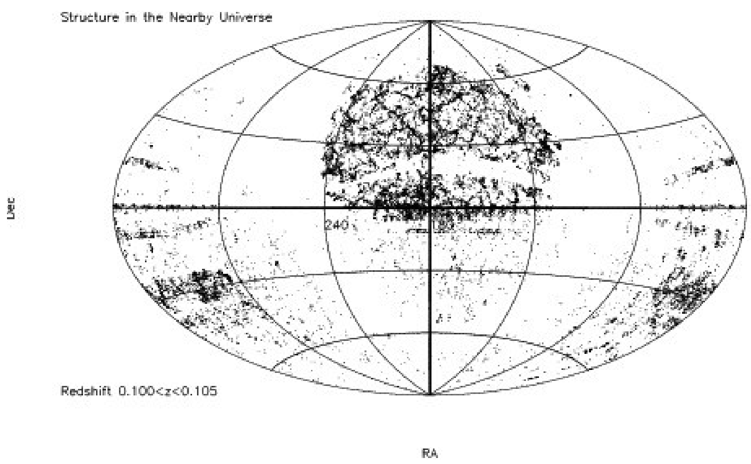

Figure 1.1
This is an all sky plot by Aitoff projection, taken from NED. The x-axis is Dec=0° and the y-axis is RA=180°. There are 30,000 galaxies in this figure, ranging red-shift between 0.1 to 0.105.

Because the data from NED are heterogeneous. Structure are not necessarily real, such as the empty strip near the center. That is why I need to use Sloan digital sky survey (SDSS) to get more homogeneous data.

1.2 SDSS

SDSS is a major multi-filter imaging and spectroscopic red-shift survey using a dedicated 2.5-m wide-angle optical telescope at Apache Point Observatory in New Mexico. SDSS provides a choice of parameters such as color, absolute magnitude in addition to spatial coordinates. I use SQL(Structured Query Language) to search the database (see Table 1.2).

Table 1.2
SQL:

```
select
z, ra, dec         //We select the parameter like red-shift z, position right ascension and declination.
from
specObj        //Get the data from this database.
where
ra BETWEEN 120 AND 240   // Right ascension between 8 hour(120 degree) and 16 hour(240 degree)
AND Dec BETWEEN -1.25 and 1.25
AND specClass = 2 AND //SpecClass = 2 denotes only galaxies can be shown.
z < 0.15 AND  //Redshift less 1.5 z
zConf > 0.35    //High confidence level in the red-shift measurements >0.35
```

Example of the SDSS data is presented in Table 1.3. There are 20,000 galaxies in this range.

z	ra	dec
0.105952	133.275050	-0.230775
0.108319	133.290050	0.984918
0.106918	133.291220	0.352864
0.107808	133.297610	0.336701
0.058980	133.298300	0.926266
0.127598	133.305600	-0.665538
0.109767	133.310040	0.342731
0.107346	133.324680	0.342489

Table 1.3
Example of the SDSS data.

Use PGPLOT to present the pie-plot(Figure 1.2)

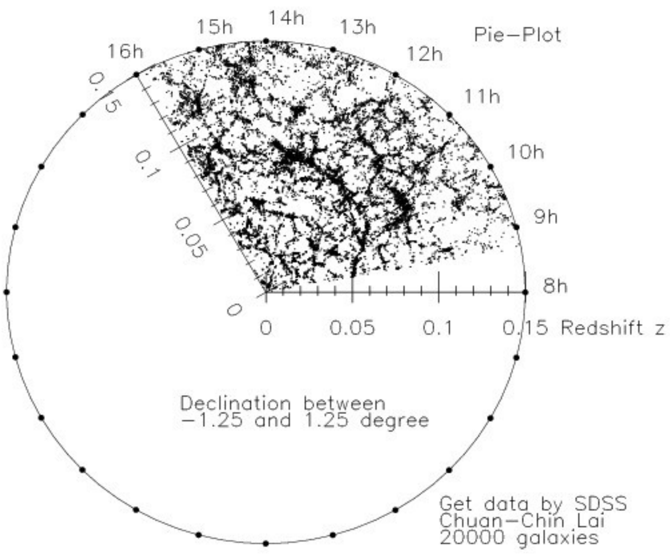

Figure 1.2
'Pie-Plot' from the SDSS data.
Declination between -1.25° to 1.25°
There are 20,000 galaxies in this figure.

We can use the SDSS data to plot the all sky structure from 3-D to 2-D by Aitoff Projection. (Figure 1.3 and Figure 1.4)

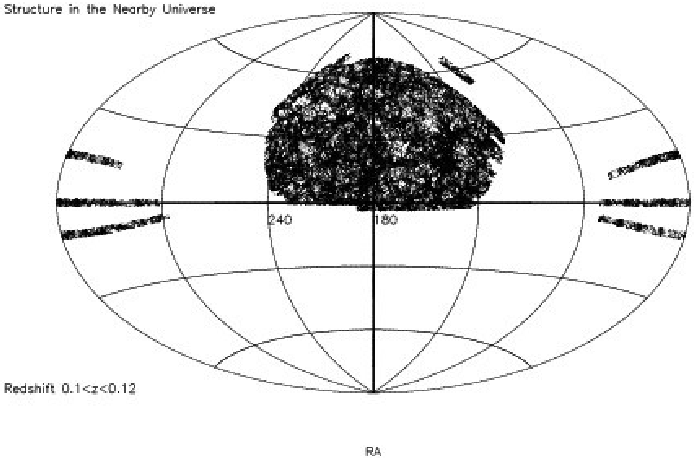

Figure1.3
All sky plot from the SDSS data. Red-shift between 0.1 to 0.12. There are~90,000 galaxies in this figure. Because the Δz is too big, the large scale structure is not obvious.

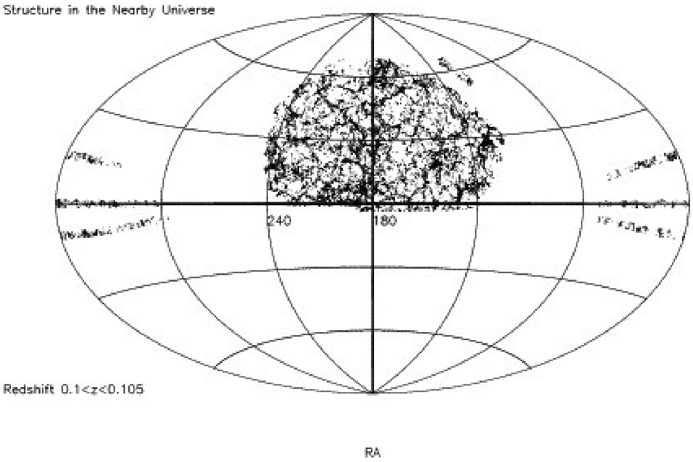

Figure1.4
All sky plot from the SDSS data. Red-shift between 0.1 to 0.105. There are ~20,000 galaxies in this figure. The large scale structure is obvious.

2 STATISTICAL ANALYSES

2.1 Outline

After we get the data of galaxy from NASA/IPAC Extra-galactic database (NED) and Sloan digital sky survey(SDSS), we can make some statistical analyses such as correlation function and power spectrum, I choose the angular correlation function which is two-dimensional correlation function. After measuring the angular correlation function, I applied the correction called "Integral Constraint", which will compensate the unwanted effect caused by the data boundary.

2.2 Two-point angular correlation function ω(θ)

If we have N points distributed over an area A, then we can express the surface density ρ by ρ = N/A, and the two-point angular correlation function ω(θ) by $\delta P = \rho^2 dV_a dV_b[1+\omega(\theta)]$, where δP is the probability to find the objects in each of two solid angle elements dV_a and dV_b separated by angle θ. In order to estimate ω(θ) from a distribution of galaxies, we need to measure the angular separation θ of all galaxy pairs and count how many galaxies at that separation (Figure 2.1) to deduce DD(θ), and then RR(θ) (Figure 2.2) by random distribution.

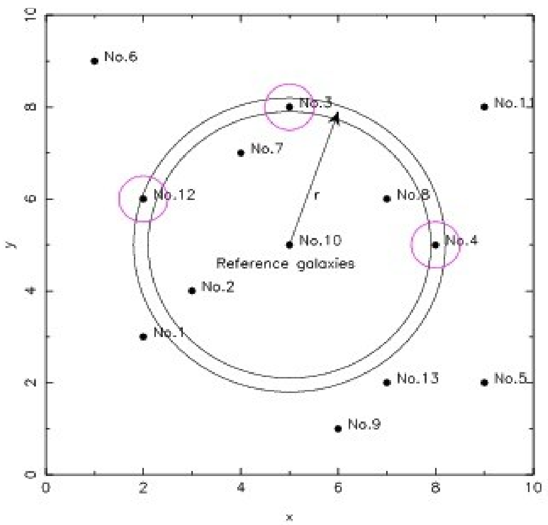

Figure 2.1
Explanation of how to find DD and RR. If figure 2.1 we have 13 points distributed over this area. First, we select the angular separation r_1 from a reference point No.10, I discover there are three points at this separation. Then, change the reference point then sum the all pairs, to deduce $DD(r_1)$. Change r_1, find the every r_i vs. $DD(r_i)$

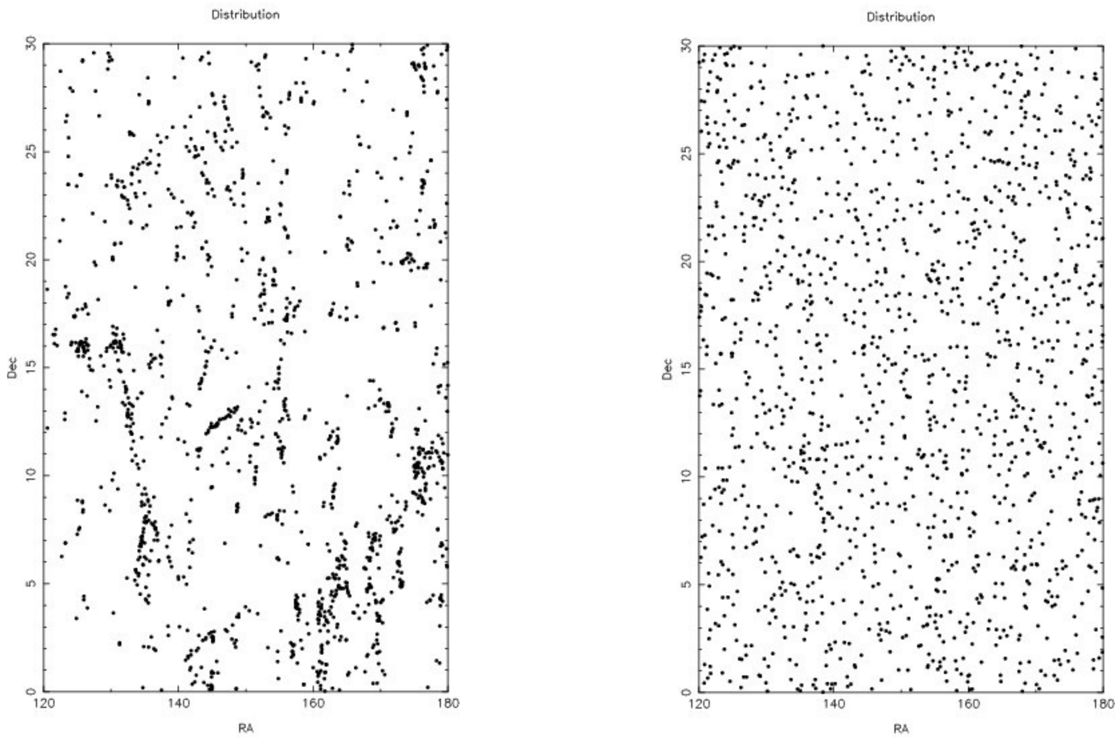

Figure 2.2

The left hand panel is the galaxy distribution over an example area, while the right panel is the simulated random distribution over the same area. Both panel contain 1,382 points in the area in this example.

There are a lot of different ways to measure the angular correlation function such as :

$\omega(\theta)= \frac{DD}{RR}-1$ (Peebles 1980)

$\omega(\theta)= \frac{DD-DR}{RR}$ (Hewett 1982)

$\omega(\theta)= \frac{DD-2\mathrm{DR}+RR}{RR}$ (Landy & Szalay 1993).

I choose the $\omega(\theta)= \frac{DD}{RR}-1$ (Peebles 1980), the simplest to measure the angular correlation function. Where DD means the number of galaxies at angular separation r counted by an observer centered on a real galaxy, RR is same mean but in random distribution. Then we statistic how many galaxies can be connected in the angular separation.

3 Result

3.1 Without integral field correction

I use the C random number generator and count DD and RR using 1.382 galaxies, The result of ω vs. θ is plotted in Figure3.1

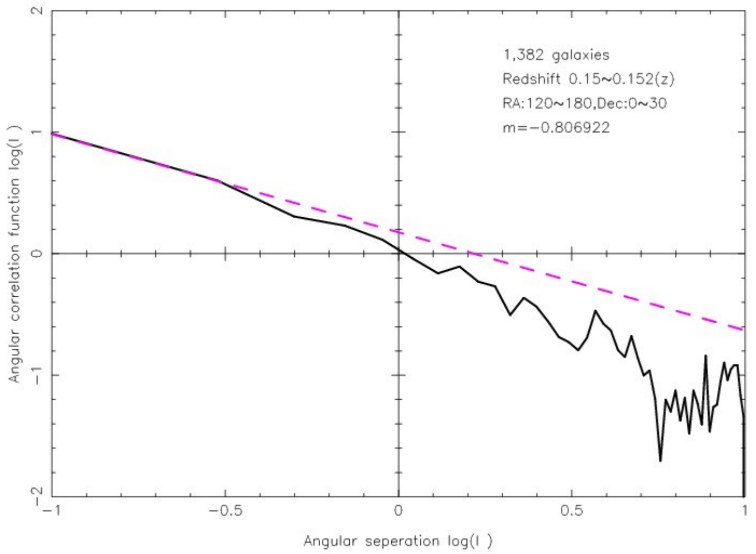

Figure 3.1
This plot is the ω vs. θ. The x-axis is log(θ) and the y-axis is log(ω). The solid line is the measured value and the dashed line is the theoretical value.

From Figure 3.1, we find out when the angular separation increases, the measured value would decrease. When the angular separation is large, some reference point near boundary of the data will miss some target points (see Figure 3.2), so we need to add a term (Integral Constraint) to correct the measured value.

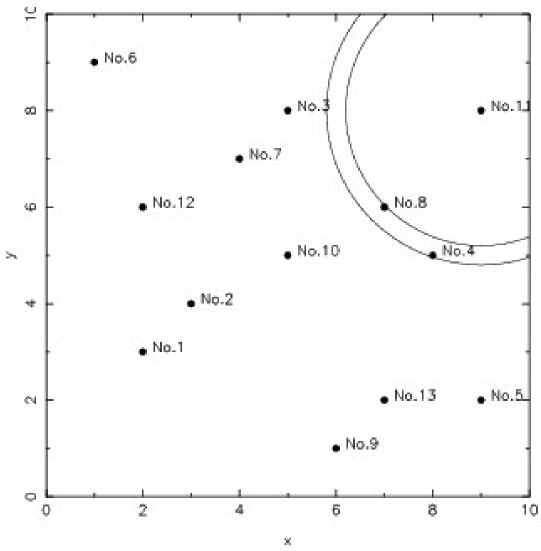

Figure3.2
When we choose No.11 point to be the point, from the plot, we know that we have 2 points in this range, but there maybe other points located outside of the boundary.

3.2 With Correction (Integral Constraint)

From Figure3.1, we know :

$\omega_{me}= A(\theta^{-\delta}-C) = \frac{DD}{RR}-1$

$\omega_{th=}\omega_{me}+AC$

So we define the integral constraint C= $\frac{\sum Nrr(\theta)\theta^{-\delta}}{\sum Nrr(\theta)}$, here Nrr=RR when the random and real galaxy distribution have the same number of galaxies. Plotting after the integral constraint, we can find the measured value very close the theoretical line (Figure 3.3).

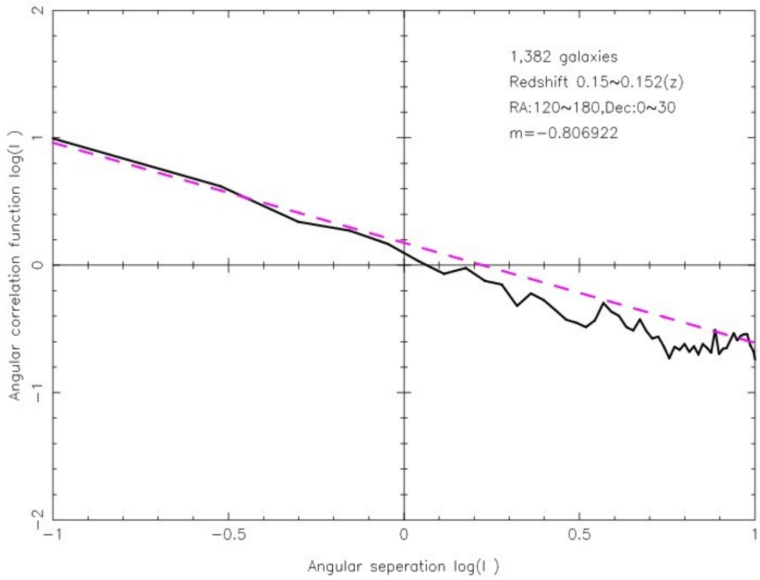

Figure 3.3
This plot is showing ω vs. θ after the Integral Field correction. The x-axis is log(θ) and the y-axis is log(ω). The solid line is the measured value and the dashed line is the theoretical value.

4 SUMMARY

Archival Data is very powerful for studying the large scale structure. From our distribution of galaxy, we know galaxies distribute in non-homogeneous way like filament, bubble, wall and void. We can present the distribution of galaxies by correlation function. Angular correlation function is useful enough to understand the distribution of galaxies. We can try further different ways to measure the ω and try more different corrections.

5 REFERENCES

Nathan Roche, Stephen A. Eales Mon Not. R. Astron. Soc. 307, 703-721 (1999)
David Woods, Gregory G. Fahlmanlthe Astrophysical Journal, 490: 11-30, 1997 November 20
Practical Statistics for Astronomers J.V Wall and C. R. Jenkins
Dark Energy; Theory and Observations Luca Amendola and Shinji Tsujikawa
NED:http://nedwww.ipac.caltech.edu/
SDSS:http://www.sdss.org/