Effect of Large Scale Structure Correlation on Cluster Counts

Michele Bianco

Munich 2017

Effect of Large Scale Structure Correlation on Cluster Counts

Michele Bianco

Dissertation
at the University Observatory Munich
Faculty of Physics at the Ludwig-Maximilians-University Munich

accomplished by
Michele Bianco
Student Number: 11379366

Munich, 22 August 2017

Wirkung der Groskala Struktur Korrelation auf Cluster Zählt

Michele Bianco

Dissertation
an der Universitäts-Sternwarte München
Fakultät für Physik der Ludwig-Maximilians-Universität München

vorgelegt von
Michele Bianco
Matrikelnummer: 11379366

München, den 22 August 2017

Supervisor: Jochen Weller

Contents

Abstract **viii**

1 Introduction **1**
1.1 Galaxy Cluster Observation 1
1.2 Current and Future Projects 3

2 Theoretical Part **5**
2.1 Cosmological Background 6
2.2 the Halo Number Counts Variance 8
2.3 Pinocchio-4.0 11

3 Analysis Part **13**
3.1 the Covariance Matrix 15
3.2 the Sample Variance and Shot Noise Ratio 17
3.3 Comparison of the Statistical Error with a Numerical Model 19
3.4 Relation with the Survey Area 23

4 Conclusion **25**

Abstract

Galaxy clusters are the largest gravitationally bound objects in the universe, they trace the highest peaks of the dark matter distribution and their abundance is therefore a considerable probe of fundamental cosmological theory.

Because of the finite size of a survey volume, the number counts of observed halos originates a *Poisson* noise, in the same way the finite size of the measured Fourier modes of the power spectrum yields to a *Halo Sample Variance* HSV, or Super-Sample variance. These two quantity contribute to the statistical error of clusters number counts: the *Halo Number Counts Variance* HNCV. At the present observations, this is currently assumed to be Poissonian, however current and upcoming surveys will increase the sample size by several orders of magnitude, from $N \sim \mathcal{O}(10^2)$ to $N \sim \mathcal{O}(10^5)$ with e.g: *eRosita*, *DES*, *Euclid*.
This change will necessitate a more sophisticated approach in the determination of the HNCV, which has to include Super-sample variance and covariance of the cluster distribution. For this reason the *Halo Number Counts Variance* will be at the center of the discussion in this dissertation and we want to test its behavior trough different mass and redshift bin for a full sky survey.

To tackle this issue we availed ourselves of a large suite of simulated cluster catalogs, using the PINpointing Orbit-Crossing Collapsed HIerarchical Objects *PINOCCHIO* code, a semi-analytic Lagrangian code that generates halo catalog on a ΛCDM cosmology.

This dissertation is organized as follows. In Chapter 1 we will briefly introduce the topic, mentioning some method of detection for galaxy cluster and summarizing the current state of play of present and future projects.
In Chapter 2 we will discuss the cosmological background on which this dissertation is based on, we will follow the theoretical development of the HNCV by Lacasa & Rosenfeld 2016 [12], Takada & Spergel 2014 [16] and Valageas et al. 2012 [17] and explain in details the simulation and the cosmology used.

Chapter 3 is where we will start our analysis, we will evaluate the *Covariance Matrix* from the mock catalog, then we will examine the relative contribution of the two aforementioned statistical fluctuation, then we will compare the data from the simulation with numerical results, then we will have a look at the behavior of the statistical error for different survey Area, changing the aperture of the past-light cone in our simulations.
In conclusion with Chapter 4 we will discuss the obtained results, the technical issues and share some ideas to improve our work.

Chapter 1

Introduction

Galaxy clusters have a long history as cosmological probes. Back in the 1933 Fritz Zwicky provided the first evidence of the presence of an unseen matter, referred as dark matter, studying the *Coma galaxy cluster* (*Abell* 1656) [1].
Another important contribution has been made in the early '70, J. Richard Gott et al. [2] studied the mass-to-light ratio of galaxy clusters, helping in a first estimation of the matter density range in the universe.

Nowadays galaxy cluster remain an important cosmological probes for testing a large variety of features of the universe. In my dissertation they will be used to study the expected number count in a survey volume and we will be especially focus on the behavior of its relative statistical error, that will be explained in detail later.

1.1 Galaxy Cluster Observation

Galaxy cluster can be observed in the optical, X-ray band or by means of the Sunyaev-Zel'dovich effect [4]. In the optical regime the observable propriety is the richness, i.e: the number of galaxies, in a specific luminosity and color range within an estimated viral radius of the halo.
On the other hand the gaseous intracluster medium (ICM), an hot and highly ionized thermal plasma which is distributed smoothly throughout the whole cluster volume, has a temperature of $10 - 100 \times 10^6\, K$ degrees, making galaxy cluster the second most luminous X-ray sources in the universe [5]. The extrapolated mass of the gas M_{gas}, the X-ray luminosity L_X and temperature T_X provide a sensible observable indicator for halo mass.
The Sunyaev-Zel'dovich effect (hereafter abbreviated SZ) consists in the scatter of CMB photons, by inverse Compton effect, with free electron in the ICM, characterizing a flux decrement Y_{SZ} so that $Y_X = T_X\, M_{gas}$ define an X-ray observable that scale with it. Observations of the SZ effect provide an almost redshift independent way to detect galaxy cluster.

Halo abundance is mainly sensitive to the amplitude of the matter power spectrum $\sigma_8(z)$ and the matter density Ω_m. The real challenge, to exploit galaxy cluster for cosmology, resides in an accurate calibration of the relation between cluster observable and cluster mass.
Following the definition in David H. Weinberg et al. 2013 [4], the mass-observable relation $P(X|M,z)$ is the probability that a halo of mass M at redshift z is detected by the galaxy cluster observable X (*e.g*: richness, Y_{SZ}, L_X), so that the expected number of galaxy clusters in a volume $V(z)$ above a threshold X_{min} is:

$$N(X_{min},z) = \int_{X_{min}}^{\infty} dX \int_0^{\infty} dM\, V(z)\, \frac{dn(z)}{dM} P(X|M,z) \tag{1.1.1}$$

where dn/dM is the halo mass function. Equation (1.1.1) conceals several potential uncertainties, the first inferred assumption is that a correct redshift measurement for each cluster is given, than in practice it occurs that the extrapolated data does not represent correctly the halo mass or redshift because of a lack of knowledge about the mass-observable relation $P(X|M,z)$. For the three above mentioned methods this main uncertainty leads to pro and con.

Up-scattered CMB photons by SZ effect does not depend on the cluster distance, so the signal does not undergo to cosmological redshift, but in practice the size of the telescope's beam and the relative size of the observed object, reduce drastically the quality of the extrapolated catalog, to such an extend that the signal does not archive enough sensitivity to detect low mass clusters, making this method complex to implement.
The main advantages of optical surveys is the low mass threshold detection, with mass as low as $5 \times 10^{13}\, M_{\odot}$ [4]. On the other hand the main lack for optical cluster detection is projected effects, consisting in mistakenly consider multiple halos along the same line of sight as a single massive cluster. At $z \approx 1$ the spectrum of early-type galaxies shift into the near-IR, but with IR adaptations of optical cluster survey, astronomers are able to avoid this problem assuring cluster detection for higher redshift, depending on the specific survey.
The ICM is tied up by the cluster gravitational potential and therefore the thermal emission is a good tracer of the cluster's potential [6]. A cluster X-ray Luminosity L_X is proportional to the square of the temperature [5],

$$L_X \propto T_X^2$$

which enhance the contrast on the sky and therefore the X-ray background can be easily subtracted from the data, moreover the clusters X-ray brightness is mostly concentrated around their center, so that the projected effects are minimized [6]. The biggest disadvantage of X-ray detection is technological, because of earth's atmospheric X-ray opacity, observations have to be space based.

1.2 Current and Future Projects

At the present days there are several survey going on and future project planned that involve cluster detection in the optical (e.g: DES), X-ray (e.g: eROSITA) and SZ (e.g: SPTpol and ACTpol) spectrum. X-ray detection is considered the cleanest method for selecting galaxy clusters and there are numerous X-ray survey that play a major role in this domain, such as the *ROSAT-ESO Flux Limited X-ray* (REFLEX) that sampled 452 clusters in an area of $14'000\,deg^2$ on a redshift $z < 0.5$ [6].
The launch of *eROSITA* should detect $\sim 10^5$ objects over the full sky out to $z \approx 1$, so that the X-rays will be ensured to play a chief role in the observation and study of galaxy clusters. This examples show an increasing mass threshold with redshift, typical of flux limited survey, to better illustrate the assertions just made, we present Figure 1.1, realized by David H. Weinberg et al. 2013 [4].

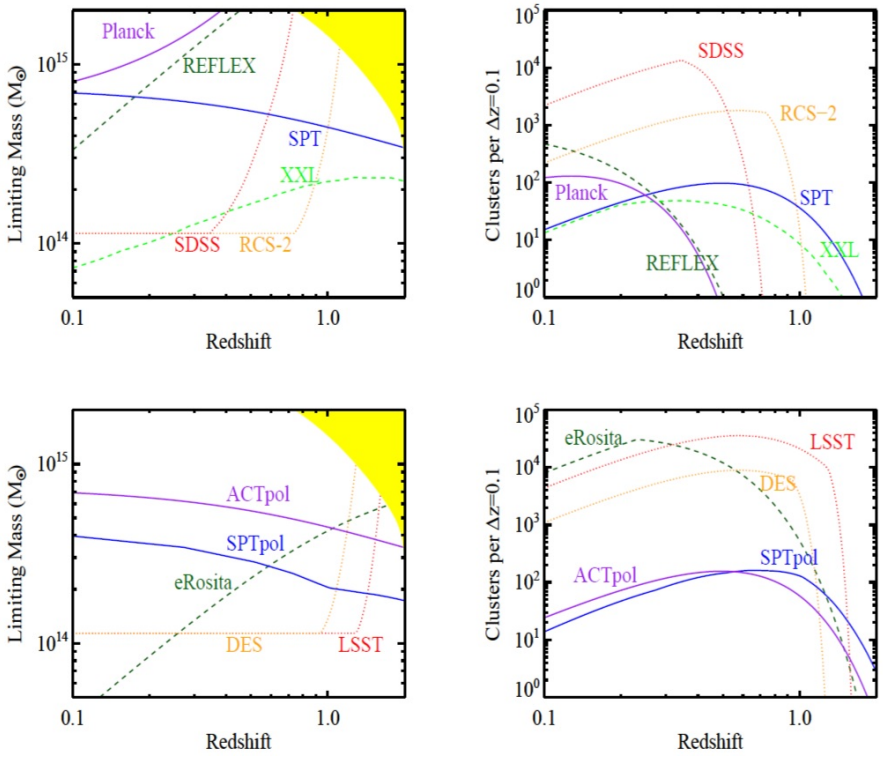

Figure 1.1: Selection of surveys
This comparisons was prepared by David H. Weinberg et al. 2013 [4], The top plots show present or currently ongoing surveys. The bottom plots show future surveys. On the left side we have the mass threshold as a function of redshift z and on the right side the number of galaxy cluster threshold in redshift bin $\Delta z = 0.1$. The yellow area corresponds to the region expected to have less than one object for the considered mass and redshift. The plots express only a roughly indicative mass and redshift limiting, in practice we never have a well defined threshold.

Also optical surveys have an important role in cluster detection, this methods is especially efficient for redshift $z < 1$, as shown in Figure 1.1, the *Dark Energy Survey* (DES) works for low mass threshold $M \geq 10^{14} M_{\odot}$ over the redshift range $0.3 < z < 1$, but for larger redshift its sensitivity declines rapidly, or as the *Sloan Digital Sky Survey* (SDSS), who reaches low mass over large sky area but only extended to $z < 0.5$, another example is the *Red-Sequence Cluster Survey 2* (RCS-2) that reaches cluster at $z \approx 1$ over a small area of approximately $1'000\, deg^2$.
Above this redshift the spectrum of early type galaxy cluster shift into near-IR, but this redshift limit can be avoided with IR adaptation, with the advent of *Euclid*, the clusters research in the IR will be capable of finding object out to $z \sim 2$. Another very important upcoming project is the *Large Synoptic Survey Telescope* (LSST), it will push the optical selection limit to $z \approx 1.5$, increasing the number of galaxy clusters by one to two order of magnitude above redshift $z = 1$.

For the actual state of play of the SZ method, there are two new project, the *South Pole Telescope* (SPTpol) and the *Atacama Cosmology Telescope* (ACTpol), they should lead to significantly lower mass thresholds for SZ cluster detection then the actual STP and ATP, with a better sensitivity to high redshift clusters $z \leq 1$ [4] and with respectively a mass thresholds of $7 \times 10^{14} M_{\odot}$ and $10^{15} M_{\odot}$. Meanwhile the *Planck* SZ survey provides the largest SZ cluster catalog present nowadays, but it is only limited to massive cluster above threshold $8 \times 10^{14} M_{\odot}$, for moderate redshift $z \leq 1$.
In the following Table 1.1, we listed the most representative survey focused on cluster detection going on at the present days and near future.

Name	Survey Area (deg^2)	Method
eROSITA	$40'000$	X-ray
Planck	$30'000$	SZ
LSST	$20'000$	optical
Euclid	$15'000$	optical/IR
REFLEX	$14'000$	X-ray
SDSS	$10'000$	optical
DES	$5'000$	optical
ACTpol	$4'000$	SZ
RCS-2	$1'000$	optical
ACT	$1'000$	SZ
SPT	600	SZ
SPTpol	600	SZ

Table 1.1: Completed or currently ongoing surveys
List of survey focused on cluster detection, we specified their survey area and the detection method. In **Bold**, we highlight the surveys that we will use later, in Chapter 3, to study the relation between the statistical error and the survey Area.

Chapter 2

Theoretical Part

Modern cosmology is based on the cosmological principle, it states that our position as observer does not play a special role in the universe. This notion assumes that the distribution of matter in the universe is homogeneous and isotropic on enough large scale.

The ΛCDM cosmological model provides a good approximation of the parametrization of the Big Bang model, based on the Einstein theory. It suggests the presence of the so called cosmological constant Λ, that represent the vacuum energy or dark energy that leads the universe expansion and it composes approximately the $\sim 70\%$ of the energy density of the universe [3], than it suggests the presence of a non-relativistic cold dark matter (CDM), the aforementioned undetectable matter that composes around the $\sim 25\%$ of the energy density [3] and other components as ordinary matter $\sim 5\%$, and a remaining small fraction of relativistic matter (CMB photons, relic neutrinos) [3].

According to the standard cosmological model ΛCDM, large-scale structure have formed through gravitational instability of a small primordial fluctuation [4], this model has survived more than a decade of precise observation, last of them the *Planck 2015* data [3]. Despite it is extremely accurate for testing large-scale structure, it suffers from intrinsic theoretical unknown, *e.g:* the understanding of cold dark matter (CDM) and the cosmological constant Λ [10].

2.1 Cosmological Background

To describe the geometry of an expanding universe, with scale factor $a(t)$, we can use the Friedmann-Robertson-Walker (FRW) metric, so that the physical distance between two points separated by a distance ds, in spherical coordinates [8], is given by

$$ds^2 = -c^2dt^2 + a^2(t)\left(\frac{dr^2}{1-Kr} + r^2d\theta^2 + r^2\sin^2(\theta)d\phi^2\right) \tag{2.1.1}$$

where K is the spatial curvature that takes the values 0 if flat and ± 1 for a open or close space. The concept of metric is completely independent from any physical application [8], the connection with general relativity comes when we relates the metric to the matter and energy in the universe, so that gravitation is the result of the curvature of the space-time by indeed mass and energy [8], this is described by the *Einstein Field Equation (EFE).*

$$G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}\mathcal{R} + \Lambda g_{\mu\nu} = 8\pi G T_{\mu\nu} \tag{2.1.2}$$

to understand the evolution of the scale factor we need to consider the zero order component [8], or the time-time component of the *Einstein Field Equation* Equation (2.1.2), the *Friedmann equation.*

$$\frac{\dot{a}^2 + k}{a^2} = \frac{8\pi G\rho + \Lambda}{3} \tag{2.1.3}$$

it is often expressed in redshift $z = \frac{1}{a} - 1$ and in terms of present values of its density components:

$$H(z) = H_0\, E(z) = H_0\sqrt{\Omega_r(1+z)^4 + \Omega_m(1+z)^3 + \Omega_K(1+z)^2 + \Omega_\Lambda} \tag{2.1.4}$$

where the quantity $\Omega_i \equiv \Omega_{i,0}$ is the total i component density at the present day, H is the Hubble parameter and H_0 the Hubble constant.
A further explanation of the equality express in Equation (2.1.4) are shown in the following four equation.

$$\sum_i \Omega_i = 1 \qquad \Omega_i \equiv \frac{\rho_i}{\rho_c}$$

$$\rho_c \equiv \frac{3H_0^2}{8\pi G} \qquad H(t) = \frac{\dot{a}}{a}$$

Where ρ_c is the critical density and G is the *Newtonian* constant of gravitation.

In cosmology there are different notion of distance, but the one that will interest us is the *comoving distance*. It approximates the spacial separation between two object considering the expansion of the universe, in the case of an observer at $z = 0$, the object of interest at redshift z [9] and with the help of the Friedmann equation (2.1.4), for a flat universe ($K = 0$) and at low redshift $0 < z < 1$, we have:

$$d_C(z) = \frac{c}{H_0} \int_0^z \frac{dz'}{E(z')} \tag{2.1.5}$$

where c is the speed of light and H_0 the aforementioned Hubble constant.

We can than take the definition of the comoving volume as the volume in which the number densities of object, that follow the Hubble flow, are constant through redshift [8]. The comoving volume element per solid angle per redshift is given by

$$\frac{dV_C}{dz\, d\Omega} = \frac{c}{H_0} \cdot \frac{d_C^2(z)}{E(z)} \tag{2.1.6}$$

as we can imagine, performing the integration for full sky the results is

$$V_C(z) = \frac{4\pi}{3} d_C^3(z)$$

In order to understand the structure formation we need to study the evolution of density inhomogeneities in the expanding universe, for small variations relative to the mean background density $\bar{\rho}$ [11], the fluctuation can be expressed as:

$$\delta_i \equiv \frac{\rho_i - \bar{\rho}}{\bar{\rho}}$$

Into non-linear regime the baryonic content nature, like for example star formation or hydrodynamical effect [11], becomes crucial to describe the distribution of galaxy that in this case maybe does not accurately reproduce the distribution of dark matter halos. The common approach is to assume number density distribution $n(\mathbf{x})$ of galaxy cluster that undergo to a simple proportion with the density fluctuation $\delta(\mathbf{x})$ [11], for sufficient large volume.

$$\frac{\delta n(\mathbf{x})}{\bar{n}} = b\, \delta(\mathbf{x}) \tag{2.1.7}$$

where b is the biasing parameter.
In that way we are able to relate galaxy cluster distribution to those of density matter fluctuation.

2.2 the Halo Number Counts Variance

In a sufficiently large comoving volume survey V_s the density fluctuation distribution is considered nearly-Gaussian [16].

$$P(\delta_m|\bar{\delta}_m, \sigma_m) = \frac{1}{\sqrt{2\pi}\sigma_m} exp\left(-\frac{\delta_m^2}{2\sigma_m^2}\right) \tag{2.2.1}$$

by construction the mean is supposed to be $\bar{\delta}_m = 0$ to assume linear perturbation, then the rms mass density fluctuation in the survey volume $\langle\delta_m^2\rangle = \sigma_m^2$ is defined as:

$$\sigma_m^2(V_s) = \frac{1}{2\pi^2}\int k^2\, P_m(k) \left|\tilde{W}(k, V_s)\right|^2 dk \tag{2.2.2}$$

here $P_m(k)$ is the linear matter power spectrum and $\tilde{W}(k, V_s)$ is the Fourier transform of the window function.

Simulation can predict the number counts of cluster as a function of their mass and comoving volume [16], to do so we followed the theoretical development explained in the paper by Lacasa & Rosenfeld 2016 [12], Takada & Spergel 2014 [16] and Valageas et al. 2012 [17].
To estimate the mean number of halos within a redshift bin $[z; z + dz]$ and mass bin of width $[M; M + dM]$, we started from the following definition [12].

$$\bar{N}(M, z) \equiv \langle N(M, z)\rangle = \int dV\, d(ln\, M)\frac{dn}{d(ln\, M)}(M, z)$$

where $dn/d(ln\, M)$ is the halo mass function. In our case we can use Equation (2.1.6) to express the volume element dV in redshift dz so that, for a full sky survey, we have:

$$\bar{N}(M, z) = 4\pi \int dz\, d(ln\, M)\frac{dV}{dz\, d\Omega} \cdot \frac{dn}{d(ln\, M)}(M, z) \tag{2.2.3}$$

Here 4π comes from the full sky integration of the solid angle and $dV/dz\, d\Omega$ can be substituted by the right hand side of Equation (2.1.6).

Of course in a practical case we have to include some hamper effects like the mass-observable relation or some detector noise, as we mentioned in Chapter 1 with Equation (1.1.1). In this dissertation such complications will not be considered for simplicity.

Then the number counts of halo within a survey volume delimited by the redshift bin limits is given by

$$N(M, z) = \int dV\, d(ln\, M)\frac{dn}{d(ln\, M)}(M, z)[1 + \bar{b}(M, z)\delta_m(V_S)] \tag{2.2.4}$$

Here $\bar{N} \equiv dV[dn/d(ln\,M)]d(ln\,M)$ should be interpreted as the ensemble average expectation of the number counts, and $\bar{b}$ is the mean halo bias.
Similarly at what we said before, simulations are able to predict the bias of cluster $b(M,z)$ in function of their mass and redshift [15], we can define the mean halo bias [16] such that

$$\bar{b}(M,z) = \frac{4\pi}{\bar{N}} \int dz\, d(ln\,M) \frac{dV}{dz\, d\Omega} \cdot\, b(M,z)\, \frac{dn}{d(ln\,M)}(M,z) \tag{2.2.5}$$

Equation (2.2.3) and Equation (2.2.5) will be used later in Chapter 3 to compare the numerical prediction and the results obtained by our simulation.

Let assume that we have a data set $\mathbf{N} = (N_1, \ldots, N_n)$, where each i^{th} index refers to the cluster number counts in that specific bin, in our case it is considered as a 2-dimensional binning, mass and redshift bin, such that $i \equiv (i_M; i_z)$. To explore the number counts $N_i \equiv N(M_{i_M}, z_{i_z})$ one can use the statistical propriety of Gaussian distribution (2.2.1), than following the theoretical construction in Valageas et al. [17], for two non overlapping bins i and j we have:

$$\begin{aligned}
\langle N_i N_j \rangle &= \delta^K_{ij}\, \langle N_i \rangle^{sn} + \langle N_i\, N_j \rangle^{sv} \\
&= \delta^K_{ij} \left\langle \bar{N}_i (1 + \bar{b}_i\, \delta_m) \right\rangle + \left\langle \bar{N}_i (1 + \bar{b}_i\, \delta_m)\, \bar{N}_j (1 + \bar{b}_j\, \delta_m) \right\rangle \\
&= \delta^K_{ij} \bar{N}_i + \delta^K_{ij}\, \bar{N}_i \bar{b}_i\, \langle \delta_m \rangle + \bar{N}_i\, \bar{N}_j + \bar{N}_i\, \bar{N}_j (\bar{b}_i + \bar{b}_j)\, \langle \delta_m \rangle + \bar{N}_i\, \bar{N}_j\, \bar{b}_i\, \bar{b}_j\, \left\langle \delta^2_m \right\rangle
\end{aligned} \tag{2.2.6}$$

Where δ^K_{ij} is the *Kronecker Delta function*, with $\delta^K_{ij} = 1$ if $i = j$, otherwise $\delta^K_{ij} = 0$.

Here in the first line, the bracket on the right evaluate the *sample variance* average without Poissonian contribution, or *shot-noise*, that is expressed separately in the bracket on the left. we obtained the second line considering $N_i \equiv \bar{N}_i[1 + \bar{b}_i \delta_m]$, relation already mentioned in Equation (2.2.4).
If we consider the last line we can see that the second and the fourth term disappears, as the density fluctuation mean $\langle \delta_m \rangle$ is equal zero, in the last term the $\langle \delta^2_m \rangle$ is the rms mass density fluctuation defined with Equation (2.2.2).
From Equation (2.2.6) we can define the *Halo Number Counts Variance* (HNCV) given by

$$\sigma^2_{HNCV} \equiv \langle N_i N_j \rangle - \langle N_i \rangle \langle N_j \rangle = \delta^K_{ij} \bar{N}_i + \bar{N}_i \bar{N}_j b_i b_j \sigma^2_m \tag{2.2.7}$$

Where the first term on the right side of the equation is the Poisson error σ^2_{Poiss}, due to the finite size of the sample halos, and the second term is the halo sample variance HSV σ^2_{HSV} contribution arising due to the super-survey modes.
In Equation (2.2.7) the quantities $\bar{N}_i$, b_i and σ^2_m can be numerically evaluated once a cosmological model is chosen, meanwhile other quantity as the *Windows function* W_i and the halo mass function can be chosen.

The origin and contribution of the Poisson error is relatively trivial to understand but less clear is the importance of the additional statistical fluctuation that the halos sample σ^2_{HSV} is generating. Moreover, as we mentioned in Chapter 1, new upcoming survey will increase the halo sample size to such a level that a more complete study of Equation (2.2.7) is necessary.
For this reasons the Halo Number Counts Variance will be at the center of the discussion in this dissertation. We want to test its behavior trough different mass and redshift bin for a full sky survey. In Chapter 3 we will start with evaluating its *Covariance Matrix*, then we will examine the relative contribution of the two statistical fluctuation, then we will compare the mock data from a simulation with some numerical results, then at the end we will have a look at the behavior of the *Halo Number Counts Variance* for different survey Area with the data obtained from the simulations.

2.3 Pinocchio-4.0

In order to generate mock catalogs in a cosmological model, we used the PINpointing Orbit Crossing Collapsed HIerarchical Objects (PINOCCHIO) code. It is a semi-Lagrangian code that use second- and third-order Lagrangian Perturbation Theory LPT to describe respectively particle displacement and halo construction. It is focused to work for the standard ΛCDM cosmology model

In the last years PINOCCHIO has been tested in several ways by comparing its extrapolated halo proprieties against the ones from N-body simulations [19], as the degeneracy between Ω_m and σ_8 from the halo mass function, the two-point correlation function, the halo power spectrum, or the clustering propriety of halos up to $k = 0.3\,h/Mpc$ showing results as the original code. Regarding some propriety that will concern our dissertation, PINOCCHIO can precisely reproduces the halo density field with the same level of precision as a N-body simulation with an agreement between the analytic fitting formula and the simulation mass function of the order $\sim 10\%$ [19] in the high mass range.

The algorithm followed by PINOCCHIO code at first creates a linear density field on a regular cubic grid in the Lagrangian space, this way of starting follows indeed the same way to create initial condition in N-body simulation. This density field is subsequently smoothed on a continuous time sampling so to predict the collapsing time, adopting a ellipsoidal collapse model, at that point these particles are expected to be grouped together into dark matter halos or the filamentary networks, mimicking their hierarchical formation [18]. Accretion has considered of two type, particle-halo or halo-halo, the two object are displaced to their expected Eulerian position at the collapsing time, here halo displacement are considered as the average displacement of all the particles that compose it, accretion or merging take place if their distance is below a threshold that depend on the largest object Lagrangian radius.

A certain freedom on the choice of which order of LPT to considered is present in PINOCCHIO code, this provides a way to test effect of increasing LPT order for displacement calculation, as a matter of fact the second-order LPT is used to implement halo displacement and third-order LPT is used to halo accretion or particle merging [18].

The PINOCCHIO code creates different outputs that can be used to perform different kind of cosmological study, the one that interests us is the past-light cone file; an object is listed in such halo catalog when the halo trajectory $x_{halo}(z)$ intersects with the light cone x_{plc}.

$$|x_{halo}(z) - x_{plc}| = d_p(z) \tag{2.3.1}$$

here d_p is the proper distance. Knowing the redshift interval in which the halo crossed the past-light cone, we can solve Equation (2.3.1) by z and so determine the time at which the detection occurred.
The setup of the parameter file consisted in the definition of the standard cosmological

parameter, shown in Table 2.1, and by defining the observer position, or origin of the past-light cone, that in our case is placed at the center of the box `PLCCenter`, to defining the comoving volume we specify the orientation `PLCAxis` of the cone, its aperture `PLCAperture` and the starting and stopping redshift `StartingzForPLC` and `LastzForPLC` in which we want to run our simulation.

Variable	Value
Ω_m	0.25
Ω_Λ	0.75
Ω_b	0.044
h	0.7
σ_8	0.8
n_s	0.96
w_0	-1
w_a	0
`BoxSize`	$1200\,Mpc/h$
`GridSize`	666
`FileWithInputSpectrum`	*no*
`MinHaloMass`	100
`StartingzForPLC`	0.19
`LastzForPLC`	0.0
`PLCCenter`	$(600, 600, 600)\,Mpc/h$
`PLCAxis`	$(1, 1, 1)\,Mpc/h$
`PLCAperture`	$180\,deg$

Table 2.1: parameter file

The first part of the Table shows the parameters to set the cosmological environment. We used the *Planck 2015* data [3] to represent a flat standard ΛCDM cosmology. The second part of the Table shows the parameter necessary to set the simulation. If not specified (`FileWithInputSpectrum`: *no*), the code consider the Eisenstein & Hu power spectra.

From the parameter file the *PINOCCHIO* code sets different physical quantity, but we have particular interest in one of them, the particle mass, calculated with the following formula.

$$M_{part} = \Omega_m\,\rho_c \left(\frac{\texttt{BoxSize}}{\texttt{GridSize}}\right)^3$$

in our case this quantity is equivalent to $M_{part} = 4.059 \times 10^{11}\,M_\odot/h$. Combining M_{part} with the initial parameter `MinHaloMass` we can assure that the smallest mass in our halo catalog is $M_{min} = 4.059 \times 10^{13}\,M_\odot/h$.

Chapter 3

Analysis Part

The simulation is set up so to emulate real galaxy cluster observation, the outputs provide a past-light cone catalog file, listing the features (mass, observed redshift, velocity, etc.) of all the galaxy clusters that satisfied the condition imposed in the parameter file. The origin of this past-light cone is placed at the center of the box, the comoving survey volume is than split into shells, this length is calculated using the comoving distance formula, Equation (2.1.5).

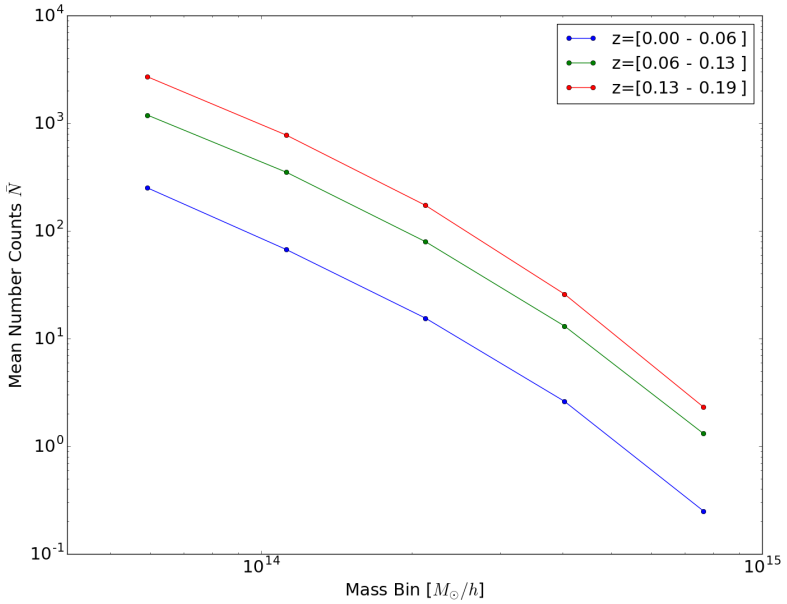

Figure 3.1: Cluster Mean Count
Comparison of the mean cluster number count between the 3 different redshift bin, blue solid line for the lowest redshift bin $z \in [0\,;\,0.06]$, green solid line for $z \in [0.06\,;\,0.13]$ and red solid line for $z \in [0.13\,;\,0.19]$. The points are situated at the center of the mass bin.

We are interested in the mass and the observed redshift of the detected galaxy clusters. We bin these observable into $n_M = 5$ mass bin, spaced in logarithmic scale $\Delta log(M) = 0.277$ between $4.1 \times 10^{13}\, M_\odot/h$ and $10^{15}\, M_\odot/h$, and $n_z = 3$ redshift bin between 0 and 0.19 with linear spacing of $\Delta z = 0.06$ (the reason for such a narrow redshift range will be explained later in the conclusions, Chapter 4).

From the binned data we arrange the number counts matrix in the form $\mathbf{N} = (N_{ij}) = (\vec{N}_1, \vec{N}_2, \ldots, \vec{N}_{n_R})$ with $i = 1, \ldots, n_R$ and $j = 1, \ldots, n_V$, where $n_R = 100$ is the number of simulations, $n_V = n_M \cdot n_z = 6$ the number of variables and $\vec{N}_i$ the data vector. Therefore the matrix element N_{ij} express the number of galaxy cluster of the i^{th} realization with the j^{th} variable.

In our case the data vector has the following form:

$$\vec{N}_i \equiv \begin{pmatrix} N_{i1} \\ N_{i2} \\ \cdots \\ N_{i\,n_V} \end{pmatrix} = \begin{pmatrix} \hat{N}_i(z_1, M_1) \\ \vdots \\ \hat{N}_i(z_1, M_{n_M}) \\ \hat{N}_i(z_2, M_1) \\ \vdots \\ \hat{N}_i(z_2, M_{n_M}) \\ \vdots \\ \hat{N}_i(z_{n_z}, M_1) \\ \vdots \\ \hat{N}_i(z_{n_z}, M_{n_M}) \end{pmatrix}$$

As a first look we show, in Figure 3.1, the mean number count calculated from the binned data for the three different redshift bin, we can see that the calculated quantity increase for higher redshift bins as these correspond to larger survey volume.

3.1 the Covariance Matrix

As a first step, we are interested in the relation between different mass bin and its evolution through different redshift, to do so we study the covariance matrix for cluster counts.
The covariance matrix is a square symmetric matrix n_V-by-n_V, with matrix elements defined as:

$$C_{jk} = \frac{1}{n_R - 1} \sum_{i=1}^{n_R} (N_{ij} - \bar{N}_j) \cdot (N_{ik} - \bar{N}_k) \tag{3.1.1}$$

where C_{jk} estimate the covariance between the j^{th} variable and the k^{th} variable. From this definition we can clearly see that the diagonal of the covariance matrix $\mathbf{C}$ is the variance of the data vector.

$$Var(N_k) = C_{kk} = \frac{1}{n_R - 1} \sum_{i=1}^{n_R} (N_{ik} - \bar{N}_k)^2 \tag{3.1.2}$$

Figure 3.2 shows the normalized cluster counts covariance matrix C_{jk}^{norm}, in absolute value. We normalized the matrix by its diagonal elements and considered the absolute value because we are only interested in the amplitude of their correlation.

$$C_{jk}^{norm} = \left| \frac{C_{jk}}{\sqrt{C_{jj} \cdot C_{kk}}} \right|$$

We can see that, at the lowest redshift, all the first three mass bins, low mass $M \leq 10^{14} M_{\odot}$, show strong correlation $C_{jk}^{norm} \geq 0.6$ amongst them, but this effect decreases for higher redshift and for far apart mass bin, keeping correlation with $C_{jk}^{norm} \geq 0.5$ only with the closes mass bin.

Due to the universe expansion, the contrast in the matter density fluctuation δ_m grows and on relatively small scale we enter into non-linear regime, for scale smaller than $\sim 10 Mpc/h$. As expected by non-linearity on small scale, the off-diagonal correlation terms increase at lower redshift.

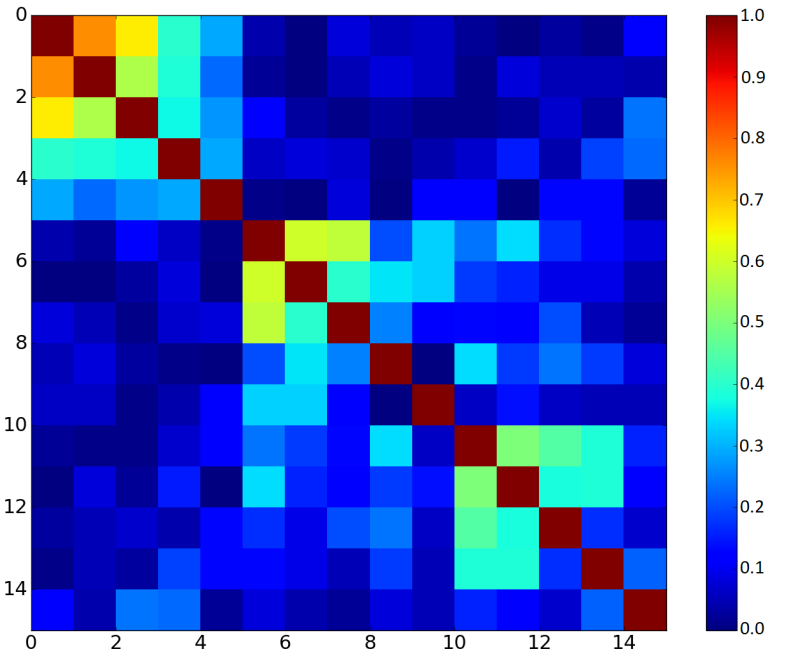

Figure 3.2: Normalized Covariance Matrix
Normalized cluster counts covariance matrix in absolute value C_{jk}^{norm}.
With 5 mass bin and 3 redshift bin.

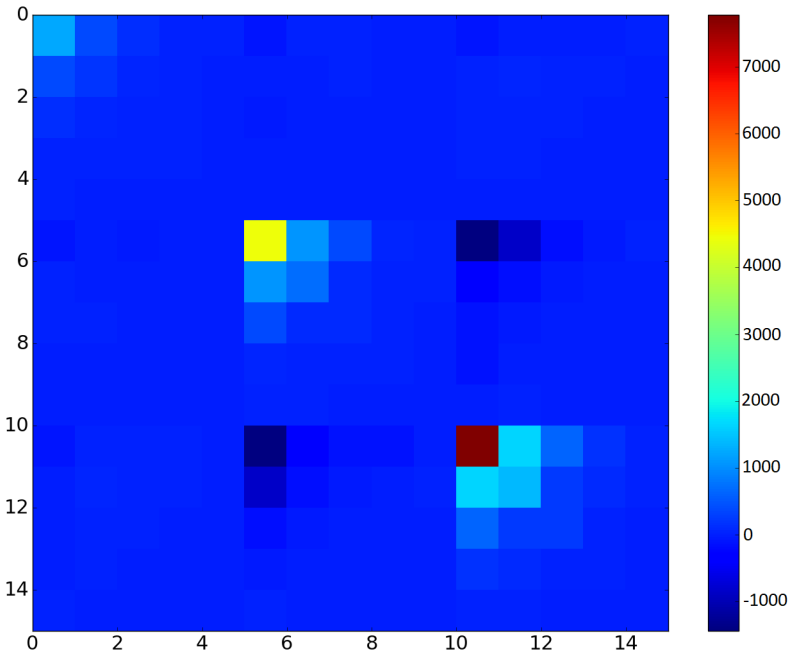

Figure 3.3: Covariance Matrix
Cluster counts covariance matrix in absolute value C_{jk}.
With 5 mass bin and 3 redshift bin.

3.2 the Sample Variance and Shot Noise Ratio

As we have seen in Chapter 2.2, there is a complete statistics to describe the variance of the halo number counts, therefore we are interested in the study of this statistical error and at first we can easily imagine that there is a dependency on the survey volume V_s of our simulation, but this section it is more about questioning on which conditions the *Poisson* error or the HSV error gives the larger contribution to the *Halo Number Counts Variance* HNCV, Equation (2.2.7).
With this purpose in mind we studied the behavior of fraction of the two above mentioned contributions.

$$\frac{\sigma^2_{HSV}}{\sigma^2_{Poiss}} = \frac{Var(N_i) - \bar{N}_i}{\bar{N}_i} \tag{3.2.1}$$

The numerical results for the value of $\sigma^2_{HSV}/\sigma^2_{Poiss}$ are shown in Table 3.1, meanwhile we plotted the results in Figure 3.4.

As we can see, there is a trend for low mass $\leq 10^{14} M_\odot/h$ to favor Sample Variance than *Poisson* Variance. In our case this effect ceases, with $\sigma_{HSV} \leq 0.8\,\sigma_{Poiss}$, at mass larger than $\sim 2 \times 10^{14} M_\odot/h$ for redshift bin Δz_1 and at the same mass bin $\Delta M_2 = [7.77\,;\,14.70] \times 10^{13} M_\odot/h$ for redshift bin Δz_2 and Δz_1.

The extreme extension of the HSV predominance takes place at the lowest mass bin $\Delta M_1 = [4.10\,;\,7.77] \times 10^{13} M_\odot/h$ and redshift bin $\Delta z_1 = [0.00\,;\,0.06]$, with a relation up to $\sigma_{HSV} \approx 2\,\sigma_{Poiss}$.

$\sigma^2_{HSV}/\sigma^2_{Poiss}$	ΔM_1	ΔM_2	ΔM_3	ΔM_4	ΔM_5
Δz_1	3.93	1.98	0.81	0.38	0.0
Δz_2	2.74	1.02	0.18	-0.01	0.04
Δz_3	1.88	0.81	0.41	0.01	-0.23

Table 3.1: HSV Test
The values in this table show the square proportion between the HSV and the Poisson variance. Results obtained using Equation (3.2.1) for the different redshift and mass bin.

The *Halo Sample Variance* σ_{HSV} seams to keep a moderately predominance over the *Poisson Variance* σ_{Poiss} for our lower Δz_1 and the higher Δz_3 redshift bin, towards the different mass bins, with $\sigma_{HSV} \approx 0.7\,\sigma_{Poiss}$, and only on the last mass bin to drop relatively with $\sigma_{HSV} \approx 0$. Instead, for Δz_2 this effect seems to drastically stop, with $\sigma_{HSV} \leq 0.4\,\sigma_{Poiss}$, once we overstep into the higher mass range.

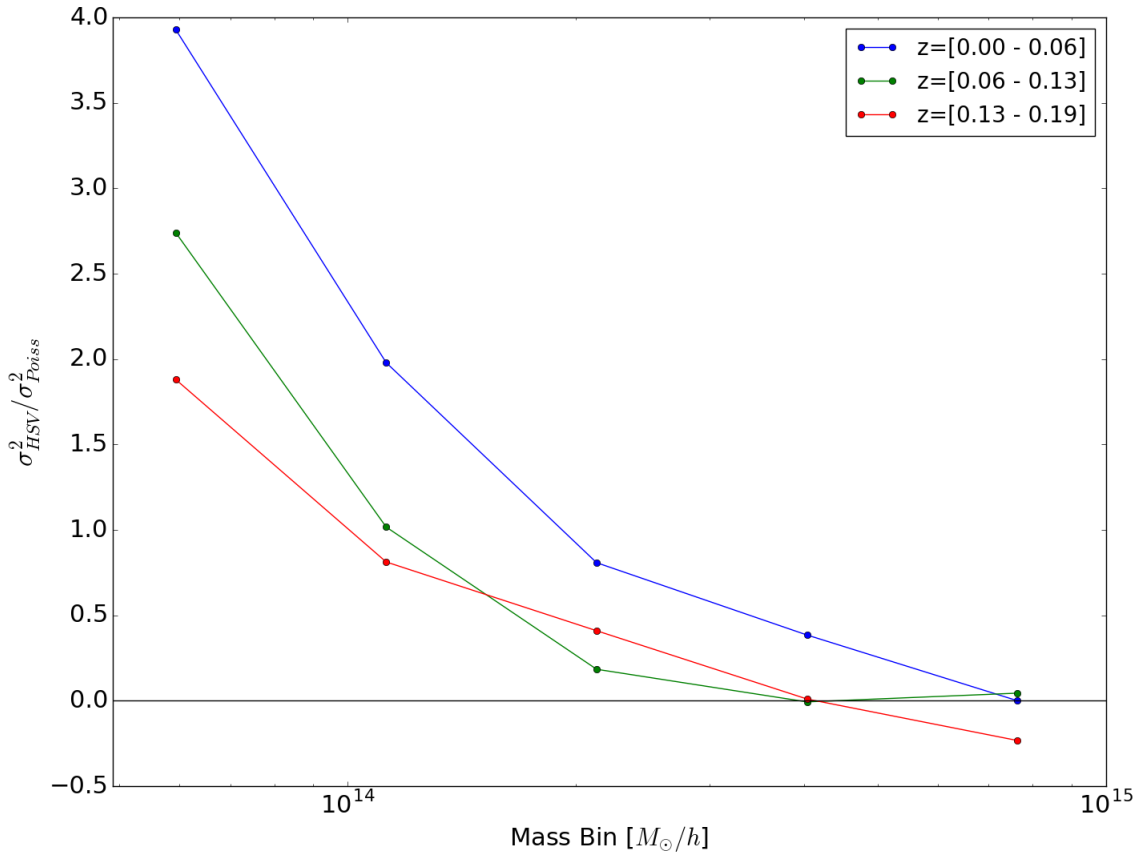

Figure 3.4: HSV Test
Results obtained from Equation (3.2.1), Ratio between the Sample Variance and Shot noise for the 3 different redshift bin, blue solid line for the lowest redshift bin $z \in [0\,;\,0.06]$, green solid line for $z \in [0.06\,;\,0.13]$ and red solid line for $z \in [0.13\,;\,0.19]$. The points are situated at the center of the mass bin.

3.3 Comparison of the Statistical Error with a Numerical Model

Our goal here is to compare our results from the simulation with a numerical prediction of Equation (2.2.7), that here we rewrite again in a more practical way:

$$\sigma^2_{HNCV} = \bar{N}_{ij} + \bar{b}^2_{ij}\bar{N}^2_{ij}\sigma^2_m(V_j) \tag{3.3.1}$$

where $\bar{N}_{ij} \equiv \bar{N}(M_i, z_j)$ and $\bar{b}_{ij} \equiv \bar{b}(M_i, z_j)$ are the mean number count and the mean bias for the relative redshift bin Δz_j and mass bin ΔM_i. Here V_j is the survey volume calculated with Equation (2.1.6).

It is important to underline that the numerical code, that we will use to evaluate the mean count $\bar{N}$ and mean bias $\bar{b}$, was developed by the doctorate and assistant of Prof. *Jochen Weller*, Mr. *Steffen Hagstotz*, we manipulated the code to fit our necessity in this dissertation.

For the construction of the rms mass density fluctuation σ_m, of the form Equation (2.2.2), we consider the windows function of a shell, with volume $V_{R,r} = \frac{4\pi}{3}(R^3 - r^3)$ where R is the external radius and r the internal radius:

$$W_{R,r}(x) = \frac{3}{4\,\pi(R^3 - r^3)}\,\Theta(R - x)\,\Theta(x - r) \tag{3.3.2}$$

here $\Theta(x)$ is the *Heaviside step function* so that $\Theta(R - x) = 1$ if $x \leq R$, zero otherwise, and $\Theta(x - r) = 1$ if $x \geq r$, again zero otherwise, the Fourier Transform then is:

$$\tilde{W}_{R,r}(k) = \frac{3}{(R^3 - r^3)k^3}\left[\sin(kR) - Rk \cdot \cos(kR) - \sin(kr) + rk \cdot \cos(kr)\right] \tag{3.3.3}$$

we can clearly see that in the case of $r = 0$, we reduce this equation to the well know spherical window function.

Then to continue the construction of the analytical model we consider the mean number counts $\bar{N}$ as defined as in Equation (2.2.3), that for simplicity we repeat here again

$$\bar{N}(M, z) = \frac{4\pi\, c}{H_0}\int dz\, d(ln\, M)\frac{d_c^2(z)}{E(z)} \cdot \frac{dn}{d(ln\, M)}(M, z) \tag{3.3.4}$$

as halo mass function we decided to use the the *Tinker 2008* halos mass function [13], defined as follow:

$$\frac{dn}{dM} = f(\sigma_m)\frac{\bar{\rho}_m}{M}\frac{d\ln(\sigma_m^{-1})}{dM} \tag{3.3.5}$$

where $\bar{\rho}_m$ is the mean mass density of the universe and the function $f(\sigma)$ express the fraction of the volume that has collapsed, it is than parametrized by the following expression

$$f(\sigma_m) = A\left[\left(\frac{\sigma_m}{b}\right)^{-a} + 1\right] e^{-c/\sigma_m^2}$$

here we have the parameter to be.

$$A = 0.175 \cdot (1+z)^{-0.012}$$

$$a = 1.53 \cdot (1+z)^{-0.04}$$

$$b = 2.55 \cdot (1+z)^{-0.194}$$

$$c = 1.19 \cdot (1+z)^{-0.021}$$

We chose these parameter because they are valid for a mass definition similar to the one used by the *PINOCCHIO* code.

In conclusion we also considered the mean bias as defined by Equation (2.2.5), that we repeat here in a more practical way,

$$\bar{b}(M,z) = \frac{1}{\bar{N}} \frac{4\pi\, c}{H_0} \int dz\, d(ln\, M) \frac{d_c^2(z)}{E(z)} \cdot\, b(M,z)\, \frac{dn}{d(ln\, M)}(M,z) \tag{3.3.6}$$

The bias b used here is the *Tinker 2010* bias [14].

In the evaluation of Equation (2.2.4) and Equation (3.3.6), we expected to have a discrepancy between the numerical result and the simulated data of the *Statistical Error* Equation (3.3.1) to be on the order of $\sim 10 - 20\%$, as suggested from the study on the halo mass function performed by *Tinker 2008* [13].
Meanwhile for our data, we consider the error on the mean number count $\delta\bar{N}$ to be the *Standard Deviation* for discrete variable, such that

$$\delta\bar{N}_j = \sqrt{\frac{1}{n_R - 1} \sum_{i=1}^{n_R} \left(N_{ij} - \bar{N}_j\right)^2} = \sqrt{C_{jj}} \tag{3.3.7}$$

As we can expect, because of the mean number count in Equation (3.3.1), there is an Error-propagation on our Cluster counts covariance matrix C_{jk}, so that we can express an error on our data δC_{jk} in the following form, for j and $k \in \{0\,;\, n_V\}$

$$\delta C_{jk} = \frac{1}{n_R - 1} \sqrt{\left(\sum_{i=1}^{n_R} (N_{ik} - \bar{N}_k) \cdot \delta\bar{N}_j\right)^2 + \left(\sum_{i=1}^{n_R} (N_{ij} - \bar{N}_j) \cdot \delta\bar{N}_k\right)^2} \tag{3.3.8}$$

We compared the numerical result obtained from Equation (3.3.1) with the diagonal of the *Covariance Matrix* C_{jk}, Figure 3.3. The results are shown in Figure 3.5.

As we can see, the the mean number count $\bar{N}$ from the simulated data, right panels Figure 3.5 in red, in general corresponds with the numerical model.
A certain degree of disagreement occurs for low mass $\leq 10^{14} M_{\odot}/h$ and this behavior is more evident when we compare the numerical and simulated *Halo Number Count Variance* σ^2_{HNCV}, left panels Figure 3.5, the discrepancy here seems to be more prominent for low redshift bin.
We notice that the simulation data seems to underestimate the number of massive halos for all the three case of redshift bin, that can maybe explained by the difficulty of accurately fit the halo mass function for this range of mass [13].

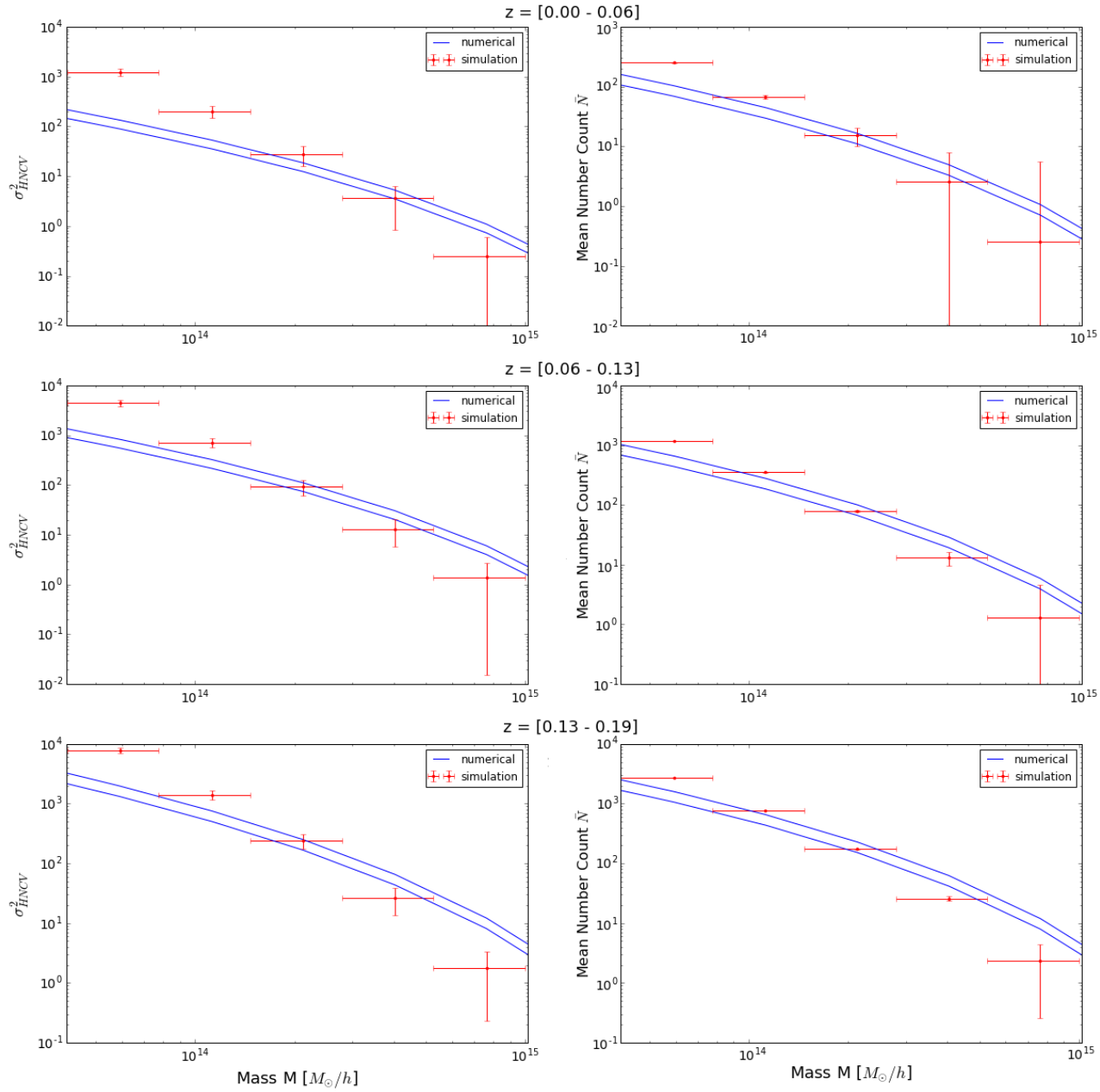

Figure 3.5: HNCV Test
The left panels show the comparison between the simulated data, in red, and the numerical evaluation, in blue, of the *Statistical Error*, Equation (3.3.1), while right panels show the comparison between the simulated data, in red, and the numerical evaluation, in blue, of the mean number count $\bar{N}$, Equation (2.2.5). For the simulation data, the x-axis error bar is the size of the mass bin, the y-axis error bar is defined, on the right by $\delta\bar{N}_j$, Equation (3.3.7), on the left by δC_{jk}, Equation (3.3.8). The two blue line correspond to the 20% discrepancy suggested by *Tinker 2008* [13].
The points are situated at the center of the mass bin.

3.4 Relation with the Survey Area

In this section we want to have a look at the statistical error and its relation with the survey area, to do so we constrained our full sky survey data by the past-light cone aperture `PLCAperture`, that results into a solid angle of the size of the wanted survey.
We took just a few out of the many present galaxy cluster survey mentioned before in Chapter 1, we chose the one considered the most promising and interesting.

To test this matter we used the *Sample Variance* and *Shot noise* Ratio, Equation (3.2.1), for a full-sky survey $40'000\,deg^2$ like *eRosita*, an half-sky survey of $20'000\,deg^2$ with *LSST*, a quarter-sky survey $15'000\,deg^2$ like *Euclid* and a narrow survey area of $5'000\,deg^2$ as *DES*.

As we can see from the results shown in Figure 3.6, the *Sample Variance* and *Shot noise* Ratio decreases for larger survey area, and this effect is more pronounced for increasing redshift. In fact we expected this trend as the study done by Valageas et al. 2012 [17] asserts that in the regime where the *Poisson* variance is dominating, the Ratio grows as the inverse square root of the total survey area $\sigma^2_{HSV}/\sigma^2_{Poiss} \propto A^{-0.5}$ [17]. Meanwhile in the regime dominated by the HSV the quantity has a more complex form that also depends on the primordial index n_s, in our case we will have the form $\sigma^2_{HSV}/\sigma^2_{Poiss} \propto A^{-0.74}$ [17].

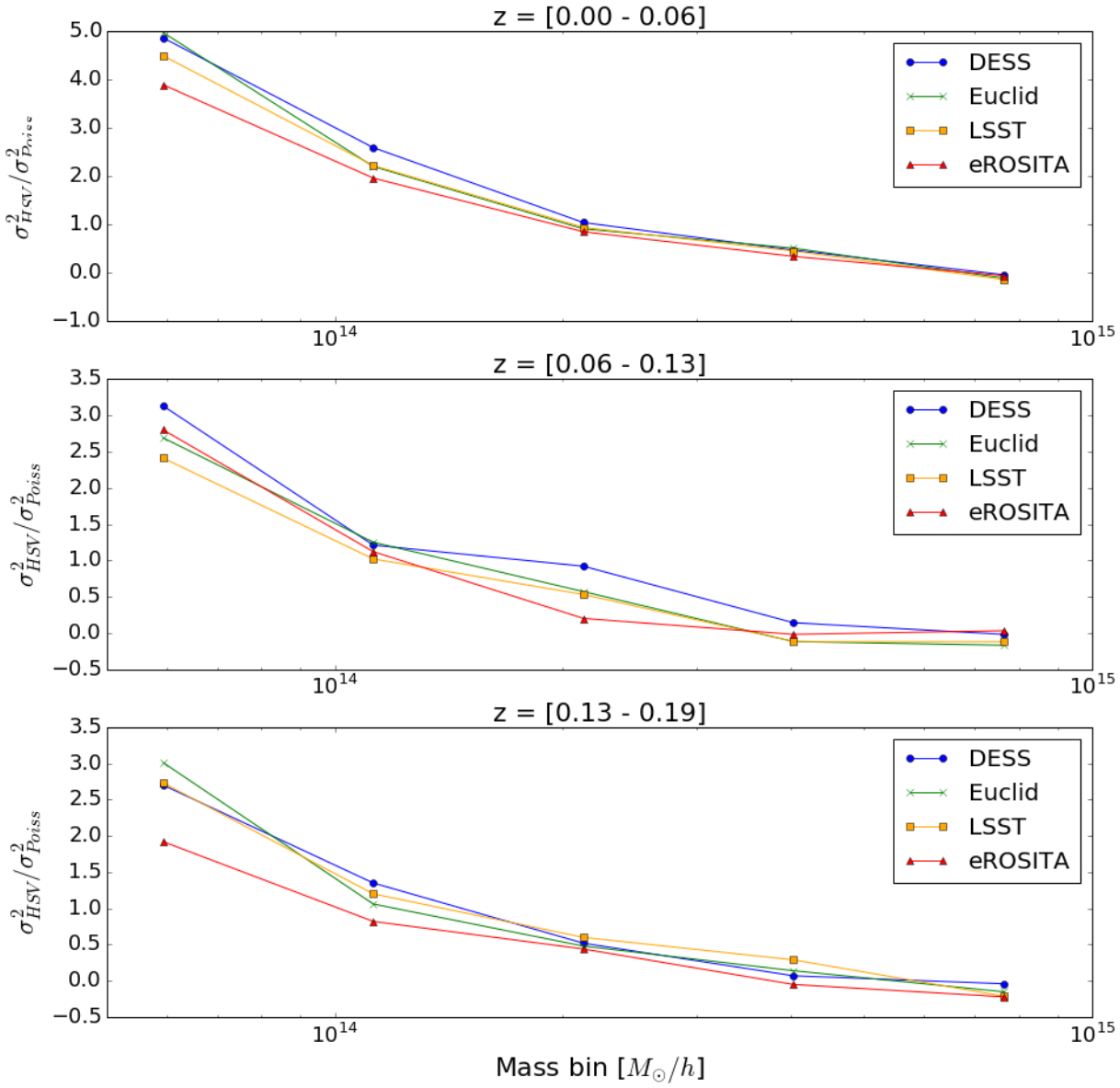

Figure 3.6: Comparison with Survey Area
The plots show the *Sample Variance* and *Shot noise* Ratio for different survey area, given in Table 1.1. Points are situated at the center of the mass bin.

Chapter 4

Conclusion

In this dissertation we have explored the effect of large scale structure correlation on the cluster counts. We first started our analysis with Chapter 3.1, we calculated the *Covariance Matrix* with the data obtained from our simulation, we noticed that strong correlation occurs for low mass $\leq 10^{14}\, M_{\odot}/h$ and that off-diagonal terms increase their correlation especially for lower redshift. In Chapter 3.2 we explored the ratio between the *Sample Variance* and *Shot noise*. We observed that the HSV prevails the *Poissonian* contribution for mass $\leq 1.5 \times 10^{14} M_{\odot}/h$ at the lowest redshift bin $\Delta z_1 = [0.00\,;\, 0.06]$, with a relation up to $\sigma_{HSV} \approx 2\sigma_{Poiss}$. For higher redshift bin this behavior acts on a smaller range of mass $\leq 7.7 \times 10^{13} M_{\odot}/h$, with a proportion of $\sigma_{HSV} \approx 1.4\sigma_{Poiss}$.
Nowadays the contribution of the *Poisson* error is well understood, less clear is the importance of the *Halos Sample Variance* contribution. Several author developed a theoretical model to describe the *Halo Number Counts Variance* σ^2_{HNCV} and that can be summarized by Equation (3.3.1). In Chapter 3.3 we concentrate our efforts to verify this relation, looking at Figure 3.5 our conclusion is that in the range where the HSV dominates, there is a discrepancy between the model and our data of some importance, becoming less and less pronounced for higher redshift. This can be generated from mismatch between the numerical and the actual halo bias measured in the simulation, or also from incompleteness in the used model, we are confident that if we include higher term of contribution, we will be able to better understand this behavior and possibly have better match with the theory prediction.
Because the advent of new galaxy cluster surveys, we wanted to mention the relation between the *Halo Number Counts Variance* and the survey area, in Chapter 3.4, we noticed that the HNCV roughly decreases for larger solid angle. This can lead to a first evaluation of the contribution of *Sample Variance* for their setup.

As we mentioned before, in this dissertation we only assumed the 1-halo term, the *Poisson* Variance, and the 2-halo term contribution, the *Halo Sample Variance*, to be relevant for the estimation of the *Halo Number Count Variance* HNCV, but we can easily see that a deeper analysis can be performed if we consider higher degree of contribution, similarly to what is proposed by Takada & Spergel 2014 [16], and this farther assumption can surely

help in a better prediction.

To conclude, it is important to clarify that our initial goal was to work with a larger redshift range $0 \leq z \leq 1$, but when we was studying the *Sample Variance* and *Shot noise* Ratio, Equation (3.2.1), some absurd results was pointed out and we later discovered that this inexplicable behavior was related to the box replication in the *PINOCCHIO* code. Despite this problematic we decided to stick on using the *PINOCCHIO* code, especially because its affordability on matter of simulation run time, and to avoid this unexpected behavior we reduced the redshift range so that our past-light cone did not exceed into the adjacent replicated box, of course this induced some technical limitation on the box size, therefore on the redshift range, imposed by the available run memory.

Bibliography

[1] Fritz Zwicky.
Die Rotverschiebung von extragalaktischen Nebeln, [1933AcHPh...6..110Z] 1933.

[2] Gott, J. R., III and Turner, E. L.
The Mean Luminosity and Mass Density in the Universe, [1976ApJ...209....1G] 1976

[3] Jonathan Aumont:, George P. Efstathiou, Antony Lewis, Alessandro Renzi, Douglas Scott et al.:
Planck 2015 results. XIII. Cosmological parameters, [arXiv:1502.01589v3] 2015.

[4] David H. Weinberg, Michael J. Mortonson, Daniel J. Eisenstein, Christopher Hirata, Adam G. Riess, Eduardo Rozo:
Observational Probes of Cosmic Accelleration, [arXiv:1201.2434v2] 2013.

[5] Joachim E. Trmper, Gnther Hsinger:
The Universe in X-Rays, [ISBN 978-3-540-34412-4] 2008.

[6] H. Bhringer, P. Schuecker, L. Guzzo, C.A. Collins, W. Voges, S. Schindle, D.M. Neumann, R.G. Cruddace, S. De Grandi, G. Chincarini, A.C. Edge, H.T. MacGillivray and P. Shaver:
The ROSAT-ESO flux limited X-ray (REFLEX) galaxy cluster survey. I. The construction of the cluster sample [AA 369, 826850] 2001.

[7] Future Projects of the High-Energy Astrophysics Group: eROSITA
URL: http://www.mpe.mpg.de/eROSITA 2017.

[8] Scott Dodelson:
Modern Cosmology, [ISBN-10 0-12-219141-2] 2001.

[9] Jochen Weller:
Cosmology and Large Scale Structure, Lecture Notes, 2012.

[10] Philip Bull et all.:
Beyond ΛCDM: Problems, solutions, and the road ahead, [arXiv:1512.05356] 2016.

[11] Peter Coles:
Large-scale Structure, Theory and Statistics, [arXiv:0103017v1] 2001.

[12] Fabien Lacasa, Rogerio Rosenfeld:
Combining Cluster Number Counts and Galaxy Clustering, [arXiv:1603.00918v2] 2016.

[13] Jeremy Tinker, Andrey V. Kravtsov, Anatoly Klypin, Kevork Abazajian, Michael Warren, Gustavo Yepes, Stefan Gottlber, Danie E. Holz:
Toward a Halo Mass Function for Precision Cosmology: The Limits of Universality, [arXiv:0803.2706v1] 2008.

[14] Jeremy L. Tinker, Brant E. Robertson, Andrey V. Kravtsov, Anatoly Klypin, Michael S. Warren, Gustavo Yepes, Stefan Gottlöber:
the Large Scale Bias of Dark Matter Halos: Numerical Calibration and Model Tests, [arXiv:1001.3162v2] 2010.

[15] Wayne Hu and J. D. Cohn:
Likelihood Methods for Cluster Dark Energy Survey, [arXiv:0602147v1] 2006.

[16] Masahiro Takada and David N. Spergel:
Joint Analysis of Cluster Number Counts and Weak Lensing Power Spectrum to Correct for the Super-Sample Covariance, [arXiv:1307.4399v2] 2014.

[17] P. Valageas, N. Clerc, F. Pacaud, and M. Pierre: *Covariance matrices for halo number counts and correlation functions*, [arXiv:1104.4015v2] 2012

[18] Emiliano Munari, Pierluigi Monaco, Emiliano Sefusatti, Emanuele Castorina, Faizan G. Mohammad, Stefano Anselmi and Stefano Borgani:
Improving fast generation of halo catalogues with higher order Lagrangian Perturbation Theory, [MNRAS:465,46584677] 2017.

[19] Luca A.Rizzo, Francisco Villaescusa-Navarro, Pierluigi Monaco, Emiliano Munari, Stefano Borgani, Emanuele Castorina, Emiliano Sefusatti:
Simulating cosmologies beyond ΛCDM with PINOCCHIO, [arXiv:1610.07624v1] 2016.