Master's thesis
Astronomy

Your Title Here

Anni Järvenpää

April 19, 2017

Tutor: Associate Professor Peter Johansson
Dr. Till Sawala

Censors: prof. Smith
doc. Smythe

UNIVERSITY OF HELSINKI
Department of Physics

PL 64 (Gustaf Hällströmin katu 2a)
00014 University of Helsinki

HELSINGIN YLIOPISTO — HELSINGFORS UNIVERSITET — UNIVERSITY OF HELSINKI

<table>
<tr><td colspan="2">Tiedekunta — Fakultet — Faculty
Faculty of Science</td><td colspan="2">Laitos — Institution — Department
Department of Physics</td></tr>
<tr><td colspan="4">Tekijä — Författare — Author
Anni Järvenpää</td></tr>
<tr><td colspan="4">Työn nimi — Arbetets titel — Title
Your Title Here</td></tr>
<tr><td colspan="4">Oppiaine — Läroämne — Subject
Astronomy</td></tr>
<tr><td>Työn laji — Arbetets art — Level
Master's thesis</td><td colspan="2">Aika — Datum — Month and year
April 19, 2017</td><td>Sivumäärä — Sidoantal — Number of pages
0 pages</td></tr>
<tr><td colspan="4">Tiivistelmä — Referat — Abstract

Abstract goes here.</td></tr>
<tr><td colspan="4">Avainsanat — Nyckelord — Keywords
Your keywords here</td></tr>
<tr><td colspan="4">Säilytyspaikka — Förvaringsställe — Where deposited</td></tr>
<tr><td colspan="4">Muita tietoja — övriga uppgifter — Additional information</td></tr>
</table>

Contents

1. Introduction

1.1 TL;DR version of prerequisite information

1. galaxies form
    - Why?
    - When?
    - How?
    - Where?
2. galaxies form in groups
3. our local group is one of these
4. something about large scale distribution of galaxies

1.2 History of Local Group Research

LG objects visible with naked eye -> realization they are something outside our galaxy -> realization they are something very much like our galaxy

First determining distance was difficult, now mass is more interesting question

1.3 Aim of This Thesis

Whatever the main results end up being, presented in somewhat coherent manner and hopefully sugar-coated enough to sound Important and Exciting.

2. Theoretical Background

Think whether LG or LCDM first

2.1 Local Group

Definition of galaxy group, our local group is one of these.

Mass estimate (Li, Yang masses for the LG and MW)

Maybe something about scale of things in our universe, what are galaxy groups made of, what do you get if you go one distance scale up, what's different in galaxy clusters

2.1.1 Structure

Galaxies that are part of LG, distribution of smaller ones around bigger ones

Current mass estimates (at least timing argument, hubble flow and maybe satellites)

2.1.2 Evolution

How have we ended up in a situation described earlier? What will happen in future?

2.2 Expanding universe

2.2.1 Discovery

Make maths, add cosmological constant, make observations, remove cosmological constant

Enough cosmology here or in other sections to make other parts of thesis to make sense and to suffice as master's thesis. How much is enough for the latter?

2.2.2 ΛCDM Cosmology

2.2.3 Hubble flow

What is, where seen, what means, how to measure, hotness/coldness

Plot: observations with fitted hubble flow

2.3 Mathematical and statistical methods

Precision of the used equipment limits accuracy of all data gathered from physical experiments, simulations or observations. Therefore the results are affected by a random process and the results have to be presented as estimates with some error, magnitude of which is affected by both number of data points and accuracy of the measurement equipment. [1]

Estimating errors for measured quantities offers a way to test hypotheses and compare different experiments[1]. This is done using different statistical methods that are too numerous to extensively cover here. Methods used in this work are shortly introduced in the following sections.

2.3.1 Regression Analysis

line fitting and other trivial things

2.3.2 Statistical testing

A common situation in scientific research is that one has to compare a sample to either a model or another sample in order to derive a conclusion from the dataset. In statistics, this is known as hypothesis testing. For example, this can mean testing hypotheses like ”these two variables are not correlated” or ”this sample is from population with mean 1.0”. [2]

Typically the process of hypothesis testing begins with forming of null hypothesis H_0 that is formatted such that the aim for the next steps is to either reject it or deduce that it cannot be rejected on chosen significance level. Negation of the null hypothesis is often called research hypothesis or alternative hypothesis and denoted as H_1. For example, this can lead to H_0 ”this dataset is sampled from a normal distribution” and H_1 ”this dataset is not sampled from a normal distribution”. Choosing the hypotheses in this manner is done because often the research hypothesis is difficult to define otherwise. [1, 2]

After setting the hypotheses one must choose an appropriate test statistic. Ideally this is chosen such that the difference between cases H_0 and H_1 is as large as possible. Then one must choose the significance level α which corresponds to the probability of rejecting H_0 in the case where H_0 actually is true. This fixes the critical region i.e. the values of test statistic that lead to the rejection of the H_0. [1, 2]

It is crucial not to look at the test results before choosing α in order to avoid intentional or unintentional fiddling with the data or shifting the goalposts. Only after these steps should the test statistic be calculated. If the test statistic falls within the critical region, H_0 should be rejected and otherwise stated that H_0 cannot be rejected at this significance level. [1, 2]

This kind of probability based decision making is always prone to error. It is easy to see that α corresponds to the chance of H_0 being rejected when it is true. This is known as error of the first kind. However, this is not the only kind of error possible. It might also occur that H_0 is false but it does not get rejected, which is known as error of the second kind. [1]

2.4 Cluster Analysis

DBSCAN

3. general simulation thingies

3.1 N-body simulations

3.1.1 Hierarchical Tree Algorithm

3.1.2 Halo Finding with Subfind

3.2 Description of actual simulations used

Volume, number of particles, compare to other simulations, where better and where maybe worse

Resimulation of interesting regions

Simulation has same parameters as EAGLE 800 Mpc volume used schaye 2015 paper DM-only parts: Volker-Springer Gadget and Gadget 2 papers 1999 and 2005 or something, gravity part is more interesting than SPH Zooms can use multiple meshes, only one is used here gravitational softening

4. Findings from DMO Halo Catalogue Analysis

4.1 Selection of Local Group analogues

criteria, how many found, what are like (some plots maybe? distributions of masses, separations, velocities or correlations between two of those?). This might be part of previous chapter too (relevant to resimulation)?

4.2 Local Anisotropy of the Hubble Flow

Hopefully there's something at least mildly interesting to report when I get to look at the new data

4.3 Statistical Estimate of the Local Group Mass

Analysis similar to Fattahi et al 2016 paper

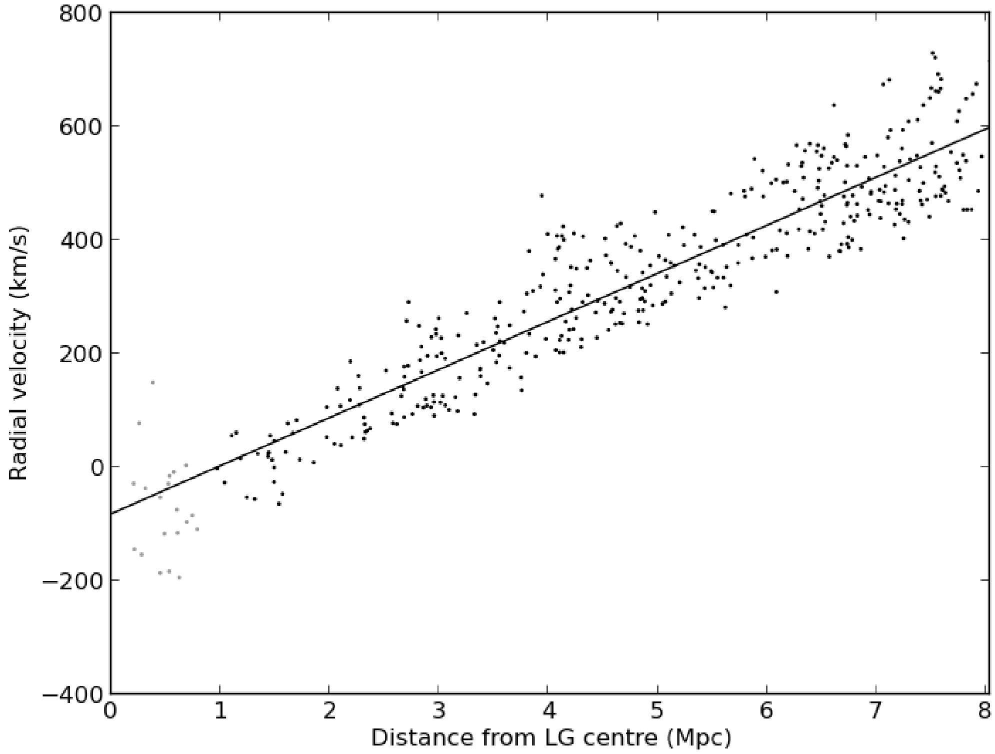

Figure 4.1: Radial velocities of haloes as a function of distance. Best fit to Hubble flow shown with solid line. Nearby points ignored when fitting shown in gray.

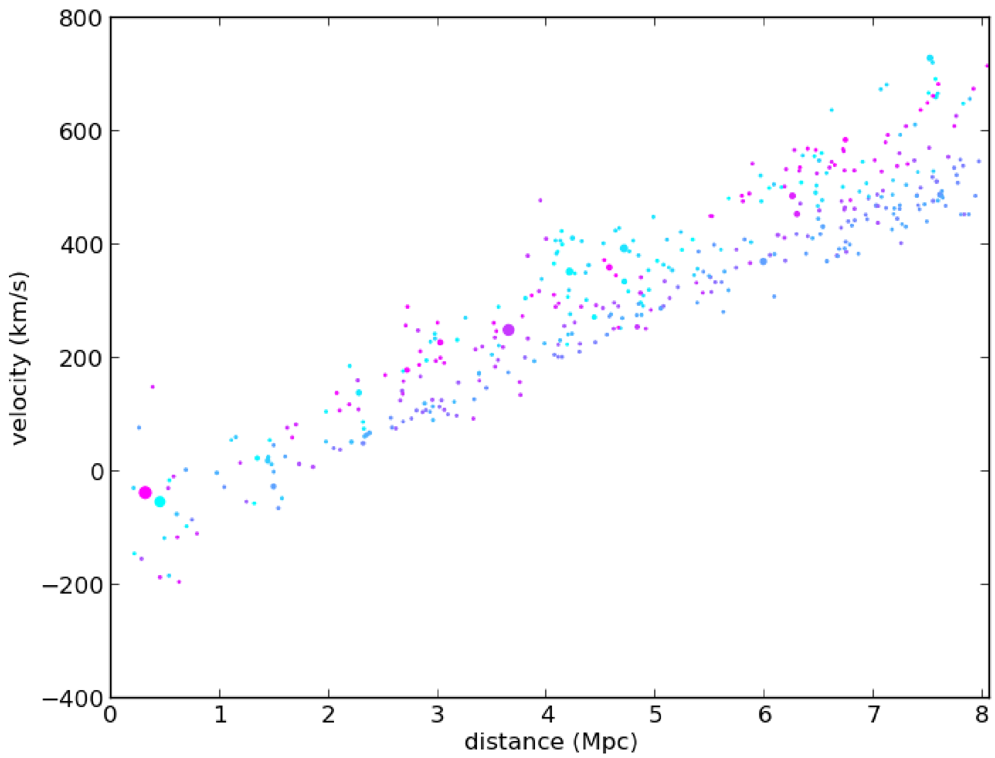

Figure 4.2: Hubble flow with colours depicting angular disstance from line connecting Milky Way and Andromeda counterparts in simulation.

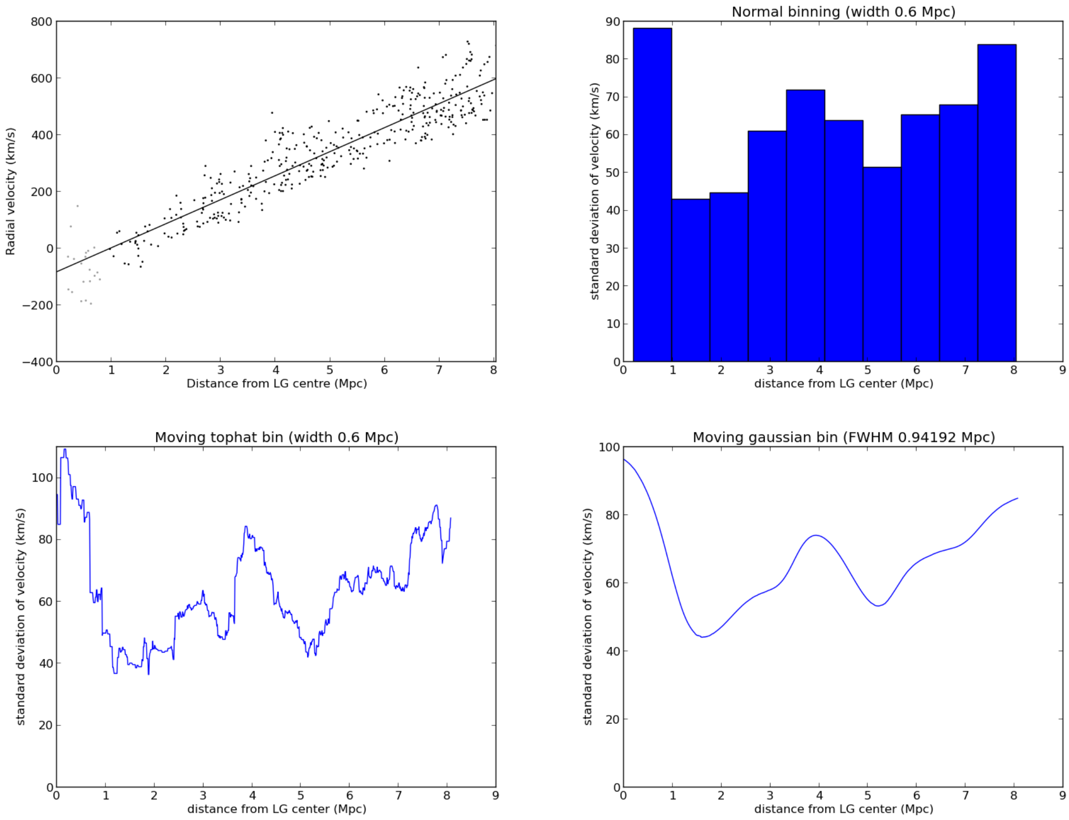

Figure 4.3: Velocity dispersion of Hubble flow.

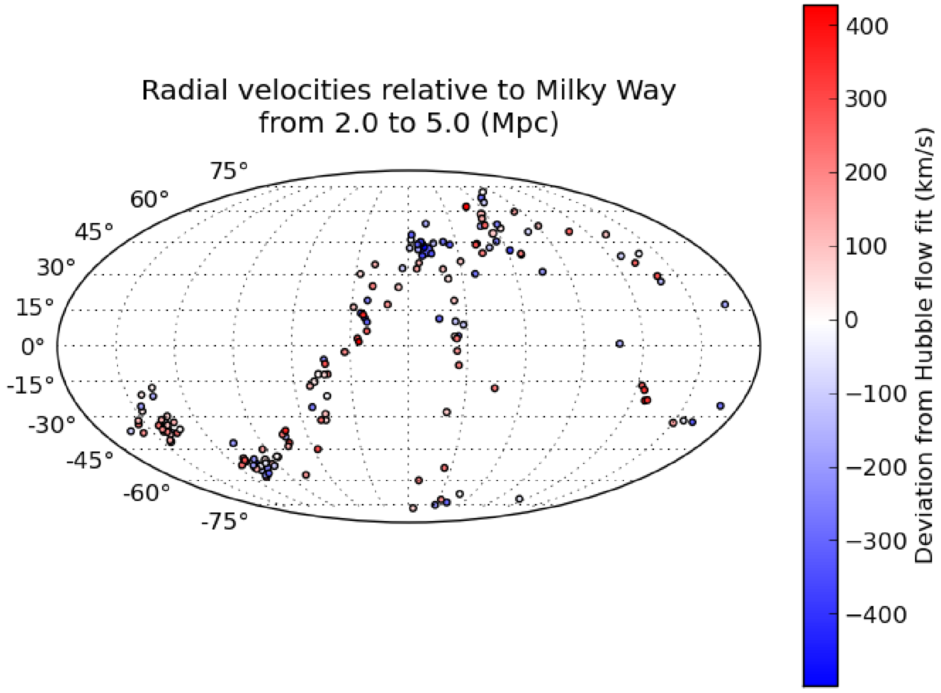

Figure 4.4: Haloes with distances between 2 and 5 Mpc as seen from Mily Way counterpart in simulation. Colours depict deviations from best linear Hubble flow fit ignoring haloes up to 2 Mpc away, blue end meaning haloes coming closer faster than expected and redder colours moving away.

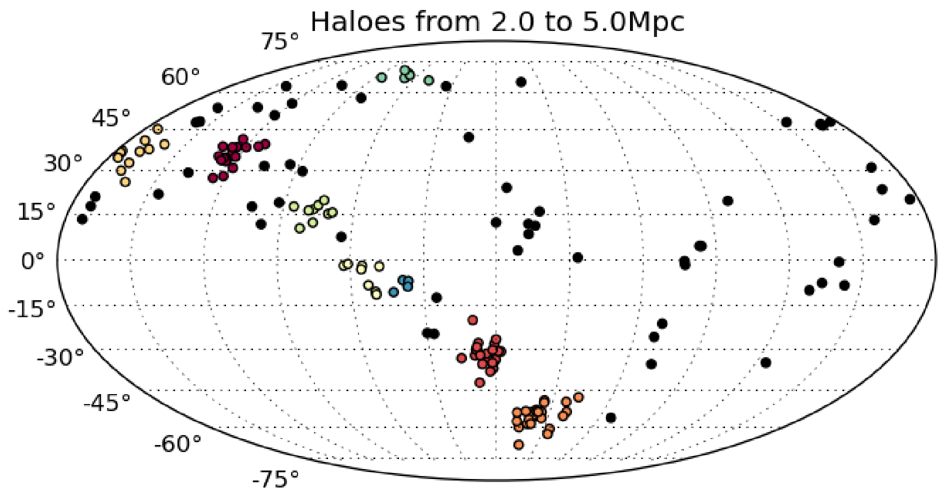

Figure 4.5: Dark matter haloes with distances from 2 to 5 Mpc grouped to clusters using DBSCAN clustering algorithm. Parameters used for this plot were ms=5 and eps=2.

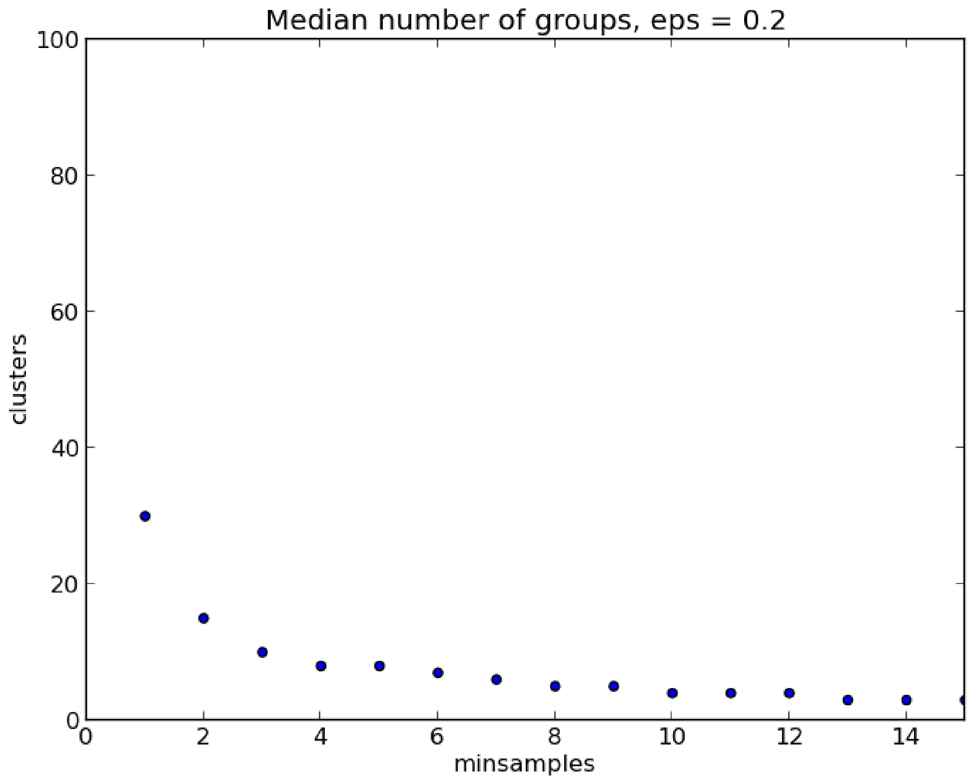

Figure 4.6: Median number of clusters found with constant eps on different minsamples.

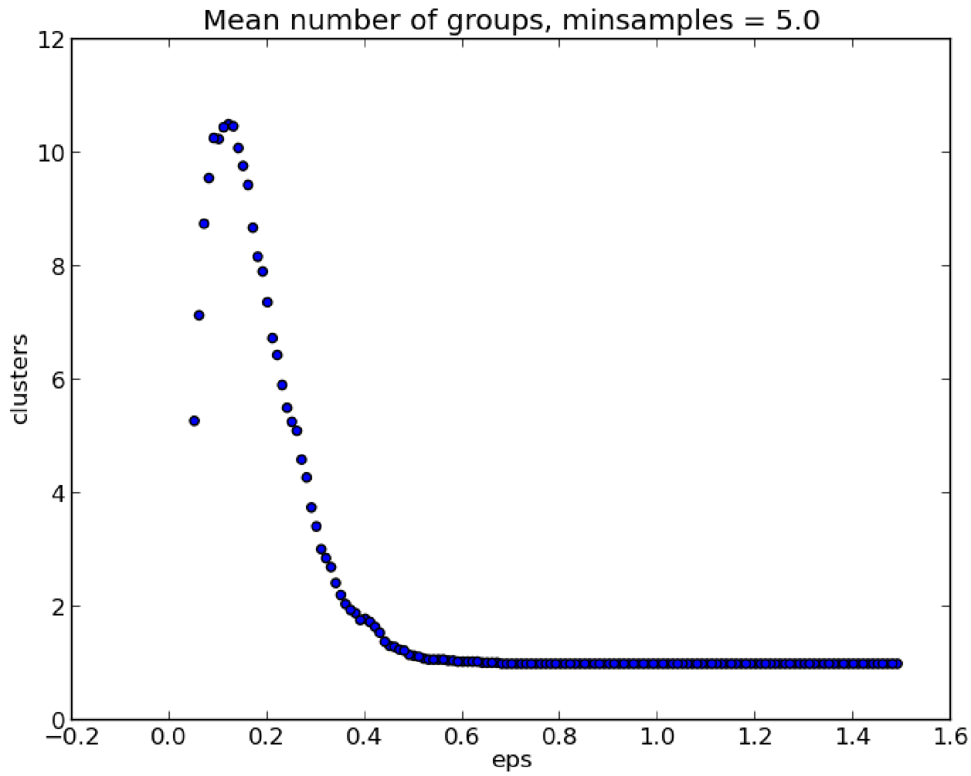

Figure 4.7: Mean number of clusters found with constant ms on different eps.

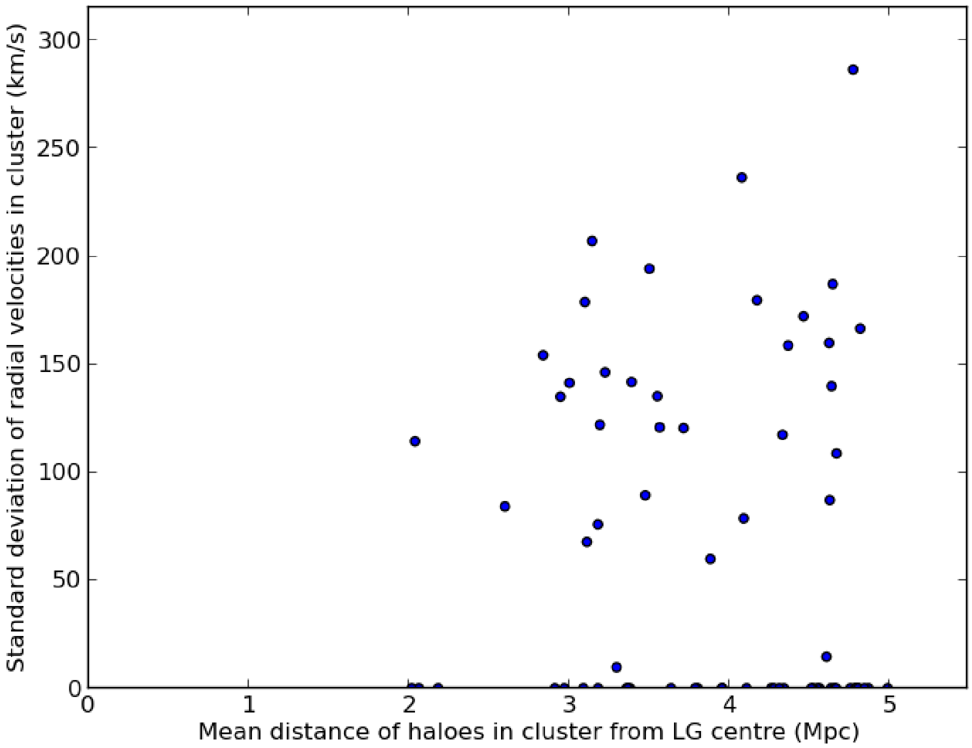

Figure 4.8: Standard deviation of velocities within cluster as a function of distance.

5. Conclusions

References

1 G. Bohm and G. Zech. *Introduction to statistics and data analysis for physicists*. DESY, 2010.

2 C. R. Jenkins J. V. Wall. *Practical Statistics for Astronomers*. Cambridge Observing Handbooks for Research Astronomers. Cambridge University Press, illustrated edition edition, 2003.