



Gravitational redshift in galaxy clusters with photometric surveys

MASTER THESIS

Ida Vægter Rasmussen

March 6th 2017

Supervised by:

Jens Hjorth

Dark Cosmology Centre

Niels Bohr Institute

University of Copenhagen

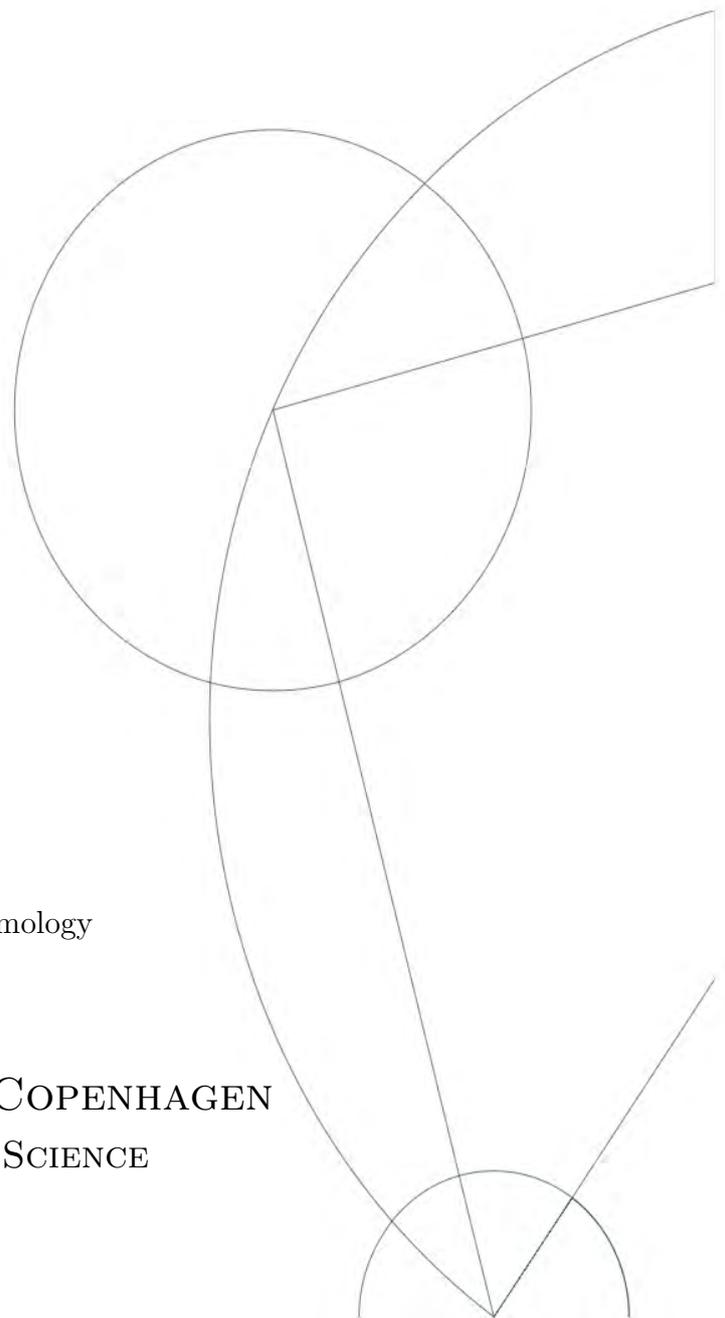
and

Radoslaw Wojtak

Kavli Institute for Particle Astrophysics and Cosmology

Stanford University

UNIVERSITY OF COPENHAGEN
FACULTY OF SCIENCE



Abstract

I will investigate whether it is possible to detect gravitational redshift in galaxy clusters using photometric data, and if this is possible with the Large Synoptic Survey Telescope (LSST). This is done by using simulated data representing different types of spectroscopic and photometric data for galaxy clusters with a mass above $10^{14}M_{\odot}h^{-1}$. The three datasets analysed represent: (1) spectroscopic data for all galaxies, (2) spectroscopic data for the Brightest Cluster Galaxies (BCG) and photometric data for all other galaxies, and (3) photometric data for all galaxies, including the BCGs.

The analysis of the spectroscopic data (1) show a clear gravitational redshift signal in clusters, and it is traced out to a radius of 6 Mpc with an average error of ± 0.8 km/s. With regards to the photometric data, (2) and (3), the results are less conclusive, as these datasets show a large scatter. The average error in the gravitational redshift for the photometric datasets are ± 3.5 km/s and ± 5.4 km/s respectively. The analysis was carried out with a sample of $9.1 \cdot 10^6$ galaxies distributed over $6 \cdot 10^4$ clusters.

The LSST is expected to observe 10^5 clusters, which would make it possible to obtain a total of $1-2 \cdot 10^7$ galaxies from these. Based on extrapolations performed, this would make it possible to bring the error on the gravitational redshift down to ± 3 km/s for the photometric datasets, (2) and (3), making it possible to detect gravitational redshift with photometry.

Contents

1	Introduction	2
2	Gravitational redshift	4
3	Data generation and processing	6
3.1	Mock data generation	7
3.2	Datasets	8
3.3	Fitting procedure	10
3.4	Bayesian Information Criterion	13
4	Analysis	13
4.1	Mock dataset	15
4.2	LSST mock 1	17
4.3	LSST mock 2	20
4.4	Errors and sample sizes	21
5	Discussion	26
6	Conclusion	30
A	Code	33
B	Additional analysis	37
B.1	Distribution of data	37
B.2	Velocity cut-off for data selection	40
B.3	Bin size	41
B.4	Convergence of fits	43
B.5	Computation time	44
B.6	Different mean of double Gaussian	45

1 Introduction

Since Einstein proposed the theory of General Relativity in 1915 describing the relationship between space, time and matter, it has been an integral part of astronomy and physics in general. It has been shown to apply to a large range of scales from time delays of GPS satellites to gravitational lensing in galaxy clusters. The most commonly used metric in cosmology, the Friedman-Robertson-Walker metric, is also derived from general relativity. As a result thereof most cosmological models today are based on general relativity.

General relativity is not the only theory capable of explaining observations. More recently different theories of modified gravity, like $f(R)$, have been introduced to explain the expansion history of the universe. Although general relativity has been shown to make accurate predictions in many areas there are still problems, like explaining singularities and the incompatibility with other forces, indicating that some modification is needed.

Testing of general relativity is a very active field within research, with the latest addition being the observation of gravitational waves (Abbott *et al.*, 2016). The approach I will take is to investigate more accurate measurements, or the prediction thereof, of an already measured effect of general relativity, namely gravitational redshift, by simulating larger dataset than what is available from observations now, as well as consider the use of photometric surveys.

Making accurate determinations of the gravitational redshift in galaxy clusters could prove to be another test of general relativity. A more accurate determination of the signal will result in tighter constraints on the theory used to describe it, and could rule out theories not consistent with the observations.

With upcoming surveys, like the Large Synoptic Survey Telescope (LSST) under way, collecting large datasets of photometric data will soon be possible. Construction of the LSST began in 2015 in Chile and is expected to make science first light in 2021 followed by a 10-year survey covering 20,000 deg^2 of the sky. From LSST the expected number of galaxies observed is 10^{10} and 10^5 clusters (LSST Science Collaboration, 2009; Tyson *et al.*, 2006), which is about twenty times the size of the entire Sloan Digital Sky Survey (SDSS) database right now. The main part of the LSST survey will be photometric observations, which is what will be the focus of this thesis by determining if it is possible to measure the gravitational redshift with this data.

Generally, redshift reflects the change in energy of the photon from emission to obser-

vation which is a sum of the contribution from different effects. The main components of redshift are effects of the Hubble flow and Doppler effect from peculiar velocities, but other smaller effects can also contribute, such as gravitational redshift (Wojtak *et al.*, 2011), transverse Doppler effect (Zhao *et al.*, 2013), relativistic beaming (Kaiser, 2013) and surface brightness modulation (Kaiser, 2013). In this thesis I will be focusing on the effect of gravitational redshift.

All these effects are small, of the order 10 km/s, compared to the typical velocity distribution of galaxies in clusters which is of the order 1000 km/s. As these effects are so much smaller than the velocity distribution, it has only recently been possible to determine them as measurements have become more accurate and the databases have become larger.

Photometric redshifts, which will be the focus in this thesis, are based on observations using multiple broad band filters. The reason photometric redshifts are not typically used for precision measurements are that they are less accurate than spectroscopic redshifts, because of the difficulty in determining the exact position of the features of the spectrum. The advantage of photometric redshift is that it is possible to obtain for a fainter number of objects. Also, observation time is much shorter, and multiple objects can be observed at the same time.

Spectroscopic redshifts have very small redshift errors compared to photometric measurements, because they are determined by observing emission or absorption lines which are very accurate. The largest disadvantage of spectroscopic measurements is the observation time.

In this thesis I will investigate whether it is possible to determine gravitational redshift, using simulated data of what can be expected to come from surveys like LSST with photometry, as well as the precision at which it can be achieved. Also, I will be looking at different samples with varying cuts on mass of the galaxy clusters and the effect of sample size.

In section 2 I will describe the theoretical background of gravitational redshift as well as present the relevant equations. I will then move on to present the simulation parameters, the datasets generated and the processing of the data in section 3 before analysing the results in section 4 and comparing the different datasets and estimating the sizes required for an accurate determination of gravitational redshift. Finally, in sections 5 and 6 I will discuss the results in a broader context, including what is possible with data available right now, and conclude on my findings.

2 Gravitational redshift

As mentioned above redshift is the effect of a photon losing energy between emission and observation. If the photon gains energy it is usually called a blueshift. Gravitational redshift, z_{grs} is caused by the emitter of photons and the observer being in potentials of different depths, $\Delta\Phi$ causing the photon to either lose or gain energy. This effect can therefore make both positive and negative contributions to the total redshift depending on the reference frame and relative potentials.

$$z_{grs} = \frac{\Delta\Phi}{c^2} \quad (1)$$

The effect is relatively small so a potential of a certain depth is required in order to detect the additional redshift, for example that of a galaxy cluster. If a galaxy is situated in a cluster, the gravitational redshift effect relative to the centre of the cluster will result in a blueshift because the orbiting galaxy is at a potential with smaller depth than the cluster centre. This effect is small, typically with a size of the order 10 km/s, compared to the Doppler redshift caused by the peculiar motion of the galaxies in the cluster, which is of the order 1000 km/s.

The theoretically predicted mean gravitational redshift of galaxies (Wojtak *et al.*, 2011) as a function of cluster centric distance, R is

$$cz_{grs}(R) = \frac{2}{c\Sigma(R)} \int_R^\infty [\Phi(0) - \Phi(r)] \frac{\rho(r)r}{\sqrt{r^2 - R^2}} dr \quad (2)$$

where Σ is the 2D surface density, Φ is the gravitational potential at distance R from the cluster centre, ρ is the 3D density profile and c is the speed of light. Here, the gravitational redshift is in units of velocity, which will be used throughout this thesis.

The 2D surface density is used as a normalization and takes the form

$$\Sigma(R) = 2 \int_R^\infty \frac{\rho(r)r}{\sqrt{r^2 - R^2}} dr \quad (3)$$

from Lokas & Mamon (2001). The 3D density is assumed to be an NFW profile (Navarro *et al.*, 1996)

$$\rho(r) \propto \frac{1}{\frac{r}{r_v} \left(1 + c_v \frac{r}{r_v}\right)^2} \quad (4)$$

where r_v is the virial radius and c_v is the concentration parameter approximated by

$$c_v = 9.60 \left(\frac{M_v}{10^{12} M_\odot h^{-1}} \right)^{-0.075} \quad (5)$$

for distinct halos at low redshift from Klypin *et al.* (2011). M_v is the virial mass of the cluster.

The virial radius is defined as the radius at which the density of the cluster is a certain multiple, given by the overdensity parameter Δ_c , of the critical density of the universe, ρ_c . It is also the radius containing the virial mass, M_v of the cluster.

$$r_v = \left(\frac{3M_v}{4\pi\Delta_c\rho_c} \right)^{1/3} \quad (6)$$

where $\rho_c = \frac{3H^2}{8\pi G}$ is the critical density of the universe today. The overdensity parameter used here takes the value $\Delta_c = 97.2$, which is a result of a matter overdensity of 360 with respect to the mean background density of $\Omega_m\rho_c$. The matter overdensity is defined as 360, because this is the value used in the cluster finder algorithm used in the simulation described in section 3.1 below.

Finally, the gravitational potential of a galaxy with an NFW density profile is

$$\Phi(r) = -\frac{GM_v}{r_v} g(c_v) \frac{\ln\left(1 + c_v \frac{r}{r_v}\right)}{\frac{r}{r_v}} \quad (7)$$

where $g(c_v)$ is defined as $g(c_v) = \frac{1}{\ln(1+c_v) - \frac{c_v}{1+c_v}}$.

From the equations above it is clear that clusters with different masses will have different gravitational redshift profiles because of differences in the depth of the potential, see figure 1. The more massive clusters have a deeper potential at the centre than the less massive, which results in a larger difference in potential between the centre of the cluster and the outer parts, resulting in different gravitational redshift profiles.

Notice the flattening of the profile beyond the virial radius ($r_v \approx 1.7$ Mpc for the profile with a mean cluster mass of $1.84 \cdot 10^{14} M_\odot h^{-1}$), but also that the profile continues to decrease at greater radius as well. Usually the virial radius is used to determine the outer limit of a cluster, but galaxies at greater radii can still be affected gravitationally by the cluster. Therefore, galaxies beyond the virial radius are included in this analysis as they contribute to the gravitational redshift signal evident by the continued decrease of the profile.

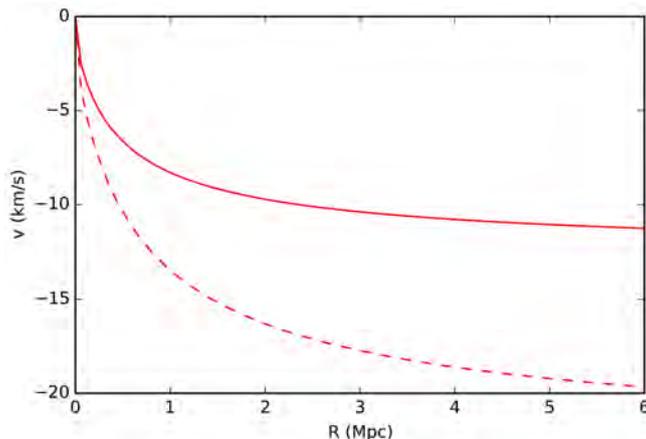


Figure 1: *The prediction of the gravitational redshift profile of two clusters with different masses; $1.84 \cdot 10^{14} M_{\odot} h^{-1}$ (full line) and $4.67 \cdot 10^{14} M_{\odot} h^{-1}$ (dashed line). $1.84 \cdot 10^{14} M_{\odot} h^{-1}$ is the typical mean mass of the full samples described in section 3.2. $4.67 \cdot 10^{14} M_{\odot} h^{-1}$ is the typical mean mass of the samples containing only massive clusters with mass above $3 \cdot 10^{14} M_{\odot} h^{-1}$. The prediction of the gravitational redshift of a cluster galaxy is relative to the centre of the cluster*

Gravitational redshift can be detected by using the velocity distribution of the galaxies in clusters. The Doppler effect from peculiar motions results in a broadening of this distribution, while the gravitational redshift will cause a shift of the centroid of the distribution. If the velocities of the galaxies are measured relative to the cluster centre as indicated in equation 2, the shift of the centroid will be negative (a blueshift).

This method of measuring the signal also means the signal will be stronger for larger distances from the centre of the cluster because this is where the largest differences in potentials are. This is consistent with the expectation from looking at the gravitational redshift profile in figure 1.

3 Data generation and processing

All data analysed in this thesis are simulated and represent galaxies in clusters. The data contain cluster-centric distance in Mpc, the velocity of each galaxy relative to the cluster centre in km/s, the cluster number a given galaxy is associated with and the total mass of the cluster in $M_{\odot} h^{-1}$.

As the focus here is on gravitational redshift, only galaxies in clusters are relevant, since the effect is small and a potential of a certain size is necessary for detection, e.g. galaxy clusters.

Galaxy clusters can vary greatly in the number of members and total mass but all have a centre where the gravitational potential is greatest, and this space is usually occupied by a massive elliptical galaxy called the Brightest Cluster Galaxy (BCG). The BCG is assumed to be at the centre of the potential, but this is an approximation, as the BCG can have a small peculiar velocity relative to the central potential (Kaiser, 2013; Kim & Croft, 2004). If this is the case, it will result in a systematic error for the cluster in question, which is not considered here.

3.1 Mock data generation

The simulated data used in this thesis was generated by Radoslaw Wojtak¹ at Kavli Institute for Particle Astrophysics and Cosmology, Stanford University, using the MDR1 simulation (Prada *et al.*, 2012) with WMAP5 cosmology (Komatsu *et al.*, 2009) from the MultiDark simulation project. The values of the cosmological parameters used for the simulation are $\Omega_m = 0.27$ as the density parameter for all matter, $\Omega_b = 0.0469$ as the density parameter for only baryonic matter, the Hubble parameter $h = 0.7$ and $\sigma_8 = 0.62$ which describes the amplitude of the mass density fluctuations in a $8 \text{ Mpc} h^{-1}$ sphere. The simulation uses 2048^3 particles in a cube with a side length of $1 \text{ Gpc} h^{-1}$ and a mass resolution of $8.721 \cdot 10^9 M_\odot h^{-1}$.

The box size with side length 1 Gpc allows for formation of a wide variety of structures. Of course structures larger than the box size are not possible to distinguish, but considering the typical galaxy cluster has a diameter of $5\text{-}30 \text{ Mpc}$ this is easily contained within the simulation.

Only clusters with a halo mass above $10^{14} M_\odot h^{-1}$ are included in the sample, and the most massive cluster obtained has a mass of $2.36 \cdot 10^{15} M_\odot h^{-1}$. In the simulation galaxies are defined as subhalos of the cluster halo, and no assumption of an NFW profile are made to define the galaxies. The halos are found using the Bound Density Maximum (BDM) algorithm (Klypin & Holtzman, 1997; Riebe *et al.*, 2013), which identifies both distinct halos and subhalos. To avoid spurious detections of subhalos a cut is applied to the minimum number of particles needed to form a subhalo, which in this case is 20, corresponding to a minimum subhalo mass of $2 \cdot 10^{11} M_\odot h^{-1}$.

Clusters are identified based on overdensities compared to the average density of the universe. The BDM algorithm defines halos using a overdensity of 360 times the background density as the outer limit of the halo.

Galaxies with halo mass above the limit of $2 \cdot 10^{11} M_\odot h^{-1}$ in the clusters are selected

¹Supervising this thesis

based on their velocity relative to the cluster centre and their cluster centric distance. The velocity cut used is ± 6000 km/s for all datasets. This cut-off in velocity ensures that all galaxies in the clusters are considered without including too many galaxies dominated by the Hubble flow as well as being wide enough to sample the full velocity distribution of all the clusters. For further discussion of this cut-off see appendix B.2. Also, galaxies must have a cluster centric distance smaller than 6 Mpc, which is the distance at which the Hubble flow starts to dominate over the local gravitational field. The virial radius is not used as a cut-off, as all gravitationally bound galaxies contribute to the gravitational redshift signal, which was also discussed in section 2.

All clusters and galaxies are combined into one final dataset. Combining clusters of different masses will result in a widening of the velocity distribution of the galaxies, because clusters have velocity distributions of different widths as the range of peculiar velocities are affected by the mass of the cluster. An advantage of combining the data into one cluster is that substructures of individual clusters are smoothed out and some of the error due to peculiar velocities of the BCGs is reduced.

As a result of the parameters of the simulation, e.g. simulation volume, number of particles and limits on halo and subhalo masses, each of the projections in the datasets described below and used in the analysis in section 4 will be of roughly the same size, with $\approx 3 \cdot 10^6$ galaxies and $2 \cdot 10^4$ clusters in each.

3.2 Datasets

Three different kinds of datasets are analysed in this thesis, each simulates a different type of data. Each type of dataset contains three different projections, i.e. each has a different line of sight in the simulation, used to test whether any of the projections contain significant statistical fluctuations and as well as combine them to a larger dataset. This gives a total of nine different projections, all with about $3 \cdot 10^6$ galaxies in each distributed over about $2 \cdot 10^4$ clusters. The exact number of galaxies vary from projection to projection, but the sample sizes are very similar with differences in size smaller than 0.5%. Therefore, all samples will be treated as of equal size.

By combining the different projections of a dataset, an even larger dataset can be obtained. The individual projections are the largest possible samples from the simulation given the parameters described in section 3.1, but these contain only 20% of the clusters which are expected to be observed with LSST, which is 10^5 clusters (Tyson *et al.*, 2006). The combination of the projections therefore allow for a closer approximation of the LSST dataset.

Using different projections does not reflect a possibility with actual observations as

we cannot change our perspective, but is used, as already mentioned, as a method of detecting the fluctuations in the measurement of the gravitational redshift from the simulation and to obtain a larger dataset.

First, the dataset (Mock dataset) is assumed to be obtained spectroscopically, meaning the error on the velocity for each galaxy is negligible. This is used to run different tests, for example to check the pipeline for the data processing as well as the simulation output. It is also used to compare the two other datasets which represent different types of observations and the errors they yield. As it is assumed this dataset contain only negligible observational error this dataset represent the best case scenario for detecting the gravitational redshift.

The second dataset (LSST mock 1) simulates data as it is expected to come from photometry using LSST. The velocity of each galaxy, except for the BCGs, is added with a random Gaussian distributed error with $\sigma = 1500$ km/s. The σ for the velocity distribution of the Mock dataset is ≈ 500 km/s, making the added distribution 3 times wider than this, which results in an overall widening of the velocity distribution of the LSST mock 1 dataset, see appendix B.1. A rough estimate for the error of the mean of the distribution, would be that it is at least 3 times larger for the LSST mock 1 dataset compared with the Mock dataset because of the widening. This is a lower limit as it does not take into account the widening of the velocity distribution from peculiar motion of the galaxies.

The reason for assuming spectroscopic data for the BCGs is that these galaxies are very bright and therefore relatively easy to obtain spectroscopic measurements for, whereas the other galaxies in a cluster can be a lot dimmer making photometry a faster way to observe them.

Finally, the third dataset (LSST mock 2), simulates data where all velocities, including the BCGs, are obtained using photometry. The same random error distribution as for the LSST mock 1 dataset was added to the velocity of each galaxy in this dataset.

The errors on the gravitational redshift of each dataset are expected to increase in the order they are listed here, as the velocity distribution becomes wider as a result of the increasing uncertainty of which the galaxies are measured.

The two LSST mocks are simulated photometric data because of the added Gaussianly distributed error, and σ was chosen to reflect the anticipated uncertainty of the photometric data from LSST. The $\sigma = 1500$ km/s chosen, corresponds to a redshift error of $\Delta z \approx 0.005$. The goal for the photometric error for the LSST survey is $\frac{\sigma_z}{1+z} \approx 0.02$, but

maybe get as low as $\frac{\sigma_z}{1+z} \approx 0.01$ for a subsample (LSST Science Collaboration, 2009). Considering this goal is for the full survey of the LSST and the kind of observations needed to detect gravitational redshift, the estimate of $\Delta z \approx 0.005$ might be optimistic, but possible. The clusters included in a sample to detect the gravitational redshift would typically have $z < 1$, and the galaxies in the clusters included would be elliptical galaxies, which are easier to determine accurately with photometry.

The typical redshift error for spectroscopy is $\Delta z \approx 0.0001$, which shows that even this optimistic estimation of the photometric redshift error, it is still about 50 times less accurate than spectroscopy. This lack of accuracy can be compensated for by compiling a larger dataset, which will be discussed further in section 4.4 below.

By using $\sigma = 1500$ km/s results in a wider velocity distribution in itself, compared to the velocity distribution from peculiar motion of the cluster galaxies, which is of the order 1000 km/s, meaning the added distribution affects the width of the overall distribution. It also indicates that the error from photometry will be the dominant source of uncertainty on the measurements.

The spectroscopic redshift error is much smaller making it negligible compared to the velocity distribution of the cluster. If observations using photometry can be made with a greater accuracy than what is assumed here, it could become comparable to or smaller than the velocity distribution of the clusters. This in turn means the photometric measurements will no longer be the dominant source of error, and data from photometry might be as useful as that from spectroscopy in this context. This will be considered further after the analysis of the data in section 4.

Another goal for the LSST is an error of 1% on photometric measurements (LSST Science Collaboration, 2009) is realistic as it has already been achieved, using a small dataset of stars from the SDSS database (Ivezic *et al.*, 2007). As also pointed out by Ivezic *et al.* (2007), similar calibration using multiple observations of the same objects should be possible with LSST, resulting in a photometric error of 1% and possibly lower, supporting the optimistic assumption of the redshift error for photometry made here.

3.3 Fitting procedure

For fitting the data, the *emcee* algorithm for Python was used (Foreman-Mackey *et al.*, 2013). The algorithm uses the Markov Chain Monte Carlo (MCMC) method to maximize the likelihood of any given input function. The MCMC method samples the posterior probability distribution for the desired parameters using walkers to probe the parameter space.

The MCMC algorithm works by having a number of walkers scattered randomly

around the parameter space indicated by the input function as the starting point. The walkers then performs random steps around the parameter space comparing the likelihood of consecutive steps and based on this either accept the new position or go back to the old position depending on which one has the largest likelihood of the two. This results in the walkers converging on the maximum of the probability distribution. To get a proper distribution of the walkers in the parameter space a certain number is required, just as the number of steps they perform must be high enough to ensure they converge on the maximum likelihood.

The advantages of using MCMC sampling are that it works well for high dimensionality of the parameter space and it marginalizes nuisance parameters automatically. The functions used to fit the data in this thesis has 4-7 free parameters, and several of these are nuisance parameters. Further, the *emcee* algorithm is very user friendly and only require tuning of a few parameters. See the full code used for setting up the *emcee* algorithm and fitting of the data in appendix A.

To fit the data maximizing of the likelihood, L is used.

$$L = \prod_{i=1}^N f(x_i|\bar{\theta}) \quad (8)$$

where i is the index running over all the data, which in this case is galaxies, x is the data for each galaxy and θ is the set of parameters to be fitted.

It is often more convenient to use the logarithmic likelihood when fitting to avoid numerical mistakes when computing.

$$\ln L = \sum_{i=1}^N \ln \left(f(x_i|\bar{\theta}) \right) \quad (9)$$

The function, f used in this thesis is a composite of multiple parts, which can vary from fit to fit. It will have a component of a linear function, $(mx + b)$, to account for background galaxies and a component of a Gaussian function which can vary between a single Gaussian

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(x - \mu)^2}{2\sigma^2}\right) p_c + (mx + b) (1 - p_c) \quad (10)$$

and a double Gaussian

$$f(x) = \left(\frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(\frac{-(x-\mu)^2}{2\sigma_1^2}\right) p_g + \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left(\frac{-(x-\mu)^2}{2\sigma_2^2}\right) (1-p_g) \right) p_c + (mx+b)(1-p_c) \quad (11)$$

where μ is the mean of the Gaussian and is the measure of gravitational redshift, p_c is the probability of a given galaxy being in the cluster or part of the background, and p_g is the normalising weight of the two Gaussian components. It is important to remember normalization of the different components, but the scaling is arbitrary. For the double Gaussian it is assumed they have the same mean value, μ , but different widths, σ_1 and σ_2 . For a discussion of this assumption see appendix B.6.

For the single Gaussian fit there are four free parameters to be fitted; μ , σ , p_c and m , where the last two are nuisance parameters. The most interesting parameter is μ which represents the gravitational redshift. The double Gaussian fit has two additional parameters, σ_2 and p_g where p_g is also a nuisance parameter. The parameter b , in the linear component, is a constant from integrating over the linear function, which should be normalized to 1 over the range of velocities used.

The linear component of the fitting function is assumed to have a slope of zero in the simulation, but if data from a survey were used, a small negative slope would be expected. This would be a result of flux limited surveys, giving the impression of more galaxies closer to the observer than further away.

For the fitting procedure certain priors are used. The linear function used to fit the background cannot assume a negative value as it would not make sense to have a negative contribution from the background. The weights used, p_c and p_g , must have a value between 0 and 1 to normalize the functions. The width of the Gaussian function, σ , is constrained as always being larger than zero. Finally, for the double Gaussian fit the width of the two Gaussian components, σ_1 and σ_2 , one is always larger than the other, which is to prevent a mixing of the two components.

The best-fit values used in all the plots in this thesis, are from the walker with the largest likelihood and the errorbars shown are the 1σ error calculated based on the distribution of the walkers in the parameter space. In some of the plots in section 4 the probability distribution of the parameters appear to be asymmetric based on the errorbars or that the fitting algorithm did not converge. However, this is explained by the

fact that the value with the maximum likelihood is not necessarily in the middle of the distribution. All fits are tested for convergence, and the probability distribution of all parameters are Gaussianly distributed. For details on convergence and distribution of the parameters see appendix B.4.

3.4 Bayesian Information Criterion

The Bayesian Information Criterion (BIC) can be used to compare models or to select a preferred model. It is calculated using the maximum likelihood, L , but also takes into account the number of parameters in the model, as it is always possible to get a better fit using more parameters, which can result in overfitting.

$$\begin{aligned} BIC &= -2\ln(L) + k\ln(N) \\ &= \chi^2 + k\ln(N) \end{aligned} \tag{12}$$

where k is the number of free parameters and N is the number of data points. The second term is a penalty for adding parameters to the fit. The second equation assumes errors are identically and independent distributed, which is typical for a Gaussian distribution.

The model with the lowest BIC-value is preferred, either because it has the fewest parameters or a better fit or both apply. By calculating the BIC for each model, it is possible to compare multiple models at once.

When analysing data in this thesis two models will be considered when calculating the BIC; one assuming no gravitational redshift, and one assuming a gravitational redshift with a mass corresponding to the mean mass of the data fitted. Comparing these two values will indicate whether a detection of the gravitational redshift is made.

4 Analysis

As mentioned in section 3.2 a total of three different datasets, each with three independent projections were used in the following analysis.

Each projection was binned into seven bins, which were chosen to be smaller in the inner regions of the clusters and larger at greater distance from the centre of the clusters. This was done because the features in the gravitational redshift profile vary the most at smaller radii within the virial radius, which is $r_v \approx 1.7$ Mpc for the full datasets, but dependent on cluster mass, see equation 6 and figure 1. Making few evenly spaced bins over the full range, or too few bins in the inner regions, could obscure the gravitational redshift signal. Therefore, the size of the bins were chosen to be uneven, with more bins

closer to the centre of the clusters and fewer at greater cluster centric distance.

There are more galaxies in the outer parts of a cluster because of a larger volume, therefore the number of galaxies in each bin vary greatly, which is reflected in the size of the errors of each bin. The number of bins can also affect the size of the errors, because a larger number of bins require the data to be distributed more thinly resulting in a smaller amount of data in each. On the other hand, too few bins can obscure the signal as already mentioned. For a full discussion of the number of bins used see appendix B.3.

The values from fitting the data of a specific bin is plotted using the value of gravitational redshift with the largest likelihood and the median radius of the bin, and all errorbars show a 1σ confidence interval.

Originally I expected to run double Gaussian fits on all the data for the best determination of the gravitational redshift based on previous work. But a double Gaussian fit has two more parameters (width of the second Gaussian, σ_2 , and weighting of the Gaussians, p_g) than a single Gaussian fit, meaning the computational time is longer and there is a danger of adding meaningless parameters in order to get a better fit. Appendix B.5 discusses the factors affecting computation time as well as give examples of the length of the computation time for the different functions.

I ran tests fitting both functions from equations 10 and 11 to the same Mock dataset projection in order to compare the values and errors from the fits. Visually the differences in fitted values and errors seem negligible indicating the two functions fit the data equally well with regards to the gravitational redshift, see figure 2.

However, when calculating the BIC values for the different fits, I found that the single Gaussian fit obtained the smaller value, which makes it the preferred model. This difference in BIC values is in large part because of the penalty of the two additional parameters for the double Gaussian function, indicating the additional parameters of the double Gaussian function is an unnecessary addition to the fitting function.

When setting up the double Gaussian function in section 3.3, it was assumed the mean of the two components was the same, and as described in detail in appendix B.6, when this assumption is not made, the largest contribution to the mean of the double Gaussian fit is the narrow component, further supporting the adequacy of the single Gaussian fit over the double Gaussian fit.

The differences between single and double Gaussian fits was also tested for the LSST datasets with similar results. The differences between the two fitting functions becomes smaller for the LSST mocks as their velocity distribution is wider as is described in section 3.2. For spectroscopically obtained data, i.e. the Mock dataset, the velocity

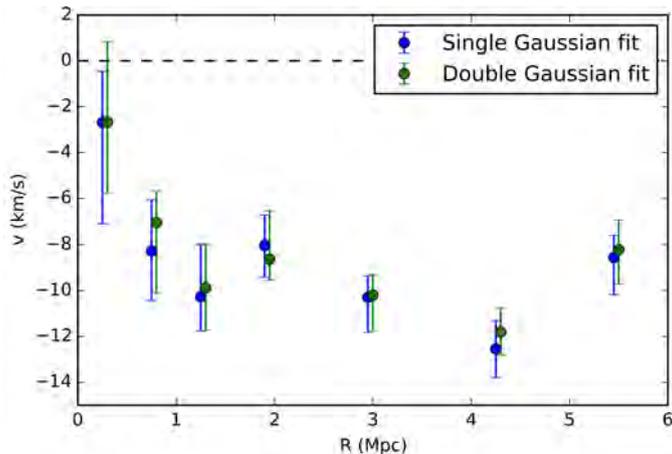


Figure 2: Profiles of the fitted gravitational redshift from one of the projections of the Mock dataset using a single Gaussian fit, equation 10 (blue) and a double Gaussian fit, equation 11 (green). Both fits show similar values of the gravitational redshift and size of errors in all bins indicating the two functions fit the data equally well with respect to the gravitational redshift

distribution is more peaked and the single Gaussian fit has difficulty fitting the wings of the distribution, but this does not affect the mean value significantly as is shown in figure 2. See appendix B.1 for histograms of the distributions of the different datasets.

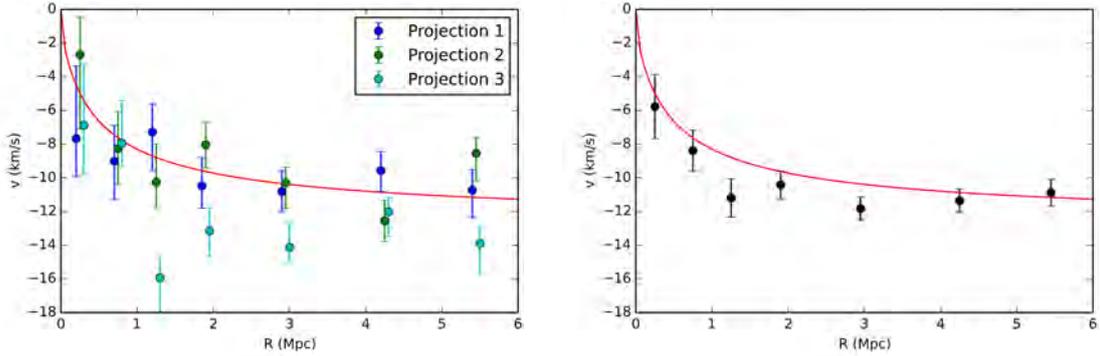
Therefore, all fits in the rest of this section are results of a single Gaussian fit of equation 10.

4.1 Mock dataset

The Mock dataset represent data obtained using spectroscopy as described previously. This dataset is expected to have a gravitational redshift signal close to the profile predicted by equation 2, and with smaller errors than the LSST mocks analysed in the following sections, because of the smaller errors on the measurements.

This dataset is similar to the data analysed in Wojtak *et al.* (2011), but with two key differences; firstly, the data analysed here are simulated, whereas the data used in the paper are from SDSS Data Release 7, and secondly the size of the dataset used here is roughly 24 times larger. Because of the smaller dataset, the data is only divided into four bins, whereas there is used seven here.

The fitted values of the gravitational redshift in all bins for the three projections of this dataset are shown in figure 3(a). There are variations between the projections as is to be expected due to statistical fluctuations, but all data agrees with the prediction. In figure 3(b) a weighted average of the combination of the three projections are shown.



(a) Profiles of the three projections of the Mock dataset. The average error over all bins is ± 1.4 km/s (b) Profile of the weighted average of the three projections of the Mock dataset

Figure 3: Gravitational redshift values from fitting the full Mock dataset (points) and the theoretical prediction of the gravitational redshift profile (line). The data from the simulation are in agreement with the theoretical prediction

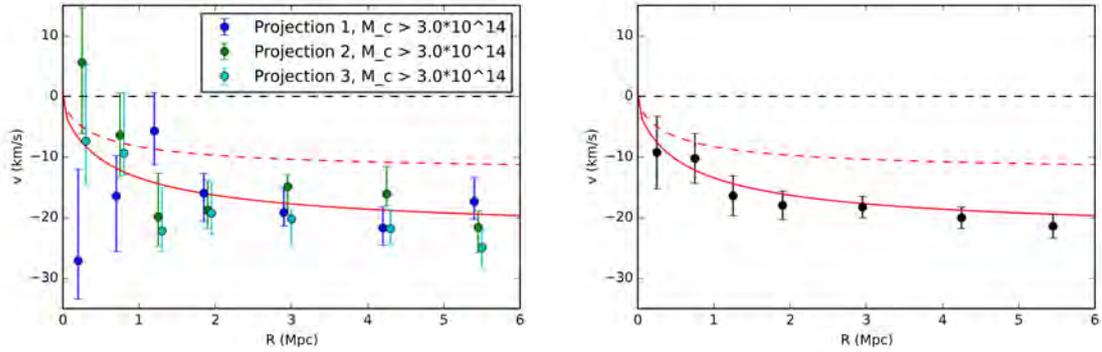
To calculate the weighted average, symmetric errors on the fitted values were assumed, see appendix B.4 for typical parameter distributions.

These results are as expected, as they have relatively small errors of ± 1.4 km/s on average over all bins for each projection, which contains roughly $3 \cdot 10^6$ galaxies distributed over $2 \cdot 10^4$ clusters each, and are close to the prediction. For the combination of the three projections the average error is ± 0.8 km/s for a dataset containing $9.1 \cdot 10^6$ galaxies and $6 \cdot 10^4$ clusters.

I also looked into the possibility of obtaining a larger and possibly clearer signal of the gravitational redshift signal by using data only from massive clusters with cluster masses $M_c > 3 \cdot 10^{14} M_\odot h^{-1}$, and the results from these fits are plotted in figure 4.

By using only the clusters with a mass above $3 \cdot 10^{14} M_\odot h^{-1}$, the sample becomes smaller, consisting of just over $5.2 \cdot 10^5$ galaxies in each projection ($\approx 17\%$ of the full sample) and $2.8 \cdot 10^3$ clusters ($\approx 14\%$ of the total number of clusters). The cut chosen, was to obtain a large enough sample that included a sufficient amount of clusters and galaxies to get useful results. If too few clusters were included, this could affect the results because substructures might not be smoothed out.

The smaller sample also affects the errors on the fit, which become roughly three times larger with the size ± 3.9 km/s for each projection. This is especially clear for the inner bin, which has very large errors because of the small number of galaxies, although the fitted values still remain consistent with theory within the uncertainties.



(a) Profiles of the three projections of the Mock dataset. The average error over all bins is ± 3.9 km/s

(b) Profile of the weighted average of the three projections of the Mock dataset showing excellent agreement with theory

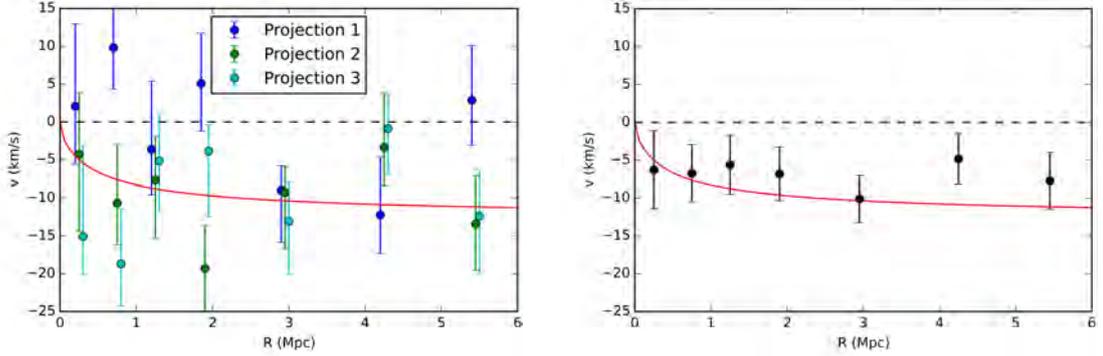
Figure 4: Plot of fits using only galaxies from clusters with a mass above $3 \cdot 10^{14} M_{\odot} h^{-1}$ from the Mock dataset (points). The solid line represents the prediction of gravitational redshift using a mean cluster mass of $4.67 \cdot 10^{14} M_{\odot} h^{-1}$ corresponding to the mean cluster mass of the massive clusters fitted, while the dashed red line is the prediction of gravitational redshift for the mean cluster mass of the entire dataset which is $1.84 \cdot 10^{14} M_{\odot} h^{-1}$

There are noticeable differences between the projections, where the first projection has large fluctuations compared to the other two especially within the inner three bins, but it is not statistically significant. I noticed through out the analysis that the bin with the largest fluctuations in values and errors was the innermost bin, which is consistent with this being the bin with the smallest number of galaxies and a smaller number of clusters contributing.

Notice the greater depth of the profile of gravitational redshift for the massive clusters compared with the full sample. The data support this, especially at larger cluster-centric distance where the differences between the predictions are greater. At small radii it is more difficult to separate the results for the massive clusters from the prediction of the entire dataset, due to a smaller sample resulting in larger errors and a smaller difference between the two predictions, figure 4(a). This becomes less noticeable when looking at the weighted mean of the three projections, figure 4(b), which shows a good agreement with the prediction.

4.2 LSST mock 1

The LSST mock 1 dataset simulates data as is expected to come with photometry from LSST. The survey will obtain both photometric and spectroscopic data, but mainly photometric data. The LSST mock 1 represents data where the BCGs are observed



(a) Profiles of the three projections of the LSST mock 1 dataset showing a large scatter. The average error over all bins is ± 5.8 km/s

(b) Profile of the weighted average of the three projections of the LSST mock 1 indicating a constant gravitational redshift signal at 6.5 km/s

Figure 5: Gravitational redshift values from fitting the LSST mock 1 (points) and the theoretical prediction of the gravitational redshift for a mean cluster mass of $1.84 \cdot 10^{14} M_{\odot} h^{-1}$ (line)

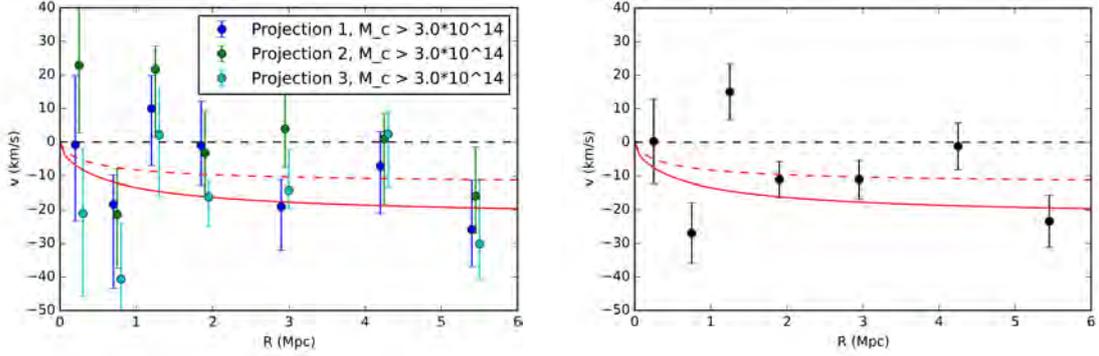
spectroscopically and all other cluster galaxies have photometric data.

First, the same fits as for the Mock dataset above were run, see figure 5. The first thing to notice is that the errors are about five times larger, ± 5.8 km/s, than for the full Mock dataset, ± 1.4 km/s for each projection. This was as expected due to the added velocity distribution, making the overall distribution wider and therefore more difficult to determine the centre of it.

The scatter is large between the projections, and it is difficult to conclude whether the gravitational redshift signal is present and what the associated mass of the cluster is. Particularly with regards to projection 1, the results indicate that no signal was detected because the values are scattered around zero. Also, the BIC value assuming no gravitational redshift is lower than the BIC value assuming that there is one. For the other two projections there were some indication of a gravitational redshift signal. This means that the weighted mean does indicate a gravitational redshift signal, figure 5(b). Although the scatter between the projections is large the results are still consistent with theory and with each other within the uncertainties which are also quite large.

When looking at the weighted mean, there appears to be a constant gravitational redshift signal around 6.5 km/s, with an average error of ± 3.5 km/s, as no drop off at small radii is indicated. This could possibly be resolved using a sample larger than the combination of the three projections consisting of $9.1 \cdot 10^6$ galaxies from $6 \cdot 10^4$ clusters, as this could bring the errors down, see section 4.4.

Again, the most massive clusters with mass above $3 \cdot 10^{14} M_{\odot} h^{-1}$ were fitted separately,



(a) Profiles of the three projections of the LSST mock 1 dataset. The average error over all bins is ± 13 km/s

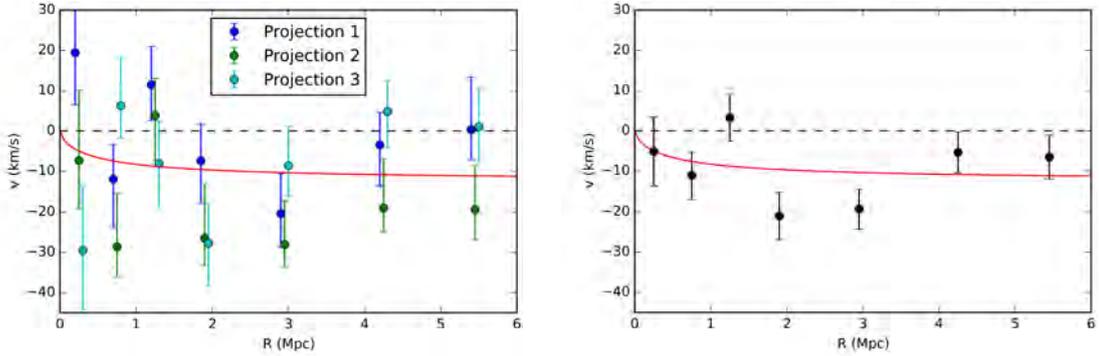
(b) Profile of the weighted average of the three projections of the LSST mock 1 dataset

Figure 6: Plot of fits using only galaxies from clusters with a mass above $3 \cdot 10^{14} M_{\odot} h^{-1}$ from the LSST mock 1 dataset (points). The solid red line represents the prediction of gravitational redshift using a mean cluster mass of $4.67 \cdot 10^{14} M_{\odot} h^{-1}$ corresponding to the mean cluster mass of the massive clusters fitted. The dashed line is the prediction of the gravitational redshift for the mean cluster mass of the entire dataset which is $1.84 \cdot 10^{14} M_{\odot} h^{-1}$

to investigate the size of the signal, see figure 6. As for the full sample, the errors are large and could be consistent with a gravitational redshift signal of zero. For the massive clusters, the second projection is the one most consistent with no signal, while none of the projections clearly differentiate between the prediction for the full sample and the prediction only for the massive clusters.

The weighted mean in figure 6(b) only has a slight preference for a gravitational redshift signal compared to no signal. A model with a lower mass, for example the mean mass for the full dataset $1.84 \cdot 10^{14} M_{\odot} h^{-1}$, has an even lower BIC making the model preferable to the model with a mean mass of the clusters used. This indicates that the dataset is only able to show whether or not a gravitational redshift is present, but it can not determine the size of it accurately.

Due to the scatter of the data, where the errors are of the same size as the signal, ± 13 km/s for a single projection, it is difficult to conclude whether or not it is possible to detect a gravitational redshift signal based on the massive clusters from a single projection alone without a significantly larger sample. The combination of the three projections indicate a gravitational redshift signal with an average error of ± 7.0 km/s, but will require a larger sample than the $1.6 \cdot 10^6$ galaxies and $8.4 \cdot 10^3$ massive clusters for a more accurate determination.



(a) Profiles of the three projections of the LSST mock 2 dataset. The average error over all bins is ± 9.5 km/s

(b) Profile of the weighted average of the three projections of the LSST mock 2

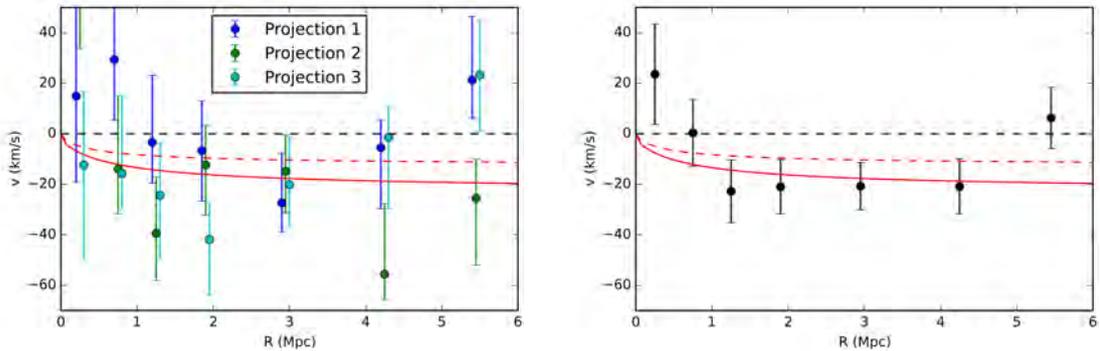
Figure 7: Gravitational redshift values from fitting the LSST mock 2 (points) and the theoretical prediction of the gravitational redshift (line) for a cluster with mass $1.84 \cdot 10^{14} M_{\odot} h^{-1}$. The scatter of the projections makes it difficult to determine whether or not a gravitational redshift signal is detectable for the individual projections

4.3 LSST mock 2

The LSST mock 2 dataset simulates photometric data for all galaxies in the sample, including the BCGs, which is expected to make the determination of the gravitational redshift even more difficult than for the LSST mock 1 sample. The results from the fitting of the full dataset is shown in figure 7. As was the case for the LSST mock 1 dataset, these projections have a large scatter and some of the projections (projections 1 and 3) are consistent with a gravitational redshift signal of zero. Calculating the BIC values for models with and without the signal gives very similar results, making them indistinguishable with regards to this dataset.

Even though the scatter is large it is still consistent with the size of the errorbars. The errors become almost twice as large, ± 9.5 km/s for a single projection, without the spectroscopic determination of the BCGs in comparison with the LSST mock 1 dataset where the average error was ± 5.8 km/s for each projection. The combination of the three projections has an average error of ± 5.4 km/s and indicate that it is possible to detect the gravitational redshift using all photometric data, if the sample is large enough.

The velocity distribution for this dataset was so wide that the ± 6000 km/s velocity cut-off used was an absolute minimum in order to get the entire distribution. A wider velocity cut-off, for example ± 7000 km/s, could potentially improve the quality of the fit, particularly with regards to the wings of the distribution. This larger range could also reduce the uncertainty of the fit. See plots of the distribution and a discussion of



(a) Profiles of the three projections of the LSST mock 2 dataset. The average error over all bins is ± 19 km/s

(b) Profile of the weighted average of the three projections of the LSST mock 2 dataset

Figure 8: Plot of fits using only galaxies from clusters with a mass above $3 \cdot 10^{14} M_{\odot} h^{-1}$ from the LSST mock 2 dataset (points). The solid line represents the prediction of gravitational redshift using a mean cluster mass of $4.67 \cdot 10^{14} M_{\odot} h^{-1}$ corresponding to the mean cluster mass of the massive clusters fitted. The dashed line is the prediction of gravitational redshift for the mean cluster mass of the entire dataset which is $1.84 \cdot 10^{14} M_{\odot} h^{-1}$

the velocity cut-off used in appendix B.2.

For this dataset it appears that the massive clusters provide a more consistent signal than the full dataset, see figure 8. It still has large errors of the same order of magnitude as the signal, ± 19 km/s for each projection, which makes it very difficult to determine the mass of the gravitational redshift profile, which was also the case for the LSST mock 1 dataset.

4.4 Errors and sample sizes

It is clear from the analysis in the sections above, that in order to determine the gravitational redshift signal accurately the type of data and the size of the dataset play important roles. By using data based on photometry a much larger sample is required compared to a sample of spectroscopic data, because, as is well known in statistics, a larger sample size can reduce the size of errors.

All the datasets used in this analysis are of the same size and contain data from an equal amount of clusters making them comparable to each other. Comparing two different types of data, for example the results from the LSST mock 1 and the results from the Mock dataset with only the massive clusters, show that the size of the errors are of the same order of magnitude, see figure 9. The LSST mock 1 sample consisting of

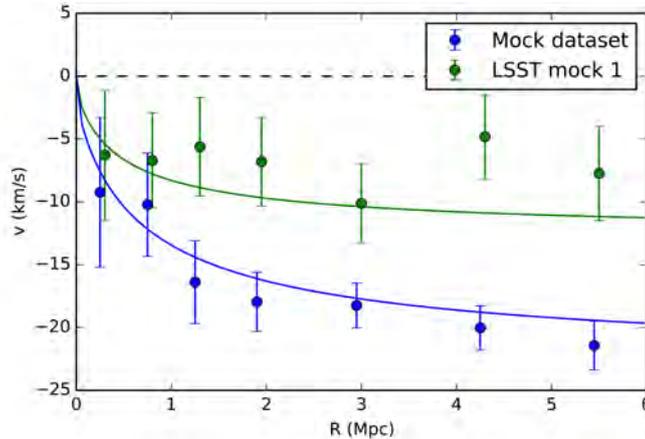


Figure 9: Profile of the full LSST mock 1 dataset (green) and the Mock dataset with cluster masses above $3 \cdot 10^{14} M_{\odot} h^{-1}$ (blue) and corresponding gravitational redshift profiles in the same colour (lines). The points are the weighted average of the three projections from the respective datasets. The LSST mock 1 dataset is about six times larger than the Mock dataset of massive clusters, but the errors are of the same order of magnitude

all three projections contain $9.1 \cdot 10^6$ galaxies from $6 \cdot 10^4$ clusters, while the Mock dataset of massive clusters from all three projections contain only $1.6 \cdot 10^5$ galaxies distributed over $8.4 \cdot 10^3$ clusters. This shows that in order to determine the gravitational redshift signal from a sample of the size that LSST is expected to generate (see below), can also be accomplished by a smaller sample containing only spectroscopic data.

An important point to keep in mind is that while the gravitational redshift signal was very clear and in agreement with prediction for the spectroscopic data, it was less accurate based on the photometric samples of the size available here. But in the following, I will only consider the size of the errors of the different datasets and not the values they are related to.

In order to investigate the effect of sample size on the errors of gravitational redshift, I drew random subsamples from the datasets and found the weighted average error on the gravitational redshift of a given sample using fits from each bin, see figure 10. The upper limit on the sample size available is the combination of the three projections of each dataset.

The data appears as a straight line in a log-log plot indicating a power law relation between the size of the error and the sample size. This is consistent with errors commonly being proportional to $\propto 1/\sqrt{N}$, where N is the number of data points. In the plot it is also evident that spectroscopic data (Mock dataset) have noticeably smaller errors than

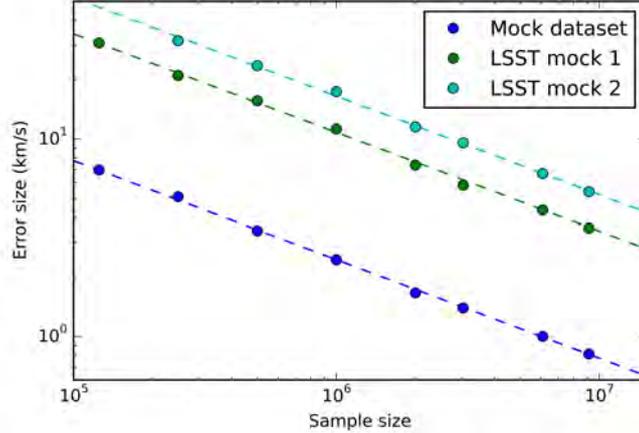


Figure 10: *Plot of the combined weighted average error in the gravitational redshift of the different datasets as a function of the size of the respective datasets. The size of the errors depend on the number of bins, and this plot is based on the use of seven bins. The errors decrease with increasing sample size at the same rate for all types of data with a proportionality $\propto 1/\sqrt{N}$ (dashed lines)*

the photometric datasets (LSST mock 1 and 2), which was as expected and clear in the analysis above.

The errors plotted in figure 10 are the weighted average of the errors in each bin for a specific dataset. By fitting the entire dataset as one bin the error will be significantly smaller for a given sample size than what is plotted in figure 10. But by doing this it will not be possible to trace the gravitational redshift profile as a function of the radius of the cluster. The errors for each dataset, as calculated here, therefore depends on the number of bins, as fewer bins will result in a larger sample in each and therefore smaller errors, while the opposite applies to more bins. The arguments for the number of bins used here are described in section 4 and appendix B.3.

The gravitational redshift value and error quoted by Wojtak *et al.* (2011), -7.7 ± 3.0 km/s, is the result of fitting all their data in a single bin. This is over half the size of error, which is shown in figure 10 for a sample the same size but binned in seven bins (± 6.9 km/s for a sample with $1.25 \cdot 10^5$ galaxies). By fitting a subsample of the Mock dataset of the size $1.25 \cdot 10^5$ galaxies in a single bin, similar results were achieved with the simulation with regard to the size of the error, -12.8 ± 3.2 km/s, showing the Mock dataset is comparable to spectroscopic data.

The average error on the gravitational redshift of a sample of spectroscopic data (Mock dataset) consisting of $2.5 \cdot 10^5$ galaxies (± 5.1 km/s) has roughly the same error as a

	± 1 km/s	± 2 km/s	± 3 km/s	± 4 km/s
Mock dataset	$6 \cdot 10^6$	$1.5 \cdot 10^6$	$7 \cdot 10^5$	$4 \cdot 10^5$
LSST mock 1	10^8	$3 \cdot 10^7$	$1.5 \cdot 10^7$	$8 \cdot 10^6$
LSST mock 2	$3 \cdot 10^8$	$7 \cdot 10^7$	$3 \cdot 10^7$	$2 \cdot 10^7$

Table 1: *Total sample sizes required to achieve a certain average error on the gravitational redshift in the different datasets analysed. The errors and sample sizes are calculated based on the use of seven bins to trace the gravitational redshift profile*

sample of photometric data for all galaxies except the BCGs (LSST mock 1) of $3 \cdot 10^6$ galaxies (± 5.8 km/s) or a sample of data with photometry for all galaxies (LSST mock 2) of $9.1 \cdot 10^6$ galaxies (± 5.4 km/s).

The relation between the size of errors are a constant scaling factor between the datasets, indicated by the identical slopes in figure 10, and the relation between the sample sizes are the square of this factor, because of the proportionality to $1/\sqrt{N}$. The scaling factor between the Mock dataset and LSST mock 1 is 4.5, between Mock dataset and LSST mock 2 is 6.5 and between LSST mock 1 and 2 is 1.5. Remembering the added random error for the LSST mock 1 and 2 had $\sigma = 1500$ km/s, which were 3 times the σ for the velocity distribution of the Mock dataset, these factors are of the same order of magnitude, showing the relation between the width of the velocity distribution and the error and sample size of the dataset.

These scaling factors are also roughly reflected in the width of the final velocity distributions of the datasets, though usually slightly smaller, see appendix B.1 for velocity distributions of the datasets. This is consistent with the fact that the errors on the mean of the Gaussian, i.e. the gravitational redshift, increases when the distribution becomes wider/flatter.

By extrapolating the LSST mock 1 and 2 datasets, beyond the sample sizes available from the simulation used in the analysis here, in order to get the error down to ± 3 km/s, sample sizes of $1.5 \cdot 10^7$ and $3 \cdot 10^7$ galaxies will be needed for LSST mock 1 and LSST mock 2 respectively. Because of the power law relation getting very low errors of the size ± 1 km/s will require much larger samples of 10^8 and $3 \cdot 10^8$ galaxies for LSST mock 1 and LSST mock 2 respectively. See table 1 for sample sizes required for different error sizes for all three datasets analysed.

The LSST is expected to observe about 10^5 clusters with a mass above $2 \cdot 10^{14} M_\odot$ (Tyson *et al.*, 2006), but will be sensitive to clusters with masses above $0.5 \cdot 10^{14} M_\odot$ (LSST Science Collaboration, 2009). Also, the LSST survey will observe around 10^{10} galaxies, but not all of these will be in clusters. Comparing these numbers to the

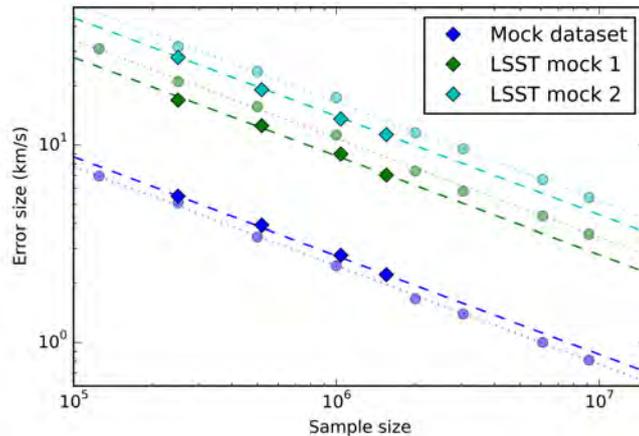


Figure 11: *Plot of the combined weighted average error of the different datasets, containing only the massive clusters with a mass above $3 \cdot 10^{14} M_{\odot} h^{-1}$, as a function of the size of the respective datasets (diamonds). The transparent points represent the full datasets (the same as in figure 10). The size of the errors depend on the number of bins, and this plot is based on the use of seven bins. The errors decrease with increasing sample size at the same rate for all types of data with a proportionality $\propto 1/\sqrt{N}$ (dashed lines)*

sample sizes extrapolated it only seem marginally possible to get errors below 3 km/s for photometric data as this would require data on over 500 galaxies per cluster, which would only be possible for the most massive clusters. The average number of galaxies in a cluster in the simulation was 153, making it likely to obtain $1-2 \cdot 10^7$ galaxies from the 10^5 clusters expected from LSST.

The largest available datasets used here would be a combination of the three projections consisting of $9.1 \cdot 10^6$ galaxies from $6 \cdot 10^4$ clusters in total, and are still smaller than what is expected by the LSST. Therefore, for photometric data for all galaxies except the BCGs the error can be expected to have an upper limit of ± 3.5 km/s, while the upper limit for photometric data for all galaxies including the BCGs is ± 5.4 km/s.

In figure 11 the weighted average errors on the gravitational redshift of the samples of massive cluster with a mass above $3 \cdot 10^{14} M_{\odot} h^{-1}$ are plotted with the results from the full samples indicated as well. It is interesting to note that for the spectroscopic data the errors for the massive clusters are slightly larger, while they for the photometric data are smaller than for the full samples.

The difference in errors for the same sample size for the spectroscopic data are small (less than 0.5 km/s), meaning that there is little to no gain in using only massive clusters for the determination of the gravitational redshift. But for the two types of

photometric data the difference is more significant, 3-4 km/s. This is probably because the velocity distribution of galaxies in the clusters are slightly more narrow for the massive clusters compared with the full sample, making it easier to determine the centre of the distribution as described in section 3.2. This is not the case for the spectroscopic sample, where the massive clusters have a wider distribution than the full dataset because of the well defined distribution, where the less massive clusters are the main contribution.

Initially this wider distribution for the massive clusters for the spectroscopic data seems counter-intuitive as adding constraints on the cluster mass, should focus the velocity distribution. This is because some of the uncertainty from the combination of different velocities for different cluster masses should be removed. When this is not the case for the massive clusters, it is because this sample span an order of magnitude in cluster mass, whereas the full sample cover only a slightly larger range. But the smaller mass range omitted from the sample of massive clusters contain the largest amount of clusters, making these the dominant contribution to the overall velocity distribution for the full sample. For the photometric datasets this apparently plays a smaller role, as the distribution is already widened because of the measurement uncertainties dominating over the width of the distribution due to peculiar motion. If the error on photometric measurements becomes smaller, this difference between massive clusters and the full sample might even out.

5 Discussion

Detection of the gravitational redshift signal, with the method described here, does not depend on observations of a specific type of galaxy (i.e. spiral or elliptical galaxies), the only requirement is that they are situated in a cluster. Elliptical galaxies are very typical in clusters, especially in the central regions, and the BCG of a cluster is almost always a large elliptical galaxy. Spectra of elliptical galaxies usually only have few features, such as the Lyman and Balmer breaks at $\sim 1000 \text{ \AA}$ and $\sim 4000 \text{ \AA}$ respectively at rest wavelengths, making the spectra relatively easy to determine and fit with a redshift.

Photometric redshifts are obtained observing in multiple broadband filters, and by using the colour differences between filters, determine the wavelength of the features in the spectrum. The redshift of an object is determined based on strong features, such as the Lyman and Balmer breaks. A drop in brightness between two bands can be an indicator of one of these breaks. Mistaking the Lyman and Balmer breaks with one another at identification can lead to large systematic errors. This can happen for a Balmer break at a low redshift or a Lyman break at a high redshift, because they would

appear in the same filter.

Because of their simple spectra with few features, elliptical galaxies have better redshift determinations with photometry than spiral galaxies. Considering the simplicity of the spectrum, it might also be possible to determine the photometric redshift for this type of galaxy with an accuracy better than the $\sigma = 1500$ km/s assumed in this thesis, see section 3.2. If this becomes the case then the errors and sample sizes for the photometric datasets calculated in section 4.4 and shown in figure 10 would become smaller, approaching the errors and sample sizes calculated for the spectroscopic data, meaning the scaling factors between the datasets will go down. Given that the sample sizes expected with an average error of ± 3 km/s for the photometric data with and without the spectroscopic determination of the BCGs are $1.5 \cdot 10^7$ and $3 \cdot 10^7$ respectively, assuming $\sigma = 1500$ km/s, any decrease in sample size would be a gain as these samples are large, even considering the size of the LSST survey.

Another way of affecting the sample sizes required for a certain error, would be to use fewer bins when binning the data with respect to cluster centric distance. If this is considered, then it will also be harder to trace the gravitational redshift profile accurately, especially within the virial radius of the cluster, see appendix B.3. On the other hand if a more accurate trace of the gravitational redshift profile is wanted than what is suggested here, the larger number of bins would result in larger sample sizes for the same size of error.

When selecting galaxies and clusters from a survey, different cuts are made in order to compile a useful dataset. Data on the BCGs are necessary, and if this needs to be spectroscopic data, a significant amount of clusters is typically disregarded. Also, a lower limit on the number of cluster members is typically set, so clusters with data on less than for example five members are disregarded, in order to have a robust determination of the cluster. These cuts can eliminate a significant amount of galaxies in a survey, so typically only 10-40% of the galaxies in a survey can be used if only spectroscopic data are considered. But if photometric data can be used, these cut will most likely be less severe as more data can be included in the samples.

As previously shown by Wojtak *et al.* (2011); Kaiser (2013); Kim & Croft (2004) and Zhao *et al.* (2013) among others, it is possible to determine the gravitational redshift using the velocity distribution of galaxy clusters, which was also confirmed here with a greater certainty, based on simulated data. In addition it also appears to be possible to determine the effect based on photometric measurements, although a much larger sample than what was used here is required for an accurate determination.

The scatter of the photometric data appear to be larger than of the spectroscopic data, so further studies of larger samples are needed to determine if this is due to drifts in the simulation or actual effects seen in observations. It appears that a larger sample would solve the problem of observing the gravitational redshift signal with greater accuracy using photometric data, considering the results from the combination of the three projections of each dataset.

It is already possible to use data from Sloan Digital Sky Survey (SDSS), as done by Wojtak *et al.* (2011), and the Dark Energy Survey (DES) to measure the effect of gravitational redshift. The SDSS has collected data from around 470 million objects with photometry and a smaller sample with spectroscopy since data collection started in 2000. The survey covers both stars, galaxies and quasars, so in order to get a sample of only galaxies from clusters of the order needed to detect the gravitational redshift using photometry, a larger sample is needed.

The DES is going to count galaxy clusters in order to determine the gravitational growth of structures in the universe. This data could also be used to determine the gravitational redshift of these clusters, but as the full DES survey is expected to include around 300 million galaxies (Dark Energy Survey Collaboration, 2016), this survey will not be large enough on its own to give an accurate determination of the gravitational redshift using only photometric data, based on the analysis carried out here.

The scope of the LSST is larger than the SDSS and DES combined and a total of 10^{10} galaxies and 10^5 clusters are expected to be observed throughout this survey (LSST Science Collaboration, 2009; Tyson *et al.*, 2006). Because of the large amount of data, it might be possible to select galaxies in clusters with the smallest error for a better determination of redshift and velocity distribution.

If the LSST is able to bring the photometric redshift error below their goal of $\frac{\sigma_z}{1+z} \approx 0.02$ and $\Delta z < 0.005$ (corresponding to $\sigma = 1500$ km/s used here), even if it is only for a subsample, this sample might be large enough to determine the gravitational redshift accurately. By lowering the assumed error on photometry the sample sizes extrapolated in section 4.4 for the LSST mock 1 and 2 will become smaller, approaching the sample sizes for the Mock dataset.

Considering the 10^5 clusters LSST is expected to observe, the sample size required for an error on gravitational redshift of ± 3 km/s for photometric data for cluster galaxies and spectroscopy for the BCGs is $1.5 \cdot 10^7$ galaxies, or data on 150 galaxies on average for all the clusters in the LSST survey. This is definitely possible, but with a lower photometric redshift error and the correspondingly smaller sample size, makes it more probable to obtain this average error.

All three surveys mentioned above could provide data for an accurate determination of the gravitational redshift signal if a combination of spectroscopic and photometric data were to be used. As shown in section 4.4 the difference in accuracy by obtaining the BCGs with spectroscopy is considerable and reduces the sample size required noticeably.

By only using data from galaxies in massive clusters, it becomes easier to detect whether or not a gravitational redshift signal is present, but in order to determine the accuracy of the measurement samples of around the same size as described above is required.

Kim & Croft (2004) have argued that the number of clusters needed for detection does not depend on the mass of the cluster once it is above $10^{14}M_{\odot}h^{-1}$, which is the lower limit on clusters selected for the simulation used here. This is consistent with my findings for spectroscopic data, but when considering photometric data there might be a gain in using massive clusters as they have a slightly more narrow velocity distribution as described in section 4.4. If the error on photometric observations becomes comparable to the velocity distribution, this difference between the full sample and samples using only massive clusters might even itself out and become similar to a spectroscopic sample, removing the gain in using massive clusters.

On the other hand, as Kim & Croft (2004) points out, there may not be enough massive clusters in the universe to make an accurate gravitational redshift detection using only these. Massive clusters are more rare than lower mass clusters, making it less probable to collect a large sample consisting only of these, especially when considering the large number of galaxies and clusters needed to make accurate determinations with photometric data. In the simulation used, only 14% of the clusters had a mass above $3 \cdot 10^{14}M_{\odot}h^{-1}$, classified as massive clusters here, which is not enough to collect a sample of the required size even considering the gain in the size of error of 3-4 km/s.

With photometric data, detection of gravitational redshift using only massive clusters does not seem possible because a large enough sample might never be available. But with spectroscopic data or a sample of photometric data using a wide range of cluster masses an accurate detection seems possible with LSST.

6 Conclusion

In this thesis I have investigated whether it is possible to determine gravitational redshift using different types of data based on simulations. The dataset containing only spectroscopic data showed a clear gravitational redshift signal consistent with the prediction.

The photometric data of two different types, where one of the datasets had spectroscopic measurement of the BCGs and the other had photometric measurements, both showed a large scatter but within errors. This made the detection of the gravitational redshift less certain based on the samples available here, where the largest were $9.1 \cdot 10^6$ galaxies distributed over $6 \cdot 10^4$ clusters. By looking at only the massive clusters from the photometric datasets, it was possible to detect a gravitational redshift signal with a larger certainty, although the uncertainties of the cluster mass associated with the gravitational redshift profile were large.

The sample size required for detection of the gravitational redshift signal, with a certain error for all datasets, was investigated and extrapolated based on the simulated data used for the analysis. To achieve an average error of ± 3 km/s over seven bins, sample sizes of $1.5 \cdot 10^7$ and $3 \cdot 10^7$ galaxies is required with photometric data with and without a spectroscopic determination of the BCG respectively. Based on what is expected to be observed with the LSST, it appears that these sample sizes of photometric data are reachable. This means that it will be possible to detect gravitational redshift with photometric data.

References

- Abbott, B.P., Abbott, R., & Abbott, T.D., et al (LIGO Scientific Collaboration and Virgo Collaboration). 2016. Properties of the binary black hole merger GW150914. *Phys.Rev.Lett.*, **116**(6). arXiv:1602.03840.
- Dark Energy Survey Collaboration. 2016. The Dark Energy Survey: more than dark energy - an overview. *MNRAS*, **460**(2), 1270–1299. arXiv:1601.00329.
- Foreman-Mackey, D., Hogg, D.W., Lang, D., & Goodman, J. 2013. emcee: The MCMC Hammer. *Publications of the Astronomical Society of the Pacific*, **125**(925), 306–312. arXiv:1202.3665.
- Ivezic, Z., Smith, J.A., Miknaitis, G., Lin, H., & Tucker, D., et.al. (SDSS Collaboration). 2007. SDSS Standard star catalog for Stripe 82: the dawn of industrial 1% optical photometry. *The Astronomical Journal*, **134**(3), 973–998.
- Kaiser, N. 2013. Measuring gravitational redshifts in galaxy clusters. *MNRAS*, **435**(2), 1278–1286. arXiv:1303.3663v2.
- Kim, Y., & Croft, R.A.C. 2004. Gravitational redshifts in simulated galaxy clusters. *ApJ*, **607**, 164–174.
- Klypin, A., & Holtzman, J. 1997. Particle-Mesh code for cosmological simulations. arXiv:9712217.
- Klypin, A.A., Trujillo-Gomez, S., & Primack, J. 2011. Dark matter halos in the standard cosmological model: results from the Bolshoi simulation. *ApJ*, **740**(2). arXiv:1002.3660.
- Komatsu, E., Dunkley, J., Nolta, M.R., Bennett, C.L., Gold, B., Hinshaw, G., Jarosik, N., D., Larson, Limon, M., Page, L., Spergel, D.N., Halpern, M., Hill, R.S., Kogut, A., Meyer, S.S., Tucker, G.S., Weiland, J.L., Wollack, E., & Wright, E.L. 2009. Five-Year Wilkinson Microwave Anisotropy Probe (WMAP) observations: cosmological interpretation. *The Astrophysical Journal Supplement*, **180**(2), 330–376. arXiv:0803.0547.
- Lokas, E.L., & Mamon, G.A. 2001. Properties of spherical galaxies and clusters with an NFW density profile. *MNRAS*, **321**(1), 155–166.
- LSST Science Collaboration. 2009. LSST Science Book, Version 2.0. arXiv:0912.0201.

- Navarro, J.F., Frenk, C.S., & White, S.D.M. 1996. The structure of cold dark matter halos. *ApJ*, **462**, 563–575.
- Prada, F., Klypin, A.A., Cuesta, A.J., Betancort-Rijo, J.E., & Primack, J. 2012. Halo concentrations in the standard Λ cold dark matter cosmology. *MNRAS*, **423**(4), 3018–3030. arXiv:1104.5130.
- Riebe, K., Partl, A.M., Enke, H., Forero-Romero, J., Gottlöber, S., Klypin, A., Lemson, G., Prada, F., Primack, J.R., Steinmetz, M., & Turchaninov, V. 2013. The MultiDark Database: Release of the Bolshoi and MultiDark cosmological simulations. *Astronomische Nachrichten*, **334**(7), 691–708.
- Tyson, J.A., Wittman, D.M., Hennawi, J.F., & Spergel, D.N. 2006. LSST: A complementary probe of dark energy. arXiv:0609516.
- Wojtak, R., Hansen, S.H., & Hjorth, J. 2011. Gravitational redshift of galaxies in clusters as predicted by general relativity. *Nature*, **477**, 567–569.
- Zhao, H., Peacock, J.A., & Li, B. 2013. Testing gravity theories, transverse Doppler and gravitational redshifts in galaxy clusters. *PhRvD*, **88**(4). arXiv:1206.5032.

A Code

Below is the Python-code used to run the fittings described in section 4. It is written using Python 2.7

```
1 # * coding: utf 8 *
2 """ author: Idavras """
3
4 import numpy as np
5 import matplotlib.pyplot as pl
6 import emcee
7 import operator
8 import corner
9
10 #Loading data
11 #data = np.loadtxt("Mock2x_new.txt")      #Mock 2x
12 #data = np.loadtxt("Mock2y_new.txt")      #Mock 2y
13 #data = np.loadtxt("Mock2z_new.txt")      #Mock 2z
14
15 #data = np.loadtxt("LSSTx_new.txt")       #LSST 1x
16 #data = np.loadtxt("LSSTy_new.txt")       #LSST 1y
17 #data = np.loadtxt("LSSTz_new.txt")       #LSST 1z
18
19 #data = np.loadtxt("LSST2x_new.txt")      #LSST 2x
20 #data = np.loadtxt("LSST2y_new.txt")      #LSST 2y
21 data = np.loadtxt("LSST2z_new.txt")      #LSST 2z
22
23 def binning(data ,xx ,lower , upper) :
24     """
25     Bins "data" in range from "lower" to "upper" in column "xx".
26     data > array
27     xx > integer
28     lower > integer
29     upper > integer
30     """
31     binned_data = []
32
33     for index in range(len(data[:,xx])):
34         if data[index,xx] < upper and data[index,xx] > lower :
35             binned_data.append(data[index,:])
36
37     binned_data = np.asarray(binned_data)
38
39     return binned_data
40
41 R05=binning(data ,0 ,0.0 ,0.5)
42 R10=binning(data ,0 ,0.5 ,1.0)
43 R15=binning(data ,0 ,1.0 ,1.5)
44 R23=binning(data ,0 ,1.5 ,2.3)
45 R36=binning(data ,0 ,2.3 ,3.6)
46 R49=binning(data ,0 ,3.6 ,4.9)
47 R60=binning(data ,0 ,4.9 ,6.0)
48
```

```

49 #%% MCMC
50 """
51 Parameters to change, when running MCMC:
52 data_used > what dataset/bin is fitted
53 velcutoff > velocity cutoff of data
54 ndim > number of parameters in fitting model
55 nwalkers > number of walkers
56 lnprior > which model is being fitted
57 lnlike > which model is being fitted
58 """
59
60 data_used = R05
61 velcutoff = 6000 #Velocity cut off
62 b = 1/(2.*velcutoff) #Constant from integration of line
63 ndim = 4 #Number of parameters in fitting function
64 nwalkers = 250 #Number of walkers must be even
65 steps = 1000 #Steps taken by the MCMC
66
67 def lnprior(theta):
68     """ Prior """
69     #Single gauss + line
70     mu, sig1, m, pc = theta
71     if sig1 > 0.0 and 0.0 <= pc <= 1.0 and (m*( velcutoff)*(1e 8) + b) > 0
72         and (m*(velcutoff)*(1e 8) + b) > 0:
73         return 0.0
74
75     #Double gauss + line
76     # mu, sig1, m, pc, sig2, pg = theta
77     # if sig1 > 0.0 and sig2 > sig1 and 0.0 <= pg <= 1.0 and 0.0 <= pc <=
78     1.0 and (m*( velcutoff)*(1e 8) + b) > 0 and (m*(velcutoff)*(1e 8) + b)
79     > 0:
80     return 0.0
81
82     return np.inf
83
84 def lnlike(theta, y):
85     """ Likelihood """
86     #Single gaussian + line
87     mu, sig1, m, pc = theta
88     return np.sum(np.log(np.exp( 0.5*(y mu)**2/(sig1**2))*(1/(np.sqrt(2*np.
89         pi)*sig1))*pc+(m*y*(1e 8)+b)*(1 pc)))
90
91     #Double gaussian + line
92     # mu, sig1, m, pc, sig2, pg = theta
93     # return np.sum(np.log((np.exp( 0.5*(y mu)**2/(sig1**2))*(1/(np.sqrt(2*
94         np.pi)*sig1))*pg+(1 pg)*np.exp( 0.5*(y mu)**2/(sig2**2))*(1/(np.sqrt(2*
95         np.pi)*sig2)))*pc+(m*y*(1e 8)+b)*(1 pc)))
96
97 def lnprob(theta, y):
98     """ Posterior probability = likelihood * prior """
99     prior = lnprior(theta)
100
101     if not np.isfinite(prior):

```

```

96         return np.inf
97
98     return prior + lnlike(theta, y)
99
100 #Initial positions between zero and one one for each walker
101 pos0=[np.random.rand(ndim) for i in xrange(nwalkers)]
102
103 #Setting up the sampler
104 sampler=emcee.EnsembleSampler(nwalkers, ndim, lnprob, args=(data_used[:,1]))
105
106 #Running MCMC
107 pos, prob, rstate =sampler.run_mcmc(pos0, steps)
108
109 print("Mean acceptance fraction:", np.mean(sampler.acceptance_fraction))
110 print("Autocorrelation time:", sampler.get_autocorr_time())
111
112 burnin = 300
113 corner.corner(sampler.chain[:, burnin:,:].reshape(1, ndim)) #Corner plot
114 pl.show()
115
116 %% Finding best fit values from walker positions
117 def bestfitvalue(fitvalues):
118     """
119     Takes walker positions from MCMC fitting, sort the values,
120     removes the largest and smallest 15.85%, the remaining 68.3% is 1 sigma
121
122     find walker with maximum likelihood (best fit value),
123     find upper and lower errors on best fit,
124     returns best fit value, upper and lower bounds
125     """
126     sorted_data=np.sort(fitvalues)
127     onesigma_values=sorted_data[round(nwalkers*0.1585):nwalkers round(
128         nwalkers*0.1585)]
129
130     #Finding walker with maximum likelihood
131     max_index, max_value = max(enumerate(prob), key=operator.itemgetter(1))
132     bestfit_value = fitvalues[max_index]
133
134     #Finding upper and lower boundary for best fit value
135     err_lower = bestfit_value - onesigma_values[0]
136     err_upper = onesigma_values[len(onesigma_values) - 1] + bestfit_value
137
138     return [bestfit_value, err_upper, err_lower]
139
140 print "mu = ", bestfitvalue(pos[:,0])[0], " +", bestfitvalue(pos[:,0])
141 [1], " ", bestfitvalue(pos[:,0])[2]
142 print "sig1 =", bestfitvalue(pos[:,1])[0], " +", bestfitvalue(pos[:,1])
143 [1], " ", bestfitvalue(pos[:,1])[2]
144 print "m = ", bestfitvalue(pos[:,2])[0], " +", bestfitvalue(pos[:,2])
145 [1], " ", bestfitvalue(pos[:,2])[2]
146 print "pc = ", bestfitvalue(pos[:,3])[0], " +", bestfitvalue(pos[:,3])
147 [1], " ", bestfitvalue(pos[:,3])[2]
148 #print "sig2 =", bestfitvalue(pos[:,4])[0], " +", bestfitvalue(pos[:,4])

```

```

    [1], "    ", bestfitvalue(pos[:,4])[2]
143 #print "pg = ", bestfitvalue(pos[:,5])[0], "  +", bestfitvalue(pos[:,5])
    [1], "    ", bestfitvalue(pos[:,5])[2]
144
145 # Parameters needed for plotting
146 mu_fit = bestfitvalue(pos[:,0])
147 sig1_fit = bestfitvalue(pos[:,1])
148 m_fit = bestfitvalue(pos[:,2])
149 pc_fit = bestfitvalue(pos[:,3])
150 #sig2_fit = bestfitvalue(pos[:,4])
151 #pg_fit = bestfitvalue(pos[:,5])
152
153 func_single = lambda xxx: (1/(np.sqrt(2*np.pi)*sig1_fit[0]))*np.exp( 0.5*(
    xxx mu_fit[0])**2/(sig1_fit[0]**2))*pc_fit[0] + (m_fit[0]*1e 8*xxx+b)
    *(1 pc_fit[0])
154 #func_double = lambda xxx: (((1/(np.sqrt(2*np.pi)*sig1_fit[0]))*np.exp
    ( 0.5*(xxx mu_fit[0])**2/(sig1_fit[0]**2))*pg_fit[0]+(1 pg_fit[0])*(1/(
    np.sqrt(2*np.pi)*sig2_fit[0]))*np.exp( 0.5*(xxx mu_fit[0])**2/(sig2_fit
    [0]**2)))*pc_fit[0] + (m_fit[0]*1e 8*xxx+b)*(1 pc_fit[0]))
155
156 #%% Plotting
157 vbinsize = 50
158 vbins = np.arange( velcutoff , velcutoff+1,vbinsize )
159 hist_data=pl.hist(data_used[:,1],bins=vbins)
160 xvalue=np.linspace( velcutoff , velcutoff ,num=100)
161 pl.plot(xvalue ,func_single(xvalue)*6.7e6 , 'r ')
162 #pl.plot(xvalue ,func_double(xvalue)*1.05e7 , 'r ')
163 pl.show()

```

B Additional analysis

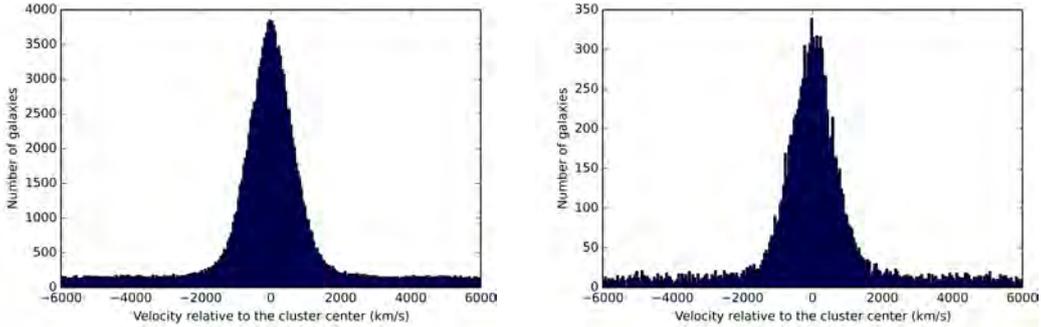
In this appendix, additional information on the data and analysis of the data presented in sections 3 and 4 is described.

B.1 Distribution of data

The datasets described in section 3.2 get a wider velocity distribution as errors on the velocity distribution becomes larger. This is simulated by adding the Gaussian distribution on the datasets LSST mock 1 and 2, and in figure 13 the wider distribution is clearly seen.

The distributions appear smooth, which is because the datasets are large. Other bins with fewer galaxies have small deviations from this, but overall a smooth distribution. When only using some of a specific dataset, as in section 4.4, the smaller sample size results in a less smooth distribution, which affects the accuracy of the fit, see figure 12. The smaller the total sample size becomes the more irregular the distributions in the individual bins becomes. This can be prevented by making fewer bins, see discussion of bin size in appendix B.3.

In figure 13 fits of a single Gaussian fit is shown and notice the difficulty of fitting the wings of the distribution for the Mock dataset in figure 13(a). This could be solved by fitting a double Gaussian, but as argued in section 4, the fitting values of the centre of the Gaussian for the two fits, and thereby the gravitational redshift, are similar with the same size of errors, and as described in section 3.4 more parameters can result in overfitting.



(a) One bin of the full Mock dataset with $3 \cdot 10^6$ galaxies

(b) One bin of a sample of $2.5 \cdot 10^5$ randomly drawn galaxies from the Mock dataset

Figure 12: Histogram of the same bin of the Mock dataset, with two different total sample sizes

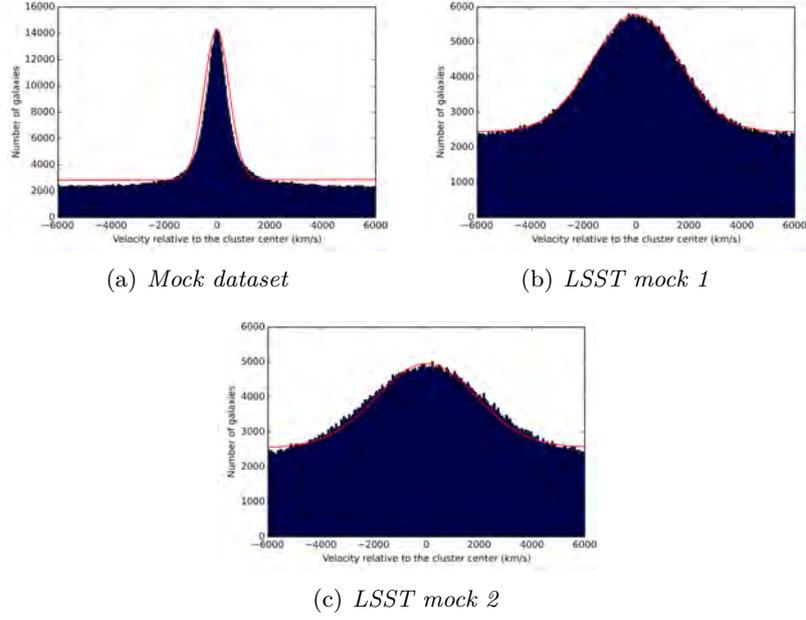


Figure 13: *Histograms of data in the outermost bin of one projection of each dataset. The red line in each plot represents the best fit of a single Gaussian function as described in equation 10*

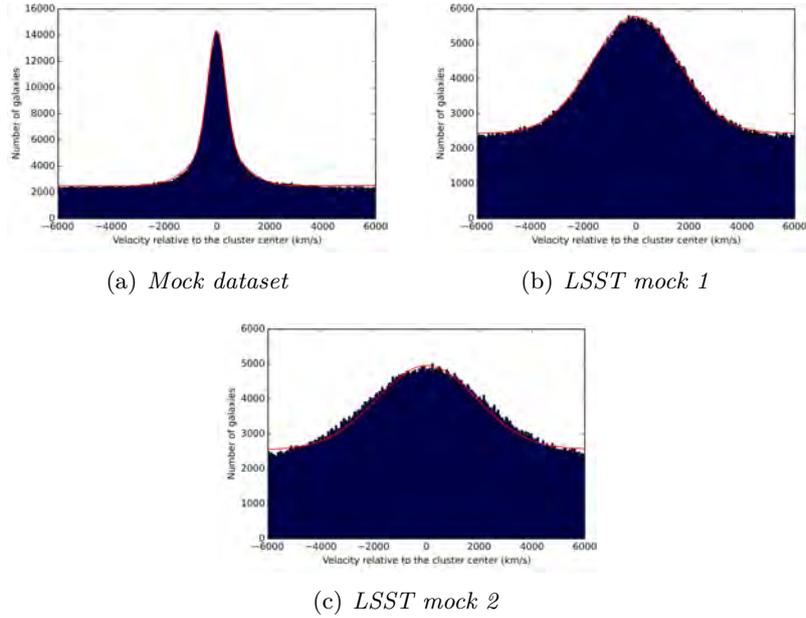


Figure 14: *Histograms of data in the outermost bin of one projection of each dataset. The red line in each plot indicate the best fit of a double Gaussian function as described in equation 11*

Visually, a double Gaussian fit is better for the Mock dataset, but indistinguishable for the LSST mock 1 and 2 datasets, As is shown in figures 13 and 14.

When the velocity distribution gets wider, as for the LSST mock 1 and LSST mock 2 datasets, the single Gaussian fit becomes better, see figure 13(b) and (c), compared with the single Gaussian fit for the Mock dataset. The reason the fit for the LSST mock 2 dataset is not as good as the LSST mock 1 dataset, is probably caused by the velocity cut-off not being wide enough for the distribution to flatten at the wings, and therefore the fit cannot determine the normalization of the Gaussian. If the velocity cut-off was set higher this single Gaussian fit would probably be similar in shape to that for the LSST mock 1 data.

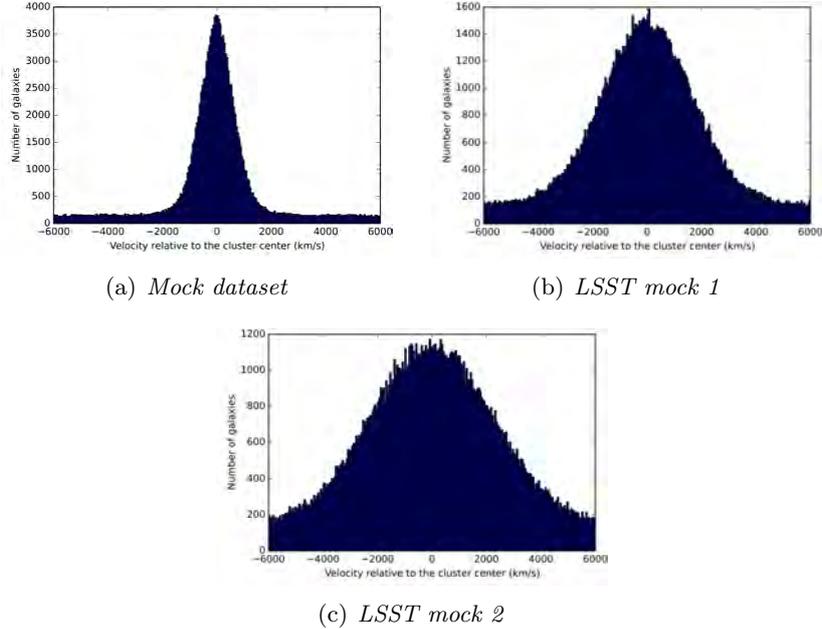


Figure 15: *Histogram of data in the inner most bin for one projection for each dataset*

B.2 Velocity cut-off for data selection

When selecting a cut-off for the data selection, it is important to make sure it covers the entire distribution, and that the cut includes the distribution flattening out at large velocities. This is necessary in order to get a good fit because of normalization of the Gaussian component.

The velocity cut-off used for the data analysed in section 4 is ± 6000 km/s for all datasets. But the different types of data could have had different cut-off velocities, see figure 15. For the Mock dataset the cut-off could easily be smaller, for example ± 4000 km/s, without affecting the results. On the other hand, the ± 6000 km/s cut-off used is suitable for the LSST mock 1 dataset, while it is too narrow for the LSST mock 2 dataset. The wings of the distribution of the LSST mock 2 dataset are cut off and a flattening of the distribution is not included, which results in less reliable fits. When the wings are cut off a different normalization of the Gaussian function would probably have resulted in a better fit. A velocity cut-off of $\pm 7-8000$ km/s would have been better for this dataset, if it were to be fitted as described in section 3.3.

When making the cut-off larger, more galaxies are included in the final dataset, compared to a more narrow cut. By having the same velocity cut-off on all the datasets used, the samples will have similar sizes and be comparable.

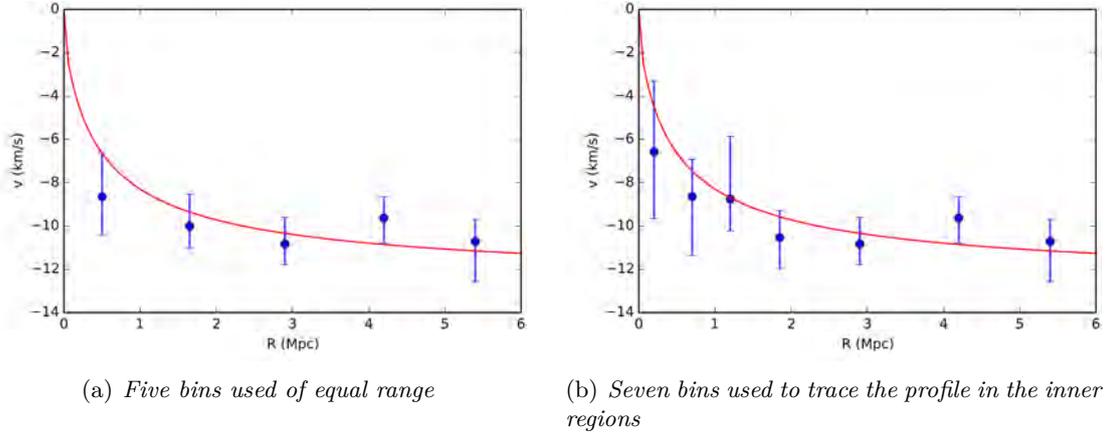


Figure 16: Plot of fit values with different number of bins for the full Mock dataset. The line represent the mean gravitational redshift signal for the full dataset with a mean mass of $1.84 \cdot 10^{14} M_{\odot} h^{-1}$. The larger number of bins in the inner regions is able to trace the gravitational redshift profile better

B.3 Bin size

All the plots shown in section 4 are plotted using seven bins to trace the gravitational redshift profile. When considering the number of bins there should be enough to trace the desired profile and its features, but dividing the data into too many bins could result in too little data in each resulting in large uncertainties for a given bin. The number of bins also depends on the amount of data available for binning, meaning less data can not be divided into as many bins as a larger dataset, as this would not yield usable results.

When binning, the data in each should also be considered. In this case, each bin should contain a representative amount of galaxies and clusters. Too few clusters in a given bin could result in a bias or substructure of the cluster could affect the value of the gravitational redshift in that bin.

At first, I used five bins of equal size to cover the range of radii from 0-6 Mpc, see figure 16(a). With this division the inner most bin covered the range from 0 Mpc to 1.1 Mpc, which is most of the virialized region of a cluster with an average mass of $1.84 \cdot 10^{14} M_{\odot} h^{-1}$, corresponding to the mean mass of the datasets analysed in section 4. As shown in figure 1 in section 2, the variation in features of the gravitational redshift profile is greatest in the inner regions within the virial radius, which has the value $r_v \approx 1.7$ Mpc for an average cluster in the datasets analysed here.

By only using five bins the variations in the signal in the inner regions of the clusters

are obscured, which is evident in comparison to the seven bins used on the same data in figure 16(b). The use of few bins in the inner regions indicate a constant signal of about 10 km/s throughout the cluster, which is clearly not the case when binning is performed slightly different and the weaker signal in the inner regions becomes distinguishable.

It is relatively safe to use large bins at larger cluster centric distances, because of the smaller variation in the size of the gravitational redshift signal for galaxies at these distances. The difference in the predicted gravitational redshift from 3.0 Mpc to 5.0 Mpc is $\Delta z_{grs} = 0.6$ km/s, which is roughly the same as the difference between the predicted signal at 0.50 Mpc and 0.65 Mpc, $\Delta z_{grs} = 0.64$ km/s. The differences in the signal within a bin at larger cluster centric distance are therefore a lot smaller than the error on the fitted value.

An argument could be made for making even smaller bins when the variations in the signal is great, as in the inner regions of the clusters. But as already mentioned above, a certain amount of data is required in each bin to obtain reliable results, and the inner regions of even the combined datasets do not contain that many galaxies, considering the sample sizes used in this analysis. Therefore, to use smaller binning a larger sample of galaxies is needed.

It is also clear from figure 16 how the number of bins affect the error in each bin. Comparing the data in the inner region being divided into two, figure 16(a), or four, figure 16(b), bins clearly affects the size of errors, as more bins equals less data in each.

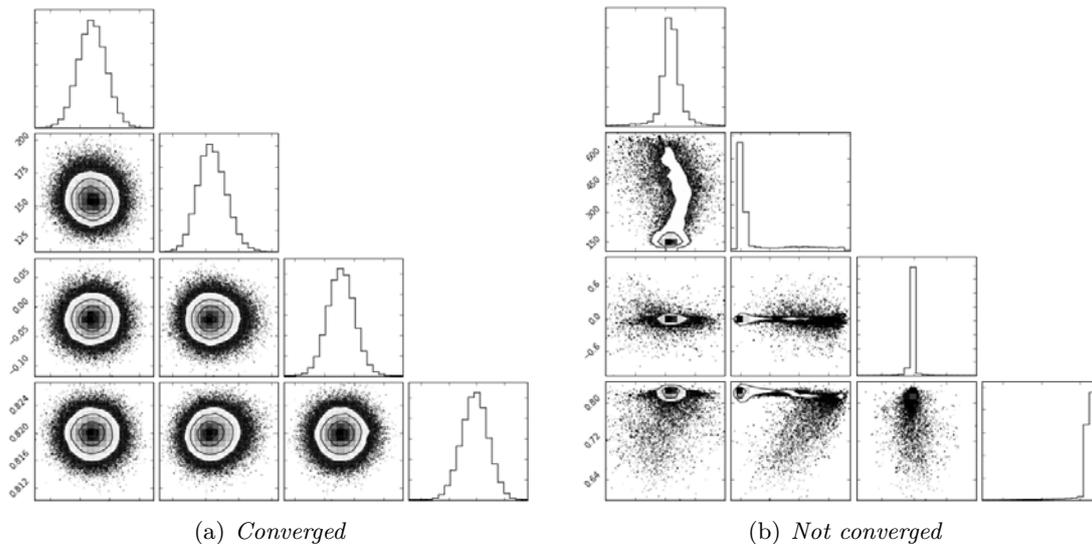


Figure 17: *Corner plots of parameters of the same fitting function, but where one is converged whereas the other is not. The difference between the two are the number of steps the MCMC algorithm uses, where (a) is with 1500 steps and (b) is with 1000 steps*

B.4 Convergence of fits

When running an MCMC routine, it can be difficult to determine the time of convergence as it, in principle, can run indefinitely. The number of steps need to be large enough to be sure of convergence, but not too large as to lengthen the computation time needlessly.

To check the convergence of the fits, I looked at the mean acceptance fraction, the autocorrelation time and corner plots of the parameters. The corner plots are the easiest way to make a fast determination of convergence, if the parameters are expected to have a Gaussian distribution, see figure 17. The difference between the two panels are the number of steps the MCMC-algorithm takes which is 50% larger for the converged run compared to the non-converged. It is clear that the parameters has a Gaussian distribution in figure 17(a), whereas the distributions are very skewed and uneven in figure 17(b).

Because the distribution of the parameters is Gaussianly distributed, an argument could be made for choosing the value to plot to be the median. But as mentioned in section 3.3, I use the value with the maximum likelihood in the plots, which can results in asymmetric errorbars that indicate a skewed distribution, even though that is not the case.

B.5 Computation time

There are multiple factors affecting the computation time of an MCMC fitting routine. Such factors are the number of walkers used, the number of steps the walkers have to take, the number of parameters of the function being fitted and of course the size of the dataset. Three of these are parameters set by the fitting routine.

The number of walkers should be large enough to probe the parameter space and give a reasonable distribution of the parameter values. Too few walkers can result in the walkers "getting stuck" in an area of low logarithmic likelihood which is not the global minimum. I used 250 walkers for all the fits.

The number of steps needed require some testing and can vary from fit to fit. As already described in appendix B.4 the convergence of a fit needs to be ensured, and a non-converged fit can converge by running longer. For the single Gaussian fits 1000 steps was enough to converge all the fits. For double Gaussian fits 1500-2000 steps were sometimes needed for a fit to converge.

When parameters are added to the fitting function the dimensionality of the parameter space goes up. The MCMC algorithm is good at handling this, but it still affects the total computation time, where more parameters results in longer computation. For the single Gaussian function there are four free parameters as described in section 3.3, while the double Gaussian function has six. If the mean (μ) of the double Gaussian function become independent for each Gaussian component as discussed in appendix B.6, then the fitting function has seven free parameters. As argued in section 4 the typical fit is carried out using a single Gaussian fitting function.

The last factor affecting the computation time is the size of the dataset. This is not a factor which can be manipulated once the data has been obtained, but results in different computation time for different bins depending on the amount of data in each.

For a typical fit using 250 walkers, taking 1000 steps, fitting a single Gaussian function to a dataset of about $3 \cdot 10^5$ galaxies the computation time was usually between four and six hours. For the same setup but fitting a double Gaussian function the computation time would double to between eight and twelve hours.

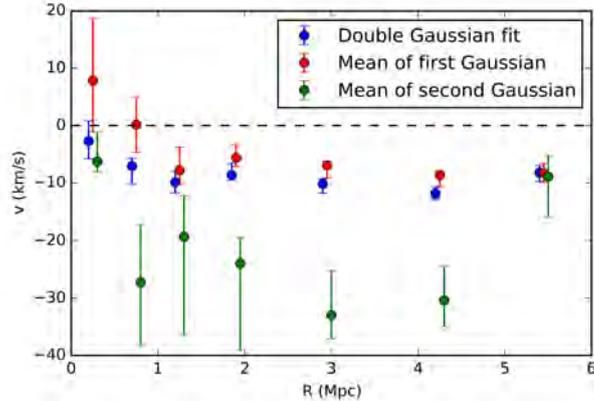


Figure 18: *Fit values from two separate fits; one with mean of the double Gaussian fit (blue), which are the same as plotted in figure 2 and one with a fit where the means of the Gaussians are independent, μ_1 (red) and μ_2 (green). μ_1 represents the narrow component of the double Gaussian function, while μ_2 represents the wide component. The data used are a projection of the Mock dataset*

B.6 Different mean of double Gaussian

Fits with the double Gaussian function, equation 11, was not carried out consistently, as described in section 4, but was initially considered in section 3.3. As described, the double Gaussian function assumes the mean value, μ , to be the same for both Gaussian components, and I ran some test to see how well this assumption holds.

Making the means of the two Gaussians independent gives another parameter to fit, which results in a longer computation time, see appendix B.5, making it less desirable and add to the risk of overfitting, see section 3.4. It also results in each mean, μ_1 and μ_2 , having larger errors compared to a common mean, see figure 18.

Notice the mean of the first Gaussian, μ_1 , for all bins, except the innermost, have small errors and are in agreement with the value from the same mean fit, whereas the mean from the second Gaussian, μ_2 , generally has a more negative values and much larger errors. This is consistent with μ_1 being the mean of the narrower Gaussian component of the two (a result from the priors described in section 3.3), meaning it fits the central parts of the distribution, while μ_2 is the mean of the wider Gaussian, fitting the wings of the distribution. Fitting the wings of the distribution with a wider Gaussian function, results in the larger errors shown. This was also the case for the different datasets analysed in section 4, where the wider distribution of the photometric data, resulted in larger errors. When looking at the relative weights, p_g , of the Gaussian components, μ_1

is weighted the heaviest (60-70%) in all bins, again except the innermost bin.

This indicates that the double Gaussian fit, with the same mean, μ , for both components, is in good agreement with the independent means, μ_1 and μ_2 . Also, the comparison with the single Gaussian fit described in section 4 are consistent with this, as the single Gaussian function fit the central part of the distribution which is weighted the heaviest, resulting in similar fit values for the single and double Gaussian functions.