Spectroscopic Richness of Cardinal Galaxy Clusters

By

Aaron Hughes

_____

A Thesis Submitted to The W.A. Franke Honors College

In Partial Fulfillment of the bachelor's degree
With Honors in

Astronomy and Physics

THE UNIVERSITY OF ARIZONA

M A Y 2 0 2 4

Approved by:

Dr. Eduardo Rozo
Department of Physics

Draft version May 2, 2024
Typeset using LaTeX **twocolumn** style in AASTeX631

Spectroscopic Richness of Cardinal Galaxy Clusters

Aaron Hughes[1]

[1] *Department of Astronomy and Steward Observatory, University of Arizona, 933 N Cherry Ave, Tucson, AZ 85719, USA*

ABSTRACT

This paper describes a method to compute the spectroscopic richness of galaxy clusters. This method stems from the redMaPPer photometric cluster finder and the idea that spectroscopic data is able to better discriminate between member and non-member galaxies. Similar to redMaPPer, the method outlined here defines a membership probability for each galaxy around a central galaxy of a cluster. However instead of utilizing the position, color, and magnitude information of the galaxy, the spectroscopic method utilizes the velocity information of galaxies around a cluster to determine if galaxies with a given position and velocity are likely to be cluster members or not. Running on halos from the Cardinal mock galaxy catalog, and determining the richness of the cluster in each halo yields a mass-richness relationship. Comparing the mass-richness relation found from this spectroscopic method and the relation from redMaPPer yields a clear decrease in the scatter of 15 percent at low mass and 25 percent at high mass. This reveals the promise of incorporating a spectroscopic component into cluster finder algorithms.

1. INTRODUCTION

Recently, galaxy clusters have been identified as cosmological probes that are able to constrain many of the cosmological parameters. Galaxy clusters are hosted by massive dark matter halos, so the study of clusters can help us understand both the growth of structure in the Universe, and how galaxies populate haloes in an environment dependent way.Because of their cosmological importance, good cluster finding algorithms are important. Given a large-sky survey of galaxies, determining cluster based on only the line-of-sight information is a non-trivial task. Currently, a popular cluster-finding algorithm, redMaPPer (Rykoff et al. 2014), utilizes the photometric data from galaxies, meaning their colors and magnitudes, to determine which galaxies are part of a cluster or not. Recent developments allowing for spectroscopic data presents a new set of galaxy information that is able to better constrain galaxies that are members of a cluster and those that are not. This paper sets out to define a spectroscopic cluster finder and compare the results to the well known results from redMaPPer. Specifically, this work will outline a new method to compute richness, the number of galaxies that are defined as members of the cluster, using the velocity distribution of galaxies around the central galaxy of the cluster along the line of sight.

This outline of this paper is as follows: Section (2) describes the data used in this paper. It describes the Cardinal mock galaxy survey and describes how the redMaPPer cluster finder was run on this data. Section (3) explains the model used to describe the velocity distribution of galaxies in halos. Section (4) describes how this model of velocity distributions can then be converted into a method to compute the richness of clusters inside these halos. Section (5) provides the results from this algorithm and compares the results to what was found by redMaPPer. Section (6) finally gives some potential improvements to the algorithm and what the future of the cluster finder could look like.

2. SIMULATION DATA

2.1. *Cardinal*

Synthetic galaxy catalogs, also known as mock catalogs, provide a valuable tool for assessing systematics on real data and developing techniques to analyze galaxies. These mocks include plausible universes are able to serve as a sandbox for researchers to test and develop methods to analyze galaxy survey data. For the testing purposes presented in this work, the mock catalog used is Cardinal, which is outlined in To et al. (2023). This synthetic survey covers a quarter of the the sky and includes galaxies with redshifts up to 2.35. To develop a realistic universe, a dark matter simulation is converted into a survey of galaxies using Subhalo abundance matching. In doing so, the survey produces mock galaxies that are fit for reproducing the properties of real galaxies and galaxy clusters observed from data from the Dark Energy Survey. Each galaxy in Cardinal includes mass information from its host halo, photometric

properties, co-moving distances, and redshift information. This assortment of data included in the synthetic survey makes it an ideal candidate for many aspects of cluster cosmology. Currently, the Cardinal mock catalogs are the only mocks that yield realistic redMaPPer clusters as a function of richness and redshift, making it the perfect choice for the computations done in this work.

2.2. *RedMaPPer on Cardinal*

To develop clusters from the given synthetic galaxies, the redMaPPer algorithm is utilized. This algorithm, outlined in Rykoff et al. (2014) incorporates the photometric data contained in a galaxy catalog. First, a red-sequence model is calibrated, which is essentially a relationship between the color and magnitude of galaxies in a cluster at a given redshift. After the calibration, every photometric galaxy in the catalog is considered as a central galaxy of a cluster. If the galaxy does not seems to fit the calibrated template for a cluster at any redshift, then it is not considered a central candidate of a cluster. As long as one galaxy around the central fits the template, this is initially considered a cluster. The process is run iteratively to determine which galaxy is most likely to be the cluster central for each cluster. It then assigns all galaxies around that central within a certain distance a membership probability based on how likely it is to belong to the cluster. To find the total richness of the cluster, all the membership probabilities are summed over. RedMaPPer finally combines all the clusters into a cluster catalog and all the membership galaxies into a membership catalog - each containing properties of the central and member galaxies respectively.

For this work, redMaPPer was run slightly differently, although it was still run utilizing photometric data. Because the desired goal was to determine the mass-richness relation of the halos and the galaxy clusters, each halo in the Cardinal galaxy catalog must correlate to exactly one central galaxy of the cluster. This information is contained within the Cardinal catalog itself. For each halo, Cardinal provides a central galaxy. A separate catalog of only these central galaxies was passed into redMaPPer along with the entire catalog of all galaxies. In doing so, redMaPPer did not undergo the process to find the central galaxies since it already had this information. Instead, redMaPPer simply computed membership probabilities for galaxies around each central and summed them to find the richness of each cluster. Since each cluster corresponds to one halo, these richnesses can be related back to their initial halo mass to reveal the mass-richness relation shown in figure (1). In this figure, the orange line represents a linear best fit. This best fit line was determined using a maximum log-likelihood model. This model assumed a Gaussian scatter around a linear regression. Therefore, the likelihood of each data point is defined as the log of a Gaussian

$$lnP = -\frac{1}{2}\frac{(log(\lambda) - \mu)^2}{\sigma^2} - \frac{1}{2}ln\sigma^2 \tag{1}$$

where in this expression, lnP is the log-likelihood, λ is the richness data point, μ is the peak of the gaussian, which is defined as a linear fit described by

$$\mu = \alpha log(M) + \beta \tag{2}$$

and σ is the scatter of the data around the mean. In the linear fit, α and β are free parameters. Because the data reveals a higher scatter at low mass than at high mass, the scatter is described a a function of mass,

$$\sigma^2 = \sigma_0^2 + \frac{\sigma_1^2}{e^{\mu}(M) - 1} \tag{3}$$

Where in this expression, σ_0 and σ_1 are free parameters. The scatter exponentially decays allowing for lower scatter at higher masses. For each data point, the variance from the linear fit and the corresponding likelihood of that point is determined. Then all likelihoods are summed over. This process repeats until the four free parameters, the two from the linear fit and the two from the scatter, that maximize this likelihood are found. Figure (1) reveals the recovered linear fit in orange and the recovered expression for the scatter in blue. This scatter around the best fit mean will be compared to the scatter from the new spectroscopic method outlined in this paper.

3. VELOCITY DISTRIBUTION OF GALAXIES IN HALOS

Knowing the spectroscopic data allows for direct access to the redshift and therefore line-of-sight velocity - of every galaxy. In addition, the positions of galaxies on the sky lead to the projected radial distances from the central galaxy of a dark matter halo. The model of the velocity distribution of galaxies around a halo will use these two pieces of information from each galaxy: the relative velocity relative to the halo central and projected distance between the galaxy and the halo central. These are two known quantities from a spectroscopic survey and are included inside the Cardinal galaxy catalog. To observe the distribution of relative line-of-sight velocities of galaxies, a histogram of all galaxys' velocities across all halos can be made. For this

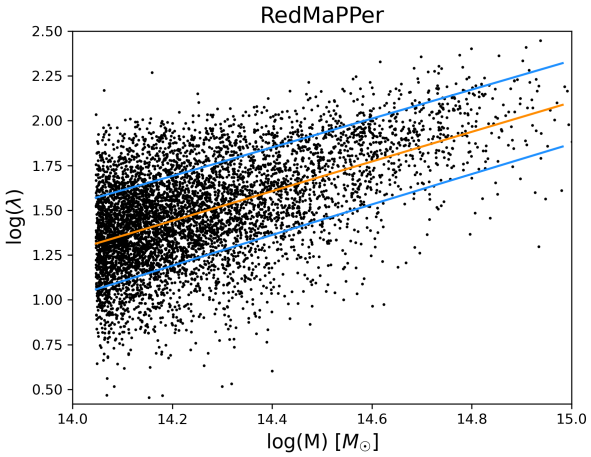


Figure 1: Figure (1) shows the mass-richness relation of Cardinal halos using the redMaPPer algorithm. The orange line is a linear best fit and the blue line is the scatter as a function of mass.

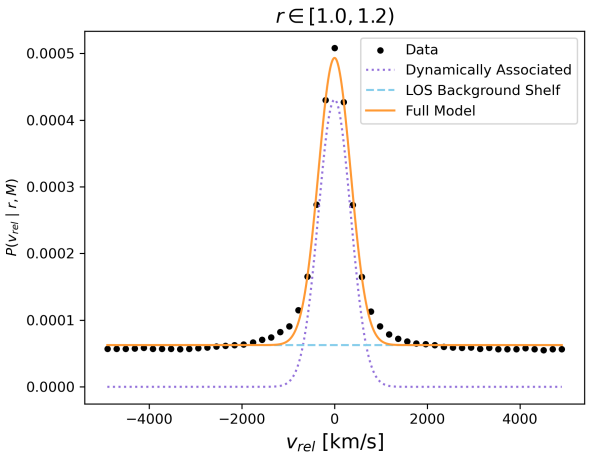


Figure 2: Figure (2) shows a histogram of relative velocities of galaxies around a cluster. This histogram is only for the lowest mass bin and between [1.0, 1.2) $[h^{-1}Mpc]$. The best fit to the data is shown in orange. It is a combination of a peak component of galaxies associated with the cluster (purple dotted) and a shelf component of galaxies not associated with the cluster but just background galaxies along the line-of-sight (blue dashed line).

model, only relative speeds up to 5000 km/s and projected distances less than 5 $[h^{-1}Mpc]$ from the central were included. The histograms were binned by their projected distances and by halo mass to observe how the distribution changes for different locations and halo sizes. The distances were binned in sizes of .2 $[h^{-1}Mpc]$ between 0 and 5 for a total of 25 bins, and the five mass bins are revealed in Figure (3). One of these histograms is shown in Figure (2).

This normalized histogram includes galaxies from the lowest mass bin and with a radial distance of 1.0 - 1.2 $[h^{-1}Mpc]$. The data reveals a peak component and a shelf component. Therefore, to fit the histogram and model the probability of each relative velocity for a given mass and radius, there must be a peak probability and a shelf probability each multiplied by their respective contributing fraction. This method parallels the work done by Tomooka et al. (2020).

$$P(v|r, M) = f_{peak}P_{peak} + f_{shelf}P_{shelf} \tag{4}$$

In this expression for the probability, f_{peak} is a free parameter for the fit. In Figure (2), the peak component of the data appears to be a Gaussian centered around zero. This peak at zero is due to the galaxies that are falling toward the central perpendicular to the line of sight, which results in a peculiar line-of-sight velocity of zero. Because there are more galaxies on the side of the central than in front or behind it, there are more galaxies with very low velocities than higher relative velocities. Therefore, a velocity distribution with a peak at zero is a good indication of a galaxy cluster. Because the peak component appears Gaussian, we set a Gaussian to be the peak probability.

$$P_{peak} = G(v|r, M) = \frac{1}{\sqrt{2\pi\sigma_{peak}^2}} e^{-\frac{1}{2}\frac{v^2}{\sigma_{peak}^2}} \tag{5}$$

For this Gaussian probability, σ_{peak} is a free parameter for the fit. In Figure (2), this peak component that corresponds to galaxies that are dynamically associated with the halo is shown in the purple dotted line.

Regarding f_{shelf}, it relates to f_{peak} since the total fraction must add to one.

$$f_{shelf} = 1 - f_{peak} \tag{6}$$

The shelf component in the histogram shows that even when observing around a galaxy cluster, there are still galaxies with very high relative velocities. This is due to galaxies that are along the line-of-sight of the halo but are not associated with the halo but instead with the Hubble Flow. Therefore, their redshifts differ greatly compared to the halo central. In Figure (2), the blue dashed line corresponds to this shelf of these background galaxies along the line of sight. This shelf can be approximated as a uniform probability distribution of background galaxy velocities. To determine the constant

value of the distribution, we can consider the normalization condition

$$\int_{-v_{max}}^{v_{max}} P_{shelf}(v)dv = 1 \tag{7}$$

Where the bounds of the integral are the upper and lower limits to the relative line-of-sight velocities chosen for the dataset. In this case, 5000 km/s was chosen as the limiting speed. Therefore, since $P_{shelf}(v)$ is a constant value, this integral simplifies to

$$P_{shelf}(v)\int_{-v_{max}}^{v_{max}} dv = P_{shelf}(v)[2v_{max}] = 1 \tag{8}$$

Solving for P_{shelf}, recovers

$$P_{shelf} = \frac{1}{2v_{max}} \tag{9}$$

Adding the peak component and the shelf component gives the total probability of each relative velocity for a given mass and radius. Overall, this completed model includes two free parameters: f_{peak} and σ_{peak}. The orange line in Figure (2) displays this combined model. All 125 histograms (5 mass bins each with 25 radial bins) can then be fit and these two parameters extracted from each. These values are all shown in Figure (3). The goal is then to determine the dependence of f_{peak} and σ_{peak} on the galaxy's distance from the central and the halo mass. Physically, we expect f_{peak} to decrease with larger r values as there are more galaxies associated with the cluster closer to the central. f_{peak} should also increase with higher halo mass as a larger halo will have more galaxies associated with it. The extracted values for f_{peak} clearly have two distinct components. From Garcia et al. (2022), there appears to be an orbiting and an infalling component. The orbiting term was found to follow an exponential decay in r. From the data recovered from the histogram fits, the infalling component appeared to be linear. Therefore, the total expression for f_{peak} is the sum of the orbiting component and the infalling component.

$$f_{peak}(r|M) = Ae^{-\frac{1}{2}(\frac{r}{r_h})^2} + B(r - r_p) + C \tag{10}$$

Where in this equation r is the galaxy's distance from the central. In this model, there are four free parameters: A, B, C, and r_h. The r_p parameter is a pivot parameter that was set to 3 $h^{-1}Mpc$. Without this parameter, the value of C is the f_{peak} value of the linear infalling component at r = 0 . However, at low r values, most galaxies are orbiting, so the exponential orbiting term dominates, so there is no linear infalling component. Therefore, by adding the r_p parameter and setting it to a high r value of 3 $h^{-1}Mpc$, C becomes the value of f_{peak} at $r = 3\ h^{-1}Mpc$. At large r values, the orbiting term is negligible so the model is only the linear infalling contribution. This prevents requiring the model to extrapolate this value of C back to the y-axis and allows for an easier initial guess for the parameter. Both of these contribute to a quicker convergence for the fitting parameter C. The four free parameters were assumed to be a power law in mass and the 5 curves for the 5 mass bins were fit simultaneously. These fits for f_{peak} are shown in Figure (3).

The σ_{peak} model is a little more complicated, but still includes an orbiting and an infalling component. Therefore, the model for the variance of the peak is the sum of the two.

$$\sigma^2_{peak} = f_{orb}\sigma^2_{orb} + f_{inf}\sigma^2_{inf} \tag{11}$$

Where in this expression, the f_{orb} is the same orbiting profile as used for f_{peak}.

$$f_{orb}(r|M) = e^{-\frac{1}{2}(\frac{r}{r_\sigma})^2} \tag{12}$$

In this expression, r_σ is a free parameter. From the data shown in Figure (3), the orbiting component appears to display a $1/r$ relationship. Therefore, the expression for σ_{orb} can be described in the same way.

$$\sigma^2_{orb}(r|M) = \frac{\sigma^2_{p,orb}}{1 + r/r_{orb}} \tag{13}$$

Where in this relationship, $\sigma_{p,orb}$ is a free parameter giving the value of σ_{peak} at $r = 0$ since the infalling component is negligible at small r. r_{orb} is also a free parameter.

Similarly to equation (6), the infalling fraction relates to the orbiting fraction since the total fraction is 1.

$$f_{inf} = 1 - f_{orb} \tag{14}$$

From the data in Figure (3), the infalling component of σ_{peak} appears to be linear, so it can be described by the following expression.

$$\sigma^2_{inf}(r|M) = \sigma^2_0 + \sigma^2_1(r^2 - r^2_{p,\sigma}) \tag{15}$$

Where in this expression σ_0 and σ_1 are free parameters. In this case, the pivot point $r_{p,\sigma}$ is set to 4 $h^{-1}Mpc$

for the same reason described previously for C to allow for a better initial guess and a quicker convergence for σ_0. With this added parameter, the value of σ_0 is now the value of σ_{peak} at r = 4 $h^{-1}Mpc$. At this large r value, $f_{orb} \approx 0$ so the only contribution is the linear infalling term. Just as with C, adding this term prevents the model from extrapolating the linear model to fit the σ_0 parameter at the y-axis and instead defines it where the linear infalling component is well-defined by the data at $r = 4\ h^{-1}Mpc$.

Similarly to f_{peak}, the free parameters were assumed to be a power law in mass and the 5 curves for the 5 mass bins were fit simultaneously. These fits for σ_{peak} are also shown in Figure (3).

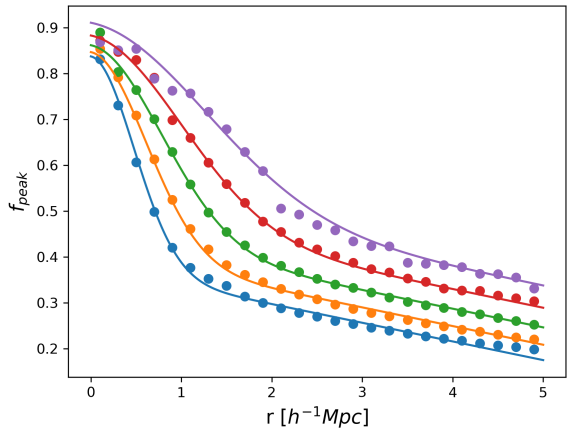
4. SPECTROSCOPIC RICHNESS

With a completed model for the velocity distribution of galaxies around a central allows for then computing the richness of each cluster - the number of galaxies that are members of the cluster. The goal here is to determine the richness using the spectroscopic information and then compare the results to the redMaPPer richness to see if the spectroscopic information made any improvements like we expect. To compute the spectroscopic richness, we first start by assigning each galaxy a membership probability, just as redMaPPer does. This membership probability expression is identical to the one used in redMaPPer (Rykoff et al. 2014), defined as

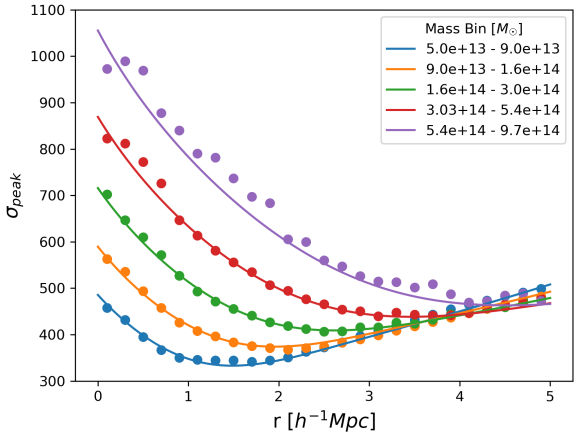
$$p = \frac{\lambda u}{\lambda u + b} \tag{16}$$

Where in this expression, λ is the richness of the galaxy cluster and u is a filter defining the probability density of associated galaxies at the given r and v. Therefore, the quantity λu is the normalized number density of galaxies associated with the cluster for a given r and v. The b value is the normalized value of background density of galaxies at the given r (galaxies not associated with the cluster).

In the case of the spectroscopic richness, the u filter includes the same radial filter as redMaPPer (NFW Profile from Navarro et al. (1995)) and a new probability of that galaxy belonging to the peak of the velocity distribution. This radial filter is the two dimensional cluster galaxy density profile ($\Sigma(r)$). After projecting all galaxies along the line of sight into a 2D surface, this profile defines the number of galaxies per area where the area is of a thin ring at radius r. Therefore, the radial probability density distribution is defined as $2\pi r\Sigma_{peak}(r)$. Including the probability density contribution from the peak of associated galaxys' velocities outlined in equation (2) the total quantity λu is defined as



(a) Figure (a) shows the model for the f_{peak} parameter. The fit is done for five mass bins simultaneously.



(b) Figure (b) shows the model for the σ_{peak} parameter. The fit is done for five mass bins simultaneously.

Figure 3: Figure (3) shows the model fits for the two parameters of the Gaussian fit to the histograms. The fits were done for five mass bins.

$$\lambda u(v, r|M) = \lambda(2\pi r\Sigma_{peak}(r))P_{peak}(v|r, M) \tag{17}$$

The background term, b, in equation (11) deals with the probability that the galaxy belongs to the shelf part of the model outlined in this paper. Galaxies with a large b will not have high membership probabilities. The expression for b is similar except the probabilities are for the shelf.

$$b(r, v|M) = 2\pi r \Sigma_b P_{shelf} \tag{18}$$

Where the P_{shelf} term is described by equation (6). This expression can be modified slightly by relating the Σ_b term to f_{peak} and Σ_{peak}. The ratio of peak NFW over the total NFW gives the ratio of number density of galaxies associated with the cluster as a function of projected radius, r, over the combined number density of galaxies associated with the cluster and associated with the background. Therefore this ratio should be the same as the fraction of galaxies associated with the peak,f_{peak}, and can be described as

$$f_{peak} = \frac{\Sigma_{peak}}{\Sigma_{peak} + \Sigma_b} \tag{19}$$

This expression can be rearranged to recover a substitution for Σ_b.

$$\Sigma_b = \Sigma_{peak} \frac{1 - f_{peak}}{f_{peak}} \tag{20}$$

Combining equations 9, 18, and 20, a final expression for b is given.

$$b(r, v) = 2\pi r \Sigma_{peak}(r) \frac{1 - f_{peak}(r|M)}{f_{peak}(r|M)} \frac{1}{2v_{max}} \tag{21}$$

Using this expression for b, a final expression for the membership probability can be found by combining equations 16, 17, and 21. For the sake of space, $\Sigma_{peak}(r)$ is simplified to Σ.

$$p_{spec} = \frac{\lambda(2\pi r\Sigma)G(v|r, M)}{\lambda(2\pi r\Sigma)G(v|r, M) + 2\pi r\Sigma \frac{1-f_{peak}}{f_{peak}} \frac{1}{2v_{max}}} \tag{22}$$

Where in this expression, λ is the richness, r is the galaxy's distance from the central, Σ_{peak} is the NFW profile for galaxies in the peak, $G(v|r, M)$ is the Gaussian probability that the galaxy is part of the peak of the velocity distribution, and $f_{peak}(r|M)$ is the peak fraction for galaxies at a given distance. Canceling terms and simplifying yields the final expression used for the membership probability of a given galaxy.

$$p_{spec} = \frac{\lambda f_{peak} G(v|r, M)}{\lambda f_{peak} G(v|r, M) + (1 - f_{peak}) \frac{1}{2v_{max}}} \tag{23}$$

To find the total richness, the membership probabilities of all galaxies within a given distance away from the central must be summed over. For the sake of this paper, all galaxies within r_h are included in the sum for each cluster, where r_h is the model parameter from the f_{peak} fit. This parameter follows a power law in mass, so the larger the halo mass, the farther out the sum will go to compute the total richness.

$$\lambda_{spec} = \sum_{r<r_h} p_{spec} \tag{24}$$

Because p_{spec} depends on λ as well, a root finder is required to solve for λ. This completes the model required to compute the spectroscopic richness of galaxy clusters around a given central.

For this paper, we utilize this framework to determine the spectroscopic richness for all halos in the Cardinal galaxy simulation. For each halo, a central was determined and all galaxies around this central within r_h were assigned a membership probability based on the model developed in section (3). Then all the probabilities were summed over and a root finder solved for λ. The results of this are revealed in section (5).

5. RESULTS

After completing the spectroscopic richness finder for all halos in Cardinal, the mass-richness relation for all Cardinal halos is shown in Figure (4). To determine the linear best fit and the model of the scatter, the same log-likelihood process done on the redMaPPer data in equations 1, 2, and 3 was done. This method allowed for a new fit to the spectroscopic data and model for the scatter as a function of mass. In Figure (4), the orange line represents the linear best fit to the data and the blue lines represent the scatter from the best fit as a function of mass.

For comparison, the spectroscopic mass-richness relation is compared to the photometric mass-richness relation from redMaPPer in Figure (5).

Finally, we will compare the scatter between the two richness methods in Figure (6). As expected, the scatter from the spectroscopic method shows an improvement from redMaPPer.

6. DISCUSSION

The results show a definite improvement in the scatter of the spectroscopic richness compared to the photometric richness. This was expected due to the ability of spectroscopic data to better discriminate between member and non-member galaxies. However, there are certainly some improvements to the method outlined in this paper. The first potential improvement would be to incorporate a scaling relation between the dispersion

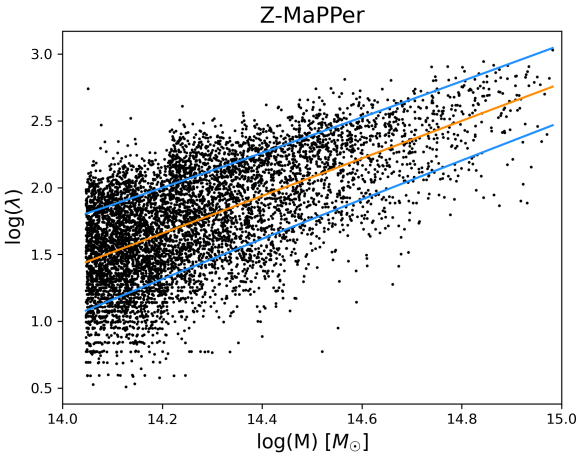
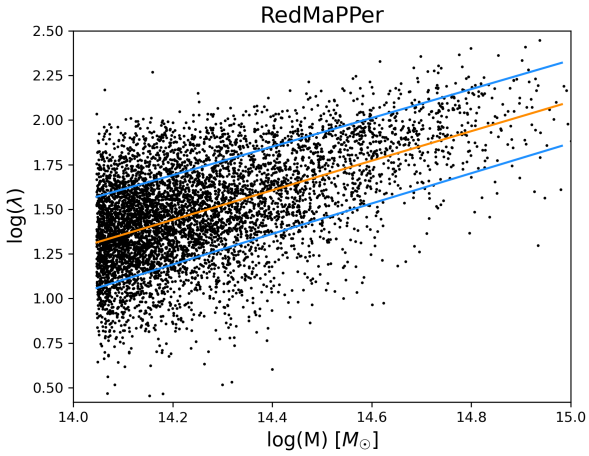


Figure 4: Figure (4) shows the mass-richness relation for the Cardinal halos using the spectroscopic method

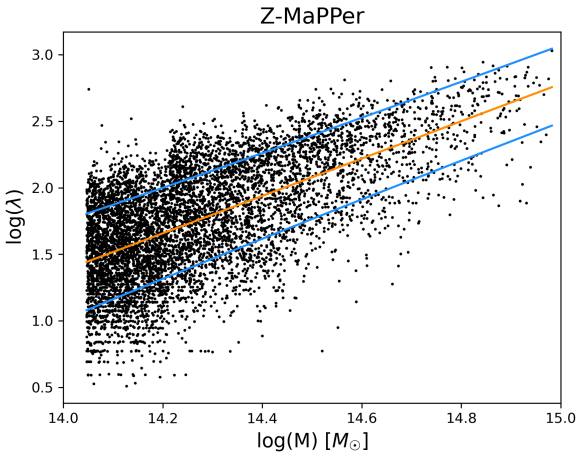
of the velocity (σ_{peak}) and the richness. From Figure (3), it is clear that a more massive halo has a higher velocity dispersion, so incorporating a direct scaling relation between the dispersion and that richness could improve precision, rather than giving all halos in a mass bin the same parameter. Another improvement that can be made from this scaling relation would be to make the richness calculation process iterative, similar to the current redMaPPer. To do so, all galaxies out to a certain distance around a central would be given a membership probability, and therefore a cluster richness just as outline in section (4). This cluster richness could be passed back to the model to determine the new velocity dispersion parameter based on the mentioned scaling relation between the velocity dispersion and the richness. The new set of model parameters is then passed into the model and a new richness for each cluster is determined using the method in section (4). This method should converge to a stable richness for each halo with only a few iterations, but this can be tested with future works.

Because this model utilizes spectroscopic data, it assumes every galaxy's redshift is known. While helpful to develop this algorithm, this is currently not possible observationally with real data. To use this spectroscopic algorithm on real data, the challenge of determining a spectroscopic redshift for every galaxy must be solved. However, for the purposes of this paper to simply determine if this spectroscopic method is even an improvement, using mock surveys works well.

While the increase in scatter is a promising result, incorporating some improvements should lower the scatter in the richness even more. After incorporating some of these procedures, a future direction for this spectroscopic method would be to combine it into redMaPPer.



(a) redMaPPer mass-richness relation



(b) Spectroscopic mass-richness relation

Figure 5: Figure (5) shows the comparison between the mass-richness relations between the two methods.

Since spectroscopic information includes all photometric data as well, combining the photometric and spectroscopic membership probabilities should also increase the algorithm's ability to determine member galaxies.

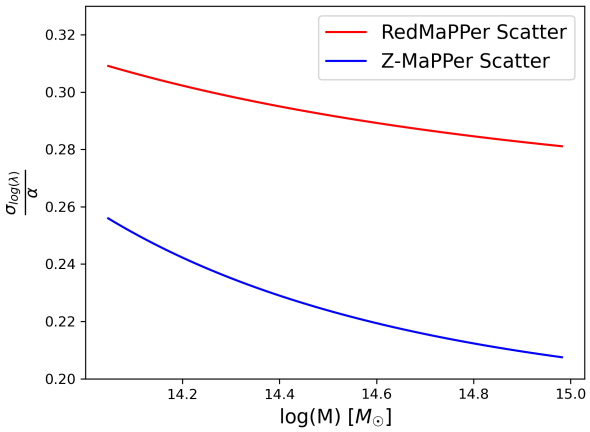


Figure 6: Figure (6) shows the comparison between the scatter as a function of mass between the two methods

7. REFERENCES

Garcia, R et al. 2022, arxiv: 2207.11827
Navarro, J. et al. 1995, arxiv: 9508025
Rykoff, E et al. 2014, arxiv: 1303.3562
To, C et al. 2023, arxiv: 2303.12104
Tomooka, P et al. 2020, arxiv: 2003.11555